ISSN : 2527-5836
e-ISSN : 2528-0074

Vol. 7 No. 3, September 2022

JISKa

Jurnal Informatika Sunan Kalijaga

Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta

Tim Pengelola JISKa Edisi September 2022

Mitra Bestari (*Reviewer*)

Reviewer Internal:

1. Mandahadi Kusuma, M.Eng. (UIN Sunan Kalijaga Yogyakarta, Indonesia)
2. Maria Ulfa Siregar, Ph.D. (UIN Sunan Kalijaga Yogyakarta, Indonesia)

Reviewer Eksternal (Mitra Bestari):

1. Ahmad Fathan Hidayatullah, M.Cs. (Universitas Islam Indonesia Yogyakarta, Indonesia)
2. Alam Rahmatulloh, M.T. (Universitas Siliwangi Tasikmalaya, Indonesia)
3. Alfan Farizki Wicaksono, Ph.D. (Universitas Indonesia, Indonesia)
4. Ardiansyah Musa Efendi, Ph.D. (Chonnam National University, Korea Selatan)
5. Dr. Aris Puji Widodo, M.T. (Universitas Diponegoro, Indonesia)
6. Dr. Cahyo Crysdian (UIN Maulana Malik Ibrahim Malang, Indonesia)
7. Dr. Enny Itje Sela (Universitas Teknologi Yogyakarta, Indonesia)
8. Dr.Eng. Ganjar Alfian (Universitas Gadjah Mada, Indonesia)
9. Muhammad Habibi, M.Cs. (Universitas Jenderal Achmad Yani Yogyakarta, Indonesia)
10. Muhammad Rifqi Maarif, M.Eng. (Universitas Jenderal Achmad Yani Yogyakarta, Indonesia)
11. Dr.Eng. M. Muhammad Syafrudin (Sejong University, Korea Selatan)
12. Dr.Eng. M. Alex Syaekhoni (Dongguk University Seoul, Korea Selatan)
13. Norma Latif Fitriyani, M.Sc. (Sejong University, Korea Selatan)
14. Nur Aini Rakhmawati, Ph.D. (Institut Teknologi Sepuluh November, Indonesia)
15. Prof. Dr. Hj. Okfalisa, S.T., M.Sc. (UIN Sultan Syarif Kasim Riau, Indonesia)
16. Oman Somantri, M.Kom. (Politeknik Negeri Cilacap, Indonesia)
17. Puji Winar Cahyo, M.Cs. (Universitas Jenderal Achmad Yani Yogyakarta, Indonesia)
18. Rischan Mafrur, M.Eng. (The University of Queensland Brisbane, Australia)
19. Dr.Eng. Sunu Wibirama, M.Eng. (Universitas Gadjah Mada, Indonesia)
20. Yudistira Dwi Wardhana Asnar, Ph.D. (Institut Teknologi Bandung, Indonesia)

ISSN : 2527-5836

e-ISSN: 2528-0074

JISKa

Vol. 7, No. 3, SEPTEMBER 2022

DAFTAR ISI

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022, Pp. 143 – 162
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Pembatasan Akses Menggunakan MAC Address dengan Metode *Access Control List*

Muhammad Aditya Rabbani [(1)*], Martanto Martanto [(2)], Yudhistira Arie Wijaya [(3)]
[1] Teknik Informatika, STMIK IKMI, Cirebon
[2] Manajemen Informatika, STMIK IKMI, Cirebon
[3] Sistem Informasi, STMIK IKMI, Cirebon
e-mail : {adityarabbani10,martantomusijo,yudhistira010471}@gmail.com.
* Penulis korespondensi.
Artikel ini diajukan 26 Februari 2022, direvisi 29 Juni 2022, diterima 29 Juni 2022, dan dipublikasikan 25 September 2022.

Abstract

The Cangehgar Cyber Command Center of the 14th Arhanud/ PWY Battalion from Cirebon City is one of the offices with IT equipment to assist the job. The servers, like the office PCs, are connected via the local network. Due to the risk of leaking secret data from within, network security concerns must be handled so that unauthorized users cannot mistakenly access the server. It seeks to limit access when there are administrative customers and employees in each room by utilizing the access control list approach using a MAC address. Access to the server is restricted to the administrator's computer, while access to the employee's PC is disallowed. Then the questionnaire was distributed to find out the respondent's assessment of the access control list. According to the results of the study on security indicators, access control lists containing MAC addresses are useful in limiting access to server computers.

Keywords: Computer, Intranet Network, MAC Address, Access Control List, Security, Questionnaire

Abstrak

Setiap kantor memiliki peralatan komputer guna menunjang pekerjaan, seperti pada kantor Cangehgar *Cyber Operation Center* di Batalyon Arhanud 14/PWY, Cirebon. Komputer yang ada di kantor tersebut saling terhubung berkat adanya jaringan intranet, begitu pula dengan komputer server. Masalah keamanan jaringan perlu ditangani, sehingga pengguna yang tidak sah tidak bisa sengaja mengakses ke server, karena risiko kebocoran data rahasia dari bagian dalam. Menggunakan metode *access control list* dengan MAC *address*, bertujuan untuk melakukan batasan akses, di mana pada setiap ruangan memiliki komputer *client* admin dan juga staf. Hanya komputer admin yang dapat berkomunikasi dengan komputer server, sementara komputer staf aksesnya ditolak. Kemudian kuesioner disebarkan untuk mengetahui penilaian responden terhadap *access control list*. Hasil penelitian terhadap indikator keamanan cenderung setuju bahwa *access control list* dengan MAC *address* berguna membatasi akses ke komputer server.

Kata Kunci: Komputer, Jaringan Intranet, MAC *Address*, *Access Control List*, Keamanan, Kuesioner

1. PENDAHULUAN

Setiap departemen di suatu perusahaan, baik pemerintah maupun swasta memiliki perangkat komputer untuk menunjang pekerjaannya (Kartini & Eko, 2019). Bagian-bagian tersebut saling berhubungan dan dapat saling berkomunikasi (Gunawan & Agung, 2019). Komunikasi dapat beroperasi melalui sistem jaringan intranet. Jaringan intranet membutuhkan sistem pertahanan yang baik untuk mencegah dan memperkecil adanya pencurian informasi yang tidak diketahui (Ardiansyah & Yudiastuti, 2021). Keamanan jaringan merupakan bagian penting dalam menjaga informasi rahasia perusahaan (Munawar & Putri, 2020). Keamanan terhadap jaringan intranet adalah bentuk perlindungan dari kemungkinan terjadinya serangan pada sistem dan pencurian data pribadi serta perusahaan yang tidak boleh diketahui orang banyak (Sulaiman & Saripurna, 2021). Perusahaan yang menggunakan jaringan intranet perlu adanya perlindungan dari segi keamanan, karena banyaknya serangan dapat menyebabkan bocornya data yang sensitif milik

perusahaan (Bustami & Bahri, 2020). Terdapat beberapa indikator yang digunakan dalam menjaga aset perusahaan, di antaranya terdiri dari *confidentiality, integrity,* dan *availability* (Kelrey & Muzaki, 2019; Riskiyadi et al., 2021). Cara yang digunakan untuk menyelesaikan kasus tersebut adalah melakukan batasan akses terhadap komputer *client* menggunakan metode *Access Control List* (ACL). Penggunaan metode *access control list* mampu menentukan sebuah paket data yang dikirim atau diterima, diberi ijin untuk melewati *router* atau tidak (Sihotang et al., 2020) guna mencegah terjadinya kebocoran informasi data dari segi internal perusahaan, yang disebabkan oleh kelalaian karena tidak adanya batasan akses menuju komputer server.

Berdasarkan penelitian terdahulu yang membahas tentang *access control list* telah dilakukan oleh Dicky Novariansyah dan Irwansyah (Irwansyah & Novariansyah, 2019) terdapat permasalahan yang terjadi berupa belum maksimalnya keamanan secara keseluruhan, yang menyebabkan akses data dapat diakses oleh siapapun serta bisa menyebabkan kebocoran aset. Hasil riset tersebut berupa penyaringan terhadap akses LAN pada arsitektur VLAN berdasarkan alamat jaringan (IP). Penelitian yang dilakukan oleh Kurniati dan Rahmat Novrianda Dasmen (Kurniati & Dasmen, 2019) memiliki masalah berupa tingkat keamanan jaringan yang masih rendah dan diperlukan sebuah upaya untuk meningkatkan keamanan dengan membatasi akses pengguna terhadap komunikasi lalu lintas jaringan. Hasil dari penelitian tersebut adalah dengan menggunakan ACL dapat membatasi akses dan hanya pengguna terdaftar yang dapat masuk. Penelitian yang dilakukan oleh Umar Hasan dan kawan-kawan (Hasan & Dewi, 2022) menyajikan permasalahan berupa banyaknya instansi dan organisasi atau kelompok besar yang tidak memperhatikan keamanan pada jaringan komputer. Sebuah celah keamanan dapat dimanfaatkan oleh pihak ketiga sebagai peluang untuk melakukan serangan siber. Hasil dari riset ini adalah aplikasi *access control list* pada *router* Cisco mampu memantau paket data yang melewati *router*, yang merupakan peningkatan dalam industri keamanan karena memiliki proses otentikasi dan verifikasi.

Riset yang telah dilakukan oleh peneliti terdahulu mengenai keamanan jaringan menggunakan metode *access control list* memiliki dampak yang baik terhadap peningkatan keamanan. Riset yang dilakukan mempunyai kontribusi dan novelty berupa bentuk visualisasi yang diterapkan langsung menggunakan *routerboard* Mikrotik, yang sebelumnya hanya dilakukan secara simulasi menggunakan Cisco Packet Tracer, serta penelitian lebih berfokus pada jaringan intranet yang ada di lingkungan kantor Cangehgar *Cyber Operation Center*. Konfigurasi *access control list* pada *routerboard* Mikrotik dalam penelitian ini menggunakan *extended access list*, yang nantinya antara komputer server mampu berkomunikasi dengan komputer *client*, serta komputer *client* khususnya komputer admin mampu berkomunikasi balik ke komputer server, kecuali komputer staf, karena MAC *address* perangkat tersebut telah diatur untuk di-*drop* agar tidak bisa melakukan akses *ping* ke komputer server.

1.1 TINJAUAN PUSTAKA

1.1.1 *Access Control List*

Access control list mampu menyaring sebuah paket data pada lalu lintas jaringan yang menentukan agar paket data tersebut cocok untuk dilewati atau dihentikan (*drop*) (Purba, 2021).

1.1.2 Jenis *Access Control List*

Metode *access control list* terdapat 2 jenis, di antaranya adalah *standard access list* dan juga *extended access list*. Keduanya memiliki perbedaan, yaitu *standard access list* hanya memperhatikan IP sumber (*resource*) dari paket yang dikirim, sementara untuk *extended access list* mempertimbangkan IP sumber (*resource*), IP tujuan (*destination*) serta protokol dan jenis yang digunakan (Hafizhan et al., 2020).

1.1.3 Mikrotik

Mikrotik merupakan alat jaringan yang digunakan untuk melakukan konfigurasi dan mengontrol sebuah jaringan (Ahmad et al., 2020). Dalam penelitian ini Mikrotik digunakan sebagai visualisasi langsung untuk menerapkan konfigurasi *access control list* dengan MAC *address*.

1.1.4 Indikator Keamanan

Sebuah sistem dapat dikatakan aman apabila terdapat beberapa indikator, di antaranya adalah *confidentiality* (kerahasiaan) data dan informasi yang tidak boleh diketahui oleh pihak lain, *integrity* (keaslian) data agar tidak bisa diakses oleh pihak yang tidak memiliki izin, *availability* (ketersediaan) data dan informasi ketika sedang dibutuhkan serta *authentication* (autentikasi) terhadap *user* yang ingin melakukan akses ke jaringan intranet (Dianta & Zusrony, 2019; Sugawara & Nikaido, 2014).

2. METODE PENELITIAN

Riset kali ini menggunakan *Network Development Life Cycle* (NDLC) untuk metode penelitiannya. Metode ini memiliki 6 langkah, mulai dari analisis (*analysis*), perancangan (design), simulasi (*simulation prototyping*), implementasi (*implementation*), pemantauan (*monitoring*), hingga pengelolaan (*management*) (Nugroho & Daniarti, 2021). Metode ini sering digunakan untuk penelitian jaringan, karena sering digunakan untuk mengembangkan jaringan sebelumnya menjadi desain jaringan baru agar lebih efisien. Berikut rincian 6 langkah tersebut.

1) *Analysis*
   Tahap ini adalah tahapan pertama yang dilakukan dalam proses penelitian. Pada tahapan ini dilakukan analisis kebutuhan, mendefinisikan masalah yang ada, dan melakukan analisis jaringan yang digunakan.
2) *Design*
   Tahap perancangan atau desain adalah tahap menampilkan gambar sebagai rancangan topologi jaringan yang akan digunakan.
3) *Simulation Prototyping*
   Pada langkah *simulation prototyping* peneliti melakukan simulasi konfigurasi yang akan dilakukan, sebelum konfigurasi tersebut diimplementasikan langsung oleh alat jaringan yang akan digunakan. Hal ini penting untuk meminimalkan kesalahan saat konfigurasi langsung.
4) *Implementation*
   Pada tahap ini dilakukan konfigurasi langsung pada perangkat jaringan yang digunakan. Dalam penelitian ini digunakan *router* Mikrotik dalam implementasinya.
5) *Monitoring*
   Langkah ini merupakan langkah setelah menyelesaikan semua konfigurasi, juga langkah untuk memeriksa dan memonitor jaringan intranet yang digunakan.
6) *Management*
   Pada tahap *management* merupakan tahap akhir di mana hal tersebut dilakukan supaya mekanisme yang dibuat berjalan lancar dan dapat bertahan lama serta mempertahankan kinerja terbaiknya.

2.1 Sumber Data dan Teknik Pengumpulan Data

Sumber data dihasilkan dari data primer dan juga sekunder. Sumber data primer merupakan hasil dari observasi lapangan, wawancara mendalam, dan juga penyebaran kuesioner. Sumber data sekunder diperoleh dari jurnal-jurnal peneliti terdahulu, studi pustaka, serta buku-buku. Berikut teknik pengumpulan data yang dilakukan.

2.1.1 Observasi Lapangan

Observasi dilaksanakan di tempat penelitian guna memperoleh data riset melalui penglihatan secara langsung terhadap kondisi yang terjadi di lapangan. Dalam tahap ini diketahui desain topologi yang sedang digunakan, yaitu topologi *ring*.

2.1.2 Wawancara Mendalam

Wawancara dilakukan kepada narasumber terkait untuk memperoleh data tambahan mengenai topik yang kita teliti. Narasumber diajukan kepada koordinator, ketua tim, dan juga anggota tim. Contoh pertanyaan dengan unsur *why* dan *how* untuk mencari informasi tambahan dapat dilihat pada Tabel 1.

Tabel 1 Contoh Pertanyaan Wawancara

No.	Pertanyaan	Jawaban
1	Mengapa jaringan intranet begitu penting digunakan di kantor Cangehgar *Cyber Operation Center*?	Karena jaringan intranet termasuk ke dalam bagian dari *Local Area Network* yang batasannya hanya mencakup komunikasi antar ruangan. Adanya jaringan intranet pun berfungsi untuk melakukan pengiriman dan penerimaan data antar komputer server dengan komputer *client* yang ada.
2	Bagaimana keamanan jaringan intranet yang saat ini digunakan?	Keamanan jaringan intranet yang ada saat ini bisa dibilang masih kurang cukup, karena belum adanya konfigurasi lanjut, masih mengandalkan kepercayaan sesama pengguna *user* lainnya yang menggunakan jaringan tersebut.

2.1.3 Kuesioner

Kuesioner disebarkan kepada responden untuk dapat menjawab sebuah pernyataan mengenai bahasan yang sedang diteliti, khususnya yang menggunakan jaringan intranet dengan ketentuan kuesioner menggunakan skala Likert. Beberapa contoh pernyataan kuesioner tersebut dapat dilihat pada Tabel 2.

Tabel 2 Pernyataan Kuesioner

No.	Indikator	Pernyataan
1	*Confidentiality*	Kerahasiaan data (*confidentiality*) pada sebuah keamanan jaringan intranet sangat penting karena meliputi dengan asset informasi yang dimiliki.
2	*Integrity*	Keaslian data (*integrity*) dalam sebuah asset yang dimiliki tidak boleh diubah saat proses pengiriman data pada jaringan intranet oleh orang yang tidak memiliki kewenangan.
3	*Availability*	Ketersediaan data (*availability*) pada saat diakses dan bisa digunakan untuk transfer *file* antar perangkat komputer.
…		…

2.2 Populasi dan Sampel Penelitian

Populasi yang digunakan sebanyak 50 orang pengguna jaringan intranet. Dari jumlah populasi tersebut ditentukan jumlah sampelnya sebagai responden penelitian. Pemilihan uji coba sampel berdasarkan rumus Isaac dan Michael pada Pers. (1) ditunjukkan pada Tabel 3.

Tabel 3 Rumus Sampel Isaac Michael

N	S			N	S		
	1%	5%	10%		1%	5%	10%
10	10	10	10	35	33	32	31
15	15	14	14	40	38	36	35
20	19	19	19	45	42	40	39
...	...	...	...	**50**	47	44	**42**

$$s = \frac{\lambda^2.N.P.Q}{d^2(N-1)+\pi^2.P.Q} \qquad (1)$$

Di mana s merupakan ukuran sampel, λ^2 yaitu chi kuadrat yang nilainya tergantung derajat kebebasan (dk) dan taraf kesalahan; dengan dk = 1, taraf kesalahan 10%, nilai chi kuadrat sebesar 2,706 (tabel chi kuadrat), N adalah total populasi, P merupakan nilai probabilitas benar (0,5) sedangkan Q menunjukkan nilai probabilitas salah (0,5), dan d adalah perbedaan sampel dan rata-rata populasi, perbedaan mencakup 0,001; 0,005; dan 0,1.

Dengan batas toleransi 10% dan nilai d = 0,05. Perhitungannya seperti pada Pers. (2).

$$s = \frac{2{,}706 \times 50 \times 0{,}5 \times 0{,}5}{0{,}05^2 \times (50-1)+2{,}706 \times 0{,}5 \times 0{,}5} = \frac{33{,}825}{0{,}799} = 42{,}33 = 42 \text{ (dibulatkan)} \qquad (2)$$

2.3 Analisa Data

Analisa data dianggap sebagai kunci pada suatu riset, karena dapat memberikan hasil penelitian sebagai suatu laporan yang bisa diambil manfaatnya (Sugawara & Nikaido, 2014). Analisa data bertujuan untuk menemukan arti dari setiap data yang telah dikumpulkan, sehingga mampu menafsirkan hubungan satu dengan lainnya yang dapat diterima oleh logika. Analisa data sendiri dilakukan terhadap hasil data yang diperoleh dari observasi, wawancara, yang kemudian dirangkum dan dikategorikan karena data tersebut berupa sebuah teks atau narasi (Sugawara & Nikaido, 2014). Kemudian perolehan data dari kuesioner dianalisa menggunakan skala Likert dengan bantuan SPSS v.26 untuk mengetahui sebuah pendapat maupun persepsi responden terhadap fenomena yang terjadi.

3. HASIL DAN PEMBAHASAN

Pada bagian hasil dan pembahasan mengulas tentang proses konfigurasi *access control list* sesuai dengan *Network Development Life Cycle (NDLC)*, mulai dari analisis, desain, simulasi, implementasi, *monitoring*, hingga manajemen. Serta menampilkan proses pengujian data yang diperoleh dari hasil kuesioner yang disebar kepada para responden, mengenai pernyataan yang berkaitan tentang *access control list* dengan MAC *address* sebagai metode keamanan jaringan intranet.

3.1 Konfigurasi Access Control List dengan MAC Address

Konfigurasi dilakukan menggunakan *routerboard* Mikrotik dengan bantuan *tool* Winbox. Proses konfigurasi mengacu pada metode *Network Development Life Cycle* (NDLC) sebagai metode yang biasa digunakan untuk perkembangan jaringan, baik dari desain sebelumnya ke desain jaringan yang baru. Berikut merupakan 6 tahap proses konfigurasi NDLC.

3.1.1 Analisis

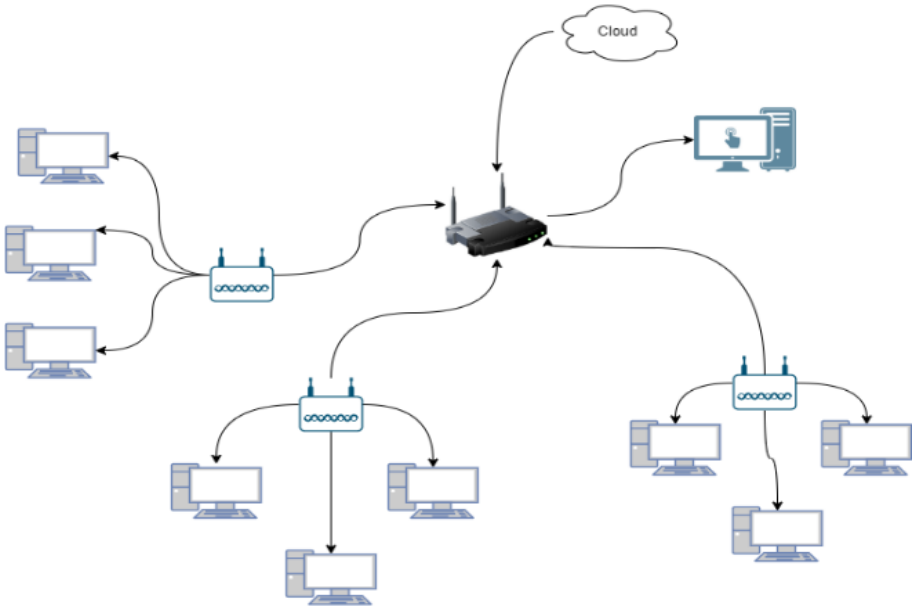
Pada tahap analisis, diperoleh hasil analisa terhadap jaringan intranet yang digunakan di lingkungan kantor cangehgar *cyber operation center* dengan detail topologi seperti yang ditunjukkan pada Gambar 1.

Hasil analisis diketahui topologi yang digunakan sebelumnya adalah topologi *ring*. Terdapat komputer server dan *access point* yang terhubung dengan *cloud*, serta memiliki perangkat *switch* pada setiap ruangan yang terhubung ke *access poin* utama yang ada di ruangan *cyber*.

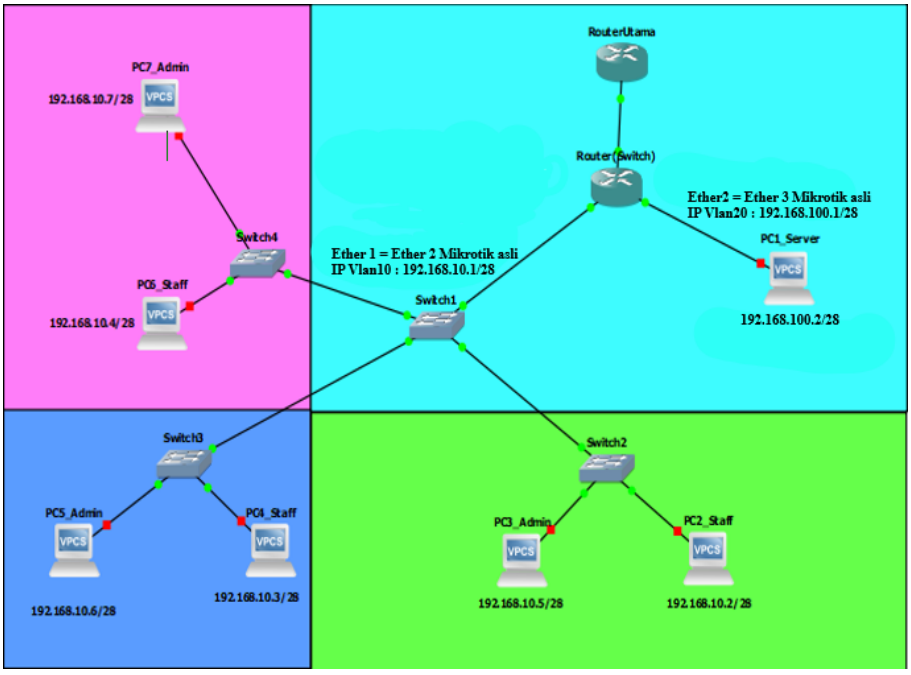
3.1.2 Desain

Dalam tahap desain ini, diusulkan sebuah jaringan baru dengan desain topologi *star*. Terdapat beberapa Mikrotik yang dijadikan sebagai pusat konfigurasi jaringan intranet. Pada topologi baru yaitu topologi *star*, terdapat Mikrotik yang digunakan sebagai pusat konfigurasi untuk jaringan intranet. Kemudian detail warna menunjukkan masing-masing divisi yang ada di lingkungan

kantor Cangehgar *Cyber Operation Center* seperti yang digambarkan pada Gambar 2. IP VLAN 20 dengan *address* 192.168.100.1/28 merupakan alamat untuk komputer server, sedangkan IP VLAN 10 dengan *address* 192.168.10.1/28 merupakan alamat untuk komputer *client,* yang terdiri dari komputer staf dan juga komputer admin.



Gambar 1 Topologi Sebelumnya



Gambar 2 Topologi Baru

3.1.3 Simulasi

Simulasi konfigurasi menggunakan *tools* jaringan yaitu GNS3, untuk konfigurasi Mikrotik. Tahap simulasi bertujuan untuk mencegah terjadinya kesalahan pada saat konfigurasi langsung

menggunakan *routerboard* Mikrotik. Contoh konfigurasi menggunakan GNS3 ditunjukkan pada Gambar 3 dan 4.

```
jan/26/2022 04:54:06 system,error,critical router was rebooted without proper shu
tdown

[admin@MikroTik] > system identity set name=RouterUtama
```

Gambar 3 Konfigurasi Identity RouterUtama

```
jan/26/2022 04:54:05 system,error,critical router was rebooted without proper shu
tdown

[admin@MikroTik] > system identity set name=RouterSwitch
[admin@RouterSwitch] >
```

Gambar 4 Konfigurasi Identity RouterSwitch

Proses konfigurasi Identity RouterUtama dan juga RouterSwitch bertujuan untuk membedakan antara kedua *router* yang saling terhubung tersebut. Pada RouterUtama lakukan setting VLAN 10 dan juga VLAN 20, dengan *address* 192.168.100.1/28 untuk VLAN 20 atau komputer server, serta 192.168.10.1/28 untuk VLAN 10 atau komputer *client.* Maka hasilnya seperti pada Gambar 5.

```
[admin@RouterUtama] > interface vlan add name=vlan10 interface=ether1 vlan-id=10
[admin@RouterUtama] > interface vlan add name=vlan20 interface=ether1 vlan-id=20
```

Gambar 5 Konfigurasi VLAN 10 dan VLAN 20

Untuk melihat *address* yang sudah terkonfigurasi, masukan perintah "*ip address print*", maka muncul *address* yang telah dikonfigurasi. Seperti ditunjukkan pada Gambar 6

```
[admin@RouterUtama] > ip address print
Flags: X - disabled, I - invalid, D - dynamic
 #   ADDRESS            NETWORK         INTERFACE
 0   192.168.10.1/28    192.168.10.0    vlan10
 1   192.168.100.1/28   192.168.100.0   vlan20
[admin@RouterUtama] >
```

Gambar 6 Konfigurasi IP VLAN 10 dan VLAN 20

Pada *router* kedua atau RouterSwitch dilakukan konfigurasi *bridge*, kepada ether1, ether2, sampai ke ether3. Buat konfigurasi *trunk* untuk ether1 yang akan menyalurkan VLAN 10 ke ether2, dan VLAN 20 ke ether3. Hasilnya seperti pada Gambar 7 dan 8.

```
[admin@RouterSwitch] > interface bridge print
Flags: X - disabled, R - running
 0 R name="bridge1" mtu=auto actual-mtu=1500 l2mtu=65535 arp=enabled arp-timeout=auto mac-address=0C:26:F1:AD:00:00
     protocol-mode=rstp fast-forward=yes igmp-snooping=no auto-mac=yes ageing-time=5m priority=0x8000 max-message-age=20s
     forward-delay=15s transmit-hold-count=6 vlan-filtering=yes ether-type=0x8100 pvid=1 frame-types=admit-all
     ingress-filtering=no dhcp-snooping=no
[admin@RouterSwitch] >
```

Gambar 7 *Interface Bridge*

```
[admin@RouterSwitch] > interface bridge port print
Flags: X - disabled, I - inactive, D - dynamic, H - hw-offload
 #     INTERFACE                  BRIDGE                 HW  PVID PRIORITY  PATH-COST INTERNAL-PATH-COST    HORIZON
 0     ether1                     bridge1                yes    1     0x80         10                 10       none
 1     ether2                     bridge1                yes   10     0x80         10                 10       none
 2     ether3                     bridge1                yes   20     0x80         10                 10       none
[admin@RouterSwitch] >
```

Gambar 8 *Interface Bridge Port*

Ether1, ether2, dan ether3 sudah tergabung kedalam *bridge1*, yang artinya semua ether sudah terhubung di dalam 1 *bridge.* Untuk melihat hasilnya ketikan perintah "*interface bridge VLAN print"*, maka hasilnya bahwa vlan10 dan vlan20 sudah berada dalam 1 *bridge*. Seperti yang ditunjukkan pada Gambar 9.

```
[admin@RouterSwitch] > interface bridge vlan print
Flags: X - disabled, D - dynamic
 #   BRIDGE                     VLAN-IDS  CURRENT-TAGGED                CURRENT-UNTAGGED
 0   bridge1                    10        ether1                        ether2
 1   bridge1                    20        ether1                        ether3
 2 D bridge1                    1                                       bridge1
                                                                        ether1
```

Gambar 9 *Interface Bridge* VLAN

Tagged adalah mode *trunk* di mana vlan-ids=10 disalurkan kepada *ports* Mikrotik ether2, serta vlan-ids=20 menyalurkan vlan20 dari *ports* ether1 kepada *ports* ether3. Hasil konfigurasi tersebut ditunjukkan pada Gambar 10.

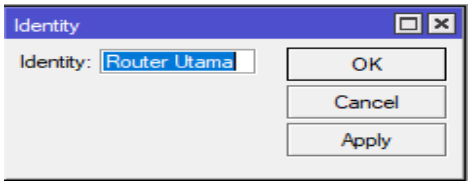
```
[admin@RouterSwitch] > interface bridge vlan add bridge=bridge1 vlan-ids=10 tagged=ether1 unta
gged=ether2
[admin@RouterSwitch] > interface bridge vlan add bridge=bridge1 vlan-id=20 tagged=ether1 untagged=ether3
[admin@RouterSwitch] > interface bridge set numbers=0 vlan-filtering=yes
```

Gambar 10 Konfigurasi *Trunk* VLAN

Pada proses yang ada di tahapan simulasi hanya sampai ke *setting* konfigurasi Mikrotiknya saja. Lebih detailnya ada pada tahapan implementasi di mana proses antar komputer *client* staf tidak bisa melakukan komunikasi *ping* (ICMP) ke komputer pusat atau server, sedangkan komputer *client* admin dapat melakukan komunikasi *ping* (ICMP) ke komputer server.

3.1.4 Implementasi

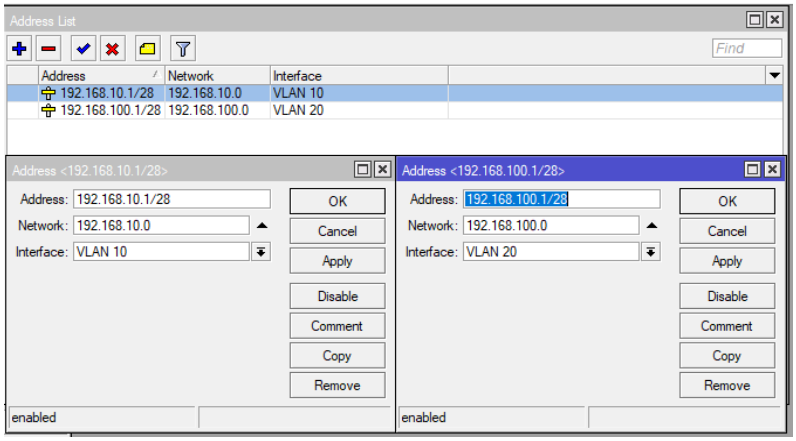
Tahap implementasi merupakan tahap implementasi langsung, di mana konfigurasi dilakukan langsung dengan menggunakan perangkat *routerboard* Mikrotik. Contoh konfigurasi *access control list* dengan MAC *address* pada Mikrotik ditunjukkan pada Gambar 11.



Gambar 11 Konfigurasi *Identity Router* Utama

Konfigurasi *identity* dilakukan untuk membedakan *router* utama dengan *router* (*switch).*

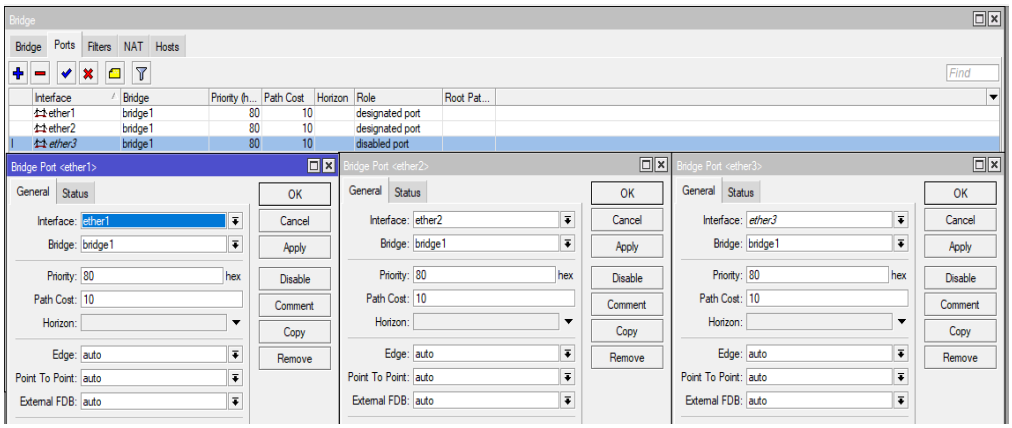
JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022: 143 – 162



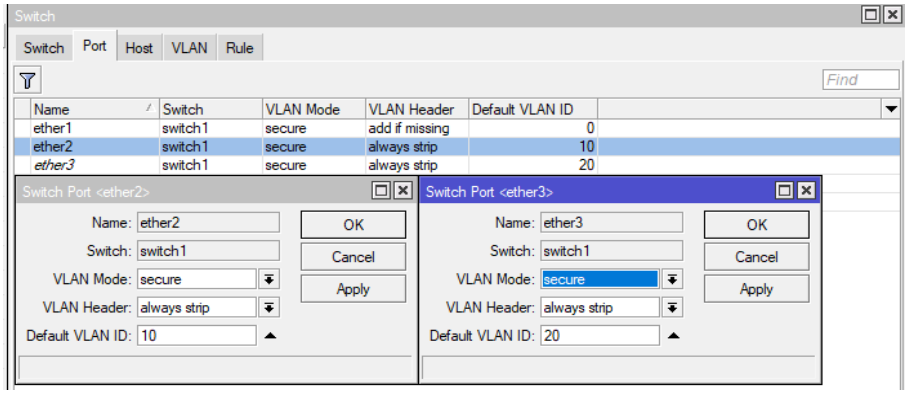
Gambar 12 Konfigurasi IP VLAN 10 dan VLAN 20

Konfigurasi *address* 192.168.10.1/28 untuk VLAN 10, seperti pada Gambar 12, yang nantinya dialokasinya untuk alamat komputer *client,* baik komputer staf maupun komputer admin. Alamat IP 192.168.100.1/28 dialokasikan untuk alamat komputer server.

Penambahan ether1 sampai dengan ether2 pada bridge1 agar dalam satu jalur atau satu jaringan seperti yang dapat dilihat pada Gambar 13 dan 14.

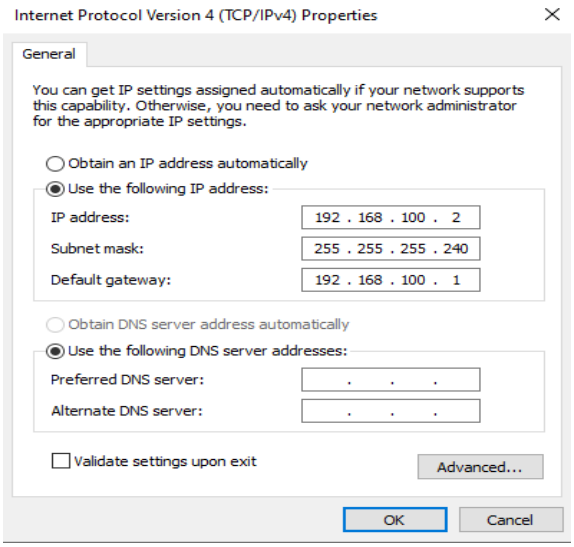


Gambar 13 Konfigurasi Bridge1 *Router* (*Switch*)

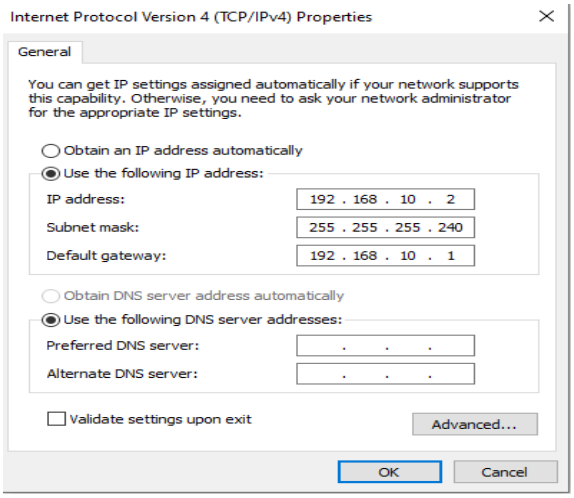


Gambar 14 Konfigurasi VLAN *Mode* dan VLAN *Header* Ether2 dan Ether3

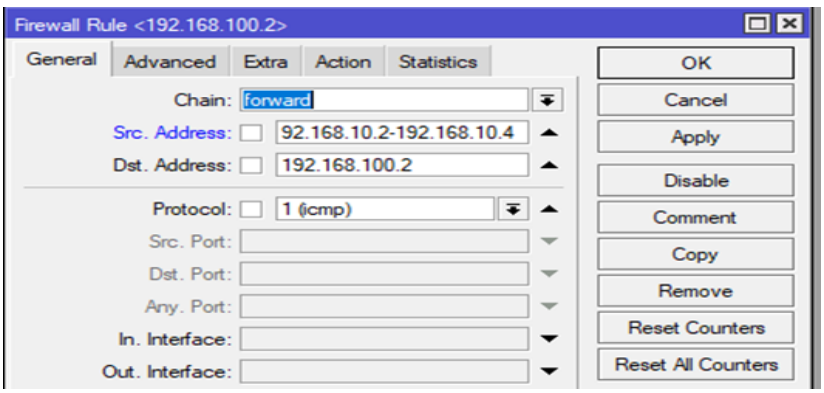
Pada bagian *IP Address* diisi dengan alamat 192.168.100.2 dengan *subnet mask* 255.255.255.240 karena menggunakan *prefix* 28 seperti yang dilakukan pada Gambar 15 dan 16.



Gambar 15 Konfigurasi IP Server



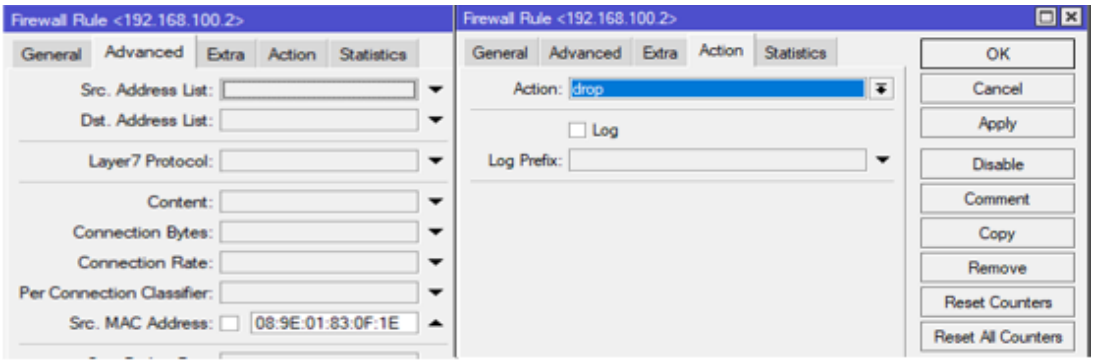
Gambar 16 Konfigurasi IP *Client* Staf



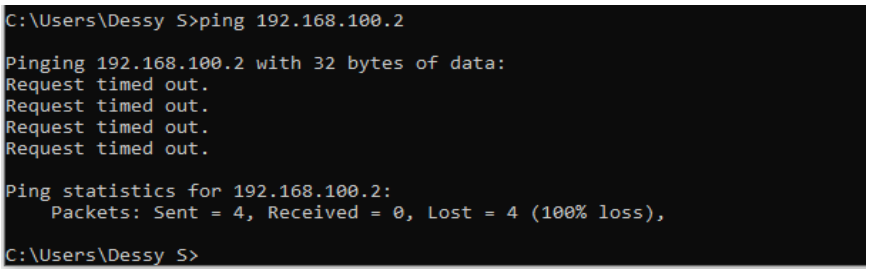
Gambar 17 Konfigurasi *Access Control List Client* Staf

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022: 143 – 162

Pada bagian *Chain* pilih *forward* dengan *Src. Address* 192.168.10.2 – 192.168.10.4 sebagai alamat dari komputer *client* Staf dengan *Dst. Address* 192.168.100.2 sebagai alamat komputer server seperti pada Gambar 17.

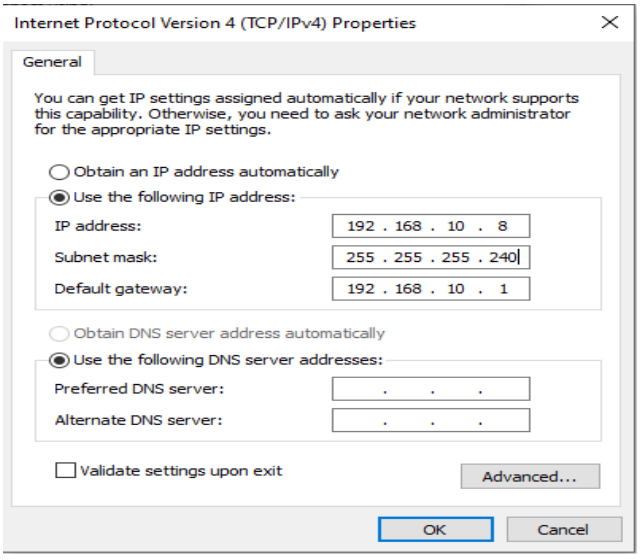


Gambar 18 Konfigurasi MAC *Address* dan *Action Drop*



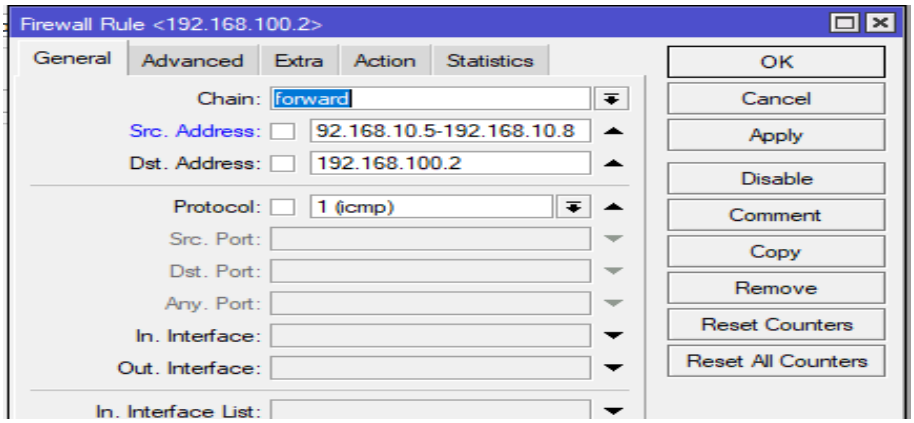
Gambar 19 Hasil Konfigurasi Setelah *Access Control List*

Komputer *client* staf tidak dapat melakukan akses kompunikasi *ping* ke komputer server, ditandai dengan *request timed out* seperti yang ditunjukkan pada Gambar 18 dan 19.



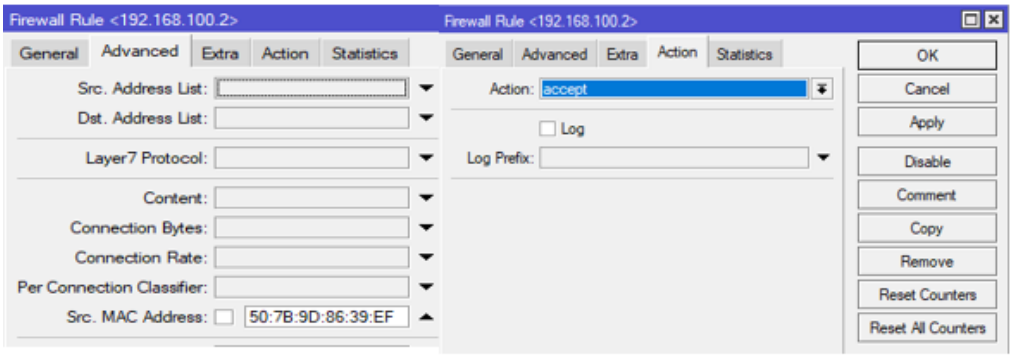
Gambar 20 Konfigurasi IP Komputer Admin

Sama dengan komputer staf, pada bagian IP *Address* diisi dengan 192.168.10.8 sebagai alamat komputer admin, dan *subnet mask* 255.255.255.240 karena menggunakan *prefix* 28 seperti pada Gambar 20.



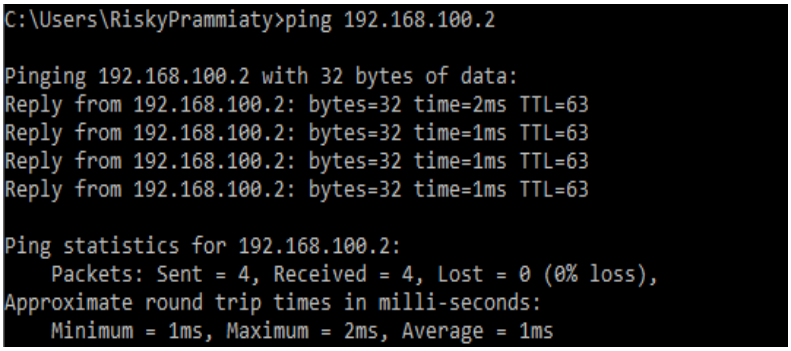
Gambar 21 Pengaturan IP *Access Control List Client* Admin

Pada bagian *Chain* pilih *forward* dengan *Src. Address* 192.168.10.5 – 192.168.10.8 sebagai alokasi alamat komputer admin, dan 192.168.100.2 sebagai alamat untuk komputer server seperti yang dicontohkan pada Gambar 21.



Gambar 22 Konfigurasi MAC *Address* dan *Action Access*

Pada bagian *Action* pilih *Accept* untuk mengizinkan komputer admin dapat melakukan komunikasi *ping* ke komputer server seperti konfigurasi yang dilakukan pada Gambar 22.



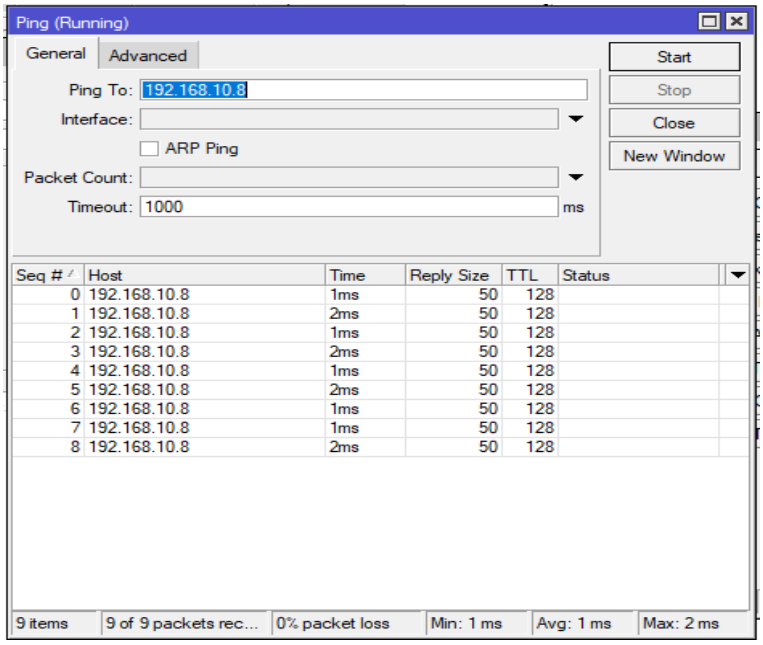
Gambar 23 Hasil *Access Control List* Komputer Admin

Hasil konfigurasi yang telah dilakukan, maka komputer *client* staf tidak dapat melakukan akses komunikasi *ping* ke komputer server, dan hanya komputer *client* admin yang dapat berkomunikasi *ping* dengan komputer server seperti hasil yang ditunjukkan pada Gambar 23. Dengan begitu pembatasan akses terhadap komputer server dapat dilakukan dan dapat meminimalkan

terjadinya kebocoran data dari pihak internal yang disebabkan oleh kelalaian anggota karena akses yang belum dibatasi.

3.1.5 *Monitoring*

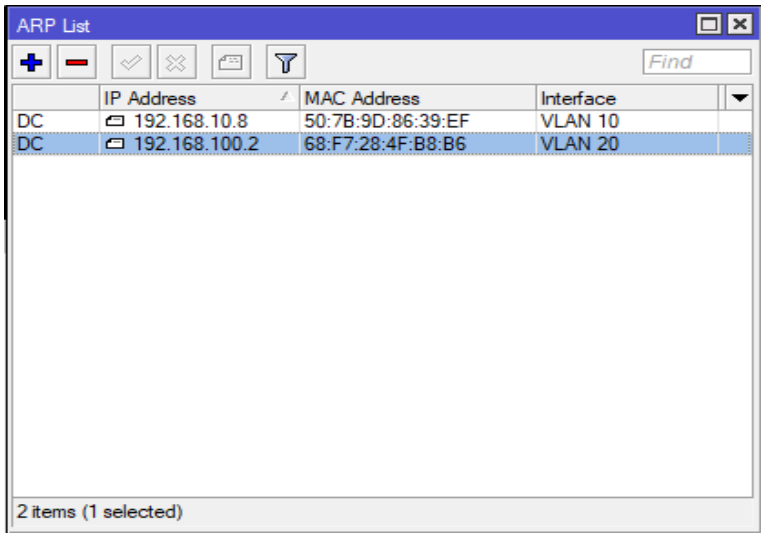
Riwayat *ping* yang dilakukan ke komputer server dapat dipantau seperti yang ditunjukkan pada Gambar 24.



Gambar 24 *Monitoring Ping* ke Komputer Server

3.1.6 Manajemen

Manajemen pada ARP (Address Resolution Protocol) List ditunjukkan pada Gambar 25.



Gambar 25 Manajemen pada ARP *List*

3.2 Proses Pengujian Data Kuesioner

Kuesioner terdiri dari total 10 pernyataan terhadap 42 responden. Pernyataan-pernyataan tersebut terdiri dari indikator keamanan, yaitu *confidentiality, integrity, accountability, availability, access control,* dan *authentication.* Hasil dari kuesioner tersebut ditunjukkan pada Tabel 4.

Tabel 4 Hasil Kuesioner Skala Likert

No. Responden	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
1	4	4	4	5	5	5	4	5	3	4
2	5	4	4	5	5	5	3	5	4	5
3	3	3	3	1	1	4	1	3	3	3
4	4	4	4	4	4	4	4	3	4	4
…	…	…	…	…	…	…	…	…	…	…
40	4	4	4	4	2	4	4	4	4	4
41	4	4	4	4	3	4	4	4	4	4
42	4	4	4	4	3	4	4	4	3	4

3.2.1 Uji Validitas Data

Pengujian validitas bertujuan untuk menguji data yang diperoleh valid (sahih) atau tidak (Erliana et al., 2019). Pengujian tersebut memakai SPSS versi 26 dengan *Pearson Correlation.* Ketentuan pengujian validitasnya adalah r hitung > r tabel. Hasil dari uji validitas data, dengan ketentuan jumlah data (N) 42 responden dengan tingkat signifikan 0,05 dan rumusnya (N-2, 0,05) = (42-2, 0,05) = 0,3044 atau 0,304. Daftar pernyataan dari uji validitas data ditunjukkan pada Tabel 5 dan 6.

Tabel 5 Pernyataan Uji Validitas Data

Indikator		Pernyataan
Confidentiality	1)	Kerahasiaan data (*confidentiality)* pada sebuah keamanan jaringan intranet sangat penting karena meliputi dengan asset informasi yang dimiliki.
Integrity	2)	Keaslian data (*integrity)* dalam sebuah asset yang dimiliki tidak boleh diubah saat proses pengiriman data pada jaringan intranet oleh orang yang tidak memiliki kewenangan
Accountability	3)	Indikator *accountability* dalam sebuah system bermanfaat untuk mengetahui data *logged user* setiap melakukan kegiatan dalam sebuah jaringan.
Availability	4)	Jaringan intranet sangat berguna untuk melakukan komunikasi antar perangkat dalam satu jaringan.
	5)	Ketersediaan data (*availability)* pada saat diakses dan bisa digunakan untuk transfer *file* antar perangkat computer.
	6)	Manfaat dari penggunaan jaringan intranet dalam melakukan pekerjaan dirasa sangat efektif.
Access Control	7)	Keamanan jaringan intranet yang digunakan di lingkungan kantor Cangehgar *Cyber Operation Center* masih minim.
	8)	Peningkatan keamanan dengan *access control list* mampu membatasi akses terhadap paket data dan user dalam melakukan akses ke jaringan intranet serta ke komputer server
Authentication	9)	Akses komunikasi antara perangkat komputer *client* dan komputer server masih terbilang bebas tanpa adanya batasan akses.
	10)	Perlu dilakukan keamanan lebih pada jaringan intranet yang digunakan

Tabel 6 Hasil Uji Validitas Data

No.	Pertanyaan	R Hitung	R Tabel	Keterangan
1	*Confidentiality* (X1)	0,681	0,304	Valid
2	*Integrity* (X2)	0,536	0,304	Valid
3	*Accountability* (X3)	0,381	0,304	Valid
4	*Availability* (X4)	0,721	0,304	Valid
5	*Availability* (X5)	0,609	0,304	Valid
6	*Availability* (X6)	0,314	0,304	Valid
7	*Access Control* (X7)	0,661	0,304	Valid
8	*Access Control* (X8)	0,515	0,304	Valid
9	*Authentication* (X9)	0,377	0,304	Valid
10	*Authentication* (X10)	0,499	0,304	Valid

3.2.2 Uji Reliabilitas Data

Pengujian reliabilitas data menggunakan *Cronbach's Alpha* dengan total 10 pernyataan terhadap 42 responden dengan tingkat signfikansi 5% atau 0,304. pengujian reliabilitas bertujuan untuk menguji data tersebut reliabel, dapat dipercaya atau juga bisa disebut konsisten (Pairingan et al., 2018). Hasil pengujian menunjukkan angka 0,745 > 0,304 seperti yang ditunjukkan pada Gambar 26.

Reliability Statistics

Cronbach's Alpha	N of Items
.745	10

Gambar 26 Hasil Uji Reliabilitas Data

3.2.3 Uji Normalitas Data

Uji normalitas adalah uji terhadap data yang dimiliki apakah berdistribusi normal atau tidak (Prasetya & Harjanto, 2020). Ketentuan pengujian data tersebut apabila nilai signifikansi > 0,05 maka disebut normal. Hasil uji dengan nilai *Kolmogorov-Smirnov* 0,200 dan *Shapiro-Wilk* 0,369 seperti yang ditunjukkan pada Gambar 27.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Hasil Responden	.079	42	.200*	.971	42	.369

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Gambar 27 Hasil Normalitas Data

Guna memperkuat hasil normalitas data, dilakukan juga pengujian *Nonparametric Test* dengan *Legacy dialogs* menggunakan metode *exact test Asymtotic Only, Monte Carlo,* dan juga *Exact* dengan masing-masing hasil seperti yang ditunjukkan pada Gambar 28 sampai 30.

Hasil dari *Asymtotic Only* pada Gambar 28 menunjukkan nilai signifikasi 0,200, dan dapat dikatakan berdistribusi normal. Selanjutnya, hasil dari Monte Carlo pada Gambar 29 menunjukkan nilai signifikasi 0,938, dan dapat dikatakan berdistribuasi normal. Lalu, hasil dari *Exact* pada Gambar 30 menunjukkan nilai signifikasi 0,940, dan dapat dikatakan berdistribusi normal.

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		42
Normal Parameters[a,b]	Mean	.0000000
	Std. Deviation	.41034226
Most Extreme Differences	Absolute	.079
	Positive	.079
	Negative	-.046
Test Statistic		.079
Asymp. Sig. (2-tailed)		.200[c,d]

a. Test distribution is Normal.
b. Calculated from data.
c. Lilliefors Significance Correction.
d. This is a lower bound of the true significance.

Gambar 28 Hasil *Asymtotic Only*

One-Sample Kolmogorov-Smirnov Test

			Unstandardized Residual
N			42
Normal Parameters[a,b]	Mean		.0000000
	Std. Deviation		.41034226
Most Extreme Differences	Absolute		.079
	Positive		.079
	Negative		-.046
Test Statistic			.079
Asymp. Sig. (2-tailed)			.200[c,d]
Monte Carlo Sig. (2-tailed)	Sig.		.938[e]
	99% Confidence Interval	Lower Bound	.932
		Upper Bound	.944

a. Test distribution is Normal.
b. Calculated from data.
c. Lilliefors Significance Correction.
d. This is a lower bound of the true significance.
e. Based on 10000 sampled tables with starting seed 334431365.

Gambar 29 Hasil Monte Carlo

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		42
Normal Parameters[a,b]	Mean	.0000000
	Std. Deviation	.41034226
Most Extreme Differences	Absolute	.079
	Positive	.079
	Negative	-.046
Test Statistic		.079
Asymp. Sig. (2-tailed)		.200[c,d]
Exact Sig. (2-tailed)		.940
Point Probability		.000

a. Test distribution is Normal.
b. Calculated from data.
c. Lilliefors Significance Correction.
d. This is a lower bound of the true significance.

Gambar 30 Hasil Exact

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022: 143 – 162

3.2.4 Uji Homogenitas Data

Pengujian homogenitas merupakan tahap di mana menguji sebuah data yang dimiliki homogen atau tidak (Dohot et al., 2020). Pengujian dilakukan terhadap variabel-variabel pernyataan. Contoh hasil akhir di mana nilai signifikan > 0,05 dan data tersebut dinyatakan homogen atau sejenis ditunjukkan pada Gambar 31.

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Responden	Based on Mean	.452	1	37	.505
	Based on Median	.510	1	37	.480
	Based on Median and with adjusted df	.510	1	33.071	.480
	Based on trimmed mean	.440	1	37	.511

Gambar 31 Hasil Uji Homogenitas

3.2.5 Uji Anova

Pengujian anova dilakukan setelah melewati beberapa tahap uji sebelumnya. Pengujian data dengan uji anova dapat dilakukan apabila data sudah berdistribusi normal dan homogen (Herawati & Irwandi, 2019). Pengujian dilakukan kepada setiap variabel pernyataan yang diajukan kepada respoden, dan hasilnya rata-rata > 0,05 dan dinyatakan lulus uji anova. Hasil uji anova terhadap masing-masing variabel ditunjukkan pada Gambar 32 sampai 41.

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16.318	2	8.159	1.332	.276
Within Groups	232.707	38	6.124		
Total	249.024	40			

Gambar 32 Uji Anova Variabel X1

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.343	2	2.171	.337	.716
Within Groups	244.682	38	6.439		
Total	249.024	40			

Gambar 33 Uji Anova Variabel X2

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11.024	2	5.512	.880	.423
Within Groups	238.000	38	6.263		
Total	249.024	40			

Gambar 34 Uji Anova Variabel X3

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	41.338	3	13.779	2.455	.078
Within Groups	207.687	37	5.613		
Total	249.024	40			

Gambar 35 Uji Anova Variabel X4

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	14.262	3	4.754	.749	.530
Within Groups	234.762	37	6.345		
Total	249.024	40			

Gambar 36 Uji Anova Variabel X5

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.142	1	5.142	.822	.370
Within Groups	243.882	39	6.253		
Total	249.024	40			

Gambar 37 Uji Anova Variabel X6

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.127	3	3.376	.523	.669
Within Groups	238.897	37	6.457		
Total	249.024	40			

Gambar 38 Uji Anova Variabel X7

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	63.928	3	21.309	4.260	.011
Within Groups	185.096	37	5.003		
Total	249.024	40			

Gambar 39 Uji Anova Variabel X8

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.167	2	4.084	.644	.531
Within Groups	240.857	38	6.338		
Total	249.024	40			

Gambar 40 Uji Anova Variabel X9

ANOVA

Responden

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12.180	2	6.090	.977	.386
Within Groups	236.844	38	6.233		
Total	249.024	40			

Gambar 41 Uji Anova Variabel X10

Hasil pengujian data yang telah dilakukan, mulai dari uji validitas, uji reliabilitas, uji normalitas, uji homogen, hingga ke uji anova, memiliki hasil nilai akhir dengan tingkat signifikansi melebihi 0,05. Data-data tersebut bisa digunakan karena telah melewati tahapan-tahapan pengujian data.

4. KESIMPULAN

Konfigurasi keamanan jaringan intranet di kantor Cangehgar Cyber Operation Center berhasil dilakukan. Hal tersebut dapat dilihat dari proses konfigurasi yang telah dilakukan. Pembatasan akses menggunakan MAC *address* dengan metode *access control list*, mampu untuk membatasi akses perangkat komputer antara *client* staf dan *client* admin. Hanya komputer admin yang dapat melakukan akses komunikasi *ping* (ICMP) ke komputer server, sedangkan komputer staf tidak bisa berkomunikasi karena aksesnya di *drop.* Pembatasan tersebut berguna untuk menghindari sembarang akses dan kebocoran informasi dari segi internal. Sejalan dengan hasil konfigurasi yang berhasil, data-data dari kuesioner yang telah diolah melewati beberapa tahap pengujian mendapatkan tingkat akhir signifikansi diatas 0,05. Hasil tersebut menandakan bahwa responden cenderung setuju perihal *access control list* dengan MAC *address* berguna untuk membatasi hak akses. Saran peneliti untuk penelitian selanjutnya adalah membuat *access control list* menggunakan alat Switch Manageable dari Mikrotik, dengan sistem operasi SWOS untuk lebih mempermudah proses konfigurasi.

UCAPAN TERIMA KASIH

Terima kasih banyak kepada Divisi Cangehgar *Cyber Operation Center* yang telah mengizinkan penulis untuk melakukan riset di tempat tersebut.

DAFTAR PUSTAKA

Ahmad, T., Imtihan, K., & Wire, B. (2020). Implementasi Jaringan Inter-VLAN Routing Berbasis Mikrotik Rb260Gs Dan Mikrotik Rb1100Ahx4. *JIRE (Jurnal Informatika & Rekayasa Elektronika)*, *3*(1), 77–84. https://doi.org/10.36595/jire.v3i1.221

Ardiansyah, A. H., & Yudiastuti, H. (2021). Perancangan Jaringan Intervlan Routing Dan Penerapan Acls Pada Pt. Sinar Alam Permai Dengan Simulasi Menggunakan Packet Tracer. *Prosiding Semhavok*, *3*(1), 210–218.

Bustami, A., & Bahri, S. (2020). Ancaman, Serangan dan Tindakan Perlindungan pada Keamanan Jaringan atau Sistem Informasi : Systematic Review. *UNISTEK*, *7*(2), 59–70. https://doi.org/10.33592/unistek.v7i2.645

Dianta, I. A., & Zusrony, E. (2019). Analisis Pengaruh Sistem Keamanan Informasi Perbankan Pada Nasabah Pengguna Internet Banking. *INTENSIF: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi*, *3*(1), 1. https://doi.org/10.29407/intensif.v3i1.12125

Dohot, S., Khairina, N., & Robin. (2020). Pembuatan Media Pembelajaran Fisika Berbasis Hots Untuk Tingkat Smp. *Pendidikan Fisika*, *9*(1), 63–67. https://doi.org/10.22611/jpf.v9i1.18173

Erliana, H., Akos, M., & Priono, S. (2019). Pengaruh Disiplin Kerja Terhadap Kinerja Dengan Kepuasan Kerja Sebagai Variabel Intervening. *Administraus*, *3*(2), 31–58. https://doi.org/10.56662/administraus.v3i2.75

Gunawan, J., & Agung, H. (2019). Implementation of PPTP and BCP with Inter-VLAN on the Topology that Uses 2 ISP as Inter-Division Connectors (Case Study: PT Kenari Djaja Prima). *Jurnal Algoritma, Logika Dan Komputasi*, *2*(1), 138–150. https://doi.org/10.30813/j-

alu.v2i1.1574

Hafizhan, M., Wahyuddin, M. I., & Komalasari, R. T. (2020). Implementasi Packet Filtering Menggunakan Metode Extended Access Control List (ACL) Pada Protokol EIGRP. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *4*(1), 185. https://doi.org/10.30865/mib.v4i1.1926

Hasan, U., & Dewi, S. (2022). Penerapan Metode Access Control List Pada Jaringan VLAN Menggunakan Router Cisco. *IMTechno: Journal of Industrial Management and Technology*, *3*(1), 37–41. https://doi.org/10.31294/imtechno.v3i1.927

Herawati, L., & Irwandi. (2019). Pengaruh Model Pembelajaran Kooperatif Tipe Jigsaw Terhadap Hasil Belajar dan Berpikir Kritis Siswa Pada Mata Pelajaran IPA di SMP Negeri 09 Lebong. *Prosiding Seminar Nasional Sains Dan Entrepreneunship Vi*, 1–9.

Irwansyah, I., & Novariansyah, D. (2019). Pengembangan Keamanan Jaringan Vlan Dan Acls Pt. Taspen (Persero) Palembang Menggunakan Simulasi Packet Tracer. *Prosiding Semhavok*, *1*(1), 95–102.

Kartini, D., & Eko, A. (2019). Upgrade Skill Komputer Perangkat Desa Pemakuan. *Jurnal Pengabdian Kepada Masyarakat MEDITEG*, *4*(2), 7–11. https://doi.org/10.34128/mediteg.v4i2.48

Kelrey, A. R., & Muzaki, A. (2019). Pengaruh Ethical Hacking Bagi Keamanan Data Perusahaan. *Cyber Security Dan Forensik Digital*, *2*(2), 77–81. https://doi.org/10.14421/csecurity.2019.2.2.1625

Kurniati, K., & Dasmen, R. N. (2019). The Simulation of Access Control List (ACLs) Network Security for Frame Relay Network at PT. KAI Palembang. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, *10*(1), 49. https://doi.org/10.24843/LKJITI.2019.v10.i01.p06

Munawar, Z., & Putri, N. I. (2020). Keamanan Jaringan Komputer pada Era Big Data. *J-SIKA|Jurnal Sistem Informasi Karya Anak Bangsa*, *2*(01), 14–20.

Nugroho, F. E., & Daniarti, Y. (2021). Rancang Bangun QoS (Quality of Service) Jaringan Wireless Local Area Network Menggunakan Metode NDLC (Network Development Life Cycle) di PT Trimitra Kolaborasi Mandiri (3KOM). *JIKA (Jurnal Informatika)*, *5*(1), 79. https://doi.org/10.31000/jika.v5i1.3970

Pairingan, A., Allo Layuk, P. K., & Pangayow, B. J. . (2018). Pengaruh Kompetensi, dan Independensi Terhadap Kualitas Audit dengan Motivasi Sebagai Variabel Pemoderasi. *Jurnal Akuntansi, Audit, Dan Aset*, *1*(1), 1–13. https://doi.org/10.52062/jurnal_aaa.v1i1.2

Prasetya, T. A., & Harjanto, C. T. (2020). Pengaruh Mutu Pembelajaran Online Dan Tingkat Kepuasan Mahasiswa Terhadap Hasil Belajar Saat Pandemi. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *17*(2), 188–197. https://doi.org/10.23887/jptk-undiksha.v17i2.25286

Purba, G. C. (2021). Implementation Of Network Packet Filtering With Extended Acl Methods On Microtik In Securing Internet Connection Office AFD IV Butong Sulfur Unit. *Infokum*, *9*(2), 287–293.

Riskiyadi, M., Anggono, A., & Tarjo. (2021). Cybercrime dan Cybersecurity pada Fintech: Sebuah Tinjauan Pustaka Sistematis. *Jurnal Manajemen Dan Organisasi*, *12*(3), 239–251. https://doi.org/10.29244/jmo.v12i3.33528

Sihotang, B. K., Sumarno, S., & Damanik, B. E. (2020). Implementasi Access Control List Pada Mikrotik dalam Mengamankan Koneksi Internet Koperasi Sumber Dana Mutiara. *JURIKOM (Jurnal Riset Komputer)*, *7*(2), 229. https://doi.org/10.30865/jurikom.v7i2.2010

Sugawara, E., & Nikaido, H. (2014). Properties of AdeABC and AdeIJK Efflux Systems of Acinetobacter baumannii Compared with Those of the AcrAB-TolC System of Escherichia coli. *Antimicrobial Agents and Chemotherapy*, *58*(12), 7250–7257. https://doi.org/10.1128/AAC.03728-14

Sulaiman, O. K., & Saripurna, D. (2021). Network Security System Analysis Using Access Control List (ACL). *International Journal of Information System & Technology Akreditasi*, *5*(2), 192–197. https://doi.org/10.30645/ijistech.v5i2.131

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022, Pp. 163 – 176
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Prototipe Alat Ukur Detak Jantung Menggunakan Sensor MAX30102 Berbasis *Internet of Things* (IoT) ESP8266 dan Blynk

Muthmainnah Muthmainnah [(1)*], Deni Bako Tabriawan [(2)]
Fisika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim, Malang
e-mail : inna@fis.uin-malang.ac.id, denibakotabriawan@gmail.com.
* Penulis korespondensi.
Artikel ini diajukan 7 Juli 2022, direvisi 30 Agustus 2022, diterima 31 Agustus 2022, dan dipublikasikan 25 September 2022.

Abstract

The heart is an important organ of the human body. The heart functions to pump blood throughout the body. Health conditions can be seen in the condition of heart function. The heart's function can be known through the beat when pumping blood. The manufacture of a heart rate device has been carried out using the PPG method. This tool uses the MAX30102 sensor as input. The measurement results are displayed on the smartphone. This tool can calculate the heart rate by sticking the surface of the fingertips for ten seconds. The light waves emitted by the sensor source will hit the surface of the finger. Changes in blood volume cause changes in light intensity according to what is received by the sensor. Based on the test results, the average standard deviation of this tool's heart rate measurement is 1.176. If considered the data from the pulse oximeter is correct, then this tool has an accuracy of 98.804%.

Keywords: Heart Rate, IoT, Blynk, MAX30102, Infrared

Abstrak

Jantung adalah organ penting bagi tubuh manusia. Jantung berfungsi memompa darah ke seluruh tubuh. Kondisi kesehatan dapat dilihat dari kondisi fungsi jantung. Fungsi jantung dapat diketahui melalui detakan pada saat memompa darah. Pembuatan alat detak jantung telah dilakukan dengan menggunakan metode PPG. Alat ini menggunakan sensor MAX30102 sebagai inputan. Hasil pengukuran ditampilkan pada *smartphone*. Alat ini dapat menghitung detak jantung dengan menempelkan permukaan ujung jari selama sepuluh detik. Gelombang cahaya yang dipancarkan oleh sumber sensor akan mengenai permukaan jari. Perubahan volume darah menyebabkan perubahan intensitas cahaya sesuai yang diterima oleh sensor. Berdasarkan hasil pengujian data, rata-rata standar deviasi pengukuran detak jantung menggunakan alat ini adalah 1,176. Jika menganggap data dari *pulse oximeter* adalah benar maka alat ini memiliki akurasi 98,804%.

Kata Kunci: Detak Jantung, IoT, Blynk, MAX30102, Inframerah

1. PENDAHULUAN

Jantung merupakan salah satu organ yang paling penting bagi manusia karena jantung bertugas memompa darah keseluruh tubuh (Hindarto et al., 2015). Darah tersebut mengandung oksigen dan sumber makanan yang dibutuhkan oleh sel-sel tubuh. Terganggunya kinerja jantung dapat menyebabkan fungsi pada organ-organ lain terhambat (Muhajirin & Ashari, 2018). Denyut jantung dapat dijadikan salah satu parameter sesorang dalam keadaan sehat atau tidak (Nurdin et al., 2015). Beberapa metode yang digunakan untuk pemeriksaan denyut jantung adalah elektrokardiogram (EKG), phonocardiogram (PCG), Auskultasi. Namun metode tersebut hanya dapat dilakukan oleh para ahli dibidangnya (Anugrah, 2016).

Sekarang mulai dikembangkan alat ukur untuk pemeriksaan denyut jantung yang portable dan dapat digunakan tanpa harus ke instansi kesehatan atau klinik. Hakim telah berhasil membuat alat pengukur detak jantung dengan memanfaatkan sms (Hakim & Nurwarsito, 2019). Saiful menggunakan elektrokardiograph (ECG) yang dikombinasi dengan Arduino dan LCD sebagai alat pengukur detak jantung (Sufri & Aswardi, 2020). Alat pengukur detak jantung juga telah dikembangkan menggunakan metode MQTT, AD8232 (Hariri et al., 2019), lora (Astutik & Bakti,

2020) dan berbasis *wireless* (Chooruang & Mangkalakeeree, 2016; Yassin et al., 2019). Ambary memanfaatkan ATmega sebagai komponen utama dalam pembuatan alat pengukur detak jantung (Ambary & Raharja, 2018). Alat *monitoring* denyut jantung telah dikembangan dengan menggunakan fotopletismograf (Sipayung et al., 2018), Arduino nirkabel (Isyanto & Jaenudin, 2018), sensor pulsa (Kumari & Victor, 2019), pulsa *heart* jari tangan (Rachmat & Ambaransari, 2018), *photodetector* (Kusuma et al., 2018), Bluetooth (Ahmad & Arfian, 2017), pulsa sensor dikombinasi dengan DS18B20(Ahmad & Arfian, 2017), Android Studio (Irawan et al., 2019), ATmega16 (Wijaya et al., 2020) menggunakan sensor suara (Faesal et al., 2020) dan sensor AD8232 berbasis *internet of things* (IoT).

Alat ukur detak jantung berbasis IoT menjadi topik penelitian yang banyak diminati sejak pandemic Covid-19. Alat ini dapat mendeteksi kondisi detak jantung tanpa harus berinteraksi langsung dengan pasien. Hasil pengukuran akan ditampilkan pada *smartphone* ataupun PC dengan jaringan internet. Beberapa penelitian tentang alat ukur detak jantung berbasis IoT telah dikembangkan. Mallick memanfaatkan perubahan volume darah yang mengalir pada ujung jari untuk mengetahui detak jantung. Alat ini menggunakan *infrared led emitting diode* (IR LED), Arduino dan *processing software* sebagai tampilan. Alfalah membuat alat ukur detak jantung yang dapat dipantau secara *real time* dengan menggunakan metode Naive Bayes. Sensor yang digunakan pada alat ini adalah AD8232. Sensor AD8232 berkerja berdasarkan elektrokardiogram. Metode Naïve Bayes diterapkan untuk mengkatagorikan pasien dalam kelompok normal dan tidak normal (Anugara, 2021).

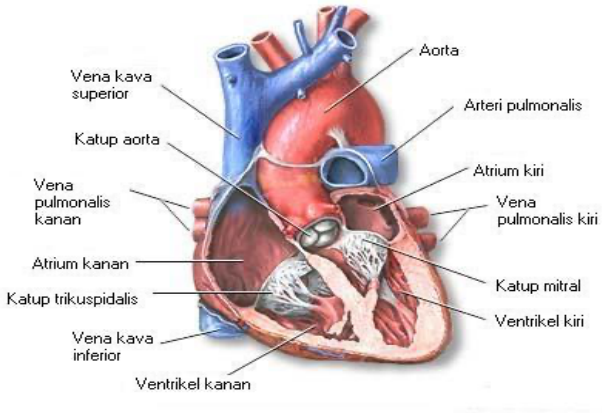
Rancang bangun alat pengukur detak jantung berbasis komunikaasi *wifi* dan android telah dilakukan oleh Yulidarti. Alat ini memanfaatkan sensor *easy pulse* dan NodeMCU sebagai IO dan kontrol. Pengukuran data dilakukan dengan menempelkan sensor di dada. Hasil pengujian menunjukkan perbandingan *error* 0-6% pada naracoba laki-laki dan perempuan (Yulidarti & Hendri, 2020). Hudhajanto membuat alat pengukur detak jantung yang *wearable*. Alat ini diletakkan pada *face shield* dan memanfaatkan ESP-WROOM-32 sebagai mikrokontroler. Dari 20 data percobaan data yang dihasilkan memiliki galad 3,2% dibandingkan dengan *pulse oximeter* (Hudhajanto et al., 2022).

Penelitian terdahulu memiliki tingkat akurasi yang belum maksimal jika dibandingkan dengan alat baku yang telah digunakan. Pada penelitian ini akan dikembangkan alat pengukur detak jantung *non-invasive* dengan metode photoplethysmograph berbasis IoT. Modul IoT yang digunakan adalah ESP8266. Modul ini dipilih karena generasinya lebih lama sehingga banyak *software* pendukung yang telah dikembangkan. Untuk proyek sederhana modul ini lebih tepat karena tidak membutuhkan perangkat berlebih seperti Bluetooth. Sensor detak jantung yang digunakan adalah sensor MAX30102. Kelebihan dari sensor ini adalah sudah *include* dengan modul I2C sehingga pada saat dihubungkan dengan perangkat mikrokontroler tidak membutuhkan banyak sambungan kabel. Pin yang digunakan hanya SCL dan SDA dari modul sensor. Generasi 30102 memiliki keunggulan dibandingkan generasi sebelumnya yaitu 30100. Selain dapat mengukur detak jantung, MAX30102 juga dapat mengukur saturasi oksigen dan dilengkapi dengan sensor suhu. Aplikasi Blynk dipilih untuk menampilkan data pada *smartphone*. Aplikasi ini mudah diunduh dan digunakan pada *smartphone* karena sudah tersedia di Google PlayStore. Untuk aplikasi-aplikasi sederhana Blynk masih menyediakan *wire* yang tidak berbayar.

1.1 Jantung

Jantung merupakan organ tubuh manusia yang berfungsi sebagai pemompa darah ke semua bagian tubuh manusia. Darah mengalir melalui pembuluh darah (arteri dan vena) dengan irama khas yang berulang. Pembuluh arteri mempunyai gelombang berdetak yang dapat dirasakan pada permukaan kulit dan disebut sebagai denyut nadi. Denyut nadi yang dilewati oleh arteri radialis terletak pada pergelangan tangan, arteri temporalis pada bagian atas tulang temporal, serta arteri dorsalis pedis pada bagian siku mata kaki (Muhajirin & Ashari, 2018). Jantung memiliki alat pacu alami yang bernama simpul Sino Atrial (SA) yang terletak pada bagian serambi kanan atas. Di dalam jantung terdapat nodus SA yang menciptakan sinyal listrik segingga membuat

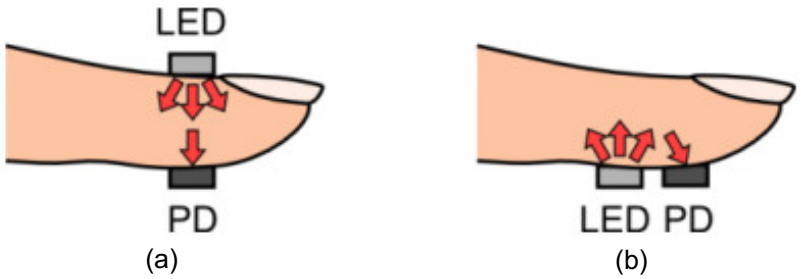
jantung dapat berdetak dan menghasilkan denyutan 60-100 bpm (Yessianto et al., 2018). Anatomi fisiologi jantung lebih jelasnya digambarkan pada Gambar 1.



Gambar 1 Anatomi Fisiologi Jantung (Karina & Thohari, 2018)

Beberapa cara pengukuran detak jantung dilakukan dengan metode elektrokardiogram (EKG), photoplethysmography (PPG), phonocardiogram (PCG), dan auskultasi (Anugrah, 2016; Yulian & Suprianto, 2017). PPG adalah pendeteksian detak jantung melalui gelombang dinamis pada kardiovaskuler. PPG mendeteksi detak jantung menggunkan metode fotoelektrik dari gelombang pembuluh darah *peripheral*. Tekanan jantung memicu gelombang dan menyebar menuju pembuluh darah yang lebih dalam melalui arteri. *Probe* stasioner pada kulit dapat mendeteksi perubahan volume darah secara dinamis terhadap waktu. Perubahan volume darah menyebabkan perubahan intensitas cahaya karena penyerapan *optic* pada jaringan kulit. Perubahan intensitas cahaya dapat dideteksi secara kualitatif dengan menggunakan sebuah sensor optik dan peralatan pengkondisian sinyal (Tamura et al., 2014).

Metode PPG menggunakan cahaya dengan panjang gelombang tertentu yang dapat disesuaikan (Brugarolas et al., 2016) seperti yang ditunjukkan pada Gambar 2. *Photodetector* mengubah pantulan intensitas gelombang cahaya menjadi setara dengan perubahan volume darah yang mengalir. IR-LED memancarkan gelombang cahaya menuju permukaan kulit. Sebagian cahaya akan diserap oleh hemoglobin yang terdapat pada darah. Sebagian gelombang cahaya akan dipantulkan dan diterima oleh *photodetector*. Gelombang pantulan akan dibandingkan dengan sumber gelombang cahaya, sehingga diketahui berapa intensitas yang diserap. Saat jantung melakukan pemompaan darah pembuluh darah akan berdetak (Rachmat & Ambaransari, 2018).



Gambar 2 (a) Transmisi PPG, (b) Refleksi PPG (Rachmat & Ambaransari, 2018)

1.2 MAX30102

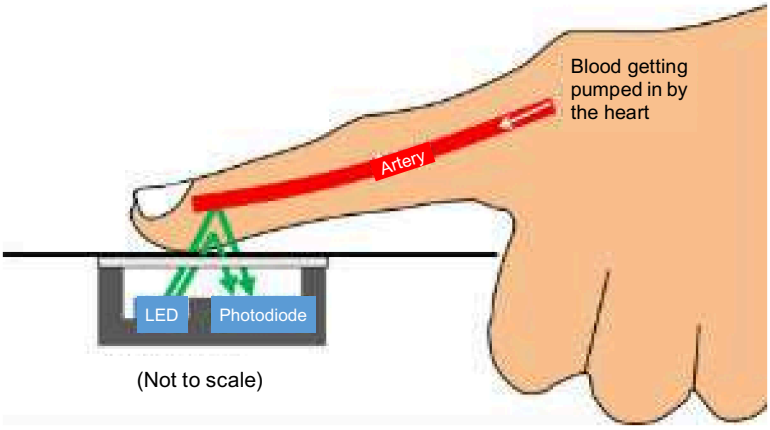
Modul MAX30102 adalah sensor yang diproduksi oleh Maxim Integrated. Sensor MAX30102 dapat mendeteksi laju detak jantung dan suhu sekaligus. Sensor ini terdiri sumber pemancar yang berupa cahaya *infrared* dan *photodetector* yang letaknya berdekatan. Sensor ini memiliki *noise* yang rendah sehingga dapat diatur dengan mudah. Pemanfaatan sensor MAX30102 lebih

pada asisten kebugaran yang dapat memantau dan memonitoring kondisi tubuh saat berolah raga melalui *interface smartphone* atau perangkat penunjang lainnya (Savitri, 2020).

MAX30102 beroprasi pada sumber tegangan tunggal sebesar 1.8 Volt dan sumber tegangan 3,3 Volt yang terpisah untuk LED internal. Modul sensor ini sudah *include* dengan I2C sebagai antar muka antara *smartphone* dan mikrokontroler. Modul ini juga dapat dikontrol menggunakan *software* (Harianto et al., 2021).Fitur dan keunggulan modul sensor MAX30102 dijabarkan pada Tabel 1.

Tabel 1 Fitur dan Keunggulan Modul Sensor MAX30102 (Karina & Thohari, 2018)

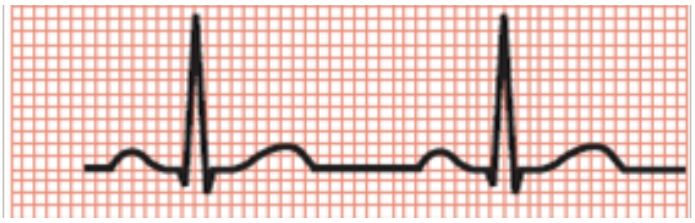
No.	Fitur dan Keunggulan
1	Monitor detak jatung dan sensor *pulse oximeter* dalam solusi refleksi LED
2	Berukuran 5,6 mm x 3,3 mm x 1,55 mm modul optic 14 pin yang dilapisi kaca penutup untuk kinerja yang kuat dan optimal
3	Operasi daya *ultra-low* untuk perangkat seluler 1) *Sample rate* yang dapat deprogram dengan arus pada LED untuk penghematan daya 2) Monitor detak jantung berdaya rendah (<1 mW) 3) Arus mati sangat rendah (0,7 µA Typ)
4	Kemampuan keluaran data yang cepat 1) *Sample rates* yang tinggi
5	Ketahanan terhadap Motion Artifact yang kuat 1) *High* SNR
6	Rentang temperatur operasi sebesar -40°C sampai dengan +85°C



Gambar 3 PPG Menggunakan Metode *Reflectance* (Karina & Thohari, 2018)

Sensor MAX30102 terdiri dari dua komponen yaitu IR-LED dan *photodiode*. Prinsip kerjanya menggunakan metode PPG seperti yang ditampilkan pada Gambar 3. Saat pertama dinyalakan LED akan memancarkan sinyal cahaya. Saat ujung jari ditempelkan, pancaran cahaya akan masuk ke pembuluh darah kapiler pada jari yang ditempelkan. Proses pemompaan darah oleh jantung akan membuat darah mengalir dari arteri yang berukuran besar ke arteri yang berukuran lebih kecil seperti jari. Perubahan intensitas cahaya disebabkan oleh terjadinya pemompaan jantung sehingga darah mengalir ke seluruh tubuh. Prubahan volume darah yang terdapat pada ujung jari akan dideteksi oleh *photodetector* melalui perubahan intensitas cahaya (Karina & Thohari, 2018).

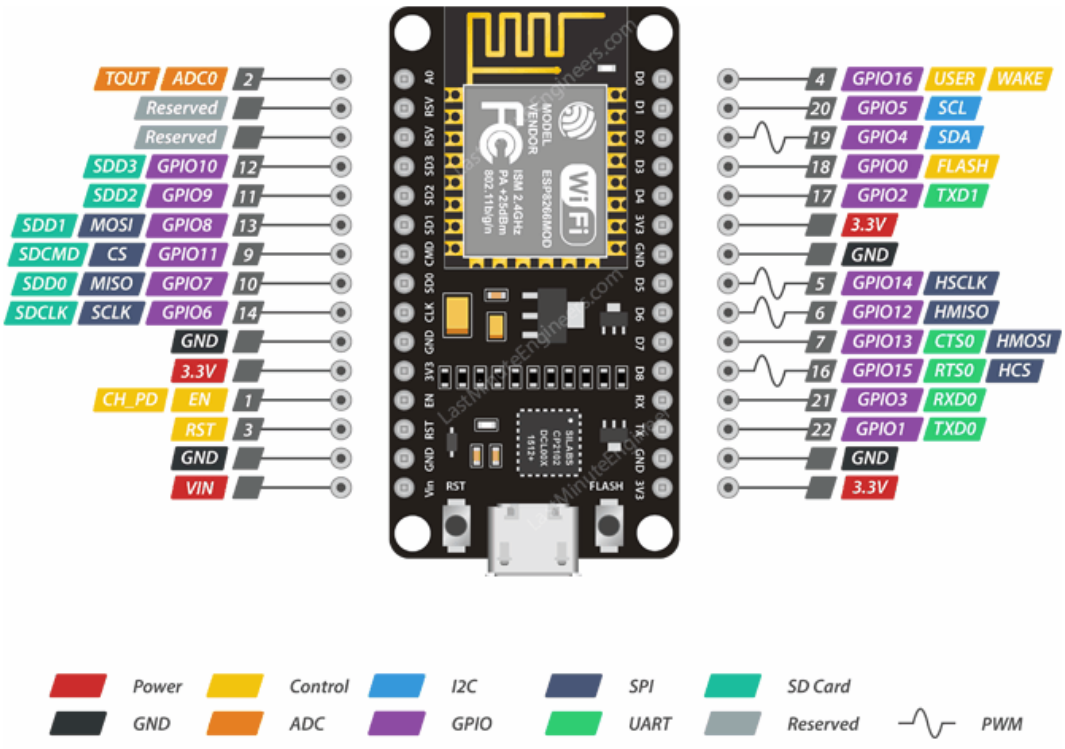
Keluaran sensor MAX30102 berupa sinyal digital yang mewakili nilai laju detak jantung. Satuan dari laju detak jantung adalah beats per minute (bpm). Pada Gambar 4 *pulse wave event* menunjukkan aktivitas depolarisasi dengan singkat dalam tekanan darah arteri yang muncul ketika katup aorta menutup (Nurdin et al., 2015).



Gambar 4 Reaksi *Heart Rate* Panjang Gelombang (Nurdin et al., 2015)

1.3 ESP8266

NodeMCU merupakan papan pengembangan produk *Internet of Things* (IoT) yang berbasiskan *Firmware eLua* dan *System on a Chip* (SoC) ESP8266-12E. ESP8266 merupakan *chip wifi* dengan protocol *stack* TCP/IP yang lengkap (Yuhefizar et al., 2019). NodeMCU dapat dianalogikan sebagai *board* Arduino-nya ESP8266. Program ESP8266 memerlukan beberapa teknik *wiring* serta tambahan modul USB to serial untuk mengunduh program. Namun NodeMCU telah mengemas ESP8266 ke dalam sebuah *board* yang kompak dengan berbagai fitur layaknya mikrokontroler ditambah dengan kapabilitas akses terhadap *wifi* juga *chip* komunikasi USB *to serial*. Sehingga untuk memprogramnya hanya diperlukan ekstensi kabel data USB yang digunakan *charging smarphone* (Rifa'i, 2016). Gambar 5 merupakan uraian kaki pin yang ada pada NodeMCU.



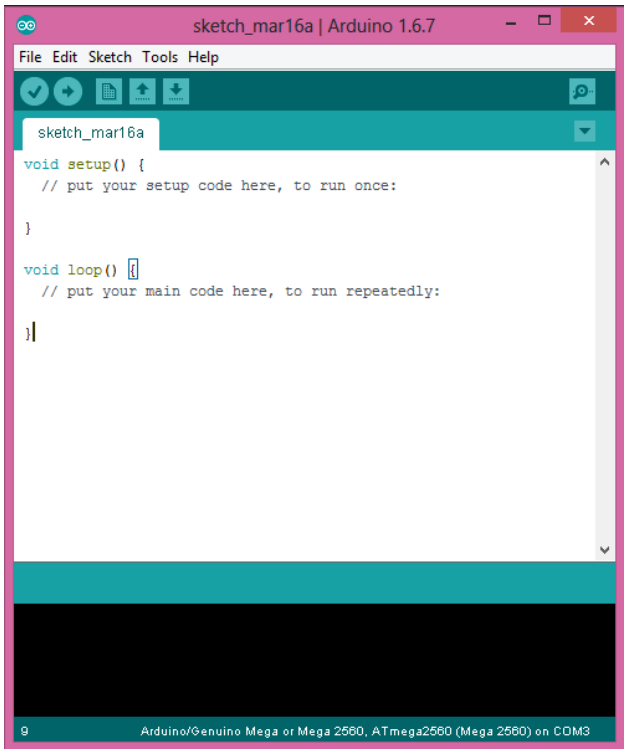
Gambar 5 NodeMCU ESP8266 (Rifa'i, 2016)

Spesifikasi dari NodeMCU sebagai berikut:

1) 10 *port* pin GPIO
2) Fungsionalitas PWM
3) Antarmuka I2C dan SPI
4) Antarmuka 1 Wire
5) ADC

1.4 Arduino IDE

IDE merupakan singkatan dari *Integrated Development Environment*. Arduino IDE merupakan aplikasi pemrograman untuk perangkat Arduino dan NodeMCU agar perangkat tersebut dapat melakukan fungsinya seperti yang diharapkan. Arduino menggunakan bahasa pemrograman sendiri yang menyerupai Bahasa C. Aplikasi Arduino IDE juga memiliki kumpulan contoh program yang berada pada *library* sehingga pemula dapat dengan mudah untuk melakukan pemrogramanan (Budi et al., 2019). Gambar 6 merupakan tampilan dari Arduino IDE serta menu pada aplikasi tersebut dijelaskan pada Tabel 2.



Gambar 6 Tampilan Arduino IDE

Tabel 2 Menu pada Arduino IDE

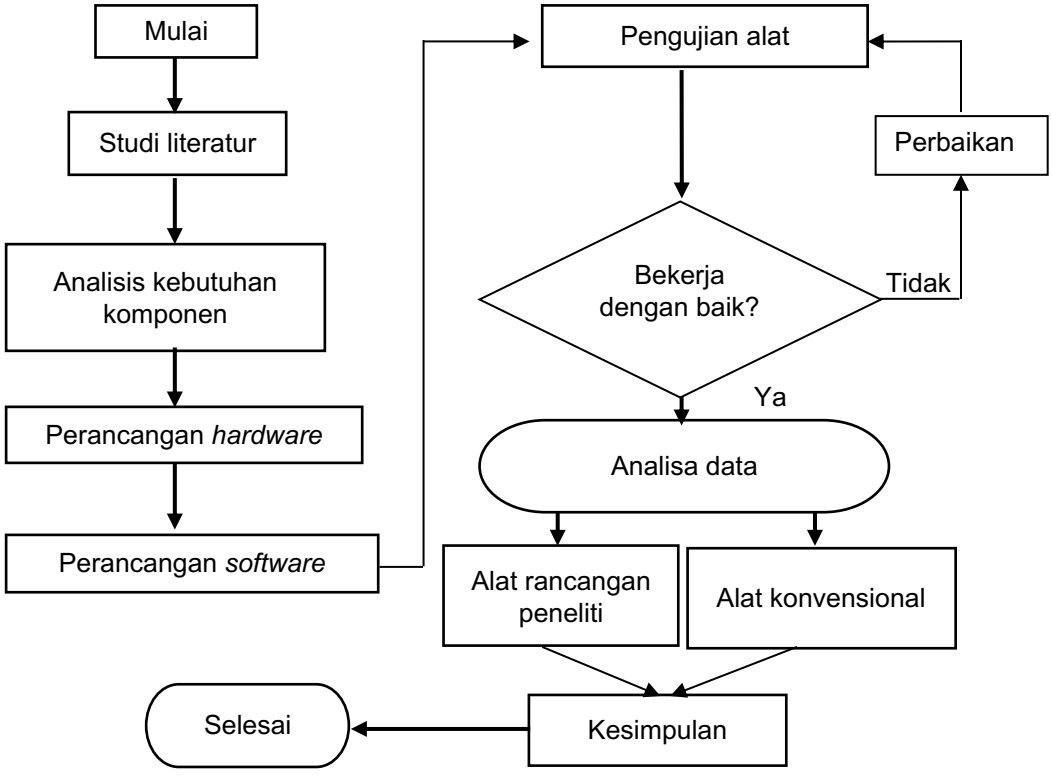
Ikon	Nama	Fungsi
	Verify	Burfungsi untuk mengecek kode program yang telah dibuat. Jika terjadi *error* maka kode program ada yang salah. Biasanya ada keterangan letak *line* dan kolom kesalahan tersebut.
	Upload	Berfungsi untuk mengkompilasi kode program ke bahasa mesin dan selanjutnya mengunggah ke mikrokontroler.
	New	Berfungsi untuk membuat program baru
	Open	Berfungsi untuk membuka *file* atau program yang telah tersimpan
	Save	Berfungsi untuk menyimpan program
	Serial Monitor	Berfungsi untuk membuka serial monitor

1.5 Blynk

Blynk merupakan aplikasi berbasis layanan yang dapat digunakan sebagai pengontrol mikrokontroler berbasis internet (Prayitno et al., 2017). Dalam aplikasi Blynk terdapat beberapa *widget* yang harus disusun sesuai dengan kebutuhan. Blynk menjadi *interface* antara mikrokontroler dan *smartphone* melalui koneksi internet. Aplikasi ini tidak berbayar jika *widget* yang digunakan sedikit. Untuk penggunaan aplikasi yang membutuhkan *widget* banyak maka ada beberapa *widget* yang berbayar. Blynk dapat merekam dan menyimpan data sesuai dengan pengaturan yang telah ditanamkan (Noar & Kamal, 2017). Blynk merupakan platform baru yang memudahkan peneliti untuk menghubungkan perangkat keras dengan tampilan pada *smartphone*. Waktu yang dibutuhkan untuk memprogram pada aplikasi Blynk relatif cepat (Serikul et al., 2018).

2. METODE PENELITIAN

Penelitian ini dimulai dengan analisis kebutuhan yaitu alat pengukur detak jantung yang ada saat ini masih menggunakan metode EKG dan PCG. Alat tersebut membutuhkan ahli medis untuk membacanya. Dibutuhkan alat detak jantung dengan metode lain lain sehingga semua kalangan dapat membaca alat tersebut. Diagram alir pada penelitian ditunjukkan pada Gambar 7.



Gambar 7 Diagram Alir Penelitian

Studi literatur dilakukan dengan mengumpulkan jurnal nasional dan jurnal internasional yang relevan terhadap topik tersebut. Jurnal-jurnal tersebut ditinjau dan dibandingkan satu dengan yang lain serta dicari metode mana yang paling mendekati dengan kebutuhan. Pengembangan alat dilakukan dengan mempertimbangkan sensor yang akan digunakan dan *output* yang diharapkan.

Alat dan bahan yang digunakan adalah sensor MAX30102, modul ESP8266, kabel power, projeck board, kabel male-fimale, Hp Android dan laptop. Pembuatan alat dan pengambilan data dilakukan di laboratorium elektronika UIN Maulana Malik Ibrahim Malang.

Perancangan alat dilakukan dengan menancapkan kaki-kaki ESP8266 pada *project board*. Kaki SDA pada sensor MAX30102 dihubungkan dengan pin D1 pada ESP8266. Kaki SCL pada MAX30102 dihubungkan dengan pin D2 pada ESP8266. Kaki SCL dan SDA pada OLED dihubungkan dengan kaki SCL dan SDA sensor MAX30102.

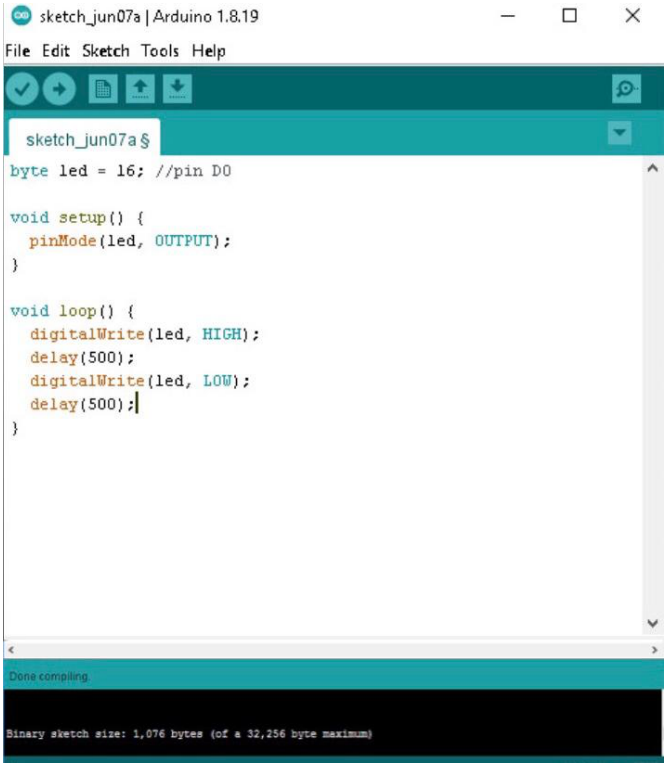
Terdapat dua perancangan *software* pada penelitian ini Arduino IDE dan Blynk. Arduino IDE diunduh pada PC atau laptop karena berfungsi untuk memprogram ESP8266. Proses pemograman pada ESP8266 dilakukan dengan menjalankan aplikasi Arduino IDE. Hubungkan ESP8266 pada laptop dengan menggunakan kabel USB. Masukkan *library* MAX30102 pada Arduino IDE. Setelah terdeteksi adanya perangkat ESP8266 maka programing dapat dilakukan. Aplikasi Blynk diunduh di *smartphone* untuk menampilkan data hasil pengukuran alat. Aplikasi Blynk yang terunduh pada *smartphone* harus diatur terlebih dahulu pada bagian GUI agar dapat menampilkan data.

Pengujian alat dilakukan dengan menguji tiap komponen dan menguji alat keseluruhan yang sudah dirancang. Pengujian komponen dilakukan satu persatu. Pengujian sensor MAX30102 dilakukan dengan menghubungkan sensor pada *board* Arduino Uno. Program pada Gambar 8 diunggah pada board Arduino dan dilihat pada serial monitor akan ada data yang keluar saat ujung jari ditempelkan.

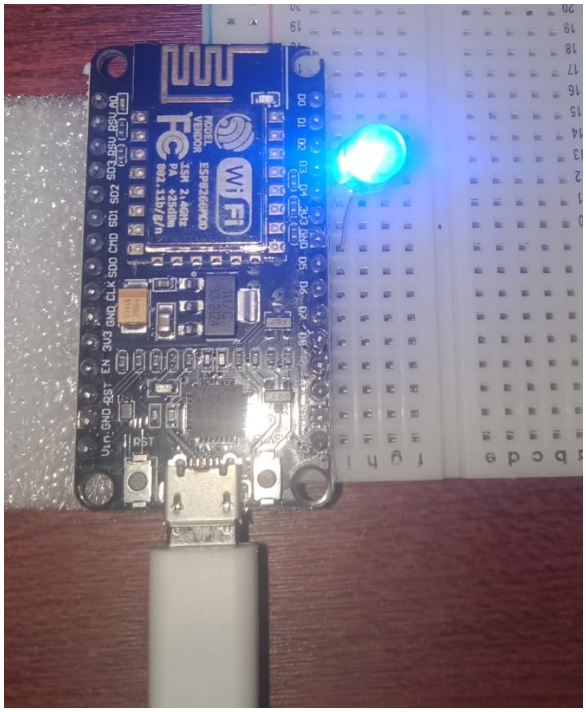
Gambar 8 Program MAX30102

ESP8266 diuji dengan menghubungkan board ke USB laptop. Saat terjadi kedip-kedip pada lampu indicator yang menandakan ESP8266 dalam kondisi menyala. Pengujian software pada ESP8266 dilakukan dengan program sederhana yaitu menyalakan LED. Program tersebut diunggah ke ESP8266 melalui software Arduino IDE seperti ditunjukkan pada Gambar 9. Tampilan ESP8266 tampak pada Gambar 10.

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022: 163 – 176



Gambar 9 Program Menghidupkan LED



Gambar 10 Pengujian LED ESP8266

Pengambilan data dilakukan dengan mengukur denyut jantung manusia. Ujung jari ditempelkan pada permukaan alat selama kurang lebih sepuluh detik. Pengukuran ini dilakukan sebanyak lima kali untuk mencari standar deviasi alat yang telah dirancang sebagaimana Pers. (1) dan (2).

Dalam waktu yang bersamaan *pulse oximeter* juga mengukur denyut jantung sebanyak lima kali. Jumlah data denyut jantung yang diolah adalah sebanyak 50.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} Xi \qquad (1)$$

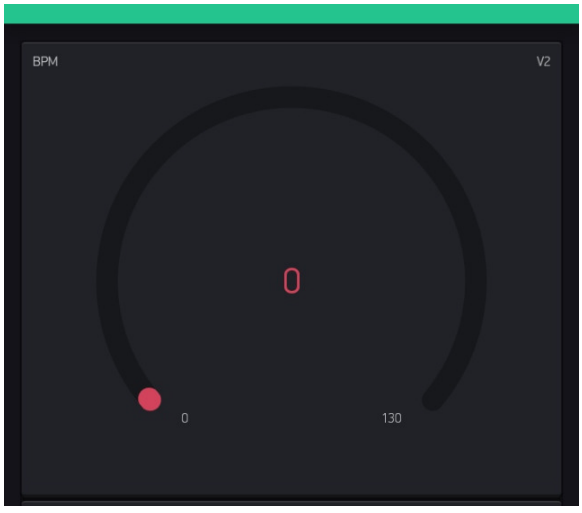
$$s = \sqrt{\frac{\sum_{i=1}^{n} (Xi - \bar{x})^2}{n-1}} \qquad (2)$$

Di mana $\bar{x}$ adalah rata-rata, s adalah standar deviasi, x_i adalah nilai sampel ke-i dan n adalah banyaknya data. Akurat merupakan titik ukur yang menunjukkan derajat kedekatan hasil analisis dengan analisa yang sebenarnya. Tingkat akurasi alat yang telah dirancang dicari dengan membandingkan nilai pengukuran alat yang telah dibuat dengan *pulse oximeter* menggunakan Pers. (3). *Pulse oximeter* yang digunakan bermerek Onemade Pulsa Oximeter Oxyone dengan kemampuan pengukuran denyut jantung 25-250 bpm. Alat ini biasa digunakan para medis untuk mengetahui kondisi awal pasien saat datang ke instansi kesehatan.

$$\%\text{Akurasi} = 100\% - \left| \frac{hasil\ alat\ kovensional - hasil\ alat\ perancangan}{hasil\ alat\ konvensional} \right| \times 100\% \qquad (3)$$

3. HASIL DAN PEMBAHASAN

Langkah pertama yang dilakukan dalam penggunaan alat ini adalah menghubungkan alat pada sumber tegangan dengan menggunakan kabel USB. Sumber tegangan dapat berupa USB laptop atau adaptor USB DC atau baterai yang memiliki kapasitas 9Vdc. Langkah kedua adalah menempelkan ujung jari pada sensor MAX30102 selama kurang lebih 10 detik. Tampilan pada *smartphone* ditunjukkan pada Gambar 11. Data pada *smartphone* adalah data berupa denyut jantung yang berasal dari sensor MAX30102. Sensor ini berkerja berdasarkan prinsip photoplethysmography (PPG) di mana volume dalam jaringan dapat terdeteksi melalui cahaya yang dipancarkan dan diterima pantulannya oleh *detector*. Saat jantung berdetak maka sebenarnya jantung sedang memompa darah keseluruh tubuh termasuk jari. Hal ini mengakibatkan volume darah pada arteri berubah. Jantung yang terus memompa akan membuat denyutan dan fluktuasi volume darah yang ada pada arteri akan terbaca dalam bentuk sinyal oleh sensor (Budi et al., 2019).



Gambar 11 Tampilan Denyut Jantung pada *Smartphone*

Sensor MAX30102 terbuat dari *infrared light emmiting diode* (IR LED) (Kumar N. et al., 2017). IR LED mentransmisikan cahaya *infrared* keujung jari yang ditempelkan pada sensor. Sebagian akan diteruskan dan sebagian akan dipantulkan kembali dari darah dalam arteri jari. *Photo diode*

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022: 163 – 176

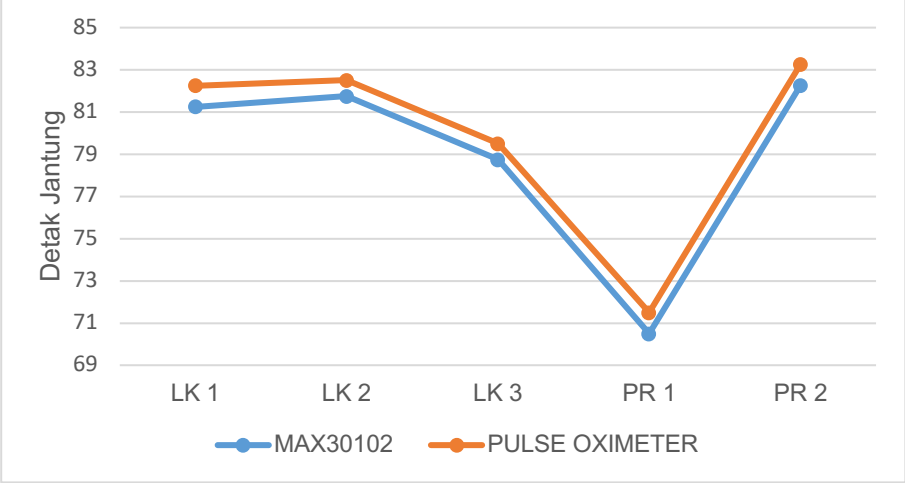
merasakan bagian cahaya yang dipantulkan dengan nilai intensitas tertentu tergantung dari volume darah yang ada pada jari. Saat jantung berdetak maka jumlah cahaya *infrared* yang diterima oleh *photo diode* juga berubah. Perubahan kecil dapat dikuatkan sehingga bisa ditangkap sebagai sinyal berdetak bagi sensor.

Hasil pengukuran alat terhadap lima naracoba ditunjukkan pada Tabel 3. Pengambilan data detak jantung untuk setiap naracoba dilakukan sebanyak 5 kali untuk alat yang dibuat dan 5 kali untuk alat pembanding yaitu *pulse oximeter*. Sehingga data keseluruhan adalah berjumlah 50. Naracoba terdiri dari tiga laki-laki dan dua perempuan. Pemilihan naracoba yang berbeda jenis kelamin dilakukan untuk mendapatkan nilai detak jantung yang bervariasi. Pegukuran pada setiap naracoba dilakukan sebanyak lima kali. Sehingga setiap data memiliki standar deviasi. Rata-rata nilai standar deviasi pada alat ukur yang telah dibuat adalah 1,716. Rata-rata nilai standar deviasi pada *pulse oximeter* adalah 1,776. Perbedaan ini tidak terlalu jauh karena hanya berbeda 0,05. Perbedaan ini terjadi karena durasi pengambilan data yang berbeda antara alat yang telah dibuat dan pulsa oksimeter. Alat membutuhkan hanya sekitar 10 detik untuk mengukur dan menampilkan data. Sedangkan *pulse oximeter* membutuhkan waktu yang relatif lebih lama. Rata-rata nilai akurasi alat adalah 98,804% jika menganggap data dari *pulse oximeter* merupakan data yang benar.

Tabel 3 Hasil Pengukuran Alat dan *Pulse Oximeter*

Naracoba	Jenis Pengolahan	MAX30102	Pulse Oximetrer
LK1	Rata-rata (bpm)	81,25	82,25
	Standar Deviasi	2,39	2,39
	Akurasi (%)	98,78	
LK2	Rata-rata (bpm)	81,75	82,5
	Standar Deviasi	2,49	2,29
	Akurasi (%)	99,09	
LK3	Rata-rata (bpm)	78,75	79,75
	Standar Deviasi	1,29	1,29
	Akurasi (%)	98,75	
LK4	Rata-rata (bpm)	70,5	71,5
	Standar Deviasi	1,12	1,12
	Akurasi (%)	98,60	
LK5	Rata-rata (bpm)	82,25	83,25
	Standar Deviasi	1,29	1,79
	Akurasi (%)	98,80	
Rata-rata Tingkat Akurasi (%)		**98,804**	

Gambar 12 merupakan gambar perbandingan hasil pengukuran yang dilakukan pada lima naracoba dengan menggunakan alat yang telah dirancang (biru) dan *pulse oximeter* (oranye). Dari gambar terlihat pengukuran *pulse oximeter* selalu lebih tinggi daripada alat ukur yang diwakili sensor MAX30102. Hal ini dapat disebabkan oleh perbedaan bagian jari yang diambil datanya. *Pulse oximeter* mengambil data pada ujung jari telunjuk tangan kiri dan alat yang telah dibuat menggukur ujung jari telunjuk tangan kanan. Perbedaan pengambilan data ini dilakukan karena pengukuran harus dalam waktu yang bersamaan. Nilai akurasi 98,804 merupakan akurasi rata-rata dari lima naracoba dan berulang sebanyak lima kali. Nilai ini cukup baik dan cukup mewakili untuk pengukuran denyut jantung awal yang relatif mengalami perubahan. Beberapa alat pengukur detak jantung telah dibuat dengan metode, komponen dan naracoba yang berbeda. Sehingga memiliki tingkat akurasi dan nilai standar deviasi yang relatif berbeda. Sensor ECG pernah diterapkan untuk mengukur denyut jantung. pengukuran dilakukan dengan menempelkan sensor pada dada. Data pengukuran ditampilkan pada LCD (Sufri & Aswardi, 2020). Berbeda dengan sensor MAX30102 yang dapat menggunakan permukaan jari untuk pengukuran denyut jantung. Alat yang dirancang ini juga dilengkapi perangkat IoT sehingga mudah diakses meskipun jaraknya berjauhan.



Gambar 12 Perbandingan Nilai Detak Jantung Alat dan *Pulse Oximeter*

4. KESIMPULAN

Alat ukur detak jantung telah dibuat dan dapat menghitung detak jantung dengan baik. Alat ini menggunakan sensor MAX30102 yang menerapkan prinsip kerja PPG. Pengukuran dilakukan dengan menempelkan jari pada alat (*non-invasive*). Data hasil pengukuran dapat diakses di *smartphone* meggunakan koneksi internet. Dari pengukuran detak jantung oleh lima naracoba dihasilkan rata-rata standar deviasi adalah 1,716. Akurasi alat ini adalah 98,804% dengan membandingkan hasil pengukurannya terhadap *pulse oximeter*.

DAFTAR PUSTAKA

Ahmad, K., & Arfian, A. (2017). Rancang Bangun Alat Pengukur Detak Jantung Antarmuka Smartphone Melalui Bluetooth. *Sinusoida*, *19*(2), 78–84.

Ambary, I. M., & Raharja, W. K. (2018). Purwarupa Alat Pendeteksi Detak Jantung Berbasis ATMEGA328. *Jurnal Ilmiah Teknologi Dan Rekayasa*, *23*(1), 38–47. https://doi.org/10.35760/tr.2018.v23i1.2449

Anugara, A. (2021). Sistem Pengukuran Detak Jantung Secara RealTime pada Platform Internet of Things Menggunakan Metode Naive Bayes. *Seminar Informatika Aplikatif Polinema*, 58–63.

Anugrah, D. (2016). Rancang Bangun Pengukur Laju Detak Jantung Berbasis PLC Mikro. *Elinvo (Electronics, Informatics, and Vocational Education)*, *1*(3), 163–170. https://doi.org/10.21831/elinvo.v1i3.10857

Astutik, R. P., & Bakti, R. F. (2020). Sistem Monitoring Detak Jantung Berbasis LoRa. *E-Link : Jurnal Teknik Elektro Dan Informatika*, *15*(1), 19. https://doi.org/10.30587/e-link.v15i1.1606

Brugarolas, R., Latif, T., Dieffenderfer, J., Walker, K., Yuschak, S., Sherman, B. L., Roberts, D. L., & Bozkurt, A. (2016). Wearable Heart Rate Sensor Systems for Wireless Canine Health Monitoring. *IEEE Sensors Journal*, *16*(10), 3454–3464. https://doi.org/10.1109/JSEN.2015.2485210

Budi, D. B. S., Maulana, R., & Fitriyah, H. (2019). Sistem Deteksi Gejala Hipoksia Berdasarkan Saturasi Oksigen dan Detak Jantung Menggunakan Metode Fuzzy Berbasis Arduino. *Journal of Information Technology Development and Computer Science*, *3*(2), 1925–1933.

Chooruang, K., & Mangkalakeeree, P. (2016). Wireless Heart Rate Monitoring System Using MQTT. *Procedia Computer Science*, *86*(March), 160–163. https://doi.org/10.1016/j.procs.2016.05.045

Faesal, A. M., Santoso, I., & Sofwan, A. (2020). Desain Stetoskop untuk Deteksi Detak Jantung Menggunakan Sensor Suara dan Penghitungan BPM (Beat Per Minute) Menggunakan Arduino. *Transmisi*, *22*(2), 44–50. https://doi.org/10.14710/transmisi.22.2.44-50

Hakim, F., & Nurwarsito, H. (2019). Sistem Pemantauan Detak Jantung dan Suhu Tubuh

menggunakan Protokol Komunikasi MQTT. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(11), 10705–10711.

Harianto, B., Hidayat, A., Hulu, F. N., Telekomunikasi, P. T., Elektro, J. T., Medan, P. N., Utara, S., Elektronika, P. T., Elektro, J. T., Medan, P. N., Utara, S., & Oksigen, S. (2021). Analisis Penggunaan Sensor MAX30100 pada Sistem. *Seminar Nasional Teknologi Sains Dan Humaniora (SemanTECH)*, *3*(1), 238–245.

Hariri, R., Hakim, L., & Lestari, R. F. (2019). Sistem Monitoring Detak Jantung Menggunakan Sensor AD8232 Berbasis Internet of Things. *Jurnal Telekomunikasi Dan Komputer*, *9*(3), 164. https://doi.org/10.22441/incomtech.v9i3.7075

Hindarto, Anshory, I., & Efiyanti, A. (2015). Aplikasi Pengukur Detak Jantung Menggunakan Sensor Pulsa. *Prosiding Simposium Nasional Teknologi Terapan (SNTT) III*, 1–5.

Hudhajanto, R. P., Mulyadi, I. H., & Sandi, A. A. (2022). Wearable Sensor Device berbentuk Face Shield untuk Memonitor Detak Jantung berbasis IoT. *Journal of Applied Informatics and Computing*, *6*(1), 87–92. https://doi.org/10.30871/jaic.v6i1.4105

Irawan, Y., Fernando, Y., & Wahyuni, R. (2019). Detecting Heart Rate Using Pulse Sensor As Alternative Knowing Heart Condition. *Journal of Applied Engineering and Technological Science (JAETS)*, *1*(1), 30–42. https://doi.org/10.37385/jaets.v1i1.16

Isyanto, H., & Jaenudin, I. (2018). Monitoring Dua Parameter Data Medik Pasien (Suhu Tubuh Dan Detak Jantung) Berbasis Aruino Nirkabel. *ELEKTUM*, *15*(1), 19–24. https://doi.org/10.24853/elektum.15.1.19-24

Karina, P., & Thohari, A. H. (2018). Perancangan Alat Pengukur Detak Jantung Menggunakan Pulse Sensor Berbasis Raspberry. *Journal of Applied Informatics and Computing*, *2*(2), 57–61. https://doi.org/10.30871/jaic.v2i2.920

Kumar N., S., Kumar B., N., A., S., D., V., & Balaji N., M. (2017). Heart Rate Monitoring System using IoT. *International Journal for Scientific Research & Development (IJSRD)*, *5*(2), 853–854.

Kumari, W. M. P., & Victor, S. P. (2019). Heart Attack Detection & Heart Rate Monitoring Using IOT Techniques. *Journal of Advanced Research in Dynamical and Control Systems*, *11*(06-Special Issue), 1471–1476.

Kusuma, R. S., Pamungkasty, M., Akbaruddin, F. S., & Fadlilah, U. (2018). Prototipe Alat Monitoring Kesehatan Jantung berbasis IoT. *Emitor: Jurnal Teknik Elektro*, *18*(2), 59–63. https://doi.org/10.23917/emitor.v18i2.6353

Muhajirin, M., & Ashari, A. (2018). Perancangan Sistem Pengukur Detak Jantung Menggunakan Arduino Dengan Tampilan Personal Computer. *Inspiration : Jurnal Teknologi Informasi Dan Komunikasi*, *8*(1). https://doi.org/10.35585/inspir.v8i2.2458

Noar, N. A. Z. M., & Kamal, M. M. (2017). The development of smart flood monitoring system using ultrasonic sensor with blynk applications. *2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, *2017-Novem*(November), 1–6. https://doi.org/10.1109/ICSIMA.2017.8312009

Nurdin, M., Aminah, N., Syahrir, Djamil, F., & Hamdani, M. F. (2015). Deteksi Denyut Jantung dengan Metode Sensor Pulsh Berbasis Ardiuno. *Prosiding Seminar Nasional Teknik Elektro & Informatika SNTEI 2015*, 201–206.

Prayitno, W. A., Muttaqin, A., & Syauqy, D. (2017). Sistem Monitoring Suhu, Kelembaban, dan Pengendali Penyiraman Tanaman Hidroponik menggunakan Blynk Android. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(4), 292–297.

Rachmat, H. H., & Ambaransari, D. R. (2018). Sistem Perekam Detak Jantung Berbasis Pulse Heart Rate Sensor pada Jari Tangan. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *6*(3), 344. https://doi.org/10.26760/elkomika.v6i3.344

Rifa'i, A. F. (2016). Sistem Pendeteksi dan Monitoring Kebocoran Gas (Liquefied Petrolum Gas) Berbasis Internet of Things. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *1*(1), 5–13. https://doi.org/10.14421/jiska.2016.11-02

Savitri, D. E. (2020). Gelang Pengukur Detak Jantung dan Suhu Tubuh Manusia Berbasis Internet of Things (IoT). In *UIN Syarif Hidayatullah Jakarta*. UIN Syarif Hidayatullah.

Serikul, P., Nakpong, N., & Nakjuatong, N. (2018). Smart Farm Monitoring via the Blynk IoT Platform : Case Study: Humidity Monitoring and Data Recording. *2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE)*, 1–6.

https://doi.org/10.1109/ICTKE.2018.8612441

Sipayung, F. H., Ramadhani, K. N., & Arifianto, A. (2018). Pengukuran Detak Jantung Menggunakan Metode Fotopletismograf. *EProceedings of Engineering*, *5*(2), 3664–3670.

Sufri, S., & Aswardi, A. (2020). Alat Pendeteksi Detak Jantung dan Kesehatan Berbasis Arduino. *JTEIN: Jurnal Teknik Elektro Indonesia*, *1*(2), 69–75. https://doi.org/10.24036/jtein.v1i2.31

Tamura, T., Maeda, Y., Sekine, M., & Yoshida, M. (2014). Wearable Photoplethysmographic Sensors—Past and Present. *Electronics*, *3*(2), 282–302. https://doi.org/10.3390/electronics3020282

Wijaya, N. H., Fauzi, F. A., T.Helmy, E., Nguyen, P. T., & Atmoko, R. A. (2020). The Design of Heart Rate Detector and Body Temperature Measurement Device Using ATMega16. *Journal of Robotics and Control (JRC)*, *1*(2), 40–43. https://doi.org/10.18196/jrc.1209

Yassin, F. M., Sani, N. A., & Chin, S. N. (2019). Analysis of Heart Rate and Body Temperature from the Wireless Monitoring System Using Arduino. *Journal of Physics: Conference Series*, *1358*(1), 012041. https://doi.org/10.1088/1742-6596/1358/1/012041

Yessianto, I., Setiawidayat, S., Effendy, D. U., & Einthoven, S. (2018). Perancangan Alat Monitoring Sinyal Jantung Menggunakan Arduino. *Conference on Innovation and Application of Science and Technology (CIASTECH 2018)*, 601–608.

Yuhefizar, Y., Nasution, A., Putra, R., Asri, E., & Satria, D. (2019). Alat Monitoring Detak Jantung Untuk Pasien Beresiko Berbasis IoT Memanfaatkan Aplikasi OpenSID berbasis Web. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *3*(2), 265–270. https://doi.org/10.29207/resti.v3i2.974

Yulian, R., & Suprianto, B. (2017). Rancang Bangun Photoplethysmograph (PPG) Tipe Gelang Tangan untuk Menghitung Detak Jantung. *Jurnal Teknik Elektro*, *6*(3), 223–231.

Yulidarti, Y., & Hendri, H. (2020). Rancang Bangun Alat Pengukur Detak Jantung Menggunakan Komunikasi Wifi dengan Android. *JTEV (Jurnal Teknik Elektro Dan Vokasional)*, *6*(1), 277. https://doi.org/10.24036/jtev.v6i1.107976

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022, Pp. 177 – 191
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Summarizing Online Customer Review using Topic Modeling and Sentiment Analysis

Muhammad Rifqi Maarif
Teknik Mesin, Fakultas Teknik, Universitas Tidar, Magelang
e-mail : rifqi@untidar.ac.id.
Artikel ini diajukan 31 Juli 2022, direvisi 5 September 2022, diterima 5 September 2022, dan dipublikasikan 25 September 2022.

Abstract

With the massive implementation of social media in various forms in various business domains, business or product owners have the opportunity to be able to take advantage of user review data that is available free of charge to evaluate the products they issue. User reviews on social media platforms, marketplaces, and e-commerce are User Generated Content (UGC) which is very useful for product owners to find out the extent of user preferences for their products. However, to be able to comprehensively read the data, the right technology is needed considering that the data is in the form of text in very large quantities. Reading one by one and then drawing conclusions is certainly not the right approach because it will take quite a lot of time. So, in this study, the researcher will use a text analysis-based approach, especially topic modeling and sentiment analysis to summarize user reviews in the comments or reviews column on the e-commerce platform. The case study used in this study is user reviews in the comments column on the Amazon site for the Lenovo K8 Note smartphone product. From the experiments carried out, the approach used can summarize the reviews written by quite many users in one summary that can be easily understood.

Keywords: User Generated Content, Online Customer Review, Text Mining, Topic Modeling, Sentiment Analysis

Abstrak

Dengan masifnya implementasi media sosial dalam berbagai bentuk di berbagai domain bisnis, pemilik bisnis ataupun produk memiliki kesempatan untuk dapat memanfaatkan data ulasan pengguna yang tersedia secara gratis untuk mengevaluasi produk yang mereka keluarkan. Ulasan pengguna di platform media sosial, *marketplace* maupun *e-commerce* merupakan *User Generated Content* (UGC) yang sangat bermanfaat bagi pemilik produk untuk mengetahui sejauh mana preferensi pengguna terhadap produk mereka. Namun, untuk dapat secara komprehensif membaca data tersebut diperlukan teknologi yang tepat mengingat data tersebut berbentuk tekstual dengan jumlah yang sangat banyak. Membaca satu per satu untuk kemudian disimpulkan tentunya bukan pendekatan yang tepat karena akan memakan cukup banyak waktu. Sehingga, dalam penelitian ini, peneliti akan menggunakan pendekatan berbasis analisis teks khususnya pemodelan topik dan analisis sentimen untuk merangkum ulasan pengguna di kolom komentar atau *review* yang terdapat di platform *e-commerce*. Studi kasus yang digunakan dalam penelitian ini adalah ulasan pengguna pada kolom komentar di situs Amazon untuk produk *smartphone* Lenovo K8 Note. Dari percobaan yang dilakukan, pendekatan yang digunakan mampu merangkum ulasan-ulasan yang dituliskan oleh pengguna yang berjumlah cukup banyak dalam satu ringkasan yang dapat dengan mudah dipahami.

Kata Kunci: *User Generated Content*, Review Pengguna Online, Penambangan Teks, Pemodelan Topik, Analisis Sentimen

1. INTRODUCTION

The ultimate goal of product development is to deliver a product that meets the customer's expectations. Hence, considering customer reviews of the product they bought and used is a key to successful product development. Integrating customer perception or opinion could be done in various ways and stages, from an early stage of product development until product release (Cui & Wu, 2017; Zhan et al., 2019). In the early stage of product development, the customer could

act as a source person for identifying and validating product requirements from their perspectives. In the middle stage of the product development cycle, the customer can also give their opinion on the concept of the product before it goes to production. In another way, the product owner can use the customers as active informants by constantly monitoring their opinion on online platforms like social media or the review section of online marketplace or e-commerce. Nowadays, by using online media, the customer is often shared their experience opinion, suggestion, and even complaint about the product they have bought and used. Those kinds of customer reviews could be potentially a source of ideas for a product owner to develop their product to the next version or even create new products (Hidayanti et al., 2018; Wang et al., 2020). With the advanced penetration of the internet, social media platforms including review sections on the online marketplace and e-commerce have changed the way customers interact and share their thoughts about the product they used. Specifically, Bashir et al. (2017) underlined that in correlation with product development, social media can be used to (1) identify user requirements and measure their review of the product, (2) recognize vocal customers that could be converted into a lead, (3) analyzing brand engagement within a certain group of society and (4) identifying the ideas for further development of a certain product.

User reviews of certain products on online platforms like social media or e-commerce are expressed in a very expressive and emotional way that makes product manufacturing companies have a more comprehensive review of products that have been marketed to the public (Elwalda et al., 2016). Thus, user comments and opinions on social media can be used as a basis for product evaluation as well as market analysis. In another study, online social media can also be used to identify potential customers by analyzing personal characteristics and their environments like friendship networks, demographic information, and other personal preference (Ramanathan et al., 2017). Furthermore, the use of social media data to obtain the voice of customers, apart from being more economical because we can get the review data for free by using the scraping technique, is also more accurate because it is sourced from a very large number of uploads from users who have very diverse characteristics (Elena, 2016).

Recently, research on social media analysis for summarizing customer reviews has been carried out with several approaches/techniques, one of which is text mining (Mahr et al., 2019). Text mining is a sub-type of data learning for textual data, so it is very suitable for analyzing customer uploaded data on social media, which is mostly in the form of text, in addition to images and videos. Techniques in text mining that are quite widely used are sentiment analysis and topic modeling. Sentiment analysis is a supervised computational technique by utilizes machine learning algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Deep Learning, and so on to extract opinion polarity from text data (Baharudin et al., 2010). The polarity of opinion is the tendency of sentiment contained in a sentence from the simplest (positive, negative, and neutral) to the more complex (happy, enjoy, sad, angry, and so on). Several studies that analyze social media for the voice of customers in product development use sentiment analysis to analyze customer perceptions on social media towards a certain product brand (Hu et al., 2020). Another study was conducted by Greco & Polli (2020) to read people's perceptions of the release of a new product. Sentiment analysis is also used to evaluate customer reviews in general for products that have been sold and used (Mirtalaie et al., 2018; Ng et al., 2021).

Another text mining approach that is widely used for customer review analysis on social media is topic modeling (Ko et al., 2018). Unlike supervised sentiment analysis, topic modeling is a machine learning approach that is unsupervised. The results of topic modeling are keywords that represent the topics/themes contained in a text. Thus, in the voice of the customer, topic modeling is widely used to explore the most frequent and influential topics discussed by customers (Jeong et al., 2019). Some examples of the voice of customer research that utilize topic modeling include identifying product features that get the most reviews by customers (Irawan et al., 2020). Further implementation of text mining is to combine sentiment analysis and topic modeling into a more integrative approach. This approach uses sentiment analysis to evaluate customer opinions for a

specific product feature. In this study, topic modeling was used to identify product features that were reviewed in the text of customer uploads on social media (Chehal et al., 2021).

Intensive monitoring of social media to analyze public opinion as users of industrial products will provide great benefits to find out the weaknesses of products that have been widely used by the public quickly and accurately, Hence, this research aims to produce a systematic and measurable summary by integrating text mining techniques for user review analysis of social media data.

2. METHODS

The proposed approach in this article is based on two theoretical backgrounds of text mining named topic modeling and sentiment analysis. First, the topic modeling approach was used for identifying the topics discussed by the user through their reviews on Amazon. The identified topics were further processed to select only topics which represent the user review statements. Second, the review value of users on each topic is measured by using a sentiment analysis approach. In this section, we will first describe the computational text mining approach by theoretically explaining topic modeling and sentiment analysis followed by an explanation of the step-by-step approach of gathering and processing user reviews for determining user review aspects and values.

2.1 Text Mining

Text mining is a transformation process of extracting structured data from unstructured text. Once extracted, those data can be further analyzed with various data analytic approaches to produce a meaningful pattern and facts. The process of text analysis employs a range of Natural Language Processing (NLP) methods. Text mining was commonly used alongside data mining techniques in which Naive Bayes, Support Vector Machine (SVM), and Artificial Neural Network (ANN) based methods including well-known deep learning were the most popular data mining techniques used in text mining. Employing advanced data mining techniques like deep learning in text processing works enables the computer to recognize any kind of insights hidden beneath the huge amount of textual documents (Suresh & Harshni, 2017).

The text mining process consists of several consecutive steps for deducting structured text data from the unstructured. The first phase of text mining is text preprocessing. In this phase, the collection text document so-called corpus would be cleaned and transformed into a standardized format. Text preprocessing is considered the phase of text mining that has the biggest portion of overall NLP activities (Vijayarani et al., 2018). This phase involves several technical works such as language identification, Part of Speech (POS) tagging, tokenization, parsing, stop word and non-standard character removal, and many more. The kind of preprocessing work was currently developing and even employees advance machine learning techniques. After the text is preprocessed, the next task is to apply text mining algorithms such as text summarization, topic modeling, text classification, sentiment analysis, named entity recognition, and many more (Maheswari; & Sathiaseelan, 2017). This research employed topic modeling and sentiment analysis consecutively for summarizing user review aspects from user reviews on Amazon.

2.2 Topic Modeling

Topic modeling is a common and widely adopted approach in text mining and machine learning for discovering semantic structures hidden within a collection of textual documents. In natural language processing, a topic model was defined as a statistical-based model for revealing a set of abstract representations of document collections. Those abstracts represent then further named as topics. In the topic model, topics are represented as a cluster of statistically similar words. Hence, a topic model discovers the topics of a set of documents based on the statistics of the words in each part of the document and estimates what the topics might be. There is some approach to topic modeling implementation (Alghamdi & Alfalqi, 2015). In our experiment, Latent Dirichlet Allocation (LDA) was used for implementing topic modeling. The use of LDA is based on the literature reviews that stated that LDA is a topic modeling algorithm that has the highest

performance, specifically when operating over a large number of textual documents (Barde & Bainwad, 2017).

LDA is a probabilistic approach proposed by (Blei et al., 2003). This approach consists of three different phases which are performed sequentially. The basic idea is LDA represents a document as a set of topics. Then, each topic was represented as probability distributions containing keywords. The architecture of LDA also shows the consecutive steps involved depicted in figure 1. LDA operates on a corpus **D** which consists of **M** number documents, each document on D denoted as ***w*** with length ***N***. In figure 1, **α** is defined as Dirichlet's parameter before the per-document topic distribution. Consecutively, **β** is a parameter of the Dirichlet that need to be defined to determine the per-topic word distribution. The next parameter is **θ** which denotes the topic distribution for the document. The sum of all **θ** is 1.0. Finally, LDA defined ***z*** as the topic for document ***w***.

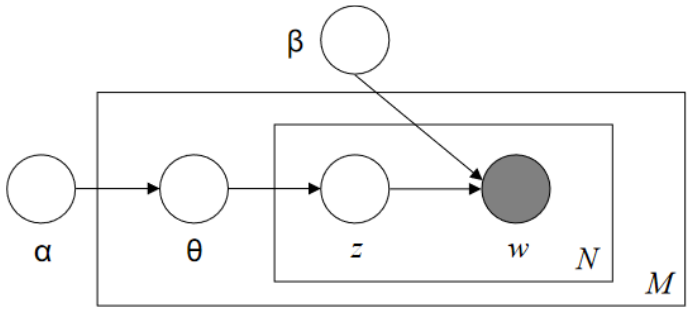


Figure 1 LDA Architecture (Blei et al., 2003)

When executing the LDA algorithm, **α** defined needs to be predefined manually as the expected number of generated topics. Then, assuming there is no prior knowledge on how many topics should be generated we can use the Elbow Method to determine the best number of topics in LDA-based topic modeling execution (Syakur et al., 2018).

2.3 Sentiment Analysis

Sentiment analysis is one of the major research fields and widely adopted techniques of NLP. It is a technique that mainly focused on recognizing the polarity of people's opinions based on their textual expressions. Sentiment analysis applies to a wide range of domains including product development which involves analyzing user perspectives. Greco & Polli (2020) performed an experiment by using sentiment analysis for measuring the popularity and user engagement of certain brands. From a technical perspective, sentiment analysis is considered a classification problem over textual data with its sentiment acting as a class. In basic form, the sentiment was categorized as positive and negative which showed the polarity of the opinions of the speaker/writer of the text data (Ahmad et al., 2017). The sentiment of a certain text written by the speaker basically could be identified in two approaches lexicon-based or text classification based.

The lexicon-based approach employed a dictionary that is already defined and contains a sentiment label with its corresponding sentiment score. One of the famous and widely used predefined dictionaries is SentiWordNet (Baccianella et al., 2010). In this approach, a sentiment of a text is identified by calculating the polarity orientation score of words or phrases which construct the document one by one. The overall sentiment value for the text is then computed by summing up those polarity scores. The polarity score could be either categorical or continuous values. When the polarity score is a categorical value, then the overall text sentiment is determined by the category label with the most appearance. When the polarity score preserves in continuous form, the overall polarity is calculated by summing up all of the scores. On the continuous polarity score, positive sentiment represents positive values, and negative sentiment represents negative values. Even though this approach yields an accurate result, constructing a dictionary for every language was considered a costly activity.

Different from the former approach, the text classification-based approach employs machine learning techniques instead of relying on a dictionary. The text classification-based sentiment analysis is a supervised machine learning task. Thus, the classifier model was built from text data labeled by their corresponding sentiments to predict the sentiment orientation (Ahmad et al., 2017). Instead of resulting sentiment in a numerical or ordinal value. The supervised approach of sentiment analysis assumes a sentiment as a discrete class label (positive or negative). Given the training data, typical machine learning algorithm like Support Vector Machine (SVM), Naive Bayes Classifier (NBC), Logistic Regression (LR), K-Nearest Neighbor (KNN), and Artificial Neural Network (ANN)-based techniques including deep learning learns from the data and build a classification model with their representations. Then, when new text is coming this classification model was employed to predict its sentiment class.

Nowadays, with the advanced research of text analytics in the industry, there are plenty of ready-to-use cloud-based analytic services to perform sentiment analysis. Those kinds of services provide a pre-trained model which builds from a large amount of textual data. Hence, those services offer high-performance sentiment analysis with no programming works on the user side. One of the most popular analytic services to perform sentiment analysis is Cloud NLP provided by Google. In this article, we used the Cloud NLP service provided by Google to perform sentiment analysis. Google NLP service use BERT (Bidirectional Encoder Representations from Transformers) algorithm for performing sentiment analysis task. BERT is a deep learning-based algorithm which particularly designed for text analytics and NLP (Sun et al., 2019). Currently, BERT gives a state-of-the-art performance in text classification among the other machine learning or deep learning techniques.

2.4 Experimental Setup

In this section, we describe the step-by-step approach in our experiment to discover the user review aspect and its corresponding sentiment expressed by the user through their comments on Google Playstore. After the data was gathered and preprocessed, we first employed topic modeling techniques to identify the key topics talked about by the users. Those key topics were then further processed and eventually formed user review aspects. Then, after the aspects were constructed, the sentiment profile of each topic was estimated by using the sentiment analysis technique. A supervised machine learning-based approach was implemented to build the classifier model for sentiment classification. Figure 2 below illustrates the consecutive process of our proposed approach.

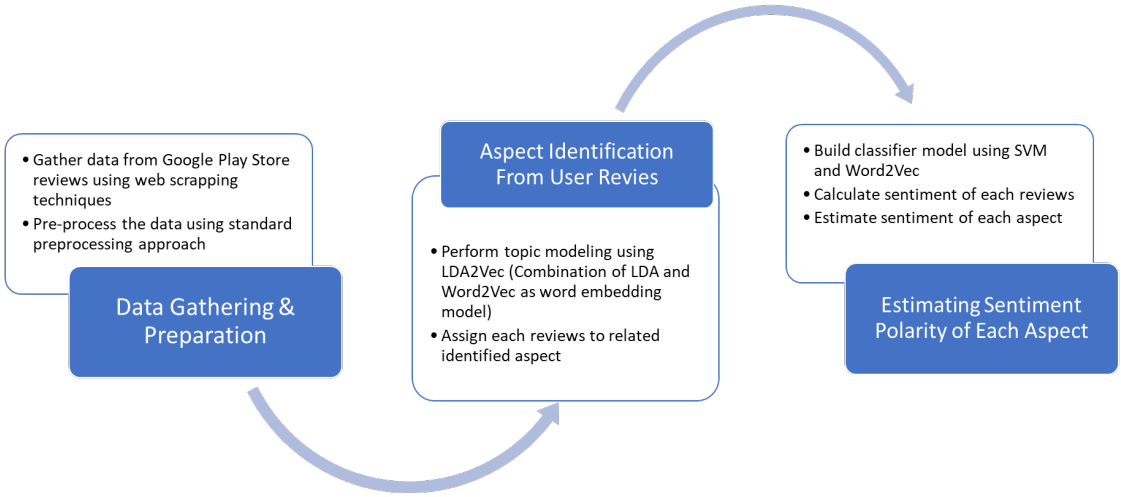


Figure 2 Proposed Framework from Identifying User Review Aspect and Its Sentiment

2.4.1 Data Preparation

The first phase of our approach is data collection. To pile up the summary of customer reviews, the data acquisition would be concentrated to acquire user reviews of one specific product named the Lenovo K80 smartphone which is available on Amazon. In this experiment, we collect data from user comments from the product review section on the e-commerce platform Amazon. Figure 3 shows the review section on a product page on Amazon. In that section, user can freely express their opinion and experience of the corresponding product they bought.

Figure 3 The Customer Reviews Section on Amazon

The next step in this phase is to preprocess the data. The data preprocessing consists of several common steps including the removal of stopwords, punctuation, and non-standard character followed by a steaming process and lemmatization. The detailed processes of the preprocessing steps are outlined below:

1) **Case folding**. Case folding is a process in text preprocessing that is done to standardize the form of characters in the data. The case folding process is the process of converting all letters to lowercase. In this process, the characters 'A'-'Z' contained in the data are converted into characters 'a'-'z'.
2) **Stop-word removal**. Stop words are common words that usually appear in large numbers and are considered meaningless. Examples of stop words for English include "of", "the" and so on. Stopword Removal is a filtering process, selecting important words from the token beyond the stopwords to represent documents.
3) **Punctuation removal**. Punctuation removal is a process where the system will remove punctuation marks or symbols that exist in the dataset. These punctuation marks or symbols are removed because they do not affect the results of text analysis.
4) **Stemming**. Stemming is the process of forming basic words. The terms obtained from the stopword disposal stage will be carried out by the steaming process. Stemming is used to reduce the form of terms to avoid mismatches that can reduce recall, where terms that are different but have the same basic meaning are reduced to a single form.
5) **Lemmatization**. Lemmatization is a process that aims to normalize text/words based on the basic form which is the lemma form. The distinction between stemming and lemmatization is while stemming changes a word into a root word without knowing the context of the word like cutting off the ends of words, lemmatization changes a word into a root word by knowing the context of the word (Balakrishnan & Ethel, 2014).
6) **Part of speech tagging**. Part-of-speech (POS) tagging or in short it can be written as tagging is the process of assigning POS tags or syntactic classes to each word in the corpus (Gimpel et al., 2011). Part of speech is a classification of words that are categorized through their

role and function in sentences of a language. When studying English, we will come across the terms noun, adjective, pronoun, and so on. These terms are part of the part of speech. Part of speech has an important role to form a sentence so that it is coherent and follows the grammar of the sentence. In this experiment, we used POS Tagging to extract nouns from each user review in the text corpus.

Afterward, once a set of sentences represents the single user reviews of the target been prepared, keywords (or key phrases) are extracted from the sentence to construct the structure of the reviews. Eventually, each user review was represented as a list of keywords and their frequency of appearance in their related review.

2.4.2 Identifying User Review Aspect

In this step, the user review aspects are defined based on topic modeling which is computed using the LDA algorithm. LDA-based topic modeling calculation requires three parameters for its operation. Those three parameters are corpus, word dictionary, and the number of topics. User reviews gathered from Google Playstore would be the corpus parameter. Then dictionary parameter was produced automatically from the corpus. Dictionary is simply defined as a set of words that appears on the corpus regardless of the frequency of its appearance. From the constructed dictionary, LDA builds a bag-of-words model for representing the documents. A major problem of the bag-of-words representation comes from the inability to capture the semantic relation of a word within the word vector. Whereas, revealing the semantically related words and using the correlation value on the word vector would greatly improve the representation model (Xun et al., 2016). Hence, in this experiment word embedding was used for empowering the document representations.

In our experiments, we implemented LDA-based topic modeling by using LDA2vec to integrate word embedding representation through the Word2Vec model. LDA2Vec generates a context vector from a document vector built around the pivot word vector. The pivot word vector itself is a result of LDA expansion by the Word2Vec model over the document vectors (Bojanowski et al., 2017). The context vector is then used for estimating the context between words within vectors. That operation enables LDA to generate more human-interpretable LDA topics The set topics from user reviews generated by LDA2vec topic modeling become the major subjects by which users are expressing their experience of using the mobile learning applications. Those set of topics is then further processed for determining the appropriate terms and phrases which closely related to user satisfaction or dissatisfaction expressions. The final operation of this phase was to generate the user review aspects in which each aspect was represented by a set of keywords (terms and/or phrases).

2.4.3 Computing User Review Sentiment on Each Aspect

In this second step, a user review of each aspect that was identified in the previous stage was calculated. Sentiment analysis was employed as a basis for user review calculation. The first step in this phase was assigning every user review in the corpus into identified aspects. These steps are done by simply matching the keywords found in the user reviews with the keywords on every topic. Then, the sentiment of every sentence was estimated by using a classifier model constructed with BERT which was provided by the Google NLP service. Prior works showed that BERT gained a high-quality performance in terms of sentiment analysis and work well even though the training data was limited. Hence, we consider this machine learning approach suitable for our work. The BERT method provided by the Google NLP service was improved by employing word embedding as feature representation. Word embedding would empower feature representation with contextual information of words in high-dimensional vectors. Thus, the sentiment classification of each user review was improved in terms of accuracy. The second step in this phase involved calculating the sentiment of each identified aspect. This calculation is simply done by taking the percentage of positive and negative sentiment.

3. RESULTS AND DISCUSSION

In this section, the results of our experiment have been outlined in line with the scenario on methods section chronologically. We will start with a discussion about the dataset and exploratory analysis of our text corpus. Afterward, the discussion goes to the extraction of the user review aspect using topic modeling. Lastly, we will discuss the result of sentiment analysis on each aspect of user review.

3.1 Dataset and Exploratory Analysis

We used approximately 14.520 textual review data gathered from amazon.com. The dataset contains a user-generated review of the Lenovo K8 smartphone product. Lenovo K8 smartphone is one of the most popular budget “high-end” smartphones. Hence, there are a lot of user reviews of this device on Amazon. We use the web scrapping technique to acquire the dataset. The tool used in this experiment for scrapping purposes is Scrappy, an Open Source and Python-based web scrapping library freely available for any use. Table 1 shows the excerpt of our dataset in a row form.

Table 1 Example of an Acquired Customer Review of Lenovo K8 Smartphone on Amazon

No.	Review
1	Superb product. A few of the features are awesome. Dual camera, front 13mp camera, back and front flash, dedicated music button, dedicated memory card slot, free transparent case, and split window for multitasking. These are some features I like about the product in my budget
2	Hello, The phone starts resetting itself randomly. This issue starts after 3 days of use, especially when you use it for a longer time like watching a youtube video for an hour or two. The issue starts to repeat after 1 or 2 days, then I tried a few options mentioned in Lenovo help APP which is pre-installed(I am feeling stupid for having done this), and after that, for 4 or 5 days it didn't show this issue. Now it is back with a bang and almost resets itself every night. Use the phone at night, then lock it and keep it aside and when you take the phone again in the morning, boom, it is dead already. And when you manually power it on there is enough juice left in the battery. Not sure what to do with this kind of issue, the phone is useless. When I approached amazon for a return, they are quoting policy and suggested that I should run pillar to post to get the problem rectified (not sure even if it is the rectifiable issue),
3	Great experience with this amazing product from Lenovo. it is equipped with almost every feature that a smartphone required. Deca core processor long-lasting battery and 64GB internal memory with 4GB RAM is just awesome. I would certainly recommend this phone for users having usage and game lovers
4	Over Heating Issue while light usage like just an internet connection, and always warm while connected to the internet. this update is After usage of 10 days. Don't prefer this one.
5	It's the jack of all trades but the king of none. Battery backup could have been better if they used some other processor. The battery drains quite fast. The camera is better than average. And I think there is no option to keep external media as your ringtone. Only custom build ringtones available to set as your ringtone. Kinda bums me out. Update after 3-day use: Battery backup is horrible, normal usage like WhatsApp and Instagram browsing consume more than 25% battery in an hour or so. Then there is a turbocharging issue, it starts with fast charging then after 10-20 minutes, depending upon mood, the rate decreases, it took 7 hours to charge it by 40% in total. Would appreciate it if Amazon could take this matter seriously and take it up with Lenovo and return the money of its customers for defective models. I would not trust my 14K bucks with Lenovo or Moto from this point on.

The acquired data from Amazon is raw and not ready for further text analysis. Thus, we perform a sequence of preprocessing tasks before feeding the data into the topic model algorithm. We

employ the removal of stopwords, and punctuation standard characters followed by a steaming process and lemmatization as a sequence of preprocessing tasks in our experiment.
After preprocessing task, we then performed POS Tagging over the preprocessed dataset. This task aims to extract nouns from each user review. We consider that the core meaning of every review written by the user is in the nouns in the sentences. Hence, we can achieve a more precise topic model by neglecting another sentence structure beyond the noun forms.

We perform a simple exploratory analysis by generating a word cloud for early investigation of the text dataset. Word cloud is a visualization approach to reveal some of the most frequent words in a text corpus. Through the word cloud, we can gain a raw insight into what the customers frequently talk about the topic. This insight can be helpful for confirmation when a topic model was constructed. Figure 4 shows the word cloud of extracted nouns from our text corpus. From the word cloud of user reviews about the Lenovo K8 smartphone product, we know that two product features, namely camera, and battery, are the most frequent object talked about by the user. Later, by using topic modeling, these two main keywords and potentially some other keywords will be revealed further using the topic modeling approach.

Figure 4 Word Cloud of Extracted Nouns from the Text Corpus

The next exploratory text analysis performed in this experiment is to reveal the most frequent pair of nouns found on the corpus. Table 2 outlines the most frequent pair of nouns found in our text corpus. In line with the word cloud depicted in figure 4, the terms battery and camera remain dominant. By analyzing the most frequent pair we can further investigate the interrelation of most frequent words even further. For instance, from table 4 we can gain an insight that the user reviews about “camera” are predominantly about the “camera quality” or “camera performance”.

Table 2 Most Frequent Topic Keyword Pair Within a Text Corpus

Rank	Topic Pair	Frequency
1	quality - camera	1110
2	battery - backup	662
3	performance - camera	545
4	quality - battery	512
5	problem - battery	499
6	battery - performance	470
7	issue - battery	437
8	battery - time	424
9	battery - hours	421
10	camera - product	416
11	mode - camera	405
12	life - battery	344
13	camera - problem	337
14	issue - camera	328
15	day - battery	325

3.2 Identified User Review Aspects

The aspect of every user review is considered as a particular topic contained in the sentences written by the users. Hence, topic modeling is the proper approach to reveal the aspects of the reviews generated by the users. In this experiment, we employ the LDA algorithm to extract the topic model from our text corpus. The LDA algorithm's main requirement is first to determine the number of topics we want to be extracted. Since we do not have prior knowledge of the data or domain, we employed the Elbow statistical approach to determine the optimal number of topics potentially generated by LDA.

In the Elbow method, the number of topics is considered optimal if the average cosine similarity between all pairs of distribution vectors of topic-words generated in topic modeling operation is minimum. Those cosine similarity values are then represented as coherence scores. The greater the coherence value, the more optimal the topic is. We then conduct some LDA algorithm experiments with various topics ranging from 4 to 10 topics. To ease the analysis process, we avoid determining the number of topics more than 10. Figure 5 shows the experimental results of determining the coherence score. From the graph in Figure 5, we can outline that the highest coherence score was achieved from our text corpus by determining the topic number as 5. Thus for the later tasks, we will use five topic clusters.

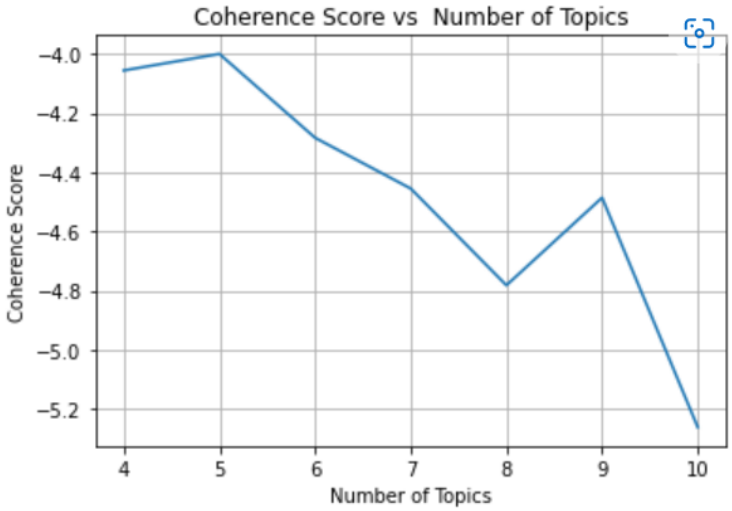


Figure 5 Coherence Score to Determine the Optimal Number of Topics

To investigate and validate whether the determined cluster number is optimum, we then visualize the cluster using an inter-topic distance map. To perform this task, we use Gensim. Gensim is an Open-Source library to perform topic modeling, including visualizing the topic clusters. We implement Gensim using Python language. Figure 6 shows the visualization of the topic cluster generated by the LDA algorithm with five clusters. The clusters are considered as good if they are well separated from each other. Hence, from Figure 6, we can see that the determined number of a cluster as five is considered an excellent or optimum number of clusters.

A cluster of topics contains a set of keywords that construct the topic model itself. Hence, after generating the topic clusters, we then reveal the set of keywords of every cluster and perform qualitative analysis to summarize those keywords into a specified label. The specified label determines the most closely related keywords within a cluster. Table 3 outline the keywords which dominantly appear on each cluster and their approximate labels. The labels outlined in table 3 were determined manually by looking at the semantic tendencies of the keywords contained in each cluster.

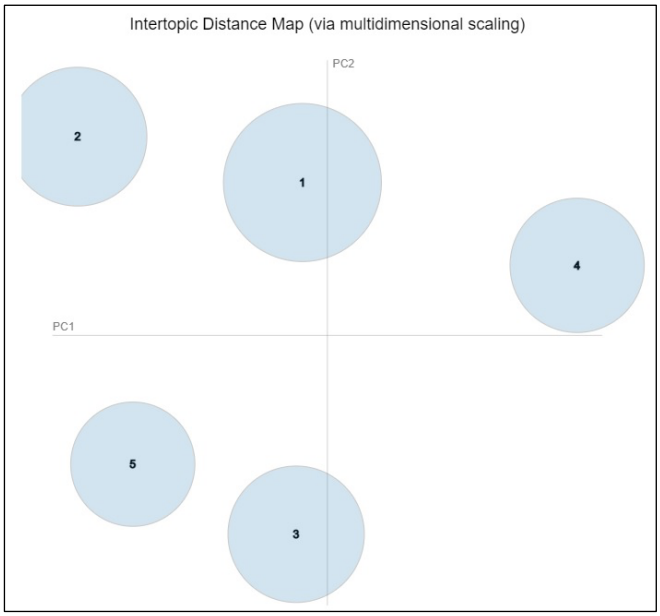


Figure 6 Topic Clusters Generated from LDA Operation

Table 3 The Keyword Generated by LDA and Its Corresponding Label

Topic Number	Keywords	Topic Label
1	camera, product, quality, performance, price, glass, range, awesome, display	camera and display quality
2	price, money, feature, screen, waste, value, option, phone	price money value
3	battery, camera, backup, quality, hour, day, life, mode	battery performance
4	the problem, issue, battery, month, heating, time, charger, heat, use	
5	product, service, amazon, call, device, return, buy, day	after-sales service

We can see in table 3 that from five topic clusters, we can derive four major topics. Topic 3 and topic 4 which shared similar semantics then be integrated as one. The four major topics derived from table 3 are "camera and display quality", "price money value", "battery performance", and "after-sales service". Those four topics are summarized as the major user concern about the product they bought on Amazon. In this case, the product is Lenovo K8 Note Smartphone. In the next section, the sentiment polarity of each major topic would be revealed.

3.3 Sentiment Analysis of User Review Aspects

The next step of online customer review summarization is to infer the sentiments of online customer reviews. Before going further on the sentiments of each topic, we will take a look at the overall sentiment distribution of the entire text corpus. Figure 7 below shows the sentiment distributions of our customer review dataset. From Figure 7, we can see that the negative sentiments slightly dominated the customer opinion about the Lenovo K8 Note smartphone on Amazon.

For better insight, we then reveal the dominant keywords in both negative and positive sentiments. Figure 8 and Figure 9 shows the predominant keywords in negative and positive review subsequently. The interesting parts of the visualizations in Figure 8 and Figure 9 are both negative and positive sentiments have similar dominant words namely "battery", "camera", and "product". Slightly different from the positive sentiment, in the negative sentiment keyword "battery" is

appear much bigger. Hence, we can conclude that the customers of the Lenovo K8 Note smartphone are mostly unhappy with the smartphone's battery performance.

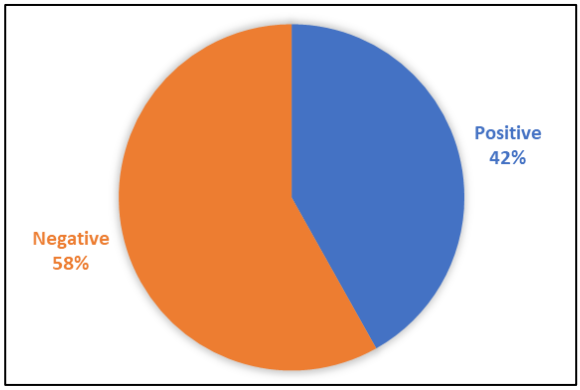


Figure 7 Sentiment Analysis Results Over the Entire Corpus

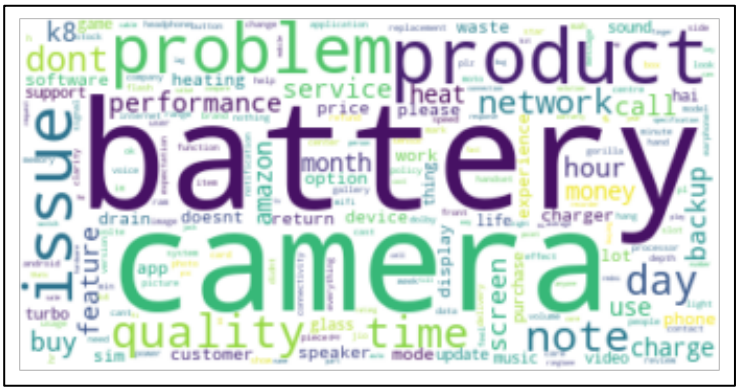


Figure 8 Word Cloud of Nouns Found in a Negative Review

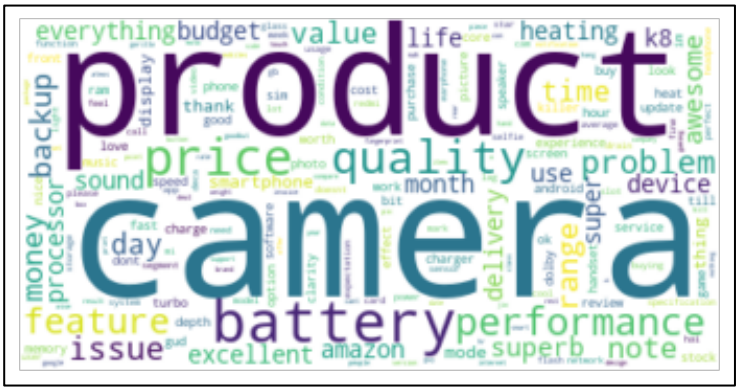


Figure 9 Word Cloud of Nouns Found in a Positive Review

The last step of this experiment is to infer the sentiment of each topic review. This step aims to provide a deeper insight into what customer likes and do not like about the Lenovo K8 Note smartphone product. Figure 10 shows the sentiment distribution of each major topic. From Figure 10, we can see that in three major aspects which are "Camera and Display", "Battery Quality" and "After Sales Service", the negative sentiments are dominant. Only one aspect named "Price and Money Value" users give mostly positive sentiments. Hence, we can conclude that even though the Lenovo K8 Note smartphone product is not meet user satisfaction in terms of its product features (camera, display, and battery) and after-sales service, the user of this product feels that

this product is worth the price. According to the customer reviews on Amazon, Lenovo K8 Smartphone is considered good as a low-budget high-end smartphone.

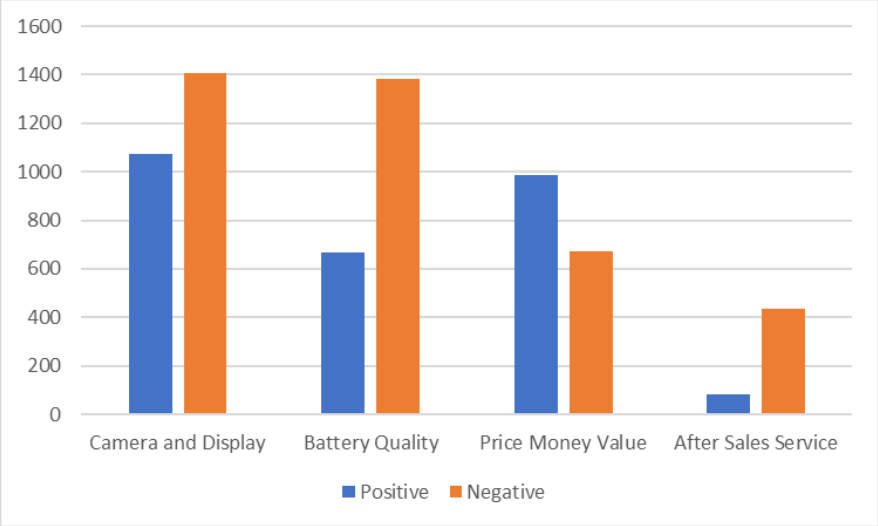


Figure 10 Sentiment Analysis Results in Each Aspect

4. CONCLUSIONS

From the conducted research we can conclude that the use of text mining specifically topic modeling and sentiment analysis was very beneficial in summarizing a large amount of text user reviews gathered from the internet. The results of running topic modeling and sentiment analysis give the product owner clear evidence quantitatively of what aspects of their product are liked by its customer and what are not. This research also gives an insight that combining two techniques of text analysis namely topic modeling and sentiment analysis gives a comprehensive result in terms of summarizing user reviews. This combination allows us to go further with user sentiment polarity on each aspect of the product instead of user sentiment in more general terms. Nevertheless, this research still has a drawback since the label of the topic was manually determined from the keywords. Therefore, this research can be further developed in terms of the automatic identification of proper labels for a cluster that contains a set of keywords.

REFERENCES

Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine Learning Techniques for Sentiment Analysis: A Review. *International Journal of Multidisciplinary Sciences and Engineering*, *8*(3).

Alghamdi, R., & Alfalqi, K. (2015). A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications*, *6*(1). https://doi.org/10.14569/IJACSA.2015.060121

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 2200–2204.

Baharudin, B., Lee, L. H., & Khan, K. (2010). A Review of Machine Learning Algorithms for Text-Documents Classification. *Journal of Advances in Information Technology*, *1*(1). https://doi.org/10.4304/jait.1.1.4-20

Balakrishnan, V., & Ethel, L.-Y. (2014). Stemming and Lemmatization: A Comparison of Retrieval Performances. *Lecture Notes on Software Engineering*, *2*(3), 262–267. https://doi.org/10.7763/LNSE.2014.V2.134

Barde, B. V., & Bainwad, A. M. (2017). An overview of topic modeling methods and tools. *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 745–750. https://doi.org/10.1109/ICCONS.2017.8250563

Bashir, N., Papamichail, K. N., & Malik, K. (2017). Use of Social Media Applications for Supporting New Product Development Processes in Multinational Corporations. *Technological Forecasting and Social Change*, *120*, 176–183. https://doi.org/10.1016/j.techfore.2017.02.028

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(4–5), 993–1022. https://doi.org/10.1016/b978-0-12-411519-4.00006-9

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Chehal, D., Gupta, P., & Gulati, P. (2021). Implementation and comparison of topic modeling techniques based on user reviews in e-commerce recommendations. *Journal of Ambient Intelligence and Humanized Computing*, *12*(5), 5055–5070. https://doi.org/10.1007/s12652-020-01956-6

Cui, A. S., & Wu, F. (2017). The Impact of Customer Involvement on New Product Development: Contingent and Substitutive Effects. *Journal of Product Innovation Management*, *34*(1), 60–80. https://doi.org/10.1111/jpim.12326

Elena, C. A. (2016). Social Media – A Strategy in Developing Customer Relationship Management. *Procedia Economics and Finance*, *39*, 785–790. https://doi.org/10.1016/S2212-5671(16)30266-0

Elwalda, A., Lü, K., & Ali, M. (2016). Perceived derived attributes of online customer reviews. *Computers in Human Behavior*, *56*, 306–319. https://doi.org/10.1016/j.chb.2015.11.051

Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., & Smith, N. A. (2011). Part-of-speech tagging for twitter: Annotation, features, and experiments. *ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, *2*, 42–47.

Greco, F., & Polli, A. (2020). Emotional Text Mining: Customer profiling in brand management. *International Journal of Information Management*, *51*, 101934. https://doi.org/10.1016/j.ijinfomgt.2019.04.007

Hidayanti, I., Herman, L. E., & Farida, N. (2018). Engaging Customers through Social Media to Improve Industrial Product Development: The Role of Customer Co-Creation Value. *Journal of Relationship Marketing*, *17*(1), 17–28. https://doi.org/10.1080/15332667.2018.1440137

Hu, S., Kumar, A., Al-Turjman, F., Gupta, S., Seth, S., & Shubham. (2020). Reviewer Credibility and Sentiment Analysis Based User Profile Modelling for Online Product Recommendation. *IEEE Access*, *8*, 26172–26189. https://doi.org/10.1109/ACCESS.2020.2971087

Irawan, M. I., Wijayanto, R., Shahab, M. L., Hidayat, N., & Rukmi, A. M. (2020). Implementation of social media mining for decision making in product planning based on topic modeling and sentiment analysis. *Journal of Physics: Conference Series*, *1490*(1), 012068. https://doi.org/10.1088/1742-6596/1490/1/012068

Jeong, B., Yoon, J., & Lee, J.-M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*, *48*, 280–290. https://doi.org/10.1016/j.ijinfomgt.2017.09.009

Ko, N., Jeong, B., Choi, S., & Yoon, J. (2018). Identifying Product Opportunities Using Social Media Mining: Application of Topic Modeling and Chance Discovery Theory. *IEEE Access*, *6*, 1680–1693. https://doi.org/10.1109/ACCESS.2017.2780046

Maheswari;, M. U., & Sathiaseelan, D. J. G. R. (2017). Text Mining: Survey on Techniques and Applications. *International Journal of Science and Research (IJSR)*, *6*(6), 1660–1664.

Mahr, D., Stead, S., & Odekerken-Schröder, G. (2019). Making sense of customer service experiences: a text mining review. *Journal of Services Marketing*, *33*(1), 88–103. https://doi.org/10.1108/JSM-10-2018-0295

Mirtalaie, M. A., Hussain, O. K., Chang, E., & Hussain, F. K. (2018). Sentiment Analysis of Specific Product's Features Using Product Tree for Application in New Product Development. In *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 8, pp. 82–95). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-319-65636-6_8

Ng, C. Y., Law, K. M. Y., & Ip, A. W. H. (2021). Assessing Public Opinions of Products Through Sentiment Analysis. *Journal of Organizational and End User Computing*, *33*(4), 125–141. https://doi.org/10.4018/JOEUC.20210701.oa6

Ramanathan, U., Subramanian, N., & Parrott, G. (2017). Role of social media in retail network operations and marketing to enhance customer satisfaction. *International Journal of Operations and Production Management*, *37*(1), 105–123. https://doi.org/10.1108/IJOPM-03-2015-0153

Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., & Jiang, P. (2019). Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. *International Conference on Information and Knowledge Management, Proceedings*, *11*, 1441–1450. https://doi.org/10.1145/3357384.3357895

Suresh, R., & Harshni, S. R. (2017). Data mining and text mining — A survey. *2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)*, *2018-Janua*, 412–420. https://doi.org/10.1109/ICCPEIC.2017.8290404

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, *336*(1), 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Vijayarani, S., Ilamathi, J., & Nithya. (2018). Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7–16. https://doi.org/10.1016/j.procs.2013.05.286

Wang, L., Jin, J. L., Zhou, K. Z., Li, C. B., & Yin, E. (2020). Does customer participation hurt new product development performance? Customer role, product newness, and conflict. *Journal of Business Research*, *109*, 246–259. https://doi.org/10.1016/j.jbusres.2019.12.013

Xun, G., Gopalakrishnan, V., Ma, F., Li, Y., Gao, J., & Zhang, A. (2016). Topic Discovery for Short Texts Using Word Embeddings. *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 1299–1304. https://doi.org/10.1109/ICDM.2016.0176

Zhan, Y., Tan, K. H., & Huo, B. (2019). Bridging customer knowledge to innovative product development: a data mining approach. *International Journal of Production Research*, *57*(20), 6335–6350. https://doi.org/10.1080/00207543.2019.1566662

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022, Pp. 192 – 210
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Evaluasi Kesuksesan Penerapan Sistem Elektronik Kinerja (E-Kinerja) Menggunakan *Enhanced Information System Success Model* di Kecamatan Benda Tangerang

Latansa Amalia [(1)*], Anik Hanifatul Azizah [(2)]
Sistem Informasi, Fakultas Ilmu Komputer, Universitas Esa Unggul, Jakarta
e-mail : latansaamalia008@student.esaunggul.ac.id, anik.hanifa@esaunggul.ac.id.
* Penulis korespondensi.
Artikel ini diajukan 22 Juli 2022, direvisi 9 September 2022, diterima 11 September 2022, dan dipublikasikan 25 September 2022.

Abstract

E-Kinerja Benda District Tangerang City is an E-Government in the form of an Employee Performance Management Information System in charge of recording and reporting the performance of all employees in the government environment. The system should aim to simplify employee performance management and measure the efficiency and creativity of employees at work, in its implementation, there are still problems that occur. Many employees complain about the system, the implementation of the E-Kinerja system either directly or indirectly is considered troublesome, the internet network is not widely connected, the presence of technology devices (computers) is limited, and the lack of employees in understanding and utilizing the E-Kinerja system, and the lack of attendance and knowledge of information about IT as part of the supporting components in the process of presenting the information. So that an evaluation of the application of E-Kinerja is carried out, with the aim of this study being able to measure evaluation to prove the success rate of implementing E-Kinerja-based performance evaluations in Benda District, Tangerang City and knowing what factors or variables influence it. This study evaluates success using the Enhanced Information System Success Model with 7 evaluation variables: information quality, system quality, service quality, user satisfaction, trust, use, and net benefits. This study will use a qualitative and quantitative data approach through observation, interviews, literature studies, and questionnaires where the total respondents obtained are 62 respondents with PLS-SEM data analysis and SmartPLS 3.3 assistance. Based on the results of the research conducted, it is proven that from the 14 hypotheses proposed there are 5 hypotheses accepted while 9 other hypotheses are rejected. This evaluation also produces recommendations from the seven variables used which contain alternatives to improve and increase the success of the implementation of E-Kinerja.

Keywords: Success Evaluation, Performance Electronic Systems (E-Kinerja), Enhanced Information System Success Model, PLS-SEM, SmartPLS

Abstrak

E-Kinerja Kecamatan Benda Kota Tangerang yakni suatu *E-Government* berbentuk Sistem Informasi Manajemen Kinerja Pegawai yang bertugas mencatat dan melaporkan kinerja semua pegawai di lingkungan pemerintahan. Sistem yang seharusnya bertujuan untuk menyederhanakan manajemen kinerja pegawai serta mengukur efisiensi dan kreativitas pegawai dalam bekerja, dalam implementasinya masih terdapat permasalahan yang terjadi. Sistem banyak dikeluhkan pegawai, penerapan sistem E-Kinerja baik secara langsung maupun tidak langsung dianggap merepotkan, jaringan internet yang tidak terhubung secara luas, kehadiran perangkat teknologi (komputer) yang terbatas, kurangnya pegawai dalam memahami dan memanfaatkan sistem E-Kinerja, serta minimnya kehadiran dan pengetahuan informasi tentang IT sebagai bagian komponen pendukung dalam proses penyajian informasi. Sehingga dilakukanlah evaluasi penerapan E-Kinerja, dengan tujuan penelitian ini dapat melakukan pengukuran evaluasi untuk membuktikan tingkat keberhasilan pelaksanaan evaluasi kinerja berbasis E-Kinerja di Kecamatan Benda Kota Tangerang serta mengetahui faktor atau variabel apa saja yang mempengaruhinya. Penelitian ini melakukan evaluasi kesuksesan menggunakan *Enhanced Information System Success Model* dengan 7 variabel evaluasi; kualitas informasi, kualitas sistem, kualitas pelayanan, kepuasan pengguna, kepercayaan, penggunaan, dan

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022: 192 – 210

manfaat bersih. Penelitian ini akan menggunakan pendekatan data kualitatif dan kuantitatif melalui observasi, wawancara, studi literatur, dan kuesioner di mana total responden yang diperoleh sebanyak 62 responden dengan analisis data PLS-SEM dan bantuan SmartPLS 3.3. Berdasarkan hasil penelitian yang dilakukan membuktikan dari 14 hipotesis yang diajukan terdapat 5 hipotesis diterima sedangkan 9 hipotesis lainnya ditolak. Evaluasi ini juga menghasilkan rekomendasi dari ketujuh variabel yang digunakan yang berisi alternatif untuk memperbaiki dan meningkatkan kesuksesan penerapan E-Kinerja.

Kata Kunci: Evaluasi Kesuksesan, Sistem Elektronik Kinerja (E-Kinerja), *Enhanced Information System Success Model*, PLS-SEM, SmartPLS

1. PENDAHULUAN

Instansi pemerintah Indonesia sekarang sedang berupaya mendorong penggunaan teknologi inovasi. Salah satunya adalah dengan menyusun perencanaan hingga pelaporan di lingkungan pemerintahan berlandas teknologi informasi dan komunikasi yang dikenal sebagai *E-Government*. Salah satu pemerintahan di Indonesia yang sudah mulai menerapkan *E-Government* di lingkungan pemerintahannya yaitu Pemerintah Kota Tangerang melalui Badan Kepegawaian. Pendidikan dan Pelatihan (BKPP) yang telah membuat dan mengembangkan sebuah Sistem Informasi Manajemen Kinerja Pegawai Pemerintah Kota Tangerang bernama Elektronik Kinerja (E-Kinerja).

Pemerintah Kota Tangerang khususnya Kecamatan Benda Kota Tangerang mulai Januari 2020 menerapkan Sistem E-Kinerja yang berguna untuk merekam, melaporkan, mengukur dan menilai presentasi kinerja seluruh pegawai negeri sipil (PNS). Sistem E-Kinerja direncanakan untuk proses optimalisasi dan penyederhanaan manajemen kepegawaian di pemerintah daerah melalui sistem pendataan kepegawaian yang terintegrasi, terkoordinasi, tertib, teratur, transparan dan aman yang juga dapat berkontribusi pada proses perencanaan, pembangunan, pemindahan atau pengangkatan, kesejahteraan, pengendalian kebijakan terkait sehubungan dengan pegawai di pemerintah daerah.

Dari pernyataan di atas, dapat disimpulkan bahwa peran dan tugas IT sebagai bagian dalam pelaksanaan sistem E-Kinerja sangat penting bagi kesinambungan sebuah organisasi pemerintah. Dengan kata lain, peran pegawai secara keseluruhan dalam menjalankan lembaga harus diperhitungkan sehingga faktor-faktor yang membantu mendukung pekerjaannya, seperti komponen TI, juga diperhitungkan. Dengan hadirnya IT juga dimungkinkan untuk mendorong pendataan di bidang kepegawaian untuk penyusunan E-Kinerja.

Dari survei sebelumnya yang peneliti lakukan, memperlihatkan bahwa penerapan sistem E-Kinerja di Kecamatan Benda Kota Tangerang tidak semuanya berjalan tanpa hambatan. Masalah khusus yang biasa terjadi ketika memasukkan laporan pekerjaan harian melalui Sistem E-Kinerja, seperti sistem banyak dikeluhkan pegawai, penerapan sistem E-Kinerja baik secara langsung maupun tidak langsung dianggap merepotkan, selain banyaknya pekerjaan dan sibuk melayani masyarakat, mereka juga harus memberikan laporan kinerja, belum lagi laporan kinerja yang harus disertai dengan bukti foto saat kegiatan dan dokumentasi hasil pekerjaan saat itu.

Masalah teknis seperti internet yang sering tidak terkoneksi juga menjadi salah satu faktor yang menyulitkan pegawai dalam mengisi laporan kerja harian melalui sistem E-Kinerja. Selain itu kehadiran perangkat teknologi (komputer) yang terbatas sehingga tidak semua pegawai mendapatkan fasilitas tersebut. Masalah lain seperti masih adanya pegawai yang belum memahami dan mengerti dalam memanfaatkan sistem E-Kinerja, serta pegawai Kecamatan dianggap kurang berkompeten dalam mengoperasikan komputer sehingga dalam mengaplikasikan sistem tersebut, pegawai meminta admin kantor untuk menginput laporan kerja hariannya hal ini menimbulkan banyak terjadi kesalahan penginputan yang berakibat fatal di laporan akhir.

Sistem E-Kinerja yang membutuhkan pendataan pegawai dalam memberikan informasi yang *up to date* dan transparan menjadi kendala bagi setiap pegawai dalam proses penyusunan maupun pencatatan. Salah satu hambatan terbesar dalam menyediakan informasi yang tepat waktu dan akurat adalah keberadaan dan informasi tentang IT sebagai bagian dari faktor pendukung dalam proses pelaporan yang masih sangat minim. Situasi ini menarik untuk dikaji lebih lanjut, jika dengan keterbatasan fasilitas dan kendala-kendala yang ada, apakah kinerja pegawai dapat dinilai secara akurat. Berdasarkan konteks latar belakang, perspektif dan pandangan kebutuhan suatu lembaga dalam mengelola sistem informasi yang terintegrasi, peneliti tertarik untuk melakukan penelitian eksplorasi tentang fenomena ini.

Untuk mengetahui dan memastikan bahwa Sistem E-Kinerja dapat mencapai tujuan yang diharapkan baik bagi para Pegawai Negeri Sipil (PNS) maupun bagi Pemerintah Daerah, maka perlu dilakukan evaluasi terhadap kesuksesan sistem tersebut dengan menggunakan model kesuksesan sistem informasi *Enhanced Information System Success Model*. Penelitian ini dilakukan dengan alasan bahwa penelitian ini akan menilai, mengukur, membuktikan tingkat keberhasilan pelaksanaan evaluasi kinerja berbasis E-Kinerja, serta mengetahui faktor atau variabel apa saja yang mempengaruhinya.

Salah satu alat dalam melakukan penelitian adalah dengan melakukan survei kepustakaan. Peneliti akan melihat beberapa penelitian yang berkaitan dengan penelitian yang sedang dilakukan. Beberapa studi yang diulas membahas tentang mengukur kesuksesan menggunakan menggunakan model DeLone dan McLean yang dimodifikasi dalam kaitannya dengan aplikasi akademik mahasiswa berbasis seluler yang diteliti oleh (Ernawati et al., 2020), hasil yang diperoleh menyimpulkan bahwa model DeLone and McLean tidak sepenuhnya empiris dalam penelitian ini, karena ada beberapa indikator variabel yang tidak valid yang harus dikeluarkan dari variabel. Dari 12 hipotesis, 5 diterima yaitu kualitas informasi berpengaruh positif dan signifikan terhadap penggunaan, kualitas informasi berpengaruh positif dan signifikan terhadap kepercayaan, kualitas sistem berpengaruh positif dan signifikan terhadap kepuasan pengguna, penggunaan berpengaruh positif dan signifikan terhadap manfaat bersih, dan kepuasan pengguna berpengaruh positif dan signifikan terhadap manfaat bersih.

Penelitian selanjutnya dilakukan (Novianto, 2020), terkait analisis faktor sukses sistem informasi akademik (Siakad) menggunakan model Delone dan McLean yang dimodifikasi, di mana hasil penelitian terhadap faktor keberhasilan dari penelitian ini dijadikan sebagai rekomendasi untuk pengembangan Siakad selanjutnya agar pelayanan kualitas Siakad meningkat. Hasil penelitian menunjukkan bahwa kualitas informasi, kualitas sistem, dan kualitas layanan mempengaruhi kepuasan pengguna, kepuasan pengguna mempengaruhi manfaat bersih yang dihasilkan oleh Siakad.

Penelitian lainnya mengenai dampak kepercayaan *(trust)* pada penggunaan media pemasaran *online* E-*Commerce* melalui evaluasi model keberhasilan sistem informasi Delone dan McLean yang dilakukan oleh (Hamid & Ikbal, 2017), yang menunjukkan bahwa struktur kepercayaan memiliki makna positif dan signifikan *(direct effect)* pada kepuasan pengguna dan *benefit*. Selain itu, struktur kepercayaan juga dapat berperan sebagai mediator yang baik antara kualitas informasi, kualitas sistem, dan struktur kualitas layanan terhadap kepuasan pengguna, yang masing-masing memiliki pengaruh tidak langsungnya *(indirect effect)* positif dan signifikan. Kemudian untuk pengaruh langsung konstruk kualitas informasi, kualitas sistem, dan kualitas layanan terhadap kepercayaan berpengaruh positif dan signifikan. Selanjutnya, struktur kepercayaan juga memiliki pengaruh positif dan tidak langsung yang signifikan terhadap manfaat yang dimediasi oleh struktur kepuasan pengguna. Hasil penelitian ini juga dapat memberikan dukungan empiris bagi keberhasilan model sistem informasi Delone dan McLean dengan memasukkan struktur kepercayaan, menunjukkan kemampuan untuk menggambarkan fenomena penggunaan sistem E-*Commerce online*.

Dikarenakan adanya perbedaan hasil studi (gap), pengembangan hasil penelitian kelompok sebelumnya dan terbatasnya penelitian sebelumnya menggunakan model keberhasilan sistem

informasi Delone dan McLean untuk mengevaluasi keberhasilan atau kegagalan peran teknologi informasi, maka penulis tertarik untuk melakukan studi evaluasi lebih lanjut untuk mengukur keberhasilan implementasi sistem informasi. Namun, dalam penelitian ini, peneliti mengubah objek menjadi aplikasi web yang digunakan oleh pemerintah untuk menganalisis beban kerja unit/satuan kerja organisasi. Penelitian ini bertujuan untuk mengukur keberhasilan penerapan sistem (E-Kinerja) serta diharapkan dapat mengetahui komponen-komponen yang mendukung atau menghambat dalam kesuksesan penggunaan E-Kinerja, sehingga dapat digunakan sebagai bahan evaluasi untuk perbaikan di masa mendatang dan untuk mengisi gap empiris dalam konteks peran sistem (E-Kinerja) sebagai sistem penilaian kinerja Pegawai Negeri Sipil (PNS) dengan menguji keberhasilan sistem informasi yang masih kurang mendapatkan pijakan pada penelitian sebelumnya.

2. METODE PENELITIAN

Penelitian ini menggunakan kerangka penelitian sistem informasi yang dikemukakan oleh (Hevner et al., 2004) yaitu metodologi *IS Research* yang telah dimodifikasi mengikut keperluan peneliti, di mana secara keseluruhan dalam kerangka penelitian terbagi atas dua sudut pandang. Sudut pandang pertama yaitu *relevance* (sesuai dengan fakta di lapangan) seperti melakukan observasi langsung, wawancara, dan penyebaran kuesioner. Sedangkan sudut pandang yang kedua yaitu *rigor* (pengetahuan) seperti menentukan studi literatur dan penelitian terdahulu.

2.1 Metode Pengumpulan Data

Penelitian ini merupakan penelitian campuran (*mix methods)*, menggabungkan dua pendekatan penelitian yaitu kualitatif dan kuantitatif. Metode ini dapat dilaksanakan pertama dengan mengumpulkan data kualitatif melalui observasi dan wawancara, kemudian data kuantitatif dalam hal ini menggunakan survei tertarget (kuesioner) ditujukan kepada responden yang menggunakan E-Kinerja di Kecamatan Benda, Kota Tangerang dengan menggunakan instrumen pernyataan kuesioner. Teknik pengumpulan data yang digunakan dalam penelitian ini adalah:

a) Observasi
   Dalam penelitian ini observasi dilakukan dengan mendatangi langsung lokasi penelitian yaitu Kantor Kecamatan Benda, Kota Tangerang, dan mencari informasi untuk memperoleh data mengenai kondisi objek yang berkaitan dengan penelitian.
b) Wawancara
   Wawancara dilakukan secara tatap muka dengan responden dan kegiatan dilakukan secara lisan untuk mengungkap permasalahan, di mana pihak yang diwawancarai dimintai pendapat dan gagasannya. Wawancara dilakukan di Kantor Kecamatan Benda, Kota Tangerang, dengan 2 responden mewakili Pegawai Negeri Sipil (PNS).
c) Studi Pustaka
   Dalam penelitian ini, studi literatur dilakukan dengan mempelajari buku-buku, dokumen-dokumen, jurnal-jurnal, referensi-referensi, dan hal-hal lain yang berhubungan dengan masalah penelitian.
d) Kuesioner
   Dalam penelitian ini digunakan kuesioner terkait dengan pengukuran keberhasilan Sistem Kinerja Elektronik (E-Kinerja) secara *online* melalui Google Form dengan link yang dikirimkan oleh peneliti dengan skala yang digunakan skala Likert.

2.2 Populasi dan Sampel Penelitian

Populasi yang dipilih berkaitan erat dengan masalah penelitian, sehingga populasi dalam penelitian ini adalah seluruh Pegawai Negeri Sipil (PNS) Kecamatan Benda yang berjumlah 55 PNS dan 7 Pegawai Admin. Sehingga didapatkan total keseluruhan populasi sebanyak 62 orang. Sampel pada penelitian ini menggunakan metode *Non-Probability Sampling* dengan teknik pengambilan sampel yang digunakan yaitu Sampling Jenuh, di mana sampel untuk penelitian ini adalah total populasi sebagai jumlah sampel yang disurvei (Sugiyono, 2013).

2.3 Analisis Data

Analisis data dengan *SEM-PLS* menggunakan bantuan sebuah program yaitu *SmartPLS* versi 3.3 melibatkan dua langkah.

a) Evaluasi Model Pengukuran *(Outer Model)*
Langkah pertama adalah menganalisis *Outer Model* di mana model pengukuran ini dibuat untuk memberikan gambaran bagaimana hubungan yang ada antara blok indikator dengan variabel latennya. Ada empat fase untuk menilai *outer model* yaitu: *Individual Item Reliability, Internal Consistency Reliability, Average Variance Extracted*, dan *Discriminant Validity*.

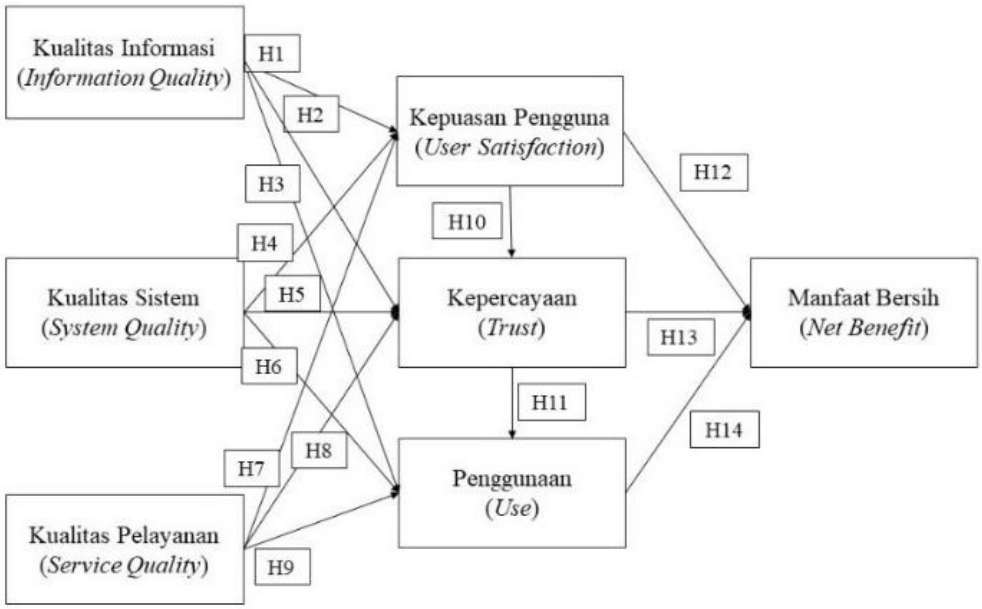
b) Evaluasi Model Struktural (*Inner Model)*
Langkah selanjutnya adalah menganalisis *Inner Model* yaitu untuk menyelidiki pengaruh hubungan antar variabel, dan hubungan variabel secara keseluruhan dalam sistem yang dibangun. Ada tiga fase dalam mengevaluasi hubungan antar konstruk, yaitu pengujian *Path Coefficient (β), T-Test* menggunakan metode *Bootstrapping,* dan *Effect Size (f^2).*

2.4 Model Penelitian

Studi ini menggunakan model evaluasi sistem *Enhanced Information System Success Model* untuk menggambarkan keberhasilan sistem informasi dari sudut pandang pengguna. *Enhanced Information System Success Model* ini merupakan model yang diperbarui, ditingkatkan, atau disempurnakan dari *Information System Success Model* DeLone & McLean (2003). Di mana dalam model DeLone sebelumnya hanya terdiri 6 variabel pengukuran yaitu Kualitas Sistem, Kualitas Informasi, Kualitas Pelayanan, Penggunaan, Kepuasan Pengguna dan Manfaat Bersih. Pada penelitian ini menambahkan variabel Kepercayaan (*Trust*) yang diadopsi dari penelitian Azizah et al. (2021) mengenai Model Kesuksesan Sistem Informasi yang diperbarui, ditingkatkan, atau disempurnakan. Berikut penjelasan variabel yang digunakan dalam penelitian ini:

1) Kualitas Informasi (*Information Quality*)
Kualitas informasi merupakan hasil dari pengguna sistem informasi. Variabel ini menggambarkan kualitas informasi yang dirasakan oleh pengguna.
2) Kualitas Sistem (*System Quality*)
Kualitas sistem adalah kinerja suatu sistem dan menunjukkan seberapa baik karakteristik perangkat keras, perangkat lunak, kebijakan, dan prosedur suatu sistem.sistem informasi dapat memenuhi kebutuhan informasi pengguna.
3) Kualitas Pelayanan (*Service Quality*)
Kualitas layanan sistem informasi adalah layanan yang diperoleh pengguna dari pengembang sistem informasi, dan layanan tersebut dapat berupa pembaruan sistem informasi dan umpan balik pengembang jika sistem bermasalah.
4) Kepuasan Pengguna (*User Satisfaction*)
Kepuasan pengguna adalah tanggapan dan umpan balik yang dihasilkan oleh pengguna setelah menggunakan sistem informasi. Sikap pengguna terhadap sistem informasi adalah ukuran subjektif dari evaluasi mereka terhadap penggunaan sistem.
5) Kepercayaan *(Trust)*
Kepercayaan diakui secara luas di semua organisasi sebagai elemen penting dalam mempromosikan kolaborasi, komunikasi, dan hubungan yang produktif. Kepercayaan merupakan salah satu faktor pendukung keberhasilan implementasi suatu sistem informasi. Kepercayaan berkaitan dengan jaminan keamanan dan privasi yang diberikan oleh sistem kepada pengguna
6) Penggunaan (*Use*)
Penggunaan digunakan untuk menunjukkan berapa kali pengguna menggunakan sistem informasi. Dalam hal ini, penting untuk membedakan apakah penggunaannya wajib atau sukarela.
7) Manfaat Bersih (*Net Benefit*)
Manfaat bersih adalah dampak dari kontribusi, keberadaan, dan penggunaan sistem informasi pada kualitas kinerja pengguna untuk individu, kelompok, dan organisasi, termasuk produktivitas, peningkatan pengetahuan, waktu pengambilan informasi yang lebih cepat, dll.

Dari penjabaran di atas, Model Konseptual penelitian digambarkan pada Gambar 1.



Gambar 1 Model Konseptual Penelitian *Enhanced Information System Success Model*

Dari model konseptual penelitian tersebut, terdapat empat belas hipotesis penelitian yang dianalisis, yaitu:

H1: Kualitas Informasi (*Information Quality*) berpengaruh positif terhadap Kepuasan Pengguna (*User Satisfaction*).
H2: Kualitas Informasi (*Information Quality*) berpengaruh positif terhadap Kepercayaan (*Trust*).
H3: Kualitas Informasi (*Information Quality*) berpengaruh positif terhadap Penggunaan (*Use*).
H4: Kualitas Sistem (*System Quality*) berpengaruh positif terhadap Kepuasan Pengguna (*User Satisfaction*).
H5: Kualitas Sistem (*System Quality*) berpengaruh positif terhadap Kepercayaan (*Trust*).
H6: Kualitas Sistem (*System Quality*) berpengaruh positif terhadap Penggunaan (*Use*).
H7: Kualitas Pelayanan (*Service Quality*) berpengaruh positif terhadap Kepuasan Pengguna (*User Satisfaction*).
H8: Kualitas Pelayanan (*Service Quality*) berpengaruh positif terhadap Kepercayaan (*Trust*).
H: Kualitas Pelayanan (*Service Quality*) berpengaruh positif terhadap Penggunaan (*Use*).
H10: Kepuasan Pengguna (*User Satisfaction*) berpengaruh positif terhadap Kepercayaan (*Trust*).
H11: Kepercayaan (*Trust*) berpengaruh positif terhadap Penggunaan (*Use*).
H12: Kepuasan Pengguna (*User Satisfaction*) berpengaruh positif terhadap Manfaat Bersih yang didapatkan (*Net Benefit*).
H13: Kepercayaan (*Trust*) berpengaruh positif terhadap Manfaat Bersih yang didapatkan (*Net Benefit*).
H14: Penggunaan (*Use*) berpengaruh positif terhadap Manfaat Bersih yang didapatkan (*Net Benefit*).

Selepas menetapkan variabel yang digunakan, berikutnya menentukan setiap indikator yang akan mewakili masing-masing variabel yang ada dalam model penelitian. Tabel 1 menyajikan indikator untuk setiap variabel yang akan digunakan dalam penelitian ini.

Tabel 1 Indikator Penelitian

Variabel	Indikator	Definisi	Kode
Kualitas Informasi (*Information Quality*)	*Completeness*	Sistem E-Kinerja memberikan data atau informasi yang Lengkap.	IQ1
	Relevance	Sistem E-Kinerja menyajikan informasi yang relevan sesuai yang saya butuhkan dan dengan data yang diinput.	IQ2
	Accurate	Sistem E-Kinerja menyajikan informasi yang akurat, padat dan jelas.	IQ3
	Timeliness	Sistem E-Kinerja dapat memberikan Informasi yang bersifat mutakhir (*up to date*).	IQ4
Kualitas Sistem (*System Quality*)	*Ease Of Use*	Sistem E-Kinerja nyaman digunakan dan mudah untuk digunakan.	SQ1
	Reliability	Sistem E-Kinerja memiliki konten atau layanan yang dapat diakses tanpa adanya masalah.	SQ2
	Flexibility	Sistem E-Kinerja memiliki tampilan yang flexible.	SQ3
	Response Time	Sistem E-Kinerja mampu merespon dengan cepat permintaan saya atas instruksi yang dibutuhkan.	SQ4
	Ease Of Learning	Sistem E-Kinerja dapat dipelajari dengan mudah oleh saya.	SQ5
Kualitas Layanan (*Service Quality*)	*Assurance*	Sistem E-Kinerja memberikan jaminan rasa aman dalam mengakses sistem.	SV1
	Empathy	Sistem E-Kinerja memberikan beberapa masukan yang mungkin berguna ketika saya mengakses konten atau layanan.	SV2
	Responsive	Sistem E-Kinerja merespon dengan cepat tanggapan sesuai dengan apa yang saya lakukan.	SV3
Kepuasan Pengguna (*User Satisfaction*)	*Efficiency*	Sistem E-Kinerja dapat membantu memberikan kepuasan terhadap solusi kaitannya dengan aktivitas saya sebagai pengguna secara efisien.	US1
	Effectiveness	Sistem E-Kinerja secara efektif mampu meningkatkan kepuasan saya terhadap sistem tersebut.	US2
	Information Satisfaction	Data dan Informasi pada Sistem E-Kinerja sangat baik dan membuat saya senang untuk mengaksesnya kembali.	US3
	Software Satisfaction	*Software* pendukung yang digunakan untuk mengakses Sistem E-Kinerja berpengaruh pada kepuasan yang saya miliki.	US4
	Overall Purchase	Saya merasa puas dengan Sistem E-Kinerja secara keseluruhan.	US5

Tabel 1 Indikator Penelitian (lanjutan)

Variabel	Indikator	Definisi	Kode
Kepercayaan (*Trust*)	*Commitment*	Sistem E-Kinerja bisa berkomitmen menyimpan dan mengelola data sehingga instansi dan pekerjaan saya tetap berjalan dengan baik dalam situasi apapun.	TR1
	Honest	Informasi yang diberikan Sistem E-Kinerja terpercaya dan dapat dipertanggungjawabkan.	TR2
	Communication	Sistem E-Kinerja menjadi media komunikasi untuk membantu menjalankan sistem pemerintahan secara lebih efisien.	TR3
Penggunaan (*Use*)	*Daily Use*	Saya sering mengunjungi Sistem E-Kinerja setiap hari.	U1
	Frequency Of Use	Sistem E-Kinerja telah digunakan secara rutin oleh saya.	U2
	Nature Of Use	Penggunaan Sistem E-Kinerja dilakukan sesuai dengan maksud yang diinginkan dan sesuai dengan pekerjaan saya.	U3
Manfaat Bersih (*Net Benefit*)	*Improve Knowledge Sharing*	Sistem E-Kinerja dapat meningkatkan pengetahuan pengguna seputar kinerja PNS.	NB1
	Speed Of Accomplishing Task	Sistem E-Kinerja dapat membantu menyelesaikan pekerjaan saya lebih cepat.	NB2
	Communication Effectiveness	Sistem E-Kinerja mempermudah saya untuk menyampaikan kritik dan saran dengan layanan yang disediakan.	NB3
	Task Productivity	Sistem E-Kinerja dapat meningkatkan produktivitas kerja saya dalam menyelesaikan tugas.	NB4

3. HASIL DAN PEMBAHASAN

Data dari kuesioner yang diisi oleh 62 responden pengguna sistem E-Kinerja kemudian diolah. Untuk pengujian yang dilakukan dengan *SEM* berbasis *PLS*, menggunakan bantuan suatu program analisis data yaitu *SmartPLS* versi 3.3. Berikut hasil dari analisis kuesioner yang dilangsungkan dengan 2 tahap.

3.1 Hasil Analisis Pengukuran *Outer Model* (*Measurement Model*)

3.1.1 Uji *Individual Item Reliability*

Pengujian pada tahap *individual item reliability* bisa dilihat dari nilai *standardized loading factor.* Berdasarkan pada Tabel 2 hasil pengujian yang menunjukkan bahwa dari 27 indikator yang digunakan dalam kuesioner, terdapat 25 indikator dengan hasil yang valid dan 2 indikator yang tidak valid. Untuk indikator dengan hasil yang tidak valid ditemukan pada indikator NB1 dan US4 karena nilai *loading factor* yang ditemukan di bawah 0,7. Oleh karena itu, 2 indikator ini perlu dihilangkan atau dilakukan penghapusan pada 2 indikator tersebut.

Setelah 2 indikator tersebut dihapus, kemudian dilakukan pengujian kembali dengan *SmartPLS 3.3*, dan didapatkan bahwa semua nilai *loading factor* sudah menangkup persyaratan (di atas 0,7). Berikut ini disajikan hasil dari pengujian *loading factor* setelah dilakukan penghapusan terhadap 2 indikator yang tidak valid pada Tabel 3.

Tabel 2 Hasil Awal Uji *Loading Factor*

Variabel	Indikator	*Loading Factor*
Information Quality	IQ1	0,833
	IQ2	0,757
	IQ3	0,858
	IQ4	0,750
Net Benefit	NB1	0,388
	NB2	0,882
	NB3	0,867
	NB4	0,703
Service Quality	SV1	0,832
	SV2	0,827
	SV3	0,758
System Quality	SQ1	0,875
	SQ2	0,729
	SQ3	0,878
	SQ4	0,707
	SQ5	0,904
Trust	TR1	0,814
	TR2	0,806
	TR3	0,739
Use	U1	0,826
	U2	0,918
	U3	0,855
User Satisfaction	US1	0,889
	US2	0,930
	US3	0,851

Tabel 3 Hasil Akhir Uji *Loading Factor*

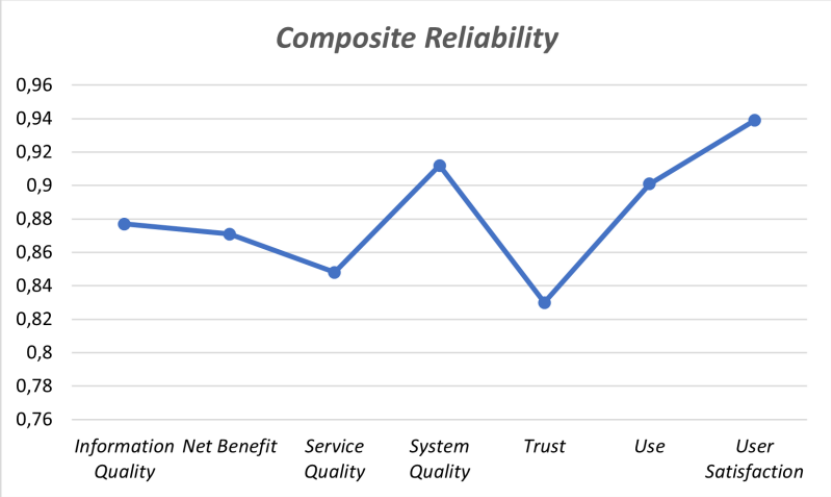
Variabel	Indikator	*Loading Factor*
Information Quality	IQ1	0,835
	IQ2	0,759
	IQ3	0,859
	IQ4	0,747
Net Benefit	NB1	0,897
	NB2	0,888
	NB3	0,700
	NB4	0,834
Service Quality	SV1	0,828
	SV2	0,756
	SV3	0,874
System Quality	SQ1	0,728
	SQ2	0,878
	SQ3	0,709
	SQ4	0,904
	SQ5	0,814
Trust	TR1	0,813
	TR2	0,731
	TR3	0,823
Use	U1	0,918
	U2	0,857
	U3	0,893
User Satisfaction	US1	0,925
	US2	0,861
	US3	0,881

3.1.2 Uji Internal *Consistency Reliability*

Setelah menguji *individual item reliability* melalui nilai *standardized loading factor*, langkah selanjutnya kita melihat *internal consistency reliability* dari nilai *Composite Reliability* (CR). Berdasarkan Tabel 3 didapatkan hasil bahwa nilai *composite reliability* dari semua variabel di atas 0,7 sehingga memenuhi syarat dan valid untuk digunakan dalam model penelitian ini serta tidak ada masalah dalam uji *composite reliability*. Berikut ini disajikan hasil dari pengujian *internal consistency reliability* pada Tabel 4.

Tabel 4 Hasil Uji *Composite Reliability* (CR)

Variabel	*Composite Reliability* (CR)	Keterangan
Information Quality	0,877	Reliabel
Net Benefit	0,871	Reliabel
Service Quality	0,848	Reliabel
System Quality	0,912	Reliabel
Trust	0,830	Reliabel
Use	0,901	Reliabel
User Satisfaction	0,939	Reliabel



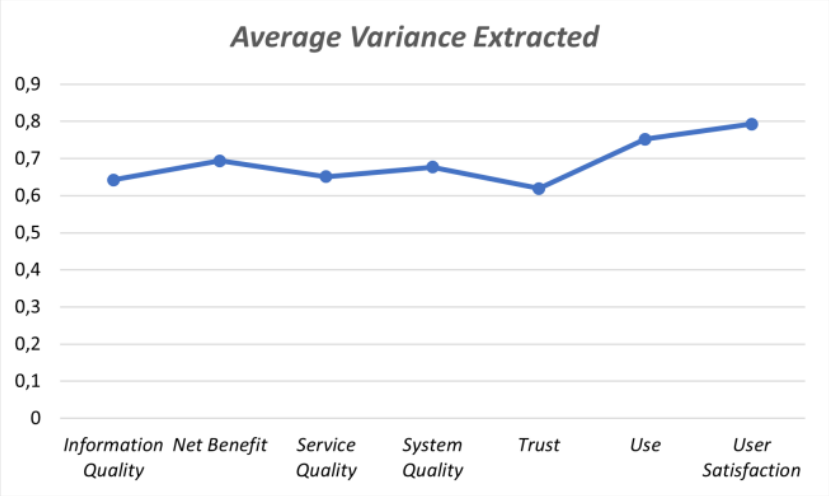
Gambar 2 Hasil Uji *Composite Reliability* (CR)

3.1.3 Uji *Average Variance Extracted*

Pengujian ini dilakukan dengan melihat nilai *average variance extracted* (AVE). Berdasarkan Tabel 5 didapatkan hasil bahwa nilai *average variance extracted* (AVE) untuk tiap variabel laten dengan indikator memiliki hubungan yang sesuai dan lebih besar dari 0,5 sehingga dapat dikatakan bahwa seluruh variabel memenuhi syarat untuk digunakan karena memiliki konstruk validitas yang baik dan tidak ada masalah dalam uji AVE.

Tabel 5 Hasil Uji *Average Variance Extracted* (AVE)

Variabel	*Average Variance Extracted* (AVE)	Keterangan
Information Quality	0,642	Valid
Net Benefit	0,694	Valid
Service Quality	0,651	Valid
System Quality	0,677	Valid
Trust	0,619	Valid
Use	0,752	Valid
User Satisfaction	0,793	Valid



Gambar 3 Hasil Uji *Average Variance Extracted* (AVE)

3.1.4 Uji *Discriminant Validity*

Hasil uji *Cross Loading* menunjukkan bahwa indikator-indikator berkorelasi dengan konstruknya lebih tinggi dibandingkan dengan nilai korelasi indikator pada blok konstruk lainnya, sehingga dapat disimpulkan bahwa setiap indikator merupakan komponen penyusun konstruk. Dapat disimpulkan bahwa seluruh variabel dapat dianggap valid dan dapat digunakan dengan nilai *loading* lebih besar dari 0,70 dan memenuhi validitas diskriminan. Oleh karena itu, dapat disimpulkan bahwa model memenuhi persyaratan untuk dilanjutkan ke tahap pengujian struktur model.

3.2 Hasil Analisis Pengukuran *Inner Model* (*Structural Model*)

3.2.1 Uji *Path Coefficient* (β)

Pengujian pada tahap pertama yaitu melakukan pengujian model struktural dengan mempertimbangkan signifikansi hubungan antara konstruk atau variabel. Hal ini bisa dilihat dari koefisien jalur (*path coefficient*) yang menunjukan kuatnya hubungan antar konstruk. Tabel 6 terlihat bahwa terdapat 10 jalur yang mendapatkan nilai lebih besar dari 0,1 yang artinya memiliki pengaruh terhadap model, sedangkan 4 jalur lainnya memiliki pengaruh yang tidak signifikan. karena berada di bawah ambang batas .0,1 yaitu SV→U, SQ→TR, SQ→U dan U→NB. Berikut ini disajikan hasil dari pengujian *path coefficient* pada Tabel 6.

3.2.2 Uji T-*Test*

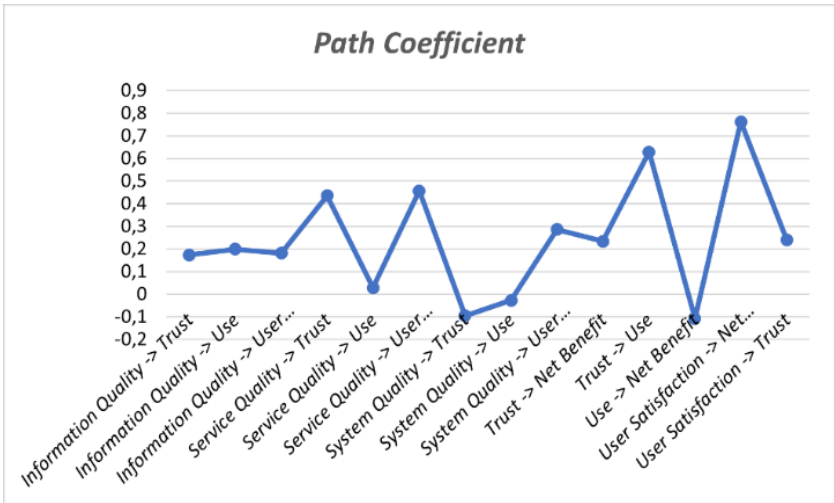
Uji T adalah salah satu uji statistik yang secara umum membandingkan nilai t hitung dengan T Tabel. Uji ini dapat dipergunakan untuk menguji hipotesis, melihat nilai *t-test* dilakukan dengan metode *bootstrapping* menggunakan uji *two-tailed*. Hipotesis akan diterima apabila nilai *t-test* lebih besar dari 1,96. Berikut ini disajikan hasil dari pengujian *t-test* pada Tabel 7.

3.2.3 Uji *Effect Size* (f^2)

Pengujian f^2 *(effect size)* untuk memprediksi pengaruh beberapa variabel terhadap variabel lain dalam struktur model. Berdasarkan Tabel 8 didapatkan hasil bahwa hubungan jalur hipotesis TR→U (0,493) memiliki nilai *effect size* yang besar terhadap struktur model. Kemudian SV→US (0,313) dan US→NB memiliki pengaruh menengah. Sedangkan 11 hipotesis lainnya memiliki pengaruh yang kecil terhadap struktur model karena nilai f^2 di bawah 0,15. Berikut ini disajikan hasil dari pengujian f^2 pada Tabel 8.

Tabel 6 Hasil Uji *Path Coefficient (β)*

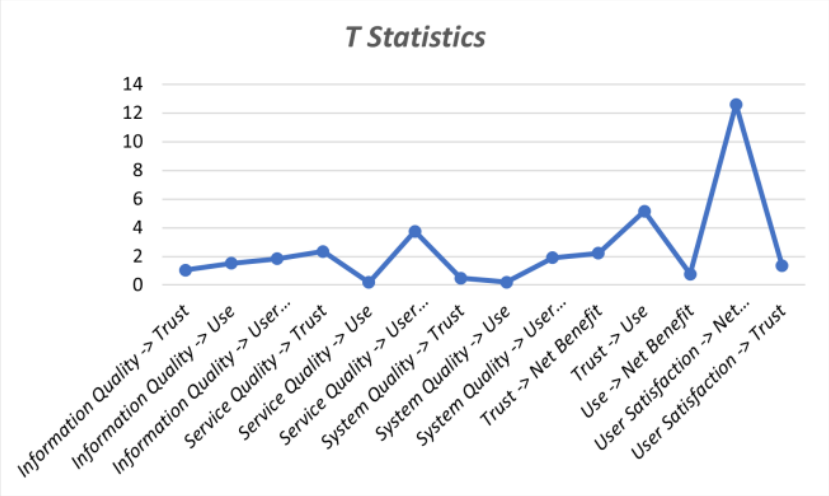
Hubungan antar Variabel	*Path Coefficient* (β)
Information Quality -> Trust	0,174
Information Quality -> Use	0,199
Information Quality -> User Satisfaction	0,181
Service Quality -> Trust	0,436
Service Quality -> Use	0,029
Service Quality -> User Satisfaction	0,457
System Quality -> Trust	-0,095
System Quality -> Use	-0,027
System Quality -> User Satisfaction	0,286
Trust -> Net Benefit	0,234
Trust -> Use	0,628
Use -> Net Benefit	-0,107
User Satisfaction -> Net Benefit	0,762
User Satisfaction -> Trust	0,241



Gambar 4 Hasil Uji *Path Coefficient (β)*

Tabel 7 Hasil Uji *T-test*

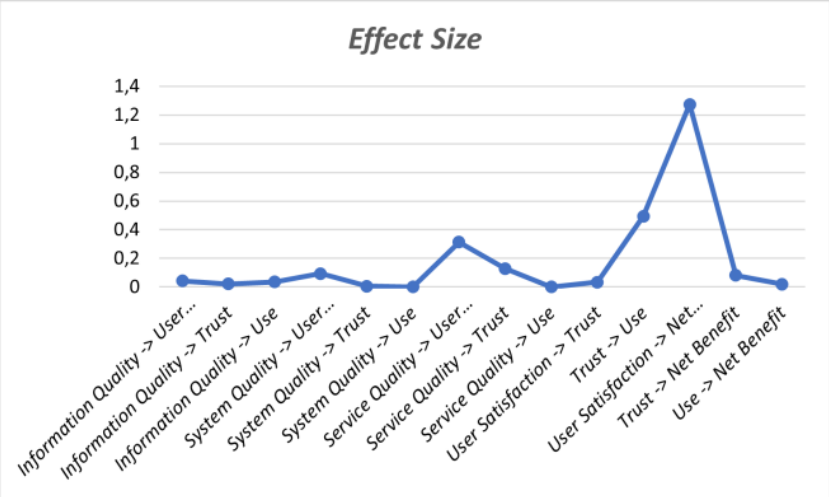
Hubungan antar Variabel	T *Statistics*	Keterangan
Information Quality -> Trust	1,044	Ditolak
Information Quality -> Use	1,525	Ditolak
Information Quality -> User Satisfaction	1,836	Ditolak
Service Quality -> Trust	2,344	Diterima
Service Quality -> Use	0,191	Ditolak
Service Quality -> User Satisfaction	3,754	Diterima
System Quality -> Trust	0,474	Ditolak
System Quality -> Use	0,192	Ditolak
System Quality -> User Satisfaction	1,909	Ditolak
Trust -> Net Benefit	2,223	Diterima
Trust -> Use	5,144	Diterima
Use -> Net Benefit	0,756	Ditolak
User Satisfaction -> Net Benefit	12,603	Diterima
User Satisfaction -> Trust	1,339	Ditolak



Gambar 5 Hasil Uji *T-test*

Tabel 8 Hasil Uji *Effect Size* (f^2)

Hipotesis	*Effect Size* (*f*2)
Information Quality -> User Satisfaction	0,043
Information Quality -> Trust	0,022
Information Quality -> Use	0,036
System Quality -> User Satisfaction	0,094
System Quality -> Trust	0,006
System Quality -> Use	0,001
Service Quality -> User Satisfaction	0,313
Service Quality -> Trust	0,129
Service Quality -> Use	0,001
User Satisfaction -> Trust	0,034
Trust -> Use	0,493
User Satisfaction -> Net Benefit	1,274
Trust -> Net Benefit	0,081
Use -> Net Benefit	0,019



Gambar 6 Hasil Uji *Effect Size* (f^2)

Hasil Pengujian Hipotesis

Berdasarkan 14 hipotesis yang diuji, diperoleh hasil bahwa terdapat 9 hipotesis yang hasilnya ditolak yakni H1, H2, H3, H4, H5, H6, H9, H10 dan H14, sedangkan 5 hipotesis lainnya diterima. Berikut penjelasan hasil analisis yang telah dilakukan sesuai dengan pertanyaan penelitian dan hipotesis yang telah dirumuskan sebelumnya.

H1: Pengaruh Kualitas Informasi *(Information Quality)* Terhadap Kepuasan Pengguna *(User Satisfaction)*
Berdasarkan hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H1 yaitu hubungan Kualitas Informasi *(Information Quality)* → Kepuasan Pengguna *(User Satisfaction)* hasilnya **ditolak**, karena memiliki nilai *t-test* sebesar 1,836 artinya <1,96, jadi bisa disimpulkan bahwa IQ tidak terdapat pengaruh positif pada US. Selanjutnya didukung pula dengan hasil nilai *effect size* (f^2) sebesar 0,043 di mana artinya IQ memiliki pengaruh kecil terhadap US. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Ernawati et al., 2020; Nuryanti, 2020) yang juga menyatakan bahwa kualitas informasi tidak mempengaruhi kepuasan pengguna.

H2: Pengaruh Kualitas Informasi *(Information Quality)* Terhadap Kepercayaan *(Trust)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H2 yaitu hubungan Kualitas Informasi *(Information Quality)* → Kepercayaan *(Trust)* hasilnya **ditolak**, karena memiliki nilai *t-test* sebesar 1,044 artinya <1,96, jadi bisa disimpulkan bahwa IQ tidak terdapat pengaruh positif pada TR. Selanjutnya didukung pula dengan hasil nilai *effect size* (f^2) sebesar 0,022 di mana artinya IQ memiliki pengaruh kecil terhadap TR.

H3: Pengaruh Kualitas Informasi *(Information Quality)* Terhadap Penggunaan *(Use)*
Berdasarkan hasil pengujian secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H3 yaitu hubungan Kualitas Informasi *(Information Quality)* → Penggunaan *(Use)* hasilnya **ditolak**, karena memiliki nilai *t-test* sebesar 1,525 artinya <1,96, jadi bisa disimpulkan bahwa IQ tidak terdapat pengaruh positif pada U. Selanjutnya didukung pula dengan hasil nilai *effect size* (f^2) sebesar 0,036 di mana artinya IQ memiliki pengaruh kecil terhadap U. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Bahesa, 2018; Nurjaya, 2017) yang juga menyatakan bahwa kualitas informasi tidak mempengaruhi penggunaan sistem.

H4: Pengaruh Kualitas Sistem *(System Quality)* Terhadap Kepuasan Pengguna *(User Satisfaction)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H4 yaitu hubungan Kualitas Sistem *(System Quality)* → Kepuasan Pengguna *(User Satisfaction)* hasilnya **ditolak**, karena mendapatkan nilai *t-test* 1,909 artinya <1,96, jadi bisa disimpulkan bahwa SQ tidak terdapat pengaruh positif pada US. Selanjutnya didukung pula dengan hasil nilai *effect size* (f^2) sebesar 0,094 di mana artinya SQ memiliki pengaruh kecil terhadap US. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Bahesa, 2018; Nurjaya, 2017) yang juga mengatakan bahwa kualitas sistem tidak mempengaruhi kepuasan pengguna.

H5: Pengaruh Kualitas Sistem *(System Quality)* Terhadap Kepercayaan *(Trust)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H5 yaitu hubungan Kualitas Sistem *(System Quality)* → Kepercayaan *(Trust)* hasilnya **ditolak**, karena memiliki nilai *t-test* sebesar 0,474 artinya <1,96, jadi bisa disimpulkan bahwa SQ tidak terdapat pengaruh positif pada TR. Selanjutnya didukung pula dengan hasil analisis jalur menggunakan nilai *path coefficient (β)* sebesar -0,095 di mana nilai ini <0,1 yang artinya SQ juga memiliki pengaruh secara negatif tidak signifikan pada TR. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Adika, 2021;

Ernawati et al., 2020) yang juga mengatakan bahwa kualitas sistem tidak mempengaruhi kepercayaan.

H6: Pengaruh Kualitas Sistem *(System Quality)* Terhadap Penggunaan *(Use)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H6 yaitu hubungan Kualitas Sistem *(System Quality)* → Penggunaan *(Use)* hasilnya **ditolak**, karena karena memiliki nilai *t-test* sebesar 0,192 artinya <1,96, jadi bisa disimpulkan bahwa SQ tidak terdapat pengaruh positif pada U. Selanjutnya didukung pula dengan hasil analisis jalur menggunakan nilai *path coefficient (β)* sebesar -0,027 di mana nilai ini <0,1 yang artinya SQ juga memiliki pengaruh secara negatif tidak signifikan pada U. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Nurjaya, 2017; Wiyati & Sarja, 2018) yang juga mengatakan bahwa kualitas sistem tidak mempengaruhi penggunaan sistem.

H7: Pengaruh Kualitas Pelayanan *(Service Quality)* Terhadap Kepuasan Pengguna *(User Satisfaction)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H7 yaitu hubungan Kualitas Pelayanan *(Service Quality)* → Kepuasan Pengguna *(User Satisfaction)* hasilnya **diterima**, karena memiliki nilai *t-test* sebesar 3,754 artinya >1,96, jadi bisa disimpulkan bahwa SV berpengaruh positif pada US. Selanjutnya didukung pula dengan hasil analisis jalur menggunakan nilai *path coefficient (β)* sebesar 0,457 di mana nilai ini >0,1 yang artinya SV juga memiliki pengaruh secara signifikan terhadap US. Hasil tersebut serupa dan didukung dengan penelitian sebelumnya dari (Nurjaya, 2017; Nuryanti, 2020) yang juga mengatakan bahwa kualitas pelayanan berpengaruh positif dan signifikan terhadap kepuasan pengguna.

H8: Pengaruh Kualitas Pelayanan *(Service Quality)* Terhadap Kepercayaan *(Trust)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H8 yaitu hubungan Kualitas Pelayanan *(Service Quality)* → Kepercayaan *(Trust)* hasilnya **diterima**, karena memiliki nilai *t-test* sebesar 2,344 artinya >1,96, jadi bisa disimpulkan bahwa SV berpengaruh positif pada TR. Selanjutnya didukung pula dengan hasil analisis jalur menggunakan nilai *path coefficient (β)* sebesar 01,436 di mana nilai ini >0,1 yang artinya SV juga memiliki pengaruh secara signifikan terhadap TR. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Adika, 2021; Pramana & Rastini, 2016) yang juga menyatakan bahwa kualitas pelayanan berpengaruh positif dan signifikan terhadap kepercayaan.

H9: Pengaruh Kualitas Pelayanan *(Service Quality)* Terhadap Penggunaan *(Use)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H9 yaitu hubungan Kualitas Pelayanan *(Service Quality)* → Penggunaan *(Use)* hasilnya **ditolak**, karena memiliki nilai *t-test* sebesar 0,191 artinya <1,96, jadi bisa disimpulkan bahwa SV tidak terdapat pengaruh positif pada U. Selanjutnya didukung pula dengan hasil analisis jalur menggunakan nilai *path coefficient (β)* sebesar 0,029 di mana nilai ini <0,1 yang berarti SV juga tidak berpengaruh terhadap U. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Ernawati et al., 2020; Nurjaya, 2017) yang juga mengatakan bahwa kualitas pelayanan tidak mempengaruhi penggunaan sistem.

H10: Pengaruh Kepuasan Pengguna *(User Satisfaction)* Terhadap Kepercayaan *(Trust)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H10 yaitu hubungan Kepuasan Pengguna *(User Satisfaction)* → Kepercayaan *(Trust)* hasilnya **ditolak**, karena memiliki nilai *t-test* sebesar 1,339 artinya >1,96, jadi bisa disimpulkan bahwa US tidak terdapat pengaruh positif pada TR. Selanjutnya didukung pula dengan hasil nilai *effect size* (f^2) sebesar 0,034 di mana artinya US memiliki pengaruh kecil terhadap TR.

H11: Pengaruh Kepercayaan *(Trust)* Terhadap Penggunaan *(Use)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H11 yaitu hubungan Kepercayaan *(Trust)* → Penggunaan *(Use)* hasilnya **diterima**, karena memiliki nilai *t-test* sebesar 5,144 artinya >1,96 jadi bisa disimpulkan bahwa TR berpengaruh positif pada U. Selanjutnya didukung pula dengan hasil analisis jalur menggunakan nilai *path coefficient (β)* sebesar 0,628 di mana nilai ini >0,1 yang artinya TR juga memiliki pengaruh secara signifikan terhadap U. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Adika, 2021; Novinda, 2011) yang juga mengatakan bahwa kepercayaan pengguna berpengaruh positif dan signifikan terhadap penggunaan.

H12: Pengaruh Kepuasan Pengguna *(User Satisfaction)* Terhadap Manfaat Bersih Yang Didapatkan *(Net Benefit)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H12 yaitu hubungan Kepuasan Pengguna *(User Satisfaction)* → Manfaat Bersih *(Net Benefit)* hasilnya **diterima**, karena memiliki nilai *t-test* sebesar 12,603 artinya >1,96, jadi bisa disimpulkan bahwa US berpengaruh positif pada NB. Selanjutnya didukung pula dengan hasil analisis jalur menggunakan nilai *path coefficient (β)* sebesar 0,762 di mana nilai ini >0,1 yang artinya US juga memiliki pengaruh secara signifikan terhadap NB. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Ernawati et al., 2020; Nuryanti, 2020) yang juga mengatakan bahwa tingkat kepuasan pengguna secara positif dan signifikan mempengaruhi manfaat bersih yang dirasakan oleh individu maupun organisasi.

H13: Pengaruh Kepercayaan *(Trust)* Terhadap Manfaat Bersih Yang Didapatkan *(Net Benefit)*
Berdasarkan hasil pengujian secara *statistic* Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H13 yaitu hubungan Kepercayaan *(Trust)* → Manfaat Bersih *(Net Benefit)* hasilnya **diterima**, karena memiliki nilai *t-test* sebesar 2,223 artinya >1,96, jadi bisa disimpulkan bahwa TR berpengaruh positif pada NB. Selanjutnya didukung pula dengan hasil analisis jalur menggunakan nilai *path coefficient (β)* sebesar 0,234 di mana nilai ini >0,1 yang artinya TR juga berpengaruh secara signifikan terhadap NB. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Hamid & Ikbal, 2017) yang juga menyatakan bahwa kepercayaan secara positif dan signifikan mempengaruhi manfaat bersih yang dirasakan oleh individu maupun organisasi.

H14: Pengaruh Penggunaan *(Use)* Terhadap Manfaat Bersih Yang Didapatkan *(Net Benefit)*
Berdasarkan pada hasil pengukuran secara *statistic* menggunakan analisis nilai *t-test* sebagaimana ditunjukan oleh Tabel 7 yang memperlihatkan bahwa H14 yaitu hubungan Penggunaan *(Use)* → Manfaat Bersih *(Net Benefit)* hasilnya **ditolak**, karena memiliki nilai *t-test* sebesar 0,756 artinya <1,96, jadi bisa disimpulkan bahwa U tidak terdapat pengaruh positif pada NB. Selanjutnya didukung pula dengan hasil analisis jalur menggunakan nilai *path coefficient (β)* sebesar -0,107 di mana nilai ini <0,1 di mana artinya U juga memiliki pengaruh secara negatif tidak signifikan pada NB. Hasil ini serupa dan didukung oleh penelitian sebelumnya dari (Bahesa, 2018; Wahyudi & Wardiyono, 2018) yang juga mengatakan bahwa penggunaan tidak mempengaruhi manfaat bersih.

4. KESIMPULAN

Berdasarkan hasil analisis data yang telah dilakukan dan pembahasan dari bab sebelumnya, berikut adalah kesimpulan dari penelitian ini:

1) Ada 14 hipotesis hasil pengukuran, penelitian ini menunjukkan bahwa:
   a) H1 ditolak, Kualitas Informasi *(Information Quality)* tidak berpengaruh signifikan terhadap Kepuasan Pengguna *(User Satisfaction)* Sistem E-Kinerja Kota Tangerang.
   b) H2 ditolak, Kualitas Informasi *(Information Quality)* tidak berpengaruh signifikan terhadap Kepercayaan *(Trust)* Sistem E-Kinerja Kota Tangerang.

c) H3 ditolak, Kualitas Informasi *(Information Quality)* tidak berpengaruh signifikan terhadap Penggunaan *(Use)* Sistem E-Kinerja Kota Tangerang.
d) H4 ditolak, Kualitas Sistem *(System Quality)* tidak berpengaruh signifikan terhadap Kepuasan Pengguna *(User Satisfaction)* Sistem E-Kinerja Kota Tangerang.
e) H5 ditolak, Kualitas Sistem *(System Quality)* tidak berpengaruh signifikan terhadap Kepercayaan *(Trust)* Sistem E-Kinerja Kota Tangerang.
f) H6 ditolak, Kualitas Sistem *(System Quality)* tidak berpengaruh signifikan terhadap Penggunaan *(Use)* Sistem E-Kinerja Kota Tangerang.
g) H7 diterima, Kualitas Pelayanan *(Service Quality)* berpengaruh signifikan terhadap Kepuasan Pengguna *(User Satisfaction)* Sistem E-Kinerja Kota Tangerang.
h) H8 diterima, Kualitas Pelayanan *(Service Quality)* berpengaruh signifikan terhadap Kepercayaan *(Trust)* Sistem E-Kinerja Kota Tangerang.
i) H9 ditolak, Kualitas Pelayanan *(Service Quality)* tidak berpengaruh signifikan terhadap Penggunaan *(Use)* Sistem E-Kinerja Kota Tangerang.
j) H10 ditolak, Kepuasan Pengguna *(User Satisfaction)* tidak berpengaruh signifikan terhadap Kepercayaan *(Trust)* Sistem E-Kinerja Kota Tangerang.
k) H11 diterima, Kepercayaan *(Trust)* berpengaruh signifikan terhadap Penggunaan *(Use)* Sistem E-Kinerja Kota Tangerang.
l) H12 diterima, Kepuasan Pengguna *(User Satisfaction)* berpengaruh signifikan terhadap Manfaat Bersih *(Net Benefit)* Sistem E-Kinerja Kota Tangerang.
m) H13 diterima, Kepercayaan *(Trust)* berpengaruh signifikan terhadap Manfaat Bersih *(Net Benefit)* Sistem E-Kinerja Kota Tangerang.
n) H14 ditolak, Penggunaan *(Use)* tidak berpengaruh signifikan terhadap Manfaat Bersih *(Net Benefit)* Sistem E-Kinerja Kota Tangerang.

2) Rekomendasi
a) Peningkatan berurutan dalam akurasi dan relevansi informasi, sarana informasi, tepat waktunya informasi, dan kejelasan informasi tertulis.
b) Administrator lebih aktif meng*update* berita dan info secepat mungkin agar informasi juga cepat diterima oleh pengguna.
c) Menambahkan fitur *Frequently Asked Questions (FAQ)* sehingga pengguna dapat menemukan pertanyaan yang sering diajukan terkait sistem E-Kinerja.
d) Peraturan terkait SOP harus segera disiapkan untuk alur mekanisme validasi, pembaruan, verifikasi, dan otorisasi data oleh sistem E-Kinerja untuk menjaga kualitas informasi dari perspektif akurasi dan pembaruan data.
e) Desain antarmuka sistem E-Kinerja perlu diperbarui agar terlihat modern, simpel, menarik, mudah digunakan, dan nyaman untuk digunakan bahkan jika pengguna mengaksesnya melalui *smartphone*.
f) Meningkatkan *bandwidth* pada server sistem E-Kinerja sehingga apabila jumlah pengunjung tinggi tidak mengalami akses yang lambat, *down*, dan dapat diakses tanpa adanya masalah. Serta selalu lakukan *maintenance* server secara berkala agar server tidak sering *crash*.
g) Menambah perangkat teknologi (komputer), sehingga semua pegawai mendapatkan fasilitas tersebut yang akan mempermudah mereka dalam menginput laporan kinerja hariannya.
h) Pelayanan bagian IT masih perlu ditingkatkan, karena pelayanan yang maksimal juga akan meningkatkan kepuasan pengguna saat menggunakan, yang juga akan mempengaruhi intensitas penggunaan sistem E-Kinerja.
i) Perihal dengan menjaga kepuasan pengguna, sistem E-Kinerja harus memberikan informasi, layanan dan perangkat lunak yang maksimal kepada semua pengguna lama atau baru, karena sistem dapat secara efektif merespon solusi yang terkait dengan aktivitas pelaporan sistem secara efisien sehingga kepuasan pengguna terhadap keseluruhan sistem bisa tercapai. Apabila kepuasan pengguna tercapai, maka hal tersebut juga akan berdampak pada tingginya tingkat kepercayaan pengguna terhadap sistem informasi yang digunakan.

j) Pengembang diharapkan harus terus berinovasi dalam proses pengembangan sistem E- Kinerja, jika semuanya telah berkualitas, kepuasan pasti akan diterima dan dirasakan oleh pengguna, yang akan mempengaruhi kepercayaan pengguna
k) Pengembang perlu menempatkan sejumlah besar staf fungsional dari organisasi TI yaitu Pranata Komputer yang melakukan kegiatan TI berbasis komputer, termasuk tata kelola dan pengelolaan lembaga komputer yang akan mendukung pengembangan Sistem E-Kinerja. Pranata Komputer adalah Pegawai Negeri Sipil (PNS) yang ditugaskan oleh otoritas yang berwenang untuk dapat fokus dalam mendukung dan membantu pengembangan sistem E-Kinerja, di mana setiap aktivitas yang terkait dengan analisis dan desain sistem informasi ini dapat memberikan nilai kredit tersendiri.
l) Pemerintah dan Kecamatan Benda perlu menerapkan sifat atau mode penggunaan yang dapat berdampak pada tingginya intensitas penggunaan sistem sehingga pengguna bisa merasakan keuntungan yang lebih tinggi dan mendapatkan manfaat yang lebih besar.

3) Saran
Untuk studi lebih lanjut, mereka yang ingin menggunakan model dasar DeLone dan McLean untuk menilai keberhasilan sistem informasi, mereka dapat langsung memilih untuk menggunakan *Enhanced Information System Success Model,* karena sudah terdapat variabel kepercayaan yang merupakan salah satu konstruk yang dapat berperan baik dalam keberhasilan sistem informasi.

DAFTAR PUSTAKA

Adika, L. A. (2021). *Pengaruh Kualitas Sistem, Kualitas Layanan, Kemudahan Pengguna, Promosi, Religiusitas Terhadap Kepuasan Pengguna Dan Keputusan Pengguna Shopee Paylater Kepercayaan Sebagai Variabel Perantara*. Universitas Muhammadiyah Surakarta.

Azizah, A. H., Sandfreni, S., & Ulum, M. B. (2021). Analisis Efektivitas Penggunaan Portal Resmi Merdeka Belajar Kampus Merdeka Menggunakan Model Delone and McLean. *Sebatik*, *25*(2), 303–310. https://doi.org/10.46984/sebatik.v25i2.1671

Bahesa, B. P. (2018). *Analisis Kesuksesan Sistem Informasi Website Pemerintah Kabupaten Pamekasan Berdasarkan Model Delone and McLean*. Institut Bisnis dan Informatika Stikom Surabaya.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, *19*(4), 9–30. https://doi.org/10.1080/07421222.2003.11045748

Ernawati, M., Hermaliani, E. H., & Sulistyowati, D. N. (2020). Penerapan DeLone and McLean Model untuk Mengukur Kesuksesan Aplikasi Akademik Mahasiswa Berbasis Mobile. *Jurnal IKRA-ITH Informatika*, *5*(18), 58–67.

Hamid, R. S., & Ikbal, M. (2017). Analisis Dampak Kepercayaan pada Penggunaan Media Pemasaran Online (E-Commerce) yang Diadopsi oleh UMKM: Perspektif Model DeLone & McLean. *Jurnal Manajemen Teknologi*, *16*(3), 310–337. https://doi.org/10.12695/jmt.2017.16.3.6

Hevner, March, Park, & Ram. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75. https://doi.org/10.2307/25148625

Novianto, R. (2020). Analysis of Success Factor Sistem Informasi Akademik (Siakad) Use the Delone and McLean Model (Case Study Stie Muhammadiyah Pringsewu Lampung). *Jurnal Technology Acceptance Model*, *11*(1), 42–47. https://doi.org/10.56327/jurnal tam.v11i1.871

Novinda, K. (2011). *Faktor–Faktor yang Mempengaruhi Kepercayaan (Trust) terhadap Partisipasi Pelanggan E-Commerce (Studi pada pelanggan e-commerce di Indonesia)*. Universitas Islam Indonesia.

Nurjaya, D. (2017). *Pengaruh kualitas sistem, informasi dan pelayanan terhadap manfaat bersih dengan menggunakan model DeLone dan McLean (studi kasus di Rumah Sakit Panti Rapih Yogyakarta)*. Sanata Dharma University.

Nuryanti. (2020). *Analisis Kesuksesan Sistem Informasi Website Pemerintah Kota Sukabumi Menggunakan Model Delone dan McLean*. Universitas Bina Sarana Informatika.

Pramana, I. G. Y., & Rastini, N. M. (2016). Pengaruh Kualitas Pelayanan Terhadap Kepercayaan

Nasabah dan Loyalitas Nasabah Bank Mandiri Cabang Veteran Denpasar Bali. *E-Jurnal Manajemen*, *5*(1), 706–733.

Sugiyono. (2013). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif dan R&D*. Alfabeta.

Wahyudi, A. S. B., & Wardiyono, W. (2018). Evaluasi Pemanfaatan Sistem Informasi Kasus dengan Model Information System Success Delone & McLean di Lembaga Bantuan Hukum Jakarta. *Bibliotech : Jurnal Ilmu Perpustakaan Dan Informasi*, *3*(2), 109. https://doi.org/10.33476/bibliotech.v3i2.914

Wiyati, R. K., & Sarja, N. L. A. K. Y. (2018). Evaluasi Kesuksesan Sistem Informasi Absensi Online Menggunakan Model Delone McLean. *Jurnal Media Aplikom*, *10*(2), 135–157. https://doi.org/10.33488/1.MA.2018.2.59

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022, Pp. 211 – 225
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Pengenalan Tulisan Tangan Huruf Latin Bersambung Menggunakan Local Binary Pattern dan K-Nearest Neighbor

Vivin Oktavia [(1)], Novan Wijaya [(2)*]
[1] Teknik Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Palembang
[2] Magister Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Palembang
e-mail : vivinoktavia@mhs.mdp.ac.id, novan.wijaya@mdp.ac.id.
* Penulis korespondensi.
Artikel ini diajukan 23 Juli 2022, direvisi 7 September 2022, diterima 11 September 2022, dan dipublikasikan 25 September 2022.

Abstract

There are 26 Latin letters in Indonesia, 5 of which are vowels and 21 consonants. This study will translate handwriting with a Latin object using the K-Nearest Neighbor method with the Local Binary Pattern extension. The research is being done with a focus on experimentation using a few methods that have already been discussed. Concatenated Latin letters have a few variations that depend on the work's author, so research will be conducted to identify cursive Latin letters based on these variations. Each of the 30 respondents wrote 26 capital letters and 26 lowercase letters on paper, which was then scanned to provide the image data. Black, blue, and red pens were used to write by every ten responders. The recognition procedure is broken into two halves, capital and non-capital letter recognition using 780 picture datasets each. In the study, k-fold cross-validation is used, with k = 6. The best value was reached at k = 7 with 29.49 percent accuracy, 33.88 percent precision, recall 33.46 percent, and F1-score 27.65 percent according to the research utilizing KNN with values k = 3, 5, and 7. and for recognizing non-capital characters, the best result was found at k=3 with accuracy, precision, recall, and F1-score of 26.28, 27.27, and 22.7%, respectively.

Keywords: Latin Letters, Cursive, K-Nearest Neighbor, Local Binary Pattern, K-Fold Cross Validation, Handwritten

Abstrak

Huruf Latin yang dimiliki Indonesia berjumlah 26 huruf, di antaranya 5 huruf vokal dan 21 huruf konsonan. Dalam penelitian ini dilakukan pengenalan tulisan tangan dengan objek huruf Latin bersambung menggunakan metode K-Nearest Neighbor dengan ekstraksi ciri Local Binary Pattern. Penelitian yang dilakukan bersifat ekperimen dengan menggunakan beberapa metode yang telah disebutkan sebelumnya. Huruf Latin bersambung memiliki beberapa variasi tergantung dari penulis itu sendiri sehingga berdasarkan variasi tersebut dilakukan penelitian dalam mengidentifikasi huruf Latin bersambung. *Dataset* citra merupakan hasil *scan* tulisan tangan di atas kertas yang ditulis oleh 30 responden yang masing-masing responden menulis 26 huruf kapital dan 26 huruf non-kapital. Tiap 10 responden menulis menggunakan pena berwarna hitam, biru, dan merah. Proses pengenalan dibagi menjadi 2 yaitu pengenalan huruf kapital dengan 780 *dataset* citra dan non kapital dengan 780 *dataset* citra. Penelitian menggunakan *k-fold cross validation* dengan nilai k yaitu 6. Hasil dari penelitian menggunakan KNN dengan nilai 3, 5, dan 7, pada pengenalan huruf kapital didapatkan nilai terbaik pada k=7 dengan *accuracy* 29.49%, *precision* 33.88%, *recall* 33.46%, dan *F1-score* 27.65% dan pada pengenalan huruf non-kapital didapatkan nilai terbaik pada k=3 dengan *accuracy* 26.28%, *precision* 27.27%, *recall* 26.27%, dan *F1-score* 22.7%.

Kata Kunci: Huruf Latin, Huruf Latin Bersambung, K-Nearest Neighbor, Local Binary Pattern, K-Fold *Cross Validation*, Tulisan Tangan

1. PENDAHULUAN

Huruf merupakan bentuk, goresan, atau simbol dari suatu sistem tulisan. Huruf Latin merupakan salah satu aksara yang paling banyak digunakan dan merupakan aksara yang pertama kali dipakai oleh orang Romawi. Huruf Latin bersambung adalah suatu bentuk tulisan tangan yang tiap hurufnya ditulis secara terhubung. Huruf Latin sering ditulis oleh anak-anak sekolah dasar. Hal ini bermanfaat untuk perkembangan kemampuan motorik anak dan perkembangan otak (Maharani et al., 2019).

Kemajuan teknologi yang semakin hari mengalami perkembangan yang pesat, dapat membawa kemajuan yang sangat berarti dalam berbagai aspek kehidupan masyarakat. Penelitian mengenai pengenalan huruf (Masrani et al., 2018), angka (Prihatiningsih et al., 2019), dan tulisan tangan (Anggraeny et al., 2020) yang beragam sudah sering dilakukan. Salah satu metode untuk ekstraksi fitur tekstur adalah Local Binary Pattern. LBP memiliki kelebihan dapat mendeskripsikan karakter tekstur pada permukaan (Rahayu et al., 2021). Lalu metode yang dapat digunakan untuk melakukan pengenalan objek adalah K-Nearest Neighbor. Beberapa penelitian yang telah dilakukan mengenai tulisan tangan ini, di antaranya tulisan tangan dengan huruf Latin, tulisan tangan aksara jawa dan sebagainya. Penelitian yang telah dilakukan sebelumnya oleh (Purbayanti, 2018) menggunakan *dataset* berupa huruf Latin, K-Nearest Neighbor, dan deteksi sobel menghasilkan tingkat akurasi tertinggi dengan jumlah data yang bernilai benar adalah 84 dari 90 data sebesar 93,334% dan akurasi terendah dengan jumlah data yang bernilai benar adalah 39 dari 72 sebesar 54,167%.

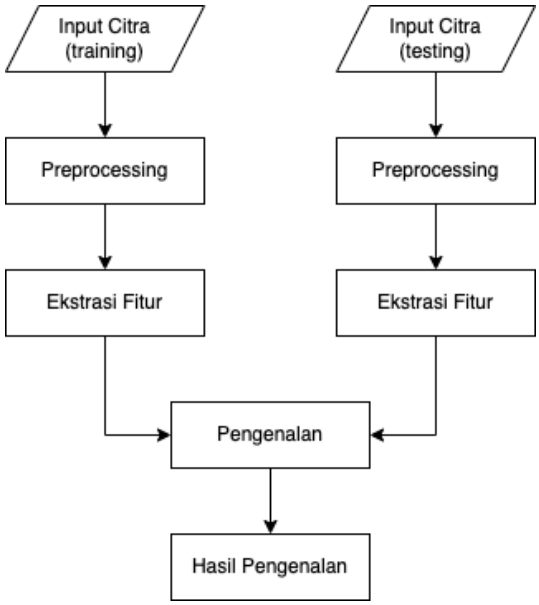
Penelitian yang dilakukan (Ilham & Rochmawati, 2020) menggunakan metode K-Nearest Neighbor menggunakan deteksi tepi Sobel untuk memperoleh bagian detail dengan perhitungan jarak dengan 3 skenario. Skenario pertama menggunakan data *training* sebanyak 40 data dan data *testing* sebanyak 20 data dengan menghasilkan akurasi sebesar 40%, skenario kedua menggunakan data *training* sebanyak 60 data dan data *testing* sebanyak 20 data dengan mencapai akurasi sebesar 55%, skenario ketiga menggunakan 100 data *training* dan 20 data *testing* lalu diperoleh akurasi sebesar 85%. Penelitian tentang unjuk kerja K-Nearest Neighbors terhadap Penggalan Karakter Jawa yang menggunakan Local Binary Pattern yang berguna untuk mengekstrak ciri unik dari citra tulisan aksara Jawa. Sementara KNN digunakan untuk menentukan kelas dari citra aksara Jawa berdasarkan hasil ekstraksi ciri LBP. Pengujian dilakukan dengan menggunakan 160 citra yang dibagi menjadi 40 citra uji dan 120 citra latih. Hasil pengujian dievaluasi menggunakan *confusion matrix* untuk menentukan akurasi dari kombinasi KNN dan LBP dengan akurasi tertinggi mencapai 82.5% di mana parameter yang digunakan adalah *cell size* berukuran 64x64 dan nilai k=3 (Bimantoro et al., 2021).

Berdasarkan studi literatur yang ada, sudah banyak penelitian mengenai pengenalan tulisan tangan dengan berbagai metode dan berbagai objek seperti huruf, angka, aksara Korea, aksara Jepang, aksara Jawa, dan lainnya. Banyaknya variasi huruf Latin bersambung yang berbeda dari penulisnya diterjemahkan ke dalam komputer dan dilakukan proses pengenalan dari huruf Latin bersambung tersebut. Belum banyak penelitian pengenalan tulisan tangan dengan objek huruf Latin terutama huruf Latin yang bersifat bersambung. Beberapa penelitian yang telah dilakukan terkait tulisan tangan dapat disimpulkan belum didapatkannya penelitian yang bersifat secara spesifik membahas mengenai tulisan tangan bersambung baik kapital dan non kapital. Penelitian yang dilakukan bersifat eksperimen menggunakan *local binary pattern* sebagai ektraksi ciri yang di mana proses untuk mendapatkan ciri unik dari pola citra yang dijadikan data input pada pengenalan citra dengan menggunakan K-Nearest Neighbor sebagai pengenalan huruf Latin bersambung kapital dan non kapital.

2. METODE PENELITIAN

Tulisan tangan adalah salah satu hal unik yang dapat dihasilkan oleh manusia selain tanda tangan (Andana et al., 2018). Beberapa penelitian menunjukkan bahwa perangkat lunak dapat mengenali tulisan tangan setiap orang. *Handwriting Recognition* atau pengenalan tulisan tangan adalah suatu proses di mana komputer menerjemahkan tulisan tangan kedalam teks komputer

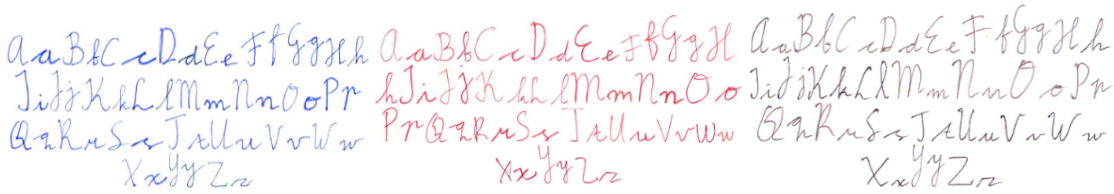
(Cahyani et al., 2018). Huruf Latin adalah huruf yang biasa digunakan dalam tulisan tangan yang ditulis berangkai-rangkai. Huruf Latin bersambung merupakan suatu bentuk atau gores dari suatu sistem tulisan di mana tulisan ditulis secara terhubung (Aranta et al., 2020). Pengenalan tulisan tangan bersambung pada penelitian ini memiliki beberapa proses seperti yang ditunjukkan pada Gambar 1.



Gambar 1 Gambaran Umum

2.1 *Dataset*

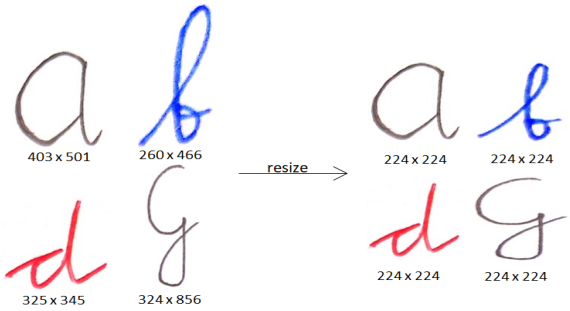
Dataset diambil dari 30 orang responden yang di mana tiap 10 responden akan menulis menggunakan pena berwarna hitam, biru, dan merah. Setiap responden akan menulis 52 huruf (26 huruf kapital dan 26 huruf non-kapital) sehingga total keseluruhan data adalah 1560 foto (Gambar 2). Proses pengumpulan dataset tulisan tangan berupa foto *scan* dan dibagi menjadi 2 kelompok yaitu 780 huruf Latin bersambung kapital dan 780 huruf Latin bersambung non-kapital. *Dataset* tersebut akan menjadi *dataset* latih dan *dataset* uji.



Gambar 2 *Dataset* Tiap Jenis Huruf Latin Bersambung

2.2 *Preprocessing*

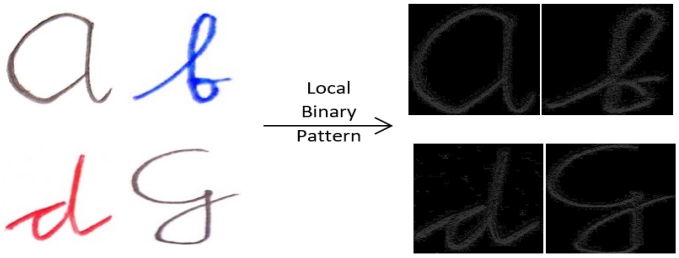
Pada tahap *preprocessing* (pengolah awal) bertujuan untuk medapatkan informasi dari citra serta meningkatkan kualitas citra yang di mana pada tahap ini meliputi *resize* dan *grayscale* dari LBP. Citra awal diberi label secara manual sesuai klasifikasi yang berukuran bervariasi, lalu semua data akan melewati proses *resize* sehingga semua ukuran citra berukuran sama yaitu 224×224 piksel. Proses *resize* dapat diilustrasikan pada Gambar 3.



Gambar 3 *Dataset Resize*

2.3 Ekstraksi Fitur

Untuk ekstraksi fitur pada penelitian ini menggunakan ektraksi ciri tekstur *Local Binary Pattern.* Local Binary Pattern merupakan metode yang dapat menghasilkan tekstur dari sebuah citra *grayscale* yang bekerja dengan memanfaatkan nilai piksel ketetanggaan yang tersebar melingkar (*circular neighborhood*) dengan berbagai jenis ukuran yang direpresentasikan dalam bentuk *matrix* (Al Rivan et al., 2020). Setelah d*ataset* di*resize,* kemudian dilakukan ekstraksi ciri *Local Binary Pattern*. Proses LBP dapat diilustrasikan pada Gambar 4.



Gambar 4 Local Binary Pattern

Dengan menerapkan operator dasar LBP yang berukuran 3×3, diambil nilai tengah sebagai ambang batas (*threshold*) dan dibandingkan dengan nilai sekelilingnya. Jika piksel sekeliling lebih besar dibandingkan dengan ambang batas maka akan bernilai 1, dan jika nilai piksel sekeliling lebih kecil dibandingkan dengan ambang batas maka akan bernilai 0. Setelah didapat nilai biner, maka akan dikali dengan nilai bobot 2^p yang adalah urutan piksel tetangga dimulai dari pojok kiri atas hingga pojok kanan bawah dengan mengelilingi piksel tengah yang dimulai dari 0 sampai 7. Hasil dari citra *grayscale* yang dilakukan ekstraksi ciri yang dapat dilihat pada Gambar 5 dan Gambar 6.

	0	1	2	3	4	5	6	7	8	9	...	49	50	51	52	53	54	55	56	57	58
0	24769.0	618.0	527.0	1654.0	799.0	48.0	1388.0	881.0	45.0	36.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11008.0
1	34526.0	353.0	340.0	1007.0	333.0	28.0	406.0	298.0	38.0	32.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8728.0
2	14705.0	1185.0	770.0	2795.0	1455.0	93.0	1660.0	1438.0	117.0	94.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17107.0
3	27493.0	630.0	501.0	1668.0	746.0	50.0	1357.0	677.0	57.0	37.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10509.0
4	10838.0	1327.0	790.0	2775.0	1558.0	134.0	1408.0	1598.0	169.0	121.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20868.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
775	39530.0	57.0	51.0	169.0	92.0	30.0	723.0	81.0	31.0	36.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6130.0
776	41644.0	72.0	56.0	188.0	64.0	31.0	600.0	73.0	31.0	29.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4110.0
777	38628.0	80.0	61.0	259.0	128.0	44.0	445.0	129.0	33.0	46.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7216.0
778	29336.0	240.0	238.0	680.0	423.0	76.0	845.0	604.0	92.0	83.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11197.0
779	36091.0	160.0	110.0	343.0	251.0	36.0	587.0	215.0	51.0	63.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8234.0

780 rows × 59 columns

Gambar 5 Hasil Ekstraksi Ciri LBP Huruf Kapital Latin Bersambung

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022: 211 – 225

	0	1	2	3	4	5	6	7	8	9	...	49	50	51	52	53	54	55	56	57	58
0	34316.0	248.0	165.0	608.0	249.0	18.0	515.0	227.0	21.0	27.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10034.0
1	29005.0	497.0	472.0	1278.0	569.0	43.0	1449.0	578.0	43.0	24.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9487.0
2	32432.0	288.0	191.0	396.0	199.0	10.0	466.0	303.0	30.0	30.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12258.0
3	30993.0	400.0	192.0	398.0	256.0	10.0	454.0	223.0	16.0	41.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13070.0
4	23782.0	851.0	634.0	2002.0	854.0	70.0	1209.0	833.0	91.0	70.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13193.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
775	37892.0	144.0	75.0	204.0	118.0	22.0	830.0	156.0	17.0	27.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6822.0
776	37652.0	134.0	132.0	307.0	193.0	46.0	670.0	167.0	33.0	47.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7084.0
777	39225.0	64.0	76.0	100.0	114.0	34.0	602.0	96.0	18.0	34.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6725.0
778	38962.0	111.0	94.0	180.0	123.0	29.0	628.0	122.0	31.0	33.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6293.0
779	37655.0	135.0	94.0	312.0	202.0	28.0	677.0	155.0	51.0	36.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6908.0

780 rows × 59 columns

Gambar 6 Hasil Ekstraksi Ciri LBP Huruf Non-Kapital Latin Bersambung

Dataset citra yang sudah dilakukan ekstraksi fitur akan dinormalisasikan terlebih dahulu sebelum dilanjutkan ke tahap pengenalan huruf. Gambar 7 dan Gambar 8 merupakan hasil normalisasi terhadap ekstraksi fitur LBP huruf kapital dan huruf non-kapital Latin bersambung.

	0	1	2	3	4	5	6	7	8	9	...	50	51	52	53	54	55	56	57	58	target
0	0.901943	0.022504	0.019190	0.060229	0.029095	0.001748	0.050543	0.032081	0.001639	0.001311	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.400847	9
1	0.968369	0.009901	0.009536	0.028244	0.009340	0.000785	0.011387	0.008358	0.001066	0.000898	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.244799	23
2	0.635821	0.051238	0.033294	0.120851	0.062912	0.004021	0.071776	0.062177	0.005059	0.004064	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.739679	19
3	0.928217	0.021270	0.016915	0.056315	0.025186	0.001688	0.045815	0.022857	0.001924	0.001249	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.354804	25
4	0.450885	0.055206	0.032866	0.115446	0.064816	0.005575	0.058576	0.066480	0.007031	0.005034	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.868156	24
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
775	0.987657	0.001424	0.001274	0.004222	0.002299	0.000750	0.018064	0.002024	0.000775	0.000899	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.153158	6
776	0.994748	0.001720	0.001338	0.004491	0.001529	0.000740	0.014332	0.001744	0.000740	0.000693	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.098175	5
777	0.982663	0.002035	0.001552	0.006589	0.003256	0.001119	0.011320	0.003282	0.000839	0.001170	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.183569	17
778	0.931225	0.007618	0.007555	0.021586	0.013427	0.002413	0.026823	0.019173	0.002920	0.002635	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.355431	1
779	0.974253	0.004319	0.002969	0.009259	0.006776	0.000972	0.015846	0.005804	0.001377	0.001701	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.222271	17

780 rows × 60 columns

Gambar 7 Hasil Ekstraksi LBP Huruf Kapital yang Telah Dinormalisasi

	0	1	2	3	4	5	6	7	8	9	...	50	51	52	53	54	55	56	57	58	target
0	0.959018	0.006931	0.004611	0.016992	0.006959	0.000503	0.014393	0.006344	0.000587	0.000755	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.280417	19
1	0.944450	0.016183	0.015369	0.041614	0.018528	0.001400	0.047182	0.018821	0.001400	0.000781	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.308912	9
2	0.934855	0.008302	0.005506	0.011415	0.005736	0.000288	0.013432	0.008734	0.000865	0.000865	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.353338	23
3	0.920691	0.011883	0.005704	0.011823	0.007605	0.000297	0.013487	0.006625	0.000475	0.001218	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.388263	22
4	0.867197	0.031031	0.023118	0.073002	0.031141	0.002553	0.044086	0.030375	0.003318	0.002553	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.481075	24
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
775	0.983452	0.003737	0.001947	0.005295	0.003063	0.000571	0.021542	0.004049	0.000441	0.000701	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.177059	16
776	0.982126	0.003495	0.003443	0.008008	0.005034	0.001200	0.017476	0.004356	0.000861	0.001226	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.184781	6
777	0.985206	0.001607	0.001909	0.002512	0.002863	0.000854	0.015120	0.002411	0.000452	0.000854	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.168910	6
778	0.986669	0.002811	0.002380	0.004558	0.003115	0.000734	0.015903	0.003090	0.000785	0.000836	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.159363	6
779	0.982850	0.003524	0.002454	0.008144	0.005272	0.000731	0.017671	0.004046	0.001331	0.000940	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.180309	1

780 rows × 60 columns

Gambar 8 Hasil Ekstraksi LBP Huruf Non-Kapital yang Telah Dinormalisasi

2.4 Tahap Pengenalan

Algoritma K-Nearest Neighbor adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Hidayat & Minati, 2019). Algoritma klasifikasi K-Nearest Neighbor termasuk algoritma *supervised learnin*g, di mana hasil dari query instance yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN (Nikmatun & Waspada, 2019). Metode KNN (K-Nearest Neighbor) melakukan pengenalan dengan objek huruf Latin bersambung untuk mengklasifikasi 26 jenis huruf Latin dengan masing-masing huruf dibagi menjadi 2 tipe yaitu huruf kapital dan huruf non-kapital. Pengujian menggunakan K-Nearest Neighbor dengan nilai k = 3, 5, dan 7.

Pada pendekatan metode *k-fold cross validation*, *dataset* dibagi menjadi sejumlah k buah partisi secara acak. Selanjutnya, dilakukan sejumlah k-kali eksperimen dengan masing-masing eksperimen menggunakan data partisi ke-k sebagai data *testing* dan menggunakan sisa partisi lainnya sebagai data *training* (Pangestu et al., 2020). Data *training* dan data *testing* dibagi menggunakan *k-fold cross validation* dengan k = 6. Hasil dari pengujian akan dihitung menggunakan *confusion matrix* yang di mana *matrix* berukuran 26×26 dikarenakan data kelas berjumlah 26 untuk mendapatkan nilai *precision, recall, F1-score* dan *accuracy*.

Confusion matrix digunakan untuk melihat performa dari suatu model yang telah dibuat yaitu *accuracy*, *precision*, *recall*, dan *F1-score* (Maulidah et al., 2020).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = \frac{2\times Recall\times Precision}{Recall+Precision} \quad (4)$$

Di mana *True Positive (TP*) merupakan jumlah data positif citra yang terklasifikasi benar oleh sistem, *True Negative (TN)* merupakan jumlah data negatif citra yang terklasifikasi benar oleh sistem, *False Positive (FP)* merupakan jumlah data positif citra yang terklasifikasi salah oleh sistem, dan *False Negative (FN)* merupakan jumlah data negatif citra yang terklasifikasi salah oleh sistem.

3. HASIL DAN PEMBAHASAN

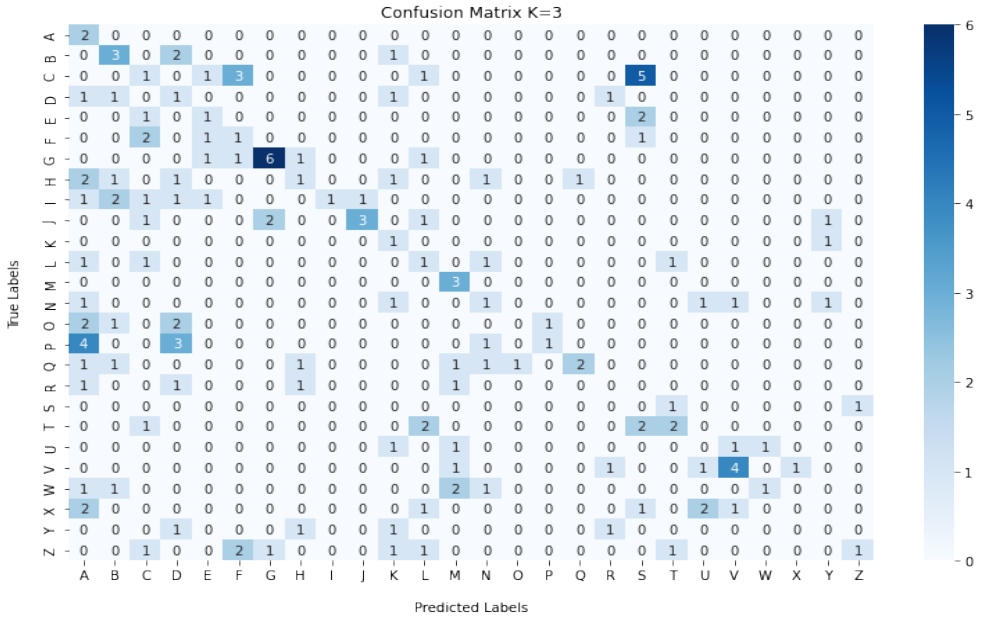
3.1 Pembahasan Pengujian KNN Huruf Kapital

Proses pengujian huruf kapital Latin bersambung menggunakan K-Nearest Neighbor di mana pada penelitian ini nilai k yang digunakan adalah 3, 5, dan 7. Hasil pengujian pengenalan huruf kapital k=3, k=5, dan k=7 dapat dilihat pada Tabel 1 sampai 3.

Gambar 9 dirincikan hasil prediksi huruf G yang dapat dilihat melalui *confusion matrix*. Kelas G dapat dikenali sebanyak 6 data yang merupakan nilai *True Positive*. Adapula data bernilai G yang diprediksi bukan G yang merupakan *False Negative* sebanyak 4 dan data yang terprediksi G yang bukan data G yang merupakan *False Positive* sebanyak 3 dengan sisa nilai lain sebagai *True* Negative. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 67%, *recall* 60%, *F1-score* 63%, dan *accuracy* 95.51%.

Tabel 1 Hasil Pengujian Pengenalan Huruf Kapital K=3

Huruf	*Precision*	*Recall*	*F1-Score*	*Accuracy*
A	11%	100%	19%	89.1%
B	30%	50%	37%	93.59%
C	11%	9%	10%	88.46%
D	8%	20%	12%	90.38%
E	20%	25%	22%	95.51%
F	14%	20%	17%	93.59%
G	67%	60%	63%	95.51%
H	20%	12%	15%	92.95%
I	100%	12%	22%	95.51%
J	75%	38%	50%	96.15%
K	12%	50%	20%	94.87%
L	12%	20%	15%	92.95%
M	33%	100%	50%	96.15%
N	17%	17%	17%	93.59%
O	0%	0%	0%	95.51%
P	50%	11%	18%	94.23%
Q	67%	25%	36%	95.51%
R	0%	0%	0%	95.51%
S	0%	0%	0%	91.67%
T	40%	29%	33%	94.87%
U	0%	0%	0%	96.15%
V	57%	50%	53%	95.51%
W	50%	17%	25%	96.15%
X	0%	0%	0%	94.87%
Y	0%	0%	0%	95.51%
Z	50%	12%	20%	94.87%
Rata-Rata per Kelas	**28.61%**	**26.04%**	**21.31%**	**94.18%**



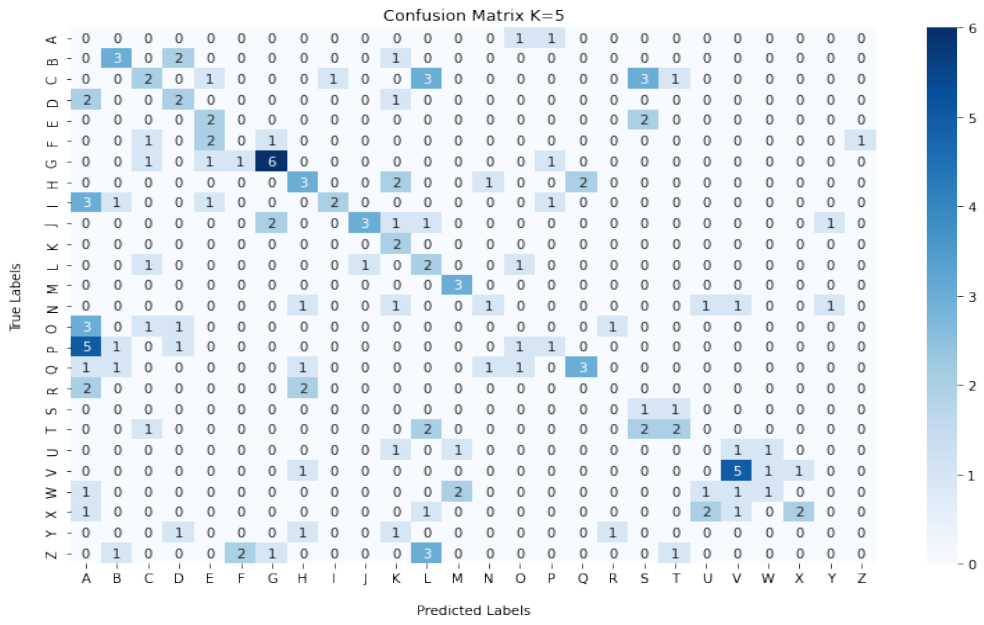
Gambar 9 *Confusion Matrix* Huruf Kapital K=3

Gambar 10 dirincikan hasil prediksi huruf G yang dapat dilihat melalui *confusion matrix*. Kelas G dapat dikenali sebanyak 6 data yang merupakan nilai *True Positive*. Adapula data bernilai G yang diprediksi bukan G yang merupakan *False Negative* sebanyak 4 dan data yang terprediksi G yang

bukan data G yang merupakan *False Positive* sebanyak 4 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision*. *recall* dan *F1-score* masing-masing 60% dan *accuracy* 94.87%.

Tabel 2 Hasil Pengujian Pengenalan Huruf Kapital K=5

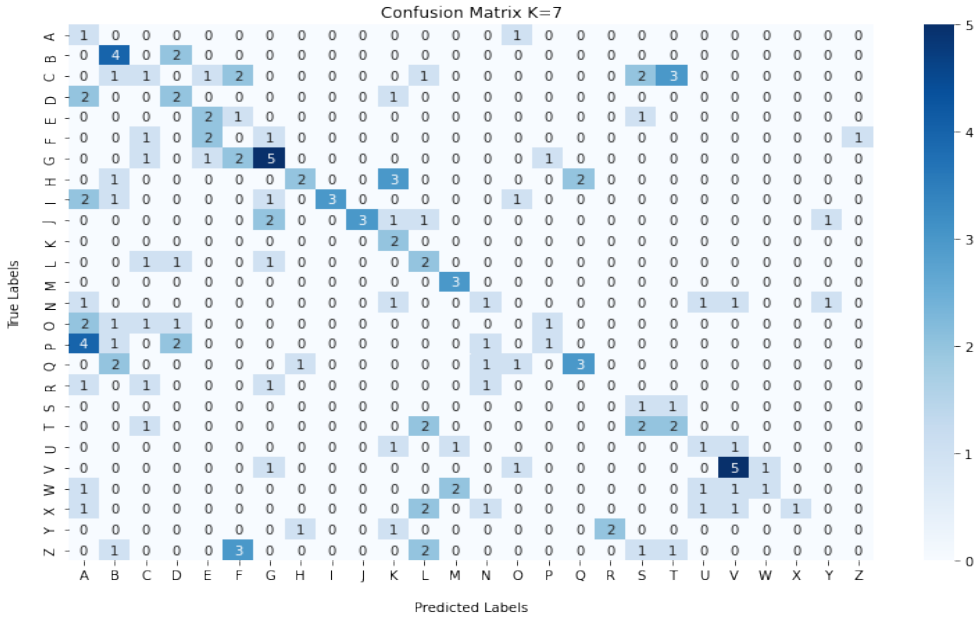
Huruf	*Precision*	*Recall*	*F1-Score*	*Accuracy*
A	0%	0%	0%	87.17%
B	43%	50%	46%	95.51%
C	29%	18%	22%	91.03%
D	29%	40%	33%	94.87%
E	29%	50%	36%	95.51%
F	0%	0%	0%	95.51%
G	60%	60%	60%	94.87%
H	33%	38%	35%	92.95%
I	67%	25%	36%	95.51%
J	75%	38%	50%	96.15%
K	20%	100%	33%	94.87%
L	17%	40%	24%	91.67%
M	50%	100%	67%	98.08%
N	33%	17%	22%	95.51%
O	0%	0%	0%	93.59%
P	25%	11%	15%	92.95%
Q	60%	38%	46%	95.51%
R	0%	0%	0%	96.15%
S	12%	50%	20%	94.87%
T	40%	29%	33%	94.87%
U	0%	0%	0%	94.87%
V	56%	62%	59%	95.51%
W	33%	17%	22%	95.51%
X	67%	29%	40%	96.15%
Y	0%	0%	0%	96.15%
Z	0%	0%	0%	94.23%
Rata-Rata per Kelas	**29.92%**	**31.23%**	**26.88%**	**94.59%**



Gambar 10 *Confusion Matrix* Huruf Kapital K=5

Tabel 3 Hasil Pengujian Pengenalan Huruf Kapital K=7

Huruf	*Precision*	*Recall*	*F1-Score*	*Accuracy*
A	7%	50%	12%	90.38%
B	33%	67%	44%	93.59%
C	14%	9%	11%	89.74%
D	25%	40%	31%	94.23%
E	33%	50%	40%	96.15%
F	0%	0%	0%	91.67%
G	42%	50%	45%	92.31%
H	50%	25%	33%	94.87%
I	100%	38%	55%	96.79%
J	100%	38%	55%	96.79%
K	20%	100%	33%	94.87%
L	20%	40%	27%	92.95%
M	50%	100%	67%	98.08%
N	20%	17%	18%	94.23%
O	0%	0%	0%	93.59%
P	33%	11%	17%	93.59%
Q	60%	38%	46%	95.51%
R	0%	0%	0%	96.15%
S	14%	50%	22%	95.51%
T	29%	29%	29%	93.59%
U	25%	25%	25%	96.15%
V	56%	62%	59%	95.51%
W	50%	17%	25%	96.15%
X	100%	14%	25%	96.15%
Y	0%	0%	0%	96.15%
Z	0%	0%	0%	94.23%
Rata-Rata per Kelas	**33.88%**	**33.46%**	**27.65%**	**94.57%**



Gambar 11 *Confusion Matrix* Huruf Kapital K=7

Gambar 11 dirincikan hasil prediksi huruf G dan V yang dapat dilihat melalui *confusion matrix*. Kelas G dapat dikenali sebanyak 5 data yang merupakan nilai *True Positive*. Adapula data bernilai G yang diprediksi bukan G yang merupakan *False Negative* sebanyak 4 dan data yang terprediksi

G yang bukan data G yang merupakan *False Positive* sebanyak 7 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 42%, *recall* 50%, *F1-score* 45%, dan *accuracy* 92.31%.

Lalu kelas V dapat dikenali sebanyak 5 data yang merupakan nilai *True Positive*. Adapula data bernilai V yang diprediksi bukan V yang merupakan *False Negative* sebanyak 3 dan data yang terprediksi V yang bukan data V yang merupakan *False Positive* sebanyak 4 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 56%, *recall* 62%, *F1-score* 59%, dan *accuracy* 95.51%.

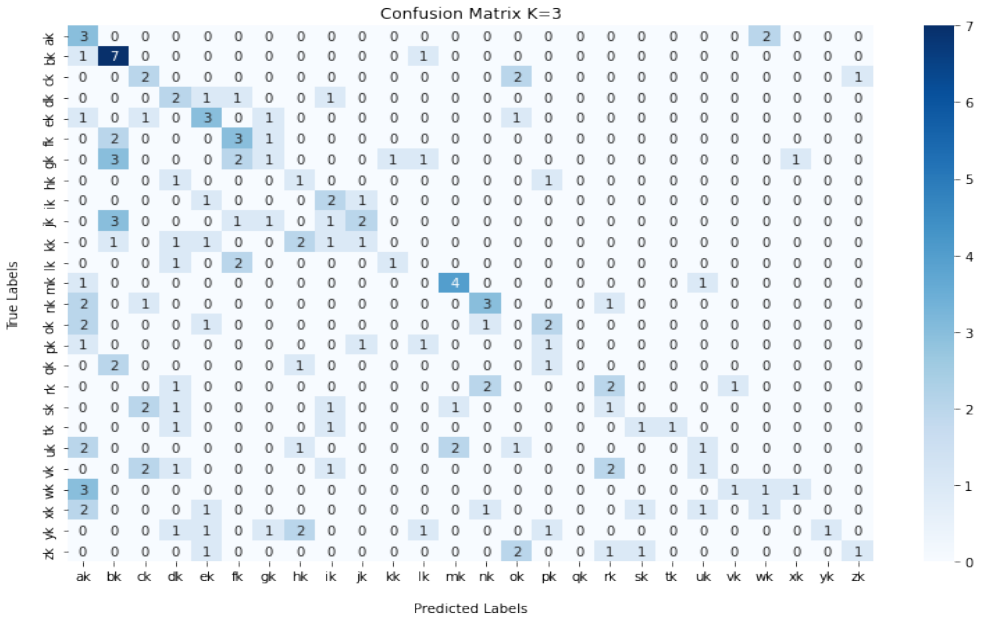
3.2 Pembahasan Pengujian KNN Non Huruf Kapital

Proses pengujian pengenalan huruf non-kapital sama seperti proses pengenalan skenario pertama di mana nilai k yang digunakan adalah 3, 5, dan 7. Hasil pengenalan huruf non-kapital k=3, k=5, dan k=7 dapat dilihat pada Tabel 4 sampai 6.

Tabel 4 Hasil Pengujian Pengenalan Non Kapital K=3

Huruf	*Precision*	*Recall*	*F1-Score*	*Accuracy*
a	17%	60%	26%	89.1%
b	39%	78%	52%	91.67%
c	25%	40%	31%	94.23%
d	20%	40%	27%	92.95%
e	30%	43%	35%	92.95%
f	33%	50%	40%	94.23%
g	20%	11%	14%	92.31%
h	14%	33%	20%	94.87%
i	25%	50%	33%	94.87%
j	40%	25%	31%	94.23%
k	0%	0%	0%	94.23%
l	0%	0%	0%	94.87%
m	57%	67%	62%	96.79%
n	43%	43%	43%	94.87%
o	0%	0%	0%	92.31%
p	17%	25%	20%	94.87%
q	0%	0%	0%	97.43%
r	29%	33%	31%	94.23%
s	0%	0%	0%	94.23%
t	100%	25%	40%	98.08%
u	25%	14%	18%	94.23%
v	0%	0%	0%	94.23%
w	25%	17%	20%	94.87%
x	0%	0%	0%	94.23%
y	100%	12%	22%	95.51%
z	50%	17%	25%	96.15%
Rata-Rata per Kelas	**27.27%**	**26.27%**	**22.7%**	**94.33%**

Gambar 12 dirincikan hasil prediksi huruf b yang dapat dilihat melalui *confusion matrix*. Kelas b dapat dikenali sebanyak 7 data yang merupakan nilai *True Positive*. Adapula data bernilai b yang diprediksi bukan b yang merupakan *False Negative* sebanyak 2 dan data yang terprediksi b yang bukan data b yang merupakan *False Positive* sebanyak 11 dengan sisa nilai lain sebagai *True Negative*. *Dengan* nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 39%, *recall* 78%, *F1-score* 52%, dan *accuracy* 91.67%.



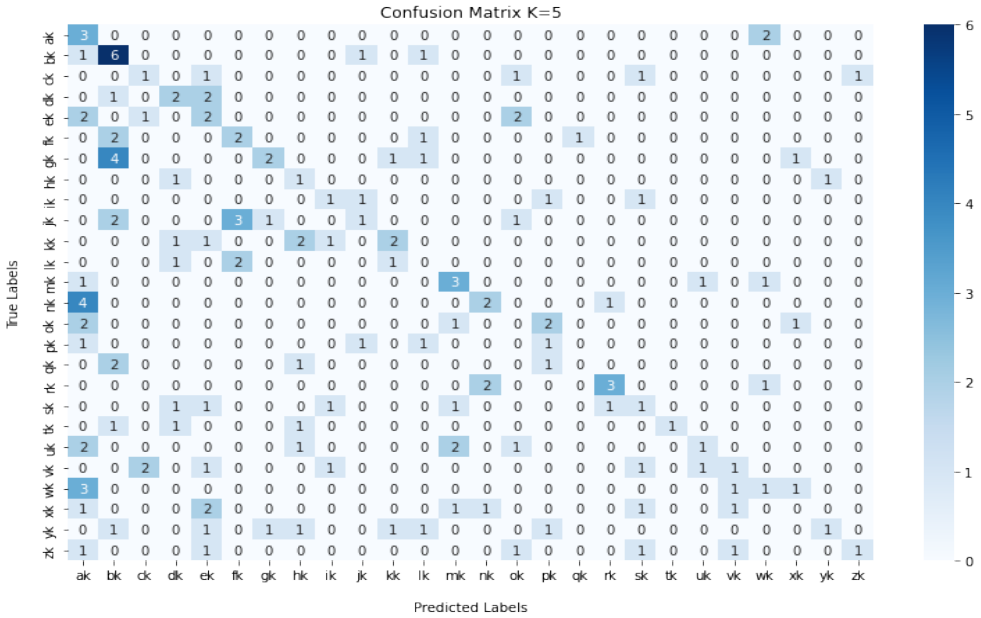
Gambar 12 *Confusion Matrix* Huruf Kapital K=3

Tabel 5 Hasil Pengujian Pengenalan Non Kapital K=5

Huruf	*Precision*	*Recall*	*F1-Score*	*Accuracy*
a	14%	60%	23%	87.18%
b	32%	67%	43%	89.74%
c	25%	20%	22%	95.51%
d	29%	40%	33%	94.87%
e	17%	29%	21%	90.38%
f	29%	33%	31%	94.23%
g	50%	22%	31%	94.23%
h	14%	33%	20%	94.87%
i	25%	25%	25%	96.15%
j	25%	12%	17%	93.59%
k	40%	29%	33%	94.87%
l	0%	0%	0%	94.23%
m	38%	50%	43%	94.87%
n	40%	29%	33%	94.87%
o	0%	0%	0%	92.31%
p	17%	25%	20%	94.87%
q	0%	0%	0%	96.79%
r	60%	50%	55%	96.79%
s	17%	17%	17%	93.59%
t	100%	25%	40%	98.08%
u	33%	14%	20%	94.87%
v	25%	14%	18%	94.23%
w	20%	17%	18%	94.23%
x	0%	0%	0%	93.59%
y	50%	12%	20%	94.87%
z	50%	17%	25%	96.15%
Rata-Rata per Kelas	**28.85%**	**24.61%**	**23.38%**	**94.23%**

Gambar 13 dirincikan hasil prediksi huruf b yang dapat dilihat melalui *confusion matrix*. Kelas b dapat dikenali sebanyak 6 data yang merupakan nilai *True Positive*. Adapula data bernilai b yang diprediksi bukan b yang merupakan *False Negative* sebanyak 3 dan data yang terprediksi b yang

bukan data b yang merupakan *False Positive* sebanyak 13 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 32%, *recall* 67%, *F1-score* 43%, dan *accuracy* 89.74%.

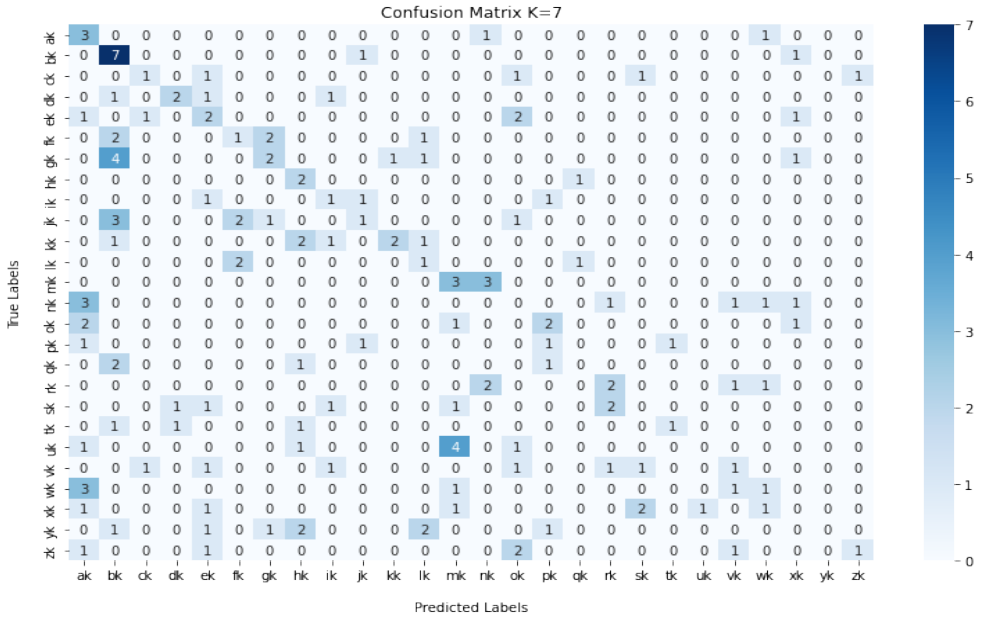


Gambar 13 *Confusion Matrix* Huruf Non Kapital K=5

Tabel 6 Hasil Pengujian Pengenalan Non Kapital K=7

Huruf	*Precision*	*Recall*	*F1-Score*	*Accuracy*
a	19%	60%	29%	90.38%
b	32%	78%	45%	89.1%
c	33%	20%	25%	96.15%
d	50%	40%	44%	96.79%
e	20%	29%	24%	91.67%
f	20%	17%	18%	94.23%
g	33%	22%	27%	92.95%
h	22%	67%	33%	94.87%
i	20%	25%	22%	95.51%
j	25%	12%	17%	93.59%
k	67%	29%	40%	96.15%
l	17%	25%	20%	94.87%
m	27%	50%	35%	92.95%
n	0%	0%	0%	91.67%
o	0%	0%	0%	91.02%
p	17%	25%	20%	94.87%
q	0%	0%	0%	96.15%
r	33%	33%	33%	94.87%
s	0%	0%	0%	93.59%
t	50%	25%	33%	97.43%
u	0%	0%	0%	94.87%
v	20%	14%	17%	93.59%
w	20%	17%	18%	94.23%
x	0%	0%	0%	92.31%
y	0%	0%	0%	94.87%
z	50%	17%	25%	96.15%
Rata-Rata per Kelas	**22.12%**	**23.27%**	**20.19%**	**94.03%**

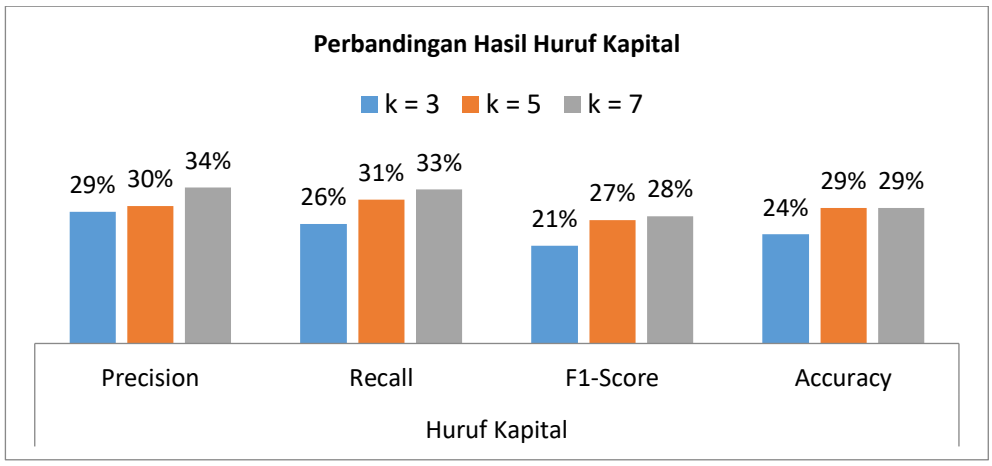
Gambar 14 dirincikan hasil prediksi huruf b yang dapat dilihat melalui *confusion matrix*. Kelas b dapat dikenali sebanyak 7 data yang merupakan nilai *True Positive*. Adapula data bernilai b yang diprediksi bukan b yang merupakan *False Negative* sebanyak 2 dan data yang terprediksi b yang bukan data b yang merupakan *False Positive* sebanyak 15 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 32%, *recall* 78%, *F1-score* 45%, dan *accuracy* 89.1%.



Gambar 14 *Confusion Matrix* Huruf Non Kapital K=7

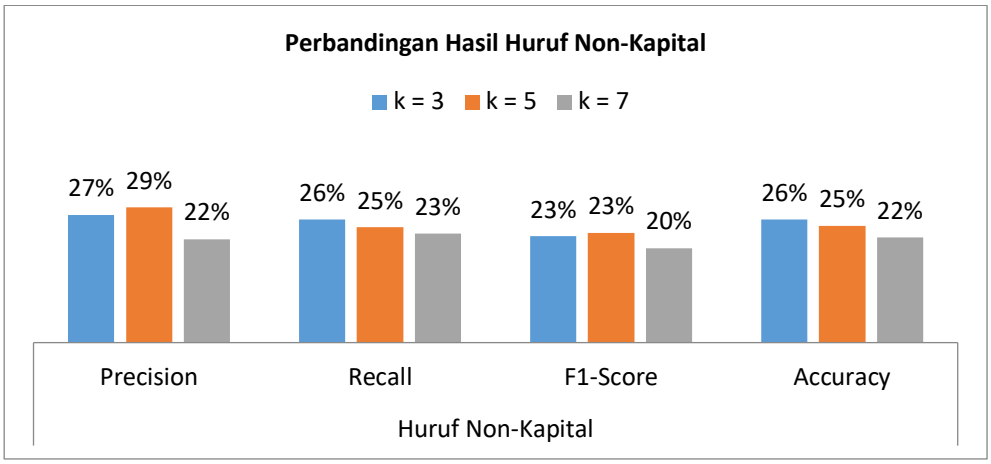
3.3 Perbandingan Huruf Kapital dan Non Kapital

Gambar 15 pada bagian huruf kapital dapat disimpulkan bahwa nilai pengujian *accuracy* tertinggi pada huruf kapital yaitu pada k = 5 dan k = 7 dengan hasil 29,49%, sedangkan pada k = 3 hanya menghasilkan 23.72%. Selanjutnya nilai rata-rata*precision*, *recall*, dan *F1-score* tertinggi yaitu pada k = 7 dengan masing-masing 33.88%, 33.46%, dan 27.65%, sedangkan Precision pada k = 5 mendapatkan hasil 29.92% dan k = 3 menghasilkan 28.61%, *recall* pada k = 5 mendapatkan hasil 31.23% dan k = 3 menghasilkan 26.04%, serta *F1-score* pada k = 5 mendapatkan hasil 26.88% dan k = 3 menghasilkan 21.31%.



Gambar 15 Grafik Perbandingan Hasil Huruf Kapital

Sementara Gambar 16 bagian huruf non kapital dapat disimpulkan bahwa nilai pengujian *accuracy* tertinggi pada huruf kapital yaitu pada k = 3 dengan hasil 26.28% sedangkan k = 5 dengan hasil 25% dan k = 7 dengan hasil 22.44%. Selanjutnya nilai rata-rata *precision* tertinggi yaitu pada k = 5 dengan hasil 28.85% sedangkan k = 3 dengan nilai 27.27% dan k = 7 dengan nilai 22.12%. Berikutnya pada nilai rata-rata *recall* tertinggi yaitu pada k = 3 dengan hasil 26.27% sedangkan k = 5 dengan nilai 24.61% dan k = 7 dengan nilai 23.27%. Dan nilai rata-rata *F1-score* tertinggi yaitu pada k = 5 dengan hasil 23.38% sedangkan k = 3 dengan nilai 22.7% dan k = 7 dengan nilai 20%.



Gambar 16 Grafik Perbandingan Hasil Huruf Non Kapital

4. KESIMPULAN

Berdasarkan pengujian pengenalan tulisan tangan huruf Latin bersambung menggunakan metode KNN dengan ekstraksi ciri Local Binary Pattern yang telah dilakukan mendapatkan kesimpulan berdasarkan hasil performa untuk k=3, k=5, dan k=7. Berdasarkan grafik perbandingan didapatkan hasil terbaik menggunakan nilai k=7 dengan tingkat *accuracy* 29% untuk huruf kapital dan nilai k=3 dengan tingkat *accuracy* 26%. Kecilnya *accuracy* yang didapatkan dikarenakan *dataset* yang digunakan masih cukup sedikit dan beberapa metode yang digunakan bersifat eksperimen sehingga dirasakan tidak pas saat melakukan proses-proses pada penelitian ini. Penelitan yang telah dilakukan masih sangat bisa dikembangkan kembali dengan penambahan *dataset* yang lebih banyak serta menggunakan metode lain dalam proses ektraksi ciri ataupun pengenalan sehingga tingkat *accuracy* yang didapatkan juga semakin tinggi.

DAFTAR PUSTAKA

Al Rivan, M. E., Devella, S., & Saputra, J. (2020). Pengenalan Iris Dengan Normalisasi Menggunakan LBP dan RBF. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, *6*(2), 122. https://doi.org/10.24014/coreit.v6i2.9685

Andana, A., Widyati, R., & Irzal, M. (2018). Pengenalan Citra Tulisan Tangan Dengan Metode Backpropagation. *Jurnal Matematika Terapan*, *2*(1), 36–44.

Anggraeny, F. T., Munir, M. S., & Purbasari, I. Y. (2020). Histogram Profil Proyeksi sebagai Metode Ekstraksi Fitur pada Pengenalan Karakter Tulisan Tangan. *Prosiding Seminar Nasional Informatika Bela Negara*, *1*, 164–168. https://doi.org/10.33005/santika.v1i0.44

Aranta, A., Bimantoro, F., & Putrawan, I. P. T. (2020). Penerapan Algoritma Rule Base dengan Pendekatan Hexadesimal pada Transliterasi Aksara Bima Menjadi Huruf Latin. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTIKA)*, *2*(1), 130–141. https://doi.org/10.29303/jtika.v2i1.96

Bimantoro, F., Aranta, A., Nugraha, G. S., Dwiyansaputra, R., & Husodo, A. Y. (2021). Pengenalan Pola Tulisan Tangan Aksara Bima menggunakan Ciri Tekstur dan KNN. *Journal of Computer Science and Informatics Engineering (J-Cosine)*, *5*(1), 60–67.

https://doi.org/10.29303/jcosine.v5i1.387

Cahyani, S., Wiryasaputra, R., & Gustriansyah, R. (2018). Identifikasi Huruf Kapital Tulisan Tangan Menggunakan Linear Discriminant Analysis dan Euclidean Distance. *Jurnal Sistem Informasi Bisnis*, *8*(1), 57. https://doi.org/10.21456/vol8iss1pp57-67

Hidayat, R., & Minati, S. (2019). Comparative Analysis of Text Mining Classification Algorithms for English and Indonesian Qur'an Translation. *IJID (International Journal on Informatics for Development)*, *8*(1), 47. https://doi.org/10.14421/ijid.2019.08108

Ilham, F., & Rochmawati, N. (2020). Transliterasi Aksara Jawa Tulisan Tangan ke Tulisan Latin Menggunakan CNN. *Journal of Informatics and Computer Science (JINACS)*, *1*(04), 200–208. https://doi.org/10.26740/jinacs.v1n04.p200-208

Maharani, D., Efendi, R., & Johar, A. (2019). Penerapan Augmented Reality Sebagai Media Pembelajaran Pengenalan Aksara Korea (Hangul). *Jurnal Rekursif*, *7*(1), 77–90. https://doi.org/10.33369/rekursif.v7i1.6320

Masrani, H., Ruslianto, I., & Ilhamsyah. (2018). Aplikasi Pengenalan Pola Pada Huruf Tulisan Tangan Menggunakan Jaringan Saraf Tiruan Dengan Metode Ekstraksi Fitur Geometri. *Jurnal Coding, Sistem Komputer Untan*, *06*(02), 69–78. https://doi.org/10.26418/coding.v6i2.26674

Maulidah, M., Windu Gata, Rizki Aulianita, & Cucu Ika Agustyaningrum. (2020). ALGORITMA KLASIFIKASI DECISION TREE UNTUK REKOMENDASI BUKU BERDASARKAN KATEGORI BUKU. *E-Bisnis : Jurnal Ilmiah Ekonomi Dan Bisnis*, *13*(2), 89–96. https://doi.org/10.51903/e-bisnis.v13i2.251

Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, *10*(2), 421–432. https://doi.org/10.24176/simet.v10i2.2882

Pangestu, R. A., Rahmat, B., & Anggraeny, F. T. (2020). Implementasi Algoritma Cnn Untuk Klasifikasi Citra Lahan Dan Perhitungan Luas. *Jurnal Informatika Dan Sistem Informasi (JIFoSI)*, *1*(1), 166–174. https://doi.org/10.33005/jifosi.v1i1.5

Prihatiningsih, S., M, N. S., Andriani, F., & Nugraha, N. (2019). Analisa Performa Pengenalan Tulisan Tangan Angka Berdasarkan Jumlah Iterasi Menggunakan Metode Convolutional Neural Network. *Jurnal Ilmiah Teknologi Dan Rekayasa*, *24*(1), 58–66. https://doi.org/10.35760/tr.2019.v24i1.1934

Purbayanti, T. S. (2018). Pengenalan Tulisan Tangan Huruf Latin Dengan Menggunakan Metode K-Nearest Neighbour. *Simki-Techsain*, *02*(02), 3–10.

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022, Pp. 226 – 236
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Sistem Pendukung Keputusan Kesesuaian Lahan Tanaman Padi Menggunakan Metode AHP dan SAW

Siti Retno Wulandari [(1)], Hamdani Hamdani [(2)*], Anindita Septiarini [(3)]
Informatika, Fakultas Teknik, Universitas Mulawarman, Samarinda
e-mail : sitiretno972@student.unmul.ac.id, {hamdani,anindita}@unmul.ac.id.
* Penulis korespondensi.
Artikel ini diajukan 9 April 2022, direvisi 7 Agustus 2022, diterima 8 Agustus 2022, dan dipublikasikan 25 September 2022.

Abstract

Rice is one of the staple foods for most of the world's population. However, in planting rice, there are several things that must be considered, namely the aspect of land suitability. Errors in land determination can cause crop failure which results in reduced rice production. This system was built based on the website using the Analytical Hierarchy Process (AHP) method to calculate the weight of the criteria and the Simple Additive Weighting (SAW) method to determine the selection of land suitability for rice plants. In determining the suitability of paddy fields using 5 criteria, namely soil type, soil pH, rainfall, temperature, and irrigation waters. The purpose of this research is to assist farmers/farmer groups in selecting land. The results of this study indicate that the AHP and SAW methods have been successfully applied in the system with the results of land recommendations in the Sungai Kunjang sub-district with a preference value of 0,989.

Keywords: Information System, Decision Support System, Rice Field, AHP, SAW

Abstrak

Padi merupakan salah satu makanan pokok yang paling penting bagi sebagian dari populasi masyarakat di dunia. Namun, dalam melakukan penanaman tanaman padi ada beberapa hal yang harus diperhatikan di antaranya aspek kesesuaian lahan. Kesalahan dalam penentuan lahan dapat menyebabkan gagal panen yang menyebabkan ketidakseimbangan antara permintaan pangan dengan lahan pertanian yang berdampak pada berkurangnya produksi padi. Sistem ini dibangun berbasis situs web menggunakan metode *Analitycal Hierarchy Process* (AHP) untuk menghitung bobot kriteria dan untuk menghitung rangking alternatif menggunakan metode *Simple Additive Weighting* (SAW) menentukan pemilihan kesesuaian lahan pada tanaman padi. Dalam penentuan kesesuaian lahan tanaman padi menggunakan 5 kriteria yaitu jenis tanah, pH tanah, curah hujan, suhu dan irigasi dan perairan. Tujuan dari penelitian ini untuk mempermudah dan membantu petani/kelompok tani dalam melakukan pemilihan lahan terbaik. Hasil dari penelitian ini menunjukkan bahwa metode AHP dan SAW berhasil diterapkan dalam sistem pendukung keputusan penentuan kesesuaian lahan tanaman padi dengan hasil rekomendasi lahan di Kecamatan Sungai Kunjang dengan nilai preferensi 0,989.

Kata Kunci: Sistem Informasi, Sistem Pendukung Keputusan, Lahan Padi, AHP, SAW

1. PENDAHULUAN

Indonesia merupakan salah satu negara yang memiliki sumber daya alam lahan terluas, di mana negara Indonesia mempunyai kepulauan yang mencakup lebih dari 17.000 pulau yang dihuni oleh sekitar 255 juta penduduk, sehingga Indonesia menjadi negara keempat dengan jumlah populasi yang terbesar di dunia (Ramadani et al., 2020). Di mana padi merupakan salah satu makanan pokok yang paling penting bagi setengah dari populasi masyarakat di dunia dan mempengaruhi beberapa miliar mata pencarian dan ekonomi (Amini et al., 2020). Salah satu daerah yang termasuk penghasil padi adalah Kalimantan Timur. Pada tahun 2020, luas panen padi di Provinsi Kalimantan Timur diperkirakan sebesar 72,25 ribu hektar dengan produksi sebesar 262,86 ribu ton Gabah Kering Giling (GKG). Jika dikonversikan menjadi beras maka, produksi beras di Provinsi Kalimantan Timur pada 2020 diperkirakan mencapai 152,11 ribu ton (Badan Pusat Statistik Provinsi Kalimantan Timur, 2020).

Namun, untuk melakukan penanaman tanaman padi ada beberapa hal yang harus diperhatikan seperti aspek kesesuaian lahan, aspek ekonomi serta operasional. Salah satu faktor utama yang menentukan keberhasilan dalam kegiatan penanaman tanaman padi adalah penentuan kesesuaian lahan (Nurkholis et al., 2020). Adapun faktor penentuan kesesuaian padi yaitu ketersediaan air menjadi perhatian utama petani (Nyamekye et al., 2018), konservasi sumber daya alam (Burra et al., 2021), perubahan iklim dengan dampaknya berupa kekeringan, kelangkaan air, curah hujan yang tidak menentu dan suhu yang tinggi (Nyamekye et al., 2021).

Kesalahan dalam penentuan lahan dapat menyebabkan gagal panen yang menyebabkan ketidakseimbangan antara permintaan pangan dengan lahan pertanian yang berdampak pada berkurangnya produksi padi (Subiyanto et al., 2018). Oleh karena itu, penentuan lahan menjadi permasalahan yang perlu dipertimbangkan oleh petani sebelum melakukan penanaman padi (Sente & Tridamayanti, 2019). Penentuan kesesuaian lahan dapat dikembangkan dengan menerapkan sistem pendukung keputusan. Di mana sistem pendukung keputusan bertujuan untuk menyediakan informasi, memberikan prediksi, dan solusi alternatif (Kamaludin et al., 2021).

Beberapa solusi permasalahan penelitian sebelumnya terkait penentuan kesesuaian lahan juga menggunakan sistem pendukung keputusan. Pada bidang pertanian digunakan untuk menentukan lahan yang sesuai atau tidak sesuai di antaranya pemetaan lahan padi seperti yang dilakukan oleh Firdiansah, (2016) menggunakan metode *Simple Additive Weighting* (SAW) sebagai penjumlahan terbobot dari kinerja setiap alternatif. Hasil dari penelitian tersebut adalah penentuan kelayakan daerah sebagai bahan pertimbangan untuk menentukan daerah yang dapat dijadikan sebagai daerah pertanian. Sistem pendukung keputusan menentukan lahan tanaman cabai yang dilakukan oleh Anwar et al., (2018) menggunakan metode SAW sebagai solusi alternatif pendukung keputusan untuk menentukan lokasi lahan cabai berupa nilai kesesuaian dan nilai pembatasnya.

Sistem pendukung keputusan pemilihan lahan pertanian yang dilakukan oleh Ramadani et al., (Anwar et al., 2018) menggunakan metode M*ulti Objective Optimization on the basis of Ratio Analysis* (MOORA) sebagai sistem multiobjektif yang mengoptimalkan dua atau lebih atribut. Hasil dari penelitian tersebut adalah memperoleh nilai tertinggi yaitu lahan pertanian yang berlokasi di Jalan Sawi Payaroba Binjai Barat dengan perolehan nilai 7 analisis sistem 0,3742. Sementara itu, penelitian yang dilakukan oleh Nurdin et al., (2020) bertujuan untuk menentukan jenis tanah yang sesuai bagi tanaman pangan menggunakan metode *Simple Multi Attribute Rating Technique Exploiting Rank* (SMARTER) dan metode SAW. Kriteria dan perhitungan bobot untuk metode SMARTER dan SAW.

Salah satu metode yang digunakan dalam penelitian ini membuat suatu sistem pendukung keputusan adalah *Analitycal Hierarchy Process* (AHP) dan SAW. Metode AHP memiliki kelebihan dalam melakukan pengambilan keputusan secara hierarki (tingkat) yang dipilih dari berbagai kriteria dan alternatif (Astari & Komarudin, 2018), sedangkan metode SAW memiliki keunggulan di mana dalam proses *ranking* yang simpel atau sederhana (Iqbalgis & Nurochman, 2019). Berdasarkan penjelasan sebelumnya maka penelitian ini mengusulkan topik yang berjudul sistem pendukung keputusan penentuan kesesuaian lahan tanaman padi menggunakan metode AHP dan SAW. Penelitian ini bertujuan untuk mempermudah petani dalam melakukan pemilihan lahan tanaman padi terbaik dengan melakukan proses pembobotan data kriteria dan *ranking* data alternatif.

1.1 Tinjauan Pustaka

Sistem pendukung keputusan pemetaan lahan padi yang dilakukan oleh Firdiansah (2016) menggunakan metode *Simple Additive Weighting* (SAW) sebagai penjumlahan terbobot dari kinerja setiap alternatif. Hasil dari penelitian tersebut adalah penentuan kelayakan daerah sebagai bahan pertimbangan untuk menentukan daerah yang dapat dijadikan sebagai daerah pertanian. Terdapat kriteria yang digunakan yaitu jenis tanah, tekstur tanah, curah hujan, suhu, dan sistem irigasi atau perairan (Wulandari et al., 2016).

Selanjutnya, sistem pendukung keputusan pemberian pinjaman pada Koperasi Warga Lingkungan (KOPWALI) yang dilakukan oleh Wahyu et al., (2020) menggunakan metode AHP dan SAW. Data yang digunakan adalah 6 alternatif dan 10 kriteria. Hasil pengujian tingkat akurasi yang didapat terhadap hasil rekomendasi pemilihan penerima dana menggunakan perhitungan *Spearman Rank Correlation Coefficient* menghasilkan nilai 0,25714.

Sistem pendukung keputusan menentukan lahan tanaman cabai yang dilakukan oleh Anwar et al., menggunakan metode SAW sebagai solusi alternatif pendukung keputusan untuk menentukan lokasi lahan cabai berupa nilai kesesuaian dan nilai pembatasnya. Hasil dari penelitian tersebut diperoleh dapat membantu para petani cabai dalam mendukung keputusan untuk menentukan lahan tanaman cabai, di mana hasilnya berupa nilai pemeringkatan yang paling tinggi yang direkomendasikan (Anwar et al., 2018).

Sistem pendukung keputusan pemilihan lahan pertanian yang dilakukan oleh Ramadani et al., (2020) menggunakan metode MOORA sebagai sistem multiobjektif yang mengoptimalkan dua atau lebih atribut. Hasil dari penelitian tersebut adalah memperoleh nilai tertinggi yaitu lahan pertanian yang berlokasi di Jalan Sawi Payaroba Binjai Barat dengan perolehan nilai analisis sistem 0,3742 dengan kriteria temperatur 31°C, curah hujan sedang, PH tanah 6, tekstur tanah kasar dan jenis tanah lempung (Kusnadi & Jaelani, 2020).

Sistem penentuan lokasi *Automatic Teller Machine* (ATM) yang dilakukan oleh Mahendra & Aryanto, (2019) menggunakan metode AHP untuk pembobotan kriteria dan SAW untuk pemeringkatan alternatif. Terdapat 7 kriteria dengan 11 sub kriteria pada pembobotan dan 76 data alternatif. Hasil pengujian yang ditampilkan dalam *confusion matrix*, pada kriteria yang tidak teruji signifikansi didapatkan 33 data *true positive*, 38 *true negative*, 5 *false negative* dan 5 *false positive* dengan akurasi sebesar 86,84%, dan pada kriteria yang teruji signifikansi didapatkan 35 data *true positive*, 35 *true negative*, 3 *false negative* dan 3 *false positive* memiliki akurasi 92,11%.

Pendukung keputusan evaluasi lahan pertanian yang dilakukan oleh Maglinets et al., (2019). Studi yang dilaporkan didanai oleh *Russian Foundation for Basic Research* (No.18-47-242002), pemerintah wilayah Krasnoyarsk, dana ilmu pengetahuan daerah Krasnoyarsk. Hasil dari penelitian tersebut adalah sistem pendukung keputusan cerdas yang memungkinkan pemecahan masalah pemeringkatan lahan pertanian berdasarkan karakteristik dan *Geographic Information System* (GIS) agronominya data. Sebagai dasar untuk menentukan indikator teknis dan ekonomi.

Sistem pendukung keputusan penilaian kesehatan tanah yang dilakukan oleh Nugroho et al., (2019) menggunakan metode SAW sebagai penjumlahan terbobot dari *rating* kinerja pada setiap alternatif pada semua atribut. Hasil dari penelitian tersebut adalah pengambil keputusan dalam menentukan status kesehatan tanah di suatu daerah setelah dilakukan penilaian kesehatan tanah secara komprehensif melalui observasi di lapangan dan pengujian indikator kesehatan tanah di laboratorium berdasarkan kriteria kesehatan tanah yang telah ditentukan.

Penelitian yang dilakukan oleh Nurdin et al., (2020) bertujuan untuk menentukan jenis tanah yang sesuai bagi tanaman pangan menggunakan metode SMARTER dan metode SAW. Kriteria dan perhitungan bobot untuk metode SMARTER dan SAW adalah kesuburan tanah, unsur hara tanah, kelembaban tanah, tekstur tanah, ketebalan gambut tanah, reaksi tanah, dan drainase tanah. Hasil penelitian penerapan metode SMARTER dan SAW menghasilkan preferensi dengan nilai tertinggi 0,824286 pada jenis tanah andosol untuk tanaman padi.

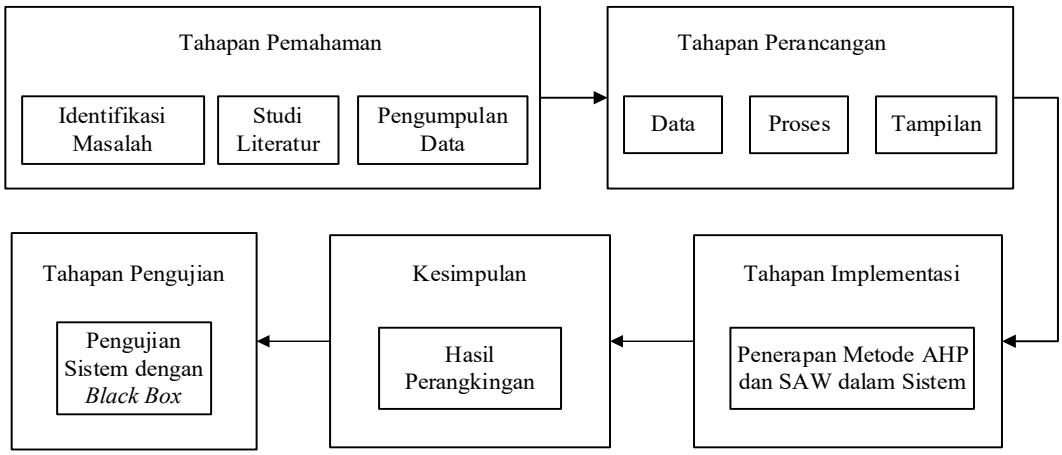
Penelitian yang dilakukan oleh Abrams et al., (2018) di Uni Emirat Arab (UEA) dan Oman. Studi tersebut melakukan pencarian air tanah di UEA utara dan Oman dengan menggambarkan potensi air tanah, kemungkinan relatif suatu lokasi. Metode yang digunakan yaitu SAW, *Probabilistic Frequency Ratios* (PFR) dan AHP. Turunan dinilai melalui validasi dengan membandingkan lokasi 645 sumur air, 49 mata air alami, dan pengamatan lapangan fitur air tanah. Hasil dari penelitian menunjukkan tertinggi terletak di Emirat Dubai/Sharjah integrasi data penginderaan jauh dengan teknik geospasial.

Sistem pendukung keputusan pemilihan varietas unggul padi yang dilakukan oleh Husein et al., (2017) menggunakan metode AHP dan *Technique for Order Preference by Similarity to Ideal Solution* (TOPSIS). Data yang digunakan yaitu data kriteria varietas yang diperoleh dari buku deskripsi varietas unggul padi edisi 2010 untuk varietas lama dan buku deskripsi varietas unggul padi edisi 2016 untuk varietas padi baru. Hasil dari sistem yang dibuat berupa peringkat alternatif mulai dari yang paling baik hingga yang paling buruk, diperoleh hasil akurasi sebesar 83.33%.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Pelaksanaan penelitian ini dilakukan dengan melalui beberapa tahapan, dimulai dari tahap pemahaman, perancangan, implementasi, kesimpulan, dan pengujian. Adapun tahapan dalam penelitian ini yang dapat dilihat pada Gambar 1.



Gambar 1 Tahapan Penelitian

Pada tahapan pemahaman dilakukan dengan mengidentifikasi masalah yang sudah diperoleh yaitu sistem pendukung keputusan kesesuaian lahan tanaman padi. Pada tahapan perancangan sistem dilakukan dengan beberapa tahap yaitu perancangan data, perancangan proses, dan perancangan tampilan. Pada tahapan implementasi dilakukan dengan membuat *flowchart* dari sistem yang dibuat dengan perancangan proses perhitungan menggunakan metode AHP dan SAW. Pada tahapan ini dilakukan penarikan kesimpulan berdasarkan tahap implementasi yang telah dilakukan pada tahap sebelumnya. Pada tahap pengujian sistem dilakukan berdasarkan kesimpulan hasil pemeringkatan yang telah dilakukan pada tahap sebelumnya dengan menggunakan proses *black box testing*.

2.2 Metode Pengumpulan Data

Pembuatan sistem pendukung keputusan terdapat beberapa metode yang dilakukan dalam proses pengumpulan data yaitu:

1) Studi Literatur
   Studi Literatur yang digunakan pada penelitian ini menggunakan metode pengumpulan data dengan melakukan pengumpulan informasi yang terhadap dari buku-buku, artikel, karya-karya ilmiah, catatan-catatan dan laporan-laporan yang berhubungan dengan penelitian.
2) Observasi
   Observasi yang dilakukan pada penelitian ini dengan mengunjungi kantor untuk mendapatkan data yang diperlukan sebagai data penelitian dan memberikan surat permohonan pengambilan data di Dinas Pertanian dan Tanaman Pangan Kota Samarinda.
3) Wawancara
   Wawancara yang dilakukan pada penelitian ini dengan melakukan tanya jawab dengan Bapak Badri, S.P., M.P di Dinas Pertanian dan Tanaman Pangan Kota Samarinda.

2.3 Analitycal Hierarchy Process (AHP)

Metode pengambilan keputusan AHP pertama kali dikembangkan oleh Thomas L. Saaty. Di mana AHP merupakan proses dalam pengambilan keputusan dengan menggunakan perbandingan berpasangan *pairwise comporisons* untuk menjelaskan faktor evaluasi dan faktor bobot dalam kondisi multi faktor (Septilia & Styawati, 2020). Berdasarkan penjelasan sebelumnya dapat disimpulkan bahwa AHP adalah metode pengambilan keputusan secara hierarki (tingkat) yang dipilih dari berbagai kriteria dan alternative (Saaty, 2008).

Secara umum, berikut langkah-langkah dalam metode AHP:
1) Mendefinisikan masalah dan menentukan solusi yang diinginkan, lalu menyusun hierarki dari permasalahan yang dihadapi.
2) Membuat matriks perbandingan pasangan antar kriteria menggunakan skala Saaty yang dapat dilihat pada Tabel 1 yaitu membandingkan elemen secara berpasangan sesuai kriteria yang diberikan Saaty (2015).

Tabel 1 Skala Penilaian Perbandingan Berpasangan Model Saaty

Tingkat Kepentingan	Definisi	Keterangan
1	Sama penting	Kedua elemen sama pentingnya
3	Sedikit lebih penting	Elemen yang satu sedikit lebih penting
5	Lebih penting	Elemen yang satu lebih penting daripada elemen lainnya
7	Sangat penting	Satu elemen jelas lebih penting daripada elemen lainnya
9	Mutlak sangat penting	Satu elemen mutlak lebih penting daripada elemen lainnya
2,4,6,8	Nilai tengah	Nilai-nilai antara dua nilai pertimbangan yang berdekatan
Kebalikan	Jika aktivitas i mendapat satu angka dibandingkan dengan aktivitas j, maka j memiliki nilai kebalikannya dibandingkan dengan i	

3) Matriks perbandingan berpasangan disintetis untuk memperoleh keseluruhan prioritas. Hal-hal yang dilakukan dalam langkah ini adalah:
   a) Menjumlahkan nilai-nilai dari setiap kolom pada matriks perbandingan berpasangan.
   b) Membagi setiap nilai dari kolom dengan total kolom yang bersangkutan untuk memperoleh normalisasi matriks.
   c) Menjumlahkan nilai-nilai dari setiap baris dan membaginya dengan jumlah elemen untuk mendapatkan nilai bobot prioritas.
4) Mengukur konsistensi matriks perbandingan berpasangan. Hal pertama yang dilakukan yaitu menghitung nilai λ maks dengan cara kalikan jumlah nilai kolom pertama perbandingan berpasangan dengan *priority vector* elemen pertama, jumlah nilai pada kolom kedua dengan *priority vector* elemen kedua dan seterusnya.
   Hitung *Consistency Index* (CI) dengan rumus yang dapat ditulis seperti Pers. (1) lalu hitung *Consistency Ratio* (CR) dengan rumus seperti pada Pers. (2).

$$CI = \frac{\lambda_{maks} - n}{n-1} \qquad (1)$$

Di mana CI merupakan *Consistency Index* (rasio penyimpangan konsistensi), $\lambda maks$ menyatakan nilai eigen terbesar dari matriks berordo n, dan n adalah jumlah elemen yang dibandingkan.

$$CR = \frac{CI}{RI} \qquad (2)$$

Di mana CR merupakan *Consistency Ratio*, CI merupakan *Consistency Index*, dan RI merupakan *Random Index*. Adapun nilai RI dipilih berdasarkan jumlah kriteria yang digunakan. Daftar Nilai RI dapat dilihat pada Tabel 2 (Saaty 2008).

Tabel 2 Nilai *Random Index*

Ukuran Matriks	Nilai RI	Ukuran Matriks	Nilai RI
1,2	0,00	9	1,45
3	0,58	10	1,49
4	0,90	11	1,51
5	1,12	12	1,48
6	1,24	13	1,56
7	1,32	14	1,57
8	1,41	15	1,59

2.4 Metode SAW

Metode SAW merupakan metode yang dikenal dengan istilah metode penjumlahan terbobot. Konsep dasar pada metode SAW adalah mencari penjumlahan terbobot dari rating kinerja pada setiap alternatif di semua atribut. Metode SAW membutuhkan proses normalisasi matriks keputusan (x) ke suatu skala yang dapat diperbandingkan dengan semua *rating* alternatif yang ada. Metode ini merupakan metode yang paling terkenal dan paling banyak digunakan dalam menghadapi situasi *Multiple Attribute Decision Making* (MADM). MADM itu sendiri merupakan suatu metode yang digunakan untuk mencari alternatif optimal dari sejumlah alternatif dengan kriteria tertentu. Metode SAW ini mengharuskan pembuat keputusan menentukan bobot bagi setiap atribut (Simanaviciene & Ustinovichius, 2010). Perhitungan akan sesuai dengan metode ini apabila alternatif yang terpilih memenuhi kriteria yang telah ditentukan (Pratiwi, 2016).

Berikut adalah langkah-langkah menggunakan metode SAW:
1) Menentukan kriteria yang akan dijadikan acuan dalam pengambilan keputusan Ci.
2) Memberikan nilai bobot untuk masing-masing kriteria sebagai W.
3) Memberikan nilai rating kecocokan setiap alternatif pada setiap kriteria.
4) Membuat matriks keputusan berdasarkan kriteria, kemudian melakukan normalisasi matriks berdasarkan persamaan yang disesuaikan dengan jenis atribut (atribut *benefit* maupun atribut *cost*) sehingga diperoleh matriks ternormalisasi R. Atribut benefit digunakan jika nilai terbesar yang terbaik dan atribut *cost* jika nilai terkecil yang terbaik.
5) Jika j adalah atribut *benefit* maka rumus dapat ditulis seperti Pers. (3).

$$r_{ij} = \frac{x_{ij}}{\underset{i}{Max}\ xij} \qquad (3)$$

Jika j adalah atribut *cost* maka rumus dapat ditulis seperti Pers. (4).

$$r_{ij} = \frac{\underset{i}{Min}\ x_{ij}}{xij} \qquad (4)$$

Di mana r_{ij} menunjukkan nilai rating kinerja ternormalisasi, X_{ij} merupakan nilai atribut yang dimiliki dari setiap kriteria, serta Max_{ij} dan Min_{ij} adalah nilai maksimum dan minimum dari setiap baris dan kolom.

Hasil akhir diperoleh dari proses pemeringkatan yaitu penjumlahan dan perkalian matriks ternormalisasi R dengan vektor bobot sehingga diperoleh nilai terbesar yang dipilih sebagai alternatif yang terbaik (Ai) sebagai solusi. Rumus nilai preferensi dapat ditulis seperti Pers. (5).

$$V_i \sum_{j=1}^{n} W_j R_{ij} \qquad (5)$$

Di mana V_i adalah nilai akhir dari alternatif, W_j adalah nilai bobot dari setiap kriteria, dan R_{ij} menunjukkan nilai rating kinerja ternormalisasi.

3. HASIL DAN PEMBAHASAN

Pengumpulan data dilakukan dengan menggunakan metode studi literatur, observasi dan wawancara. Berdasarkan hasil pengumpulan data, didapatkan data kriteria dan data alternatif yang digunakan dalam penentuan kesesuaian lahan tanaman padi. Data kriteria beserta subkriteria dapat dilihat pada Tabel 3.

Tabel 3 Data Kriteria

Kriteria	Subkriteria	Skala Nilai	Kategori
Jenis Tanah (C1)	Tanah gambut	2	Rendah
	Tanah organosol/gleyhumus	3	Cukup
	Tanah podsolik merah kuning	4	Tinggi
	Tanah aluvial	5	Sangat Tinggi
pH Tanah (C2)	<4,5	1	Rendah
	4,6 – 5,5	2	Sedang
	5,6 – 6,5	3	Cukup
	>6,6	4	Tinggi
Curah Hujan (C3)	<200 mm	1	Rendah
	201 – 400 mm	2	Sedang
	>401 mm	3	Tinggi
Suhu (C4)	<16 ^{O}C	5	Dingin
	16 – 22 ^{O}C	4	Sedang
	23 – 28 ^{O}C	3	Cukup
	29 – 34 ^{O}C	2	Panas
	35 ^{O}C	1	Sangat Panas
Irigasi Perairan	Irigasi permukaan	1	Sedang
	Perairan dengan pompa air	2	Cukup
	Irigasi tadah hujan	3	Tinggi

Data alternatif yang digunakan juga didapatkan di Dinas Pertanian Kota Samarinda merupakan data pada tahun 2020. Di mana data alternatif yang digunakan dalam penelitian ini dapat dilihat pada Tabel 4.

Tabel 4 Data Alternatif

Alternatif	C1	C2	C3	C4	C5
Palaran	Tanah aluvial	4,5	1160	28 ^{O}C	Tadah hujan
Sambutan	Tanah organosol/gleyhumus	5,5	2000	26 ^{O}C	Tadah hujan
Samarinda Sebrang	Tanah aluvial	4,0	1255	34 ^{O}C	Tadah hujan
Loa Janan Ilir	Tanah organosol/gleyhumus	6,5	3300	33 ^{O}C	Tadah hujan
Sungai Kunjang	Tanah podsolik merah kuning	6,5	3600	35 ^{O}C	Tadah hujan
Samarinda Utara	Tanah podsolik merah kuning	5,8	220	30 ^{O}C	Tadah hujan

3.1 Implementasi AHP dan SAW

Berikut langkah-langkah perhitungan metode AHP untuk mendapatkan bobot masing-masing kriteria yang digunakan dalam penentuan kesesuaian lahan tanaman padi, sebagai berikut:

1) Membuat matriks perbandingan berpasangan antara kriteria menggunakan skala Saaty yang dapat dilihat pada Tabel 5.
2) Normalisasi matriks dengan cara membagi setiap nilai dari kolom dengan total jumlah kolom yang bersangkutan. Adapun hasil normalisasi matriks dapat dilihat pada Tabel 6.
3) Menghitung nilai bobot dan konsistensi perbandingan berpasangan kriteria. Selanjutnya, hasil perhitungan bobot dan konsistensi dapat dilihat pada Tabel 7.

Tabel 5 Perbandingan Berpasangan Kriteria

	C1	C2	C3	C4	C5
C1	1	0,143	0,333	0,143	1
C2	7	1	9	3	5
C3	3	0,111	1	0,5	1
C4	7	0,333	2	1	3
C5	1	0,2	1	0,333	1
Jumlah	**19**	**1,787**	**13,333**	**4,976**	**11**

Tabel 6 Normalisasi Perbandingan Kriteria

	C1	C2	C3	C4	C5
C1	0,053	0,080	0,025	0,029	0,091
C2	0,368	0,560	0,675	0,603	0,455
C3	0,158	0,062	0,075	0,100	0,091
C4	0,368	0,187	0,150	0,201	0,273
C5	0,053	0,112	0,075	0,067	0,091

Tabel 7 Hasil Perhitungan Bobot Kriteria

Kriteria	Nama Kriteria	Bobot Kriteria
C1	Jenis Tanah	0,055
C2	pH Tanah	0,532
C3	Curah Hujan	0,097
C4	Suhu	0,236
C5	Irigasi Perairan	0,079
λ maks		5,247
Consistency Index (CI)		0,061
Consistency Ratio (CR)		0,054

Setelah memperoleh bobot kriteria selanjutnya melakukan pemeringkatan alternatif menggunakan metode SAW. Berikut adalah langkah-langkah perhitungan metode SAW:

1) Memberikan nilai *rating* kecocokan dari semua alternatif. Sebelum melakukan perhitungan, dilakukan terlebih dahulu pencocokan nilai antara setiap alternatif dengan setiap subkriteria yang telah memiliki masing-masing skala nilai. *Rating* kecocokan atau matriks keputusan dapat dilihat pada Tabel 8.
2) Normalisasi matriks berdasarkan jenis atribut kriteria *benefit* maupun *cost*. Kriteria yang digunakan semua berjenis *benefit*, maka normalisasi matriks dilakukan dengan membagi antara x_{ij} dengan nilai maksimal kolom *j* berdasarkan pada Pers. (3). Hasil normalisasi matriks disusun ke dalam matriks ternormalisasi R seperti pada Pers. (4).

$$R = \begin{bmatrix} 1 & 0{,}33 & 0{,}33 & 0{,}33 & 1 \\ 0{,}6 & 0{,}67 & 0{,}33 & 0{,}33 & 1 \\ 1 & 0{,}33 & 1 & 0{,}5 & 1 \\ 0{,}6 & 1 & 1 & 0{,}5 & 1 \\ 0{,}8 & 1 & 1 & 1 & 1 \\ 0{,}8 & 1 & 0{,}67 & 0{,}5 & 1 \end{bmatrix} \quad (4)$$

3) Selanjutnya, untuk menghitung nilai preferensi dari setiap alternatif menggunakan Pers. (5). Di mana matriks ternormalisasi (R) dikalikan dengan nilai bobot kriteria (W) untuk mendapatkan nilai preferensi. Nilai bobot kriteria yang digunakan merupakan hasil dari perhitungan metode AHP. Setelah memperoleh nilai preferensi masing-masing alternatif selanjutnya dilakukan pemeringkatan yang dapat dilihat pada Tabel 9.

Tabel 8 *Rating* Kecocokan

	C1	C2	C3	C4	C5
A1	5	1	1	3	3
A2	3	2	1	3	3
A3	5	1	3	2	3
A4	3	3	3	2	3
A5	4	3	3	1	3
A6	4	3	2	2	3

Tabel 9 Hasil Pemeringkatan Alternatif

Ranking	Alternatif	Nama Alternatif	Nilai Preferensi
1	A5	Sungai Kunjang	0,989
2	A4	Loa Janan Ilir	0,860
3	A6	Samarinda Utara	0,839
4	A2	Sambutan	0,578
5	A3	Samarinda Sebrang	0,527
6	A1	Palaran	0,423

Setelah melakukan proses perhitungan manual, proses selanjutnya adalah uji coba dengan tujuan mengetahui bahwa hasil perancangan sesuai hasil yang ditampilkan pada web dalam sistem pendukung keputusan untuk penentuan kesesuaian lahan tanaman padi. Berikut tampilan hasil perhitungan pembobotan dengan metode AHP pada Gambar 2 dan pemeringkatan dengan metode SAW dalam Gambar 3.

Hasil Perhitungan Bobot Kriteria

Kriteria	Jenis Tanah	pH Tanah	Curah Hujan	Suhu	Irigasi dan Perairan	Jumlah Per Baris	Bobot Kriteria
Jenis Tanah	0.053	0.08	0.025	0.029	0.091	0.277	0.055
pH Tanah	0.368	0.56	0.675	0.603	0.455	2.66	0.532
Curah Hujan	0.158	0.062	0.075	0.1	0.091	0.486	0.097
Suhu	0.368	0.187	0.15	0.201	0.273	1.179	0.236
Irigasi dan Perairan	0.053	0.112	0.075	0.067	0.091	0.397	0.079
Principe Eigen Vector (λ maks)							5.349
Consistency Index							0.087
Consistency Ratio							0.078

Gambar 2 Hasil Perhitungan Bobot Kriteria

Hasil Perangkingan Metode SAW

Rangking	Nama Alternatif	Nilai Preferensi	Aksi
1	Sungai Kunjang	0.989	Lihat Detail
2	Loa Janan Ilir	0.86	Lihat Detail
3	Samarinda Utara	0.839	Lihat Detail
4	Sambutan	0.578	Lihat Detail
5	Samarinda Sebrang	0.527	Lihat Detail
6	Palaran	0.423	Lihat Detail

Gambar 3 Hasil Pemeringkatan Alternatif

Hasil pemeringkatan pada Gambar 3, menunjukkan hasil berdasarkan pembobotan yang dilakukan secara subjektif oleh *decision makers* (DMs) menggunakan perbandingan berpasangan sesuai model Saaty. Sementara itu, pengujian dalam penelitian ini menggunakan *black box testing* untuk mengtahui keberhasilan sistem dalam uji efektivitas sistem dalam pengambilan keputusan apakah sudah sesuai dan berjalan seperti yang dilakukan secara tradisional tanpa menggunakan sistem pendukung keputusan.

4. KESIMPULAN

Berdasarkan hasil penelitian sistem pendukung keputusan ini digunakan untuk penentuan kesesuaian lahan tanaman padi, sehingga dihasilkan penerapan metode AHP dan SAW untuk sistem pendukung keputusan menggunakan 5 kriteria seperti jenis tanah, pH tanah, curah hujan, suhu, dan irigasi perairan. Hasil dari sistem pendukung keputusan dengan metode AHP dan SAW sistem mampu melakukan pembobotan nilai kriteria dan pemeringkatan nilai alternatif sebagai hasil pemilihan kesesuaian lahan pada tanaman padi pada alternatif pertama di Sungai Kunjang dengan nilai preferensi 0,989. Metode AHP dan SAW dapat digunakan serta diterapkan dalam melakukan pembobotan kriteria dan pemeringkatan alternatif pada kasus pemilihan lahan tanaman padi.

Rencana penelitian selanjutnya, diperlukan pengembangan model secara *hybrid* dalam sistem pendukung keputusan kelompok untuk pemilihan lahan padi yang mampu menggabungkan berdasarkan hasil *ranking* masing-masing DMs dalam menghasilkan alternatif secara bersama.

DAFTAR PUSTAKA

Abrams, W., Ghoneim, E., Shew, R., LaMaskin, T., Al-Bloushi, K., Hussein, S., AbuBakr, M., Al-Mulla, E., Al-Awar, M., & El-Baz, F. (2018). Delineation of groundwater potential (GWP) in the northern United Arab Emirates and Oman using geospatial technologies in conjunction with Simple Additive Weight (SAW), Analytical Hierarchy Process (AHP), and Probabilistic Frequency Ratio (PFR) techniques. *Journal of Arid Environments*, *157*(February), 77–96. https://doi.org/10.1016/j.jaridenv.2018.05.005

Amini, S., Rohani, A., Aghkhani, M. H., Abbaspour-Fard, M. H., & Asgharipour, M. R. (2020). Assessment of land suitability and agricultural production sustainability using a combined approach (Fuzzy-AHP-GIS): A case study of Mazandaran province, Iran. *Information Processing in Agriculture*, *7*(3), 384–402. https://doi.org/10.1016/j.inpa.2019.10.001

Anwar, D. S., Rohpandi, D., & Indriyanti, I. (2018). Sistem Pendukung Keputusan Untuk Menentukan Lahan Tanaman Cabai Dengan Menggunakan Metode Simple Additive Weighting. *Seminar Nasional Sistem Informasi Dan Teknologi Informasi*, 657–660. https://doi.org/10.30700/pss.v1i1.374

Astari, A. P., & Komarudin, R. (2018). Sistem Pendukung Keputusan Pemilihan Karyawan Terbaik Dengan Metode Fuzzy Tahani. *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, *6*(2), 169–178. https://doi.org/10.33558/piksel.v6i2.1507

Badan Pusat Statistik Provinsi Kalimantan Timur. (2020). *Luas Panen dan Produksi Padi di Provinsi Kalimantan Timur*. Badan Pusat Statistik. https://kaltim.bps.go.id/indicator/53/318/1/luas-panen-padi-menurut-kabupaten-kota.html

Burra, D. D., Parker, L., Than, N. T., Phengsavanh, P., Long, C. T. M., Ritzema, R. S., Sagemueller, F., & Douxchamps, S. (2021). Drivers of land use complexity along an agricultural transition gradient in Southeast Asia. *Ecological Indicators*, *124*(April 2020), 107402. https://doi.org/10.1016/j.ecolind.2021.107402

Husein, M. R., Roisdiansyah, Widodo, A. W., & Hidayat, N. (2017). Sistem Pendukung Keputusan Untuk Pemilihan Penanaman Varietas Unggul Padi Menggunakan Metode AHP dan TOPSIS. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *1*(10), 2548–2964.

Hussain, M., Ajmal, M. M., Khan, M., & Saber, H. (2015). Competitive priorities and knowledge management. *Journal of Manufacturing Technology Management*, *26*(6), 791–806. https://doi.org/10.1108/JMTM-03-2014-0020

Iqbalgis, H., & Nurochman, N. (2019). Aplikasi Metode Simple Additive Weighting (SAW) Dalam

Pengembangan Sistem Pencarian Toko Batik Berbasis Android. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *4*(2), 51. https://doi.org/10.14421/jiska.2019.42-07

Kelik Nugroho, A., Permadi, I., Nofiyati, N., & Hayyu Naufal Ulfa, S. (2019). Sistem Pendukung Keputusan Penilaian Kesehatan Tanah Dengan Metode Simple Additive Weighting. *Jurnal Informatika: Jurnal Pengembangan IT*, *4*(1), 61–69. https://doi.org/10.30591/jpit.v4i1.1034

Maglinets, Y., Raevich, K., & Tsibulsky, G. (2019). The Intelligent Managerial Decision Support System for Agricultural Land Evaluation. *E3S Web of Conferences*, *75*, 03005. https://doi.org/10.1051/e3sconf/20197503005

Mahendra, G. S., & Ernanda Aryanto, K. Y. (2019). SPK Penentuan Lokasi ATM Menggunakan Metode AHP dan SAW. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *5*(1), 49–56. https://doi.org/10.25077/TEKNOSI.v5i1.2019.49-56

Nurdin, N., Fahrozi, F., Ula, M., & . M. (2020). Sistem Pendukung Keputusan Penentuan Jenis Tanah yang Sesuai untuk Tanaman Pangan Menggunakan Metode Smarter dan SAW. *Informatika Pertanian*, *29*(2), 83. https://doi.org/10.21082/ip.v29n2.2020.p83-94

Nurkholis, A., Muhaqiqin, & Susanto, T. (2020). Algoritme Spatial Decision Tree untuk Evaluasi Kesesuaian Lahan Padi Sawah Irigasi. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(5), 978–987. https://doi.org/10.29207/resti.v4i5.2476

Nyamekye, A. B., Dewulf, A., Van Slobbe, E., Termeer, K., & Pinto, C. (2018). Governance arrangements and adaptive decision-making in rice farming systems in Northern Ghana. *NJAS: Wageningen Journal of Life Sciences*, *86–87*(1), 39–50. https://doi.org/10.1016/j.njas.2018.07.004

Nyamekye, A. B., Nyadzi, E., Dewulf, A., Werners, S., Van Slobbe, E., Biesbroek, R. G., Termeer, C. J. A. M., & Ludwig, F. (2021). Forecast probability, lead time and farmer decision-making in rice farming systems in Northern Ghana. *Climate Risk Management*, *31*, 100258. https://doi.org/10.1016/j.crm.2020.100258

Pratiwi, H. (2016). *Buku Ajar Sistem Pendukung Keputusan* (1st ed.). Deepublish.

Ramadani, S., Khair, H., & Bangun, S. D. (2020). Sistem Pendukung Keputusan Pemilihan Lahan Pertanian yang Tepat untuk Meningkatkan Hasil Panen Cabai Menggunakan Metode MOORA. *Jurnal Informatika Kaputama*, *4*(2), 241–252. https://doi.org/10.1234/jik.v4i2.296

Sente, U., & Tridamayanti, H. C. (2019). Peningkatan Pengetahuan Petani Melalui Keefektifan Demonstrasi Plot Penangkaran Padi Di Kabupaten Barito Timur Kalimantan Tengah. *Prosiding Temu Teknis Jabatan Fungsional Non Peneliti*, 611–619.

Septilia, H. A., & Styawati, S. (2020). Sistem Pendukung Keputusan Pemberian Dana Bantuan Menggunakan Metode AHP. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, *1*(2), 34–41. https://doi.org/10.33365/jtsi.v1i2.369

Simanaviciene, R., & Ustinovichius, L. (2010). Sensitivity Analysis for Multiple Criteria Decision Making Methods: TOPSIS and SAW. *Procedia - Social and Behavioral Sciences*, *2*(6), 7743–7744. https://doi.org/10.1016/j.sbspro.2010.05.207

Subiyanto, Hermanto, Arief, U. M., & Nafi, A. Y. (2018). An accurate assessment tool based on intelligent technique for suitability of soybean cropland: case study in Kebumen Regency, Indonesia. *Heliyon*, *4*(7), e00684. https://doi.org/10.1016/j.heliyon.2018.e00684

Wahyu, I., Suparni, S., & Pohan, A. B. (2020). Sistem Pendukung Keputusan Pemberian Pinjaman Pada KOPWALI Tangerang Dengan Metode AHP dan SAW. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *5*(1), 21–30. https://doi.org/10.31294/ijcit.v5i1.6559

Wulandari, W., Mustofa, A., Ponidi, P., Muslihudin, M., & Firdiansah, F. A. (2016). Decision Support System Pemetaan Lahan Pertanian yang Berkualitas untuk Meningkatkan Hasil Produksi Padi Menggunakan Metode Simple Additive Weighting (SAW). *Seminar Nasional Teknologi Informasi Dan Multimedia 2016*, 1.3-19-1.3-24.

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 7, No. 3, SEPTEMBER, 2022, Pp. 237 – 255
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Optimasi Seleksi Fitur *Information Gain* pada Algoritma Naïve Bayes dan K-Nearest Neighbor

Muhammad Norhalimi [(1)*], Taghfirul Azhima Yoga Siswa [(2)]
Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur, Samarinda
e-mail : muhammadnorhalimi1999@gmail.com, tay758@umkt.ac.id.
* Penulis korespondensi.
Artikel ini diajukan 25 Juli 2022, direvisi 23 September 2022, diterima 23 September 2022, dan dipublikasikan 25 September 2022.

Abstract

There was an increase in the number of late payments of tuition fees by 3,018 from a total of 5,535 students at the end of 2020. This study uses the Python library which requires data to be of numeric type, so it requires data transformation according to the type of data in the study, data that has a scale is transformed using an ordinal encoder, and data that does not have a scale is transformed using one-hot encoding. The purpose of this study was to evaluate the performance of the Naïve Bayes algorithm and K-Nearest Neighbor with a confusion matrix in predicting late payment of tuition fees at UMKT. The dataset used in this study was sourced from the financial administration bureau as many as 12,408 data with a distribution of 90:10. Based on the results of the calculation of the selection of information gain features, the best 4 attributes that influence the research are obtained, namely faculty, study program, class, and gender. The results of the evaluation of the confusion matrix that have the best performance using the Naïve Bayes with information gain algorithm obtain an accuracy of 55.19%, while the K-Nearest Neighbor with information gain only obtains an accuracy of 50.76%. Based on the accuracy results obtained in the prediction of late payment of tuition fees by using attributes derived from information gain, it influences increasing the accuracy of Naïve Bayes, but the use of the information gain attribute on the K-Nearest Neighbor algorithm makes the accuracy obtained decrease.

Keywords: Prediction, Naïve Bayes, K-Nearest Neighbor, Information Gain, Confusion Matrix

Abstrak

Terjadi kenaikan angka keterlambatan pembayaran biaya kuliah sebanyak 3.018 dari total 5.535 mahasiswa pada periode akhir 2020. Penelitian ini menggunakan *library* Python yang mengharuskan data bertipe numerik, sehingga memerlukan transformasi data yang sesuai dengan jenis data pada penelitian, pada data yang memiliki skala dilakukan transformasi menggunakan *ordinal encoder,* pada data yang tidak memiliki skala ditransformasi menggunakan *one-hot encoding*. Tujuan penelitian ini adalah untuk mengevaluasi kinerja algoritma Naïve Bayes serta K-Nearest Neighbor dengan *confusion matrix* dalam memprediksi keterlambatan pembayaran biaya kuliah di UMKT. *Dataset* yang digunakan dalam penelitian ini bersumber dari biro administrasi keuangan sebanyak 12.408 data dengan pembagian 90:10. Berdasarkan hasil perhitungan seleksi fitur *information gain* memperoleh 4 atribut terbaik yang berpengaruh dalam penelitian yaitu fakultas, prodi, angkatan, dan gender. Hasil evaluasi *confusion matrix* yang memiliki kinerja terbaik menggunakan algoritma Naïve Bayes *with information gain* memperoleh *accuracy* sebesar 55,19%, sedangkan K-Nearest Neighbor *with information gain* hanya memperoleh *accuracy* 50,76%. Berdasarkan hasil akurasi yang diperoleh dalam prediksi keterlambatan pembayaran biaya kuliah dengan menggunakan atribut yang berasal dari *information gain* mempunyai pengaruh dalam meningkatkan akurasi Naïve Bayes, namun penggunaan atribut *information gain* terhadap algoritma K-Nearest Neighbor membuat akurasi yang diperoleh menjadi menurun.

Kata Kunci: Prediksi, Naïve Bayes, K-Nearest Neighbor, *Information Gain, Confusion Matrix*

1. PENDAHULUAN

Universitas Muhammadiyah Kalimantan Timur (UMKT) merupakan salah satu lembaga pendidikan dibawah naungan Muhammadiyah sebagai perguruan tinggi swasta yang pembiayaan kuliahnya dibebankan kepada mahasiswa melalui SPP. Berdasarkan data dari bagian Keuangan terdapat angka kenaikan dan penurunan dalam pembayaran SPP pada periode 2017 hingga 2020. Akan tetapi pada periode akhir 2020 justru mengalami kenaikan yang sangat drastis, bahkan sampai melewati jumlah batas mahasiswa yang membayar SPP tepat waktu. Jika dilihat dari jumlah mahasiswa bisa mencapai 3.018 dari total 5.535 mahasiswa Universitas Muhammadiyah Kalimantan Timur. Sehingga keterlambatan pembayaran kuliah tentu sangat berpengaruh terhadap operasional akademik dan menghambat pembangunan sarana dan prasarana penunjang pembelajaran. Untuk dapat menganalisis permasalahan tersebut, maka perlu dilakukan prediksi agar keterlambatan pembayaran biaya kuliah dapat dilakukan pencegahan dan penanganan sedini mungkin.

Penelitian tentang prediksi keterlambatan pembayaran biaya pendidikan pernah dilakukan sebelumnya seperti pada penelitian (Muqorobin et al., 2020) tentang sistem estimasi keterlambatan pembayaran SPP sekolah menggunakan algoritma Naïve Bayes, data pada penelitian ini bersumber dari dapodik 2017 hingga 2018, data yang digunakan berjumlah 236 data yang bertipe ordinal, data tersebut memiliki 6 atribut seperti, penghasilan orang tua, tanggungan keluarga, pendidikan ayah, usia ayah, pendidikan ibu, dan usia ibu. Hasil akurasi yang diperoleh sebesar 67%, kemudian pada penelitian (Rohmayani, 2020) tentang analisis prediksi keterlambatan pembayaran biaya siswa menggunakan algoritma Naïve Bayes dengan *optimasi particle swarm*, pada penelitian ini menggunakan 115 *dataset* yang bersumber dari angket yang diisi oleh siswa, *dataset* tersebut bertipe ordinal dan nominal, di dalamnya terdapat 8 atribut yang digunakan seperti, pendapatan ayah, jumlah tanggungan orang tua, uang saku/bulan, jasa keuangan, layanan akademik, program studi, cara pembayaran SPP, dan pekerjaan ibu. Hasil akurasi yang diperoleh sebesar 73,94%. Pada penelitian (Akhmad & Siswa, 2022) tentang prediksi pembayaran biaya kuliah mahasiswa di UMKT, pada penelitian ini menggunakan 12.408 data keuangan mahasiswa pada tahun 2017 sampai 2021, *dataset* tersebut bertipe ordinal, adapun atributnya seperti penghasilan ayah, penghasilan ibu, pendidikan ayah, dan pendidikan ibu, hasil akurasi yang diperoleh sebesar 52,82%. Dari beberapa penelitian sebelumnya yang menggunakan Naïve Bayes dan K-Nearest Neighbor dalam prediksi pembayaran biaya kuliah hasil akurasi yang diperoleh masih kurang maksimal sehingga pada penelitian ini menambahkah fitur seleksi information gain untuk meningkatkan kinerja dari algoritma Naïve Bayes dan K-Nearest Neighbor serta menambahkan *metode one-hot-encoding* untuk mengubah atribut kategorikal ke dalam bentuk numerik sehingga dapat bekerja lebih baik dengan algoritma klasifikasi Naïve Bayes dan *K-Nearest.* Menurut (Saputro & Sari, 2020) beberapa algoritma tidak dapat menggunakan variabel kategori sebagai masukannya, sehingga dibutuhkan perubahan terhadap variabel kategori tersebut agar dapat digunakan oleh suatu algoritma dalam proses komputasi.

Pada penelitian prediksi keterlambatan biaya kuliah ini menggunakan metode pendekatan *data mining* yaitu dengan mengkomparasi Algoritma Naïve Bayes dan K-Nearest Neighbor. Algoritma Naïve Bayes bekerja lebih baik dibanding model klasifikasi lainnya seperti *decision trees, neural networks* (Wanto et al., 2020). Algoritma Naïve Bayes juga telah banyak digunakan dalam penelitian sebelumnya pada bidang pendidikan. Seperti prediksi tingkat kelulusan mahasiswa tepat waktu pada UIN Syarif Hidayatullah Jakarta dengan hasil akurasi sebesar 80,72% (Salmu & Solichin, 2017). Prediksi tingkat kelulusan mahasiswa tepat waktu pada Fakultas Ekonomi dan Bisnis Universitas Pendidikan Nasional dengan hasil akurasi sebesar 98% (Suardika, 2019). Prediksi tingkat kelulusan peserta sertifikasi *Microsoft Office Specialist* (MOS) dengan akurasi sebesar 99.24% (Rifai et al., 2019). Prediksi masa studi mahasiswa, hasil akurasi yang diperoleh sebesar 85.17% (Amelia et al., 2017).

Algoritma K-Nearest Neighbor juga pernah dilakukan pada penelitian sebelumnya. Menurut (Suntoro, 2019) K-Nearest Neighbor memiliki kelebihan sehingga sering dipakai oleh para peneliti

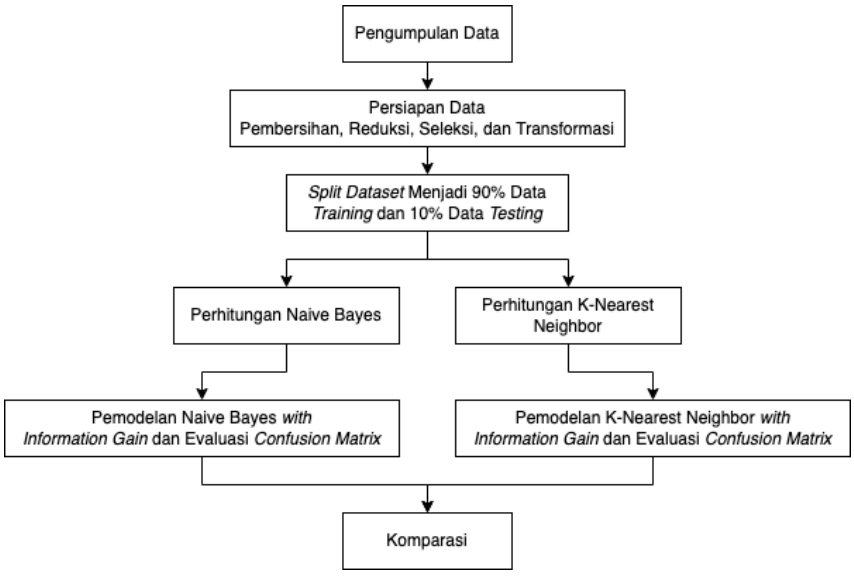
karena dapat memperoleh nilai akurasi yang tinggi dan tidak ada asumsi pada data. Penelitian yang menggunakan K-Nearest Neighbor seperti (Mustakim & Oktaviani, 2015) tentang prediksi prestasi mahasiswa, hasil akurasi yang diperoleh sebesar 82%. Prediksi tingkat kelulusan tepat waktu mendapatkan akurasi sebesar 85% pada algoritma Naïve Bayes sedangkan algoritma K-Nearest Neighbor menghasilkan 68.89% (Rahmatullah, 2019). Selanjutnya pada penelitian (Widaningsih, 2019) tentang perbandingan algoritma C4.5, Naïve Bayes, KNN, dan SVM terhadap prediksi nilai dan waktu kelulusan prodi TI menghasilkan akurasi sebesar 76,79% pada Naïve Bayes, SVM mendapatkan akurasi 74,04%, KNN dengan K=3 memperoleh akurasi 68,05%, dan C4.5 menghasilkan akurasi sebesar 75,96%.

Penelitian ini menggunakan seleksi fitur *information gain, Information gain* adalah perolehan informasi menggunakan *entropy* untuk menentukan atribut terbaik, *entropy* merupakan parameter untuk mengukur ketidakpastian yang di mana semakin tinggi *entropy*, maka semakin tinggi pula ketidakpastian (Suyanto, 2017). Seleksi fitur *information gain* telah banyak digunakan seperti pada penelitian (Muqorobin et al., 2019) tentang prediksi keterlambatan pembayaran sumbangan pembinaan pendidikan sekolah.

Penelitian ini juga menerapkan 2 metode transformasi yaitu *ordinal encoding* dan *one-hot encoding*. *Ordinal encoding* merupakan proses pemeringkatan yang dimulai dari skala terkecil 0 hingga ke n, *ordinal encoding* digunakan untuk mentrasformasi data yang memiliki tingkatan atau data yang bertipe ordinal. *One-Hot Encoding* merupakan proses untuk membuat suatu kolom baru dari variabel kategorikal, di mana setiap kategori menjadi kolom baru dengan nilai 0 atau 1 yang di mana nilai 0 menandakan tidak mewakili ada dan 1 mewakili ada (Daqiqil Id, 2021). *One-Hot Encoding* digunakan untuk mentrasformasi data yang tidak memiliki tingkatan dan datanya bertipe nominal, metode *One-Hot Encoding* pernah dilakukan pada penelitian sebelumnya seperti pada penelitian (Kinoto et al., 2020) tentang prediksi *employee churn* dengan *uplift* modeling. Perbedaan dengan penelitian ini adalah perpaduan transformasi data menggunakan *Ordinal Encoding* dan *One-Hot Encoding* serta penambahan fitur *selection information gain*. Penelitian ini bertujuan untuk mengidentifikasi atribut yang berpengaruh menggunakan fitur *selection information gain* dan mengevaluasi kinerja algoritma Naïve Bayes serta K-Nearest Neighbor dengan *confusion matrix* dalam memprediksi keterlambatan pembayaran biaya kuliah di Universitas Muhammadiyah Kalimantan Timur.

2. METODE PENELITIAN

Penelitian ini akan menggunakan 5 tahapan, adapun tahapannya seperti pada Gambar 1.



Gambar 1 Tahapan Penelitian

Berdasarkan pada Gambar 1 langkah pertama dalam penelitian adalah pengumpulan data, Langkah selanjutnya persiapan data terdiri dari pembersihan data, reduksi data, seleksi data menggunakan *information gain,* transformasi data *One-Hot Encoding*, dan *Ordinal Encoder*. Setelah persiapan data tahap berikutnya yaitu membagi dataset menjadi 90% *data training* dan 10% *data testing,* kemudian dilakukan perhitungan Naïve Bayes dan K-Nearest Neighbor secara manual. Pada tahap ini dilakukan permodelan algoritma Naïve Bayes *with Information gain* dan K-Nearest Neighbor *with Information gain,* setelah dilakukan permodelan, tahap berikutnya yaitu evaluasi menggunakan *confusion matrix* untuk melihat nilai akurasi. Tahap terakhir yaitu komparasi antara hasil *confusion matrix* dari permodelan algoritma Naïve Bayes *with Information gain* dan K-Nearest Neighbor *with Information gain.*

2.1 *Information Gain*

Information gain adalah perolehan informasi menggunakan *entropy* untuk menentukan atribut terbaik, *entropy* merupakan parameter untuk mengukur ketidakpastian yang di mana semakin tinggi *entropy*, maka semakin tinggi pula ketidakpastian (Suyanto, 2017). Rumus *feature selection information gain* dapat dilihat pada Pers. (1) dan (2) (Suntoro, 2019).

$$Entropy(S) \equiv \sum_{i=1}^{n} -pi * log_2\, p \qquad (1)$$

Di mana S merupakan himpunan kasus, n adalah jumlah partisi S, dan pi menunjukkan proporsi himpunan kasus ke-i.

$$Gain(S, A) \equiv Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si) \qquad (2)$$

Di mana S adalah himpunan kasus, A adalah atribut, n menunjukkan jumlah partisi atribut A, |S| adalah jumlah kasus dalam S, dan $|S_i|$ jumlah kasus pada partisi ke-i.

2.2 Algoritma Naïve Bayes

Algoritma Naïve Bayes dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas (Rajaraman & Ullman, 2011). Langkah-langkah permodelan sebagai berikut (Suntoro, 2019):

1) Membaca data *training*
2) Menghitung jumlah kelas target pada data training
3) Perhitungan menggunakan data numerik
   Langkah awal dalam perhitungan data numerik yaitu mencari nilai *mean* dan nilai standar deviasi dari masing-masing atribut yang menggambarkan data angka. Adapun rumus yang digunakan untuk menghitung nilai *mean* dapat dilihat pada Pers. (3).

$$\mu = \sum_{i}^{n} =_1\ x_i \quad atau\ \mu = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} \qquad (3)$$

   Di mana μ menyatakan nilai rata-rata hitung (*mean*), Xi merupakan nilai sampel ke -I, dan n menunjukkan jumlah sampel.

   Langkah selanjutnya menghitung nilai standar deviasi, adapun rumus perhitungan standar deviasi dapat dilihat pada Pers. (4).

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n-1}} \qquad (4)$$

   Di mana σ merupakan standar deviasi, Xi menyatakan nilai x ke -I, μ adalah nilai rata-rata hitung, dan n adalah jumlah sampel.

4) Nilai Distribusi *Gaussian*
Selanjutnya menghitung nilai probabilitas untuk fitur *data testing* yang memiliki data numerik. Rumus perhitungan distribusi Gaussian dapat dilihat pada Pers. (5).

$$P = (X_i = x_i \,|^{Y=yj}) = \frac{1}{\sqrt{2\pi\sigma ij}} \times e^{-\frac{(xi-\mu ij)^2}{2\sigma^2 ij}} \quad (5)$$

5) Menentukan Nilai Akhir
Setelah mendapatkan nilai distribusi *Gaussian*, langkah selanjutnya yaitu mengkalikan semua nilai *Gaussian* yang memiliki label tepat dan terlambat seperti pada Pers. (6).

$$(Kelas\,|X) = P(Kelas) \times P(X) \quad (6)$$

Kemudian setelah mendapatkan nilai akhir, langkah selanjutnya adalah membandingkan nilai antara Probabilitas keterangan "Tepat" dan Probabilitas keterangan "Terlambat".

2.3 Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor merupakan salah satu algoritma machine learning, Algoritma K-Nearest Neighbor bekerja dengan cara melakukan pencarian terhadap nilai k objek atau pola pada data training yang tersedia yang paling mendekati dengan pola masukan dan memilih kelas dengan jumlah pola terbanyak diantara nilai k pola tersebut (Suyanto, 2017). Adapun rumus Euclidien distance dapat dilihat pada Pers. (7).

$$Euclidien\ distance = \sqrt{\sum_{i=1}^{p}(a_k - b_k)^2} \quad (7)$$

Di mana a_k merupakan sampel data, b_k merupakan data uji *testing*, P menyatakan dimensi data, dan i adalah variabel data.

2.4 *Confusion Matrix*

Confusion matrix merupakan sebuah teknik yang bisa dipakai untuk mengetahui seberapa akurat model klasifikasi menggunakan tabel *confusion matrix* (Primartha, 2021). *Confusion matrix* dapat dihitung menggunakan Pers. (8) (Id, 2021).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

Di mana TP adalah *True Positive*, TN adalah *True Negative*, FP adalah *False Positive*, dan FN adalah *False Negative*.

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Data yang digunakan pada penelitian ini bersumber dari Biro Administrasi Keuangan berupa data pembayaran kuliah mahasiswa pada periode 2017 hingga 2020. Data yang diperoleh dari Bagian Administrasi Keuangan berjumlah 39.644, data tersebut terdiri dari 8.833 mahasiswa terlambat membayar biaya kuliah dan 30.811 mahasiswa tepat waktu dalam melakukan pembayaran kuliah. Data tersebut memiliki 9 atribut seperti fakultas, prodi, angkatan, gender, pendapatan ayah, pendapatan ibu, pendidikan ayah, pendidikan ibu, dan label target (tepat atau terlambat). Data biro administrasi keuangan dapat dilihat pada Tabel 1.

Tabel 1 Data Biro Administrasi Keuangan

No.	Fakultas	Prodi	Angkatan	Gender	Penghasilan Ayah	Penghasilan Ibu	Pend. Ayah	Pend. Ibu	Label
1	Ilmu Keperawatan	Keperawatan	2018	L	Rp. 2,000,000 - 4,999,999	Kurang dari Rp. 500,000		Tidak sekolah	Tepat
2	Ilmu Keperawatan	Keperawatan	2018	L	Kurang dari Rp. 500,000	Rp. 2,000,000 - 4,999,999	SMA	S1	Tepat
3	Ilmu Keperawatan	Keperawatan	2018	L	Rp. 2,000,000 - 4,999,999	Rp. 2,000,000 - 4,999,999		Tidak sekolah	Tepat
…	…	…	…	…	…	…	…	…	…
39644	Sains Dan Teknologi	Teknik Sipil	2017	L	Rp. 1,000,000 - 1,999,999	Kurang dari Rp. 500,000	SMP	SD	Terlambat

3.2 Persiapan Data

Pada tahapan ini terdapat beberapa persiapan data seperti pembersihan data, reduksi data, seleksi data, serta proses transformasi data untuk mendapatkan data yang berkualitas agar dapat mempermudah dalam proses *data mining* dan dijadikan masukan dalam tahap permodelan (*modeling*). Adapun prosesnya sebagai berikut.

3.2.1 Pembersihan Data

Pada tahap ini dilakukan pembersihan terhadap data yang tidak memiliki nilai atribut lengkap (*missing value*) dan eror. Adapun data awal berjumlah 39.644, terdapat 10.099 data yang perlu di bersihkan karena memiliki nilai atribut yang tidak lengkap dan eror, pada data ini tidak dapat dilakukan imputation terhadap data dikarenakan data yang kosong tidak dapat diterka seperti penghasilan orang tua dan pendidikan orang tua. Setelah dilakukan proses *cleaning* data menjadi 29.545. Data dapat dilihat pada Tabel 2.

Tabel 2 Data Setelah Melalui Proses Pembersihan

No	Fakultas	Prodi	Angkatan	Gender	Penghasilan Ayah	Penghasilan Ibu	Pend. Ayah	Pend. Ibu	Label
1	Ilmu Keperawatan	Keperawatan	2018	L	Kurang dari Rp. 500,000	Rp. 2,000,000 - 4,999,999	SMA	S1	Tepat
2	Ilmu Keperawatan	Keperawatan	2018	L	Rp. 500,000 - 999,999	Kurang dari Rp. 500,000	SD	SD	Tepat
3	Ilmu Keperawatan	Keperawatan	2018	L	Rp. 5,000,000 - 20,000,000	Rp. 2,000,000 - 4,999,999	S2	S1	Tepat
…	…	…	…	…	…	…	…	…	…
29545	Sains Dan Teknologi	Teknik Sipil	2017	L	Rp. 1,000,000 - 1,999,999	Kurang dari Rp. 500,000	SMP	SD	Terlambat

3.2.2 Reduksi Data

Pada tahap ini dilakukan penyeimbangan terhadap *dataset* yang ada dikarenakan data awal memiliki data tepat sebanyak 23.341 sedangkan data terlambat 6.204 sehingga data tersebut tidak seimbang. Maka perlu dilakukan penyeimbangan dengan cara mengambil secara acak data mahasiswa yang melakukan pembayaran kuliah secara tepat sebanyak 6.204 dan terlambat sebanyak 6.204 data. Proses reduksi data merujuk pada penelitian terdahulu seperti pada penelitian Ali *dkk,* (2019) di mana *dataset* yang memiliki kelas target yang paling banyak akan menyebabkan pemodelan menjadi bias terhadap kelas target yang sedikit. Sehingga dilakukan pengurangan terhadap *dataset* yang memiliki label target tepat menjadi 6.204 agar seimbang dengan label kelas target terlambat 6.204. peneliti menggunakan undersampling karena menurut (Kurniawan) teknik undersampling lebih menguntungkan dikarenakan hanya dengan sebagian data yang digunakan dalam penelitian proses komputasi bisa lebih hemat dan waktu proses menjadi lebih singkat, adapun syaratnya yaitu data yang diambil dapat mewakili dari seluruh data

populasi, dalam pengambilan data tidak ada patokan yang terkhusus, akan tetapi kita harus menghindari *sampling bias*. Adapun hasil reduksi dapat dilihat pada Tabel 3.

Tabel 3 Data Setelah Reduksi

No	Fakultas	Prodi	Angkatan	Gender	Penghasilan Ayah	Penghasilan Ibu	Pend. Ayah	Pend. Ibu	Label
1	Ekonomi Bisnis dan Politik	Manajemen	2019	P	Rp. 500,000 - 999,999	Kurang dari Rp. 500,000	SMA	SMP	Terlambat
2	Ilmu Keperawatan	Keperawatan	2019	P	Rp. 2,000,000 - 4,999,999	Kurang dari Rp. 500,000	SMA	SMA	Tepat
3	Ilmu Keperawatan	Keperawatan	2020	P	Rp. 1,000,000 - 1,999,999	Kurang dari Rp. 500,000	SMA	Tidak sekolah	Tepat
4	Sains Dan Teknologi	Teknik Informatika	2018	L	Rp. 2,000,000 - 4,999,999	Kurang dari Rp. 500,000	SMA	SMP	Tepat
…	…	…	…	…	…	…	…	…	…
12408	Ilmu Keperawatan	Keperawatan	2017	L	Rp. 2,000,000 - 4,999,999	Kurang dari Rp. 500,000	D3	SMA	Tepat

3.2.3 Seleksi Data

Pada tahap ini dilakukan penyeleksian data terhadap atribut-atribut yang akan digunakan dengan merujuk pada penelitian Muqorobin *dkk*, (2019) dengan menggunakan *feature selection information gain* agar mendapat atribut yang lebih informatif sehingga dapat meningkatkan akurasi serta efisien terhadap algoritma Naïve Bayes. Perhitungan *information gain* menggunakan *dataset* pada Tabel 3, terdapat 12.408 data yang di dalamnya memiliki 9 atribut. Adapun rumus perhitungan *information gain* dapat dilihat pada Pers. (1) dan (2).

1) Menghitung Nilai *Entropy*
Jumlah data kelas Tepat = 6204
Jumlah data kelas Terlambat = 6204
Jumlah data total = 12408

$$Entropy(total) = \left(-\frac{6204}{12408} * log_2\left(\frac{6204}{12408}\right)\right) + \left(-\frac{6204}{12408} * log_2\left(\frac{6204}{12408}\right)\right) = 1$$

Tahap selanjutnya menghitung nilai *entropy* setiap atribut, perhitungan nilai *entropy* setiap atribut sama seperti pada perhitungan *entropy* total. Berikut adalah hasil dari perhitungan *entropy* dapat dilihat pada Tabel 5.

2) Menghitung Nilai *Gain*
Tahap ini akan menghitung nilai *gain* dari masing masing atribut, adapun rumus perhitungan nilai *gain* dapat dilihat pada Pers. (2). Berikut adalah hasil dari perhitungan *gain* dapat dilihat pada Tabel 4.

Tabel 4 Perhitungan Nilai *Gain*

No.	Atribut	Gain
1	Fakultas	0,029837
2	Prodi	0,039215
3	Angkatan	0,071045
4	Gender	0,005528
5	Penghasilan Ayah	0,001513
6	Penghasilan Ibu	0,001077
7	Pendidikan Ayah	0,000974
8	Pendidikan Ibu	0,000974

Berdasarkan hasil *gain* pada Tabel 4 maka dipilih empat atribut yang memiliki nilai *gain* tertinggi untuk digunakan dalam implementasi Naïve Bayes *with information gain*, Atribut yang digunakan adalah fakultas, prodi, angkatan, dan gender.

Tabel 5 Perhitungan *Entropy*

Fakultas	**Jumlah**	**Tepat**	**Terlambat**	**Entropy**
Ekonomi Bisnis dan Politik	3362	1677	1685	0,999996
Farmasi	930	431	499	0,99614
Hukum	611	218	393	0,939988
Ilmu Keperawatan	2074	1433	641	0,892091
Keguruan Dan Ilmu Pendidikan	561	226	335	0,972595
Kesehatan Masyarakat	2021	1076	945	0,996967
Psikologi	688	302	386	0,98922
Sains Dan Teknologi	2161	841	1320	0,964263
Prodi	**Jumlah**	**Tepat**	**Terlambat**	**Entropy**
Farmasi	930	431	499	0,99614
Hubungan Internasional	356	162	194	0,994164
Hukum	611	218	393	0,939988
Keperawatan	1922	1281	641	0,918469
Kesehatan Lingkungan	447	205	242	0,995052
Kesehatan Masyarakat	1574	871	703	0,991767
Manajemen	3006	1515	1491	0,999954
Ners	152	152	0	0
Pendidikan Bahasa Inggris	287	109	178	0,957894
Pendidikan Olah Raga	274	117	157	0,984572
Psikologi	688	302	386	0,98922
Teknik Informatika	1038	461	577	0,990972
Teknik Mesin	404	128	276	0,900903
Teknik Sipil	719	252	467	0,934502
Angkatan	**Jumlah**	**Tepat**	**Terlambat**	**Entropy**
2017	1468	1320	148	0,471581
2018	2315	938	1377	0,973902
2019	5553	2524	3029	0,994026
2020	3072	1422	1650	0,996023
Gender	**Jumlah**	**Tepat**	**Terlambat**	**Entrophy**
L	5163	2314	2849	0,992241
P	7245	3890	3355	0,996063
Penghasilan Ayah	**Jumlah**	**Tepat**	**Terlambat**	**Entropy**
Kurang dari Rp. 500,000	3481	1800	1681	0,999157
Lebih dari Rp. 20,000,000	43	20	23	0,996486
Rp. 1,000,000 - Rp. 1,999,999	2871	1347	1524	0,997257
Rp. 2,000,000 - Rp. 4,999,999	3936	2016	1920	0,999571
Rp. 5,000,000 - Rp. 20,000,000	610	329	281	0,995529
Rp. 500,000 - Rp. 999,999	1467	692	775	0,99769
Penghasilan Ibu	**Jumlah**	**Tepat**	**Terlambat**	**Entropy**
Kurang dari Rp. 500,000	9086	4582	4504	0,999947
Lebih dari Rp. 20,000,000	12	7	5	0,979869
Rp. 1,000,000 - Rp. 1,999,999	966	461	505	0,998503
Rp. 2,000,000 - Rp. 4,999,999	1371	721	650	0,998065
Rp. 5,000,000 - Rp. 20,000,000	141	64	77	0,993859
Rp. 500,000 - Rp. 999,999	832	369	463	0,990773

Tabel 5 Perhitungan *Entropy* (lanjutan)

Pendidikan Ayah	Jumlah	Tepat	Terlambat	Entropy
Tidak Sekolah	2836	1429	1407	0,999957
TK	1	1	0	0
SD	1500	796	704	0,997285
SMP	1219	616	603	0,999918
SMA	4681	2302	2379	0,999805
D1	45	15	30	0,918296
D2	41	18	23	0,989245
D3	242	124	118	0,999557
S1	1512	751	761	0,999968
S2	310	142	168	0,99492
S3	21	10	11	0,998364
Pendidikan Ibu	**Jumlah**	**Tepat**	**Terlambat**	**Entropy**
Tidak Sekolah	2872	1443	1429	0,999983
TK	2	2	0	0
SD	2264	1148	1116	0,999856
SMP	1830	927	903	0,999876
SMA	3668	1797	1871	0,999706
D1	48	25	23	0,998747
D2	19	9	10	0,998001
D3	260	121	139	0,99654
S1	1352	681	671	0,999961
S2	84	47	37	0,989753
S3	9	4	5	0,991076

3.2.4 Transformasi Data

Pada tahap ini dilakukan transformasi data, yaitu merubah data yang bertipe kategori menjadi numerik. Menurut Kurniawan (2020) algoritma *machine learning* membutuhkan data numerik untuk melakukan *training dataset*, dalam melakukan tranformasi menggunakan metode *One-Hot Encoding* dan *Ordinal Encoding. One-Hot Encoding* merupakan proses untuk membuat suatu kolom baru dari variabel kategorikal, di mana setiap kategori menjadi kolom baru dengan nilai 0 atau 1 yang di mana nilai 0 menandakan tidak mewakili ada dan 1 mewakili ada (Id, 2021).

Alasan peneliti menggunakan transformasi pada penelitian ini adalah karena data yang digunakan dalam penelitian terdapat 2 jenis data yaitu kategorikal dan numerik, algoritma *Gaussian* Naïve Bayes tidak dapat digunakan jika terdapat data kategorikal (ibnu daqiqil). Sehingga perlu adanya transformasi data dari kategorikal menjadi numerik, transformasi dilakukan terhadap 2 tipe data yaitu nominal dan ordinal, sehingga pada data yang tidak memiliki tingkatan seperti fakultas, prodi dan gender dilakukan menggunakan transformasi *One-Hot Encoding*. Sedangkan pada data yang memiliki skala seperti penghasilan ayah, penghasilan ibu, pendidikan ayah, dan pendidikan ibu dilakukan menggunakan *Ordinal Encoding* dengan cara data Ordinal yang memiliki skala paling kecil diubah ke dalam bentuk numerik dengan bilangan 0 hingga n sebanyak kategori perfitur. Hasil dari transformasi *One-Hot Encoding* mendapatkan 25 atribut seperti pada tabel 6. Sedangkan hasil dari *Ordinal Encoding* dapat dilihat pada Tabel 7 dan 8. Adapun data setelah proses transformasi *One-Hot Encoding* dan *Ordinal Encoding* dapat dilihat pada Tabel 9

Tabel 6 Data Setelah Proses Transformasi *One-Hot Encoding*

No.	F. Ekonomi Bisnis dan Politik	F. Farmasi	F. Hukum	F. Ilmu Keperawatan	F. Keguruan Dan Ilmu Pendidikan	…	Label
1	1	0	0	0	0	…	Terlambat
2	0	0	0	1	0	…	Tepat
3	0	0	0	1	0	…	Tepat
4	0	0	0	0	0	…	Tepat
…	…	…	…	…	…	…	…
12408	0	0	0	1	0	…	Tepat

Tabel 7 Pembobotan Penghasilan Ayah dan Ibu

Pembobotan Penghasilan Ayah dan Ibu	
Kurang dari Rp. 500,000	0
Rp. 500,000 - Rp. 999,999	1
Rp. 1,000,000 - Rp. 1,999,999	2
Rp. 2,000,000 - Rp. 4,999,999	3
Rp. 5,000,000 - Rp. 20,000,000	4
Lebih dari Rp. 20,000,000	5

Tabel 8 Pembobotan Pendidikan Ayah Dan Ibu

Pembobotan Pendidikan Ayah Dan Ibu	
Tidak Sekolah	0
TK	1
SD	2
SMP	3
SMA	4
D1	5
D2	6
D3	7
D4 / S1	8
S2	9
S3	10

Tabel 9 Hasil Data Setelah Proses Transformasi

No.	F. Ekonomi Bisnis dan Politik	F. Farmasi	F. Hukum	F. Ilmu Keperawatan	F. Keguruan Dan Ilmu Pendidikan	…	Label
1	1	0	…	2	8	4	Terlambat
2	0	0	…	0	4	3	Terlambat
3	0	0	…	0	4	2	Terlambat
4	0	0	…	1	4	3	Terlambat
…	…	…	…	…	…	…	…
12408	0	0	…	2	8	4	Tepat

3.3 *Split* Data

Penelitian ini menggunakan *Split* data agar mendapatkan hasil estimasi akurasi yang baik dari proses *dataset*. *Dataset* yang digunakan kemudian dibagi menjadi 90% *data training* dan 10% *data testing*.

Tabel 10 Pembagian *Data Training* dan *Testing*

Hasil Pembagian Data	
Jumlah Data Training	11.167
Jumlah Data Testing	1.241

Tabel 10 merupakan hasil pembagian data 90% data *training* dan 10% data *testing*. Didapatkan hasil jumlah data *training* sebanyak 11.167 dan data *testing* 1.241.

3.4 Perhitungan Algoritma Naïve Bayes

Pada tahap ini dilakukan pemilihan model yang sesuai agar dapat mengoptimalkan hasil akurasi. Adapun algoritma yang digunakan dalam permodelan ini adalah algoritma Naïve Bayes, data yang digunakan pada permodelan adalah data *training* berjumlah 11.167 di dalamnya terdapat 30 atribut. Adapun data dapat dilihat pada Tabel 11.

Tabel 11 Data *Training*

No.	F. Ekonomi Bisnis dan Politik	F. Farmasi	F. Hukum	F. Ilmu Keperawatan	F. Keguruan Dan Ilmu Pendidikan	…	Label
1	1	0	0	0	0	…	Terlambat
2	0	1	0	0	0	…	Terlambat
3	0	0	1	0	0	…	Terlambat
…	…	…	…	…	…	…	…
11167	0	0	0	0	0	…	Tepat

Tabel 12 Data *Testing*

No.	F. Ekonomi Bisnis dan Politik	F. Farmasi	F. Hukum	F. Ilmu Keperawatan	F. Keguruan Dan Ilmu Pendidikan	…	Label
1	0	0	0	0	0	…	Terlambat
2	0	0	0	1	0	…	Tepat
3	0	0	1	0	0	…	Tepat
…	…	…	…	…	…	…	…
1241	1	0	0	0	0	…	Terlambat

Setelah membagi data, langkah selanjutnya melakukan perhitungan menggunakan algoritma Naïve Bayes adapun langkahnya sebagai berikut.

3.4.1 Menghitung nilai probabilitas

Langkah pertama adalah menghitung nilai probabilitas dari kelas label "Tepat" dan "Terlambat". Di mana jumlah data yang memiliki kelas label "Tepat" sebanyak 5.568 data dan jumlah kelas label "Terlambat" sebanyak 5.599 data kemudian dibagi dengan jumlah data label target sebanyak 11.167 data.

$$P(label = Tepat) = \frac{5.568}{11.167} = 0{,}5$$

$$P(label = Terlambat) = \frac{5.599}{11.167} = 0{,}5$$

3.4.2 Menghitung nilai *mean*

Langkah kedua adalah menghitung nilai *mean* pada setiap atribut yang ada pada data *training*. Perhitungan *mean* akan dilakukan menggunakan Pers. (3) pada Microsoft Excel, adapun hasil perhitungan dapat dilihat pada Tabel 13.

Tabel 13 Perhitungan *Mean* Tepat dan Terlambat

	F. Hukum	F. Ilmu Keperawatan	Keguruan dan Ilmu Pendidikan	Kesehatan Masyarakat	…	Pendidikan Ibu Tidak Sekolah
Tepat	0,036099	0,231142	0,36099	0,175287	…	0,230603
Terlambat	0,064666	0,102715	0,051804	0,151661	…	0,230260

3.4.3 Menghitung nilai standar deviasi

Setelah mencari nilai *mean*, langkah selanjutnya adalah menghitung nilai standar deviasi menggunakan Pers. (4) pada setiap atribut *data training* seperti berikut.

Atribut “Fakultas Ekonomi Bisnis dan Politik”

$$\sigma_{label} = Tepat \sqrt{\frac{\sum(0-0{,}270115)^2 + \ldots + (0-0{,}270115)^2 + (0-0{,}270115)^2}{5.568}} = 0{,}412470$$

$$\sigma_{label} = terlambat \sqrt{\frac{\sum(0-0{,}272419)^2 + \ldots + (1-0{,}272419)^2 + (0-0{,}272419)^2}{5.599}} = 0{,}488042$$

Perhitungan standard deviasi selanjutnya akan dilakukan menggunakan Microsoft Excel, adapun hasilnya dapat dilihat pada Tabel 14.

Tabel 14 Standard Deviasi Tepat dan Terlambat

	F. Hukum	F. Ilmu Keperawatan	Keguruan dan Ilmu Pendidikan	Kesehatan Masyarakat	…	Pendidikan Ibu Tidak Sekolah
Tepat	0,186537	0,421563	0,186537	0,380213	…	0,421221
Terlambat	0,235935	0,303587	0,221632	0,358692	…	0,420999

3.4.4 Menghitung nilai *Gaussian*

Langkah selanjutnya adalah menghitung nilai *Gaussian* menggunakan Pers. (5) pada data *testing*. Perhitungan nilai *gaussian* tepat dan terlambat akan dilakukan menggunakan Microsoft Excel, hasil perhitungan dapat dilihat pada Tabel 15 dan 16.

Tabel 15 Perhitungan Nilai *Gaussian* Tepat

No.	F. Hukum	F. Ilmu Keperawatan	Keguruan dan Ilmu Pendidikan	Kesehatan Masyarakat	…	Pendidikan Ibu Tidak Sekolah
1	0,906787	0,52882	0,906787	0,581907	…	0, 529276
2	0, 906787	0, 52882	1,47E-06	0, 581907	…	0, 529276
3	0, 906787	0, 52882	0,906787	0, 581907	…	0,529276
4	0, 906787	0, 52882	0, 906787	0, 581907	…	0,115948
…	…	…	…	…	…	…
1241	0, 906787	0,116489	0, 906787	0, 581907	…	0,529276

Tabel 16 Perhitungan Nilai *Gaussian* Terlambat

No.	F. Hukum	F. Ilmu Keperawatan	Keguruan dan Ilmu Pendidikan	Kesehatan Masyarakat	…	Pendidikan Ibu Tidak Sekolah
1	0,777314	0,683945	0,824785	0,609311	…	0,529567
2	0,777314	0,683945	8,99E-05	0,609311	…	0,529567
3	0,777314	0,683945	0,824785	0,609311	…	0,529567
4	0,777314	0,683945	0,824785	0,609311	…	0,115605
…	…	…	…	…	…	…
1241	0,777314	0,009182	0,824785	0,609311	…	0,529567

3.4.5 Menentukan nilai akhir

Pada penentuan nilai akhir, langkah yang harus dilakukan adalah mengkalikan semua nilai *Gaussian* yang memiliki label tepat dan terlambat dengan probabilitas kelas target. Selanjutnya membandingkan nilai antara Probabilitas keterangan "Tepat" dan Probabilitas keterangan "Terlambat". Maka didapatkan hasil prediksi data *testing* pertama adalah Terlambat. Perhitungan data *testing* menggunakan Pers. (6) dilakukan menggunakan *tools* tambahan yaitu Microsoft Excel seperti pada Tabel 17.

Tabel 17 Penentuan Nilai Akhir Tepat dan Terlambat

No.	Nilai Akhir Tepat	Nilai Akhir Terlambat
1	2,2E-14	0,62775E-14
2	1,67E-29	4,93229E-23
3	3,31E-13	4,8002E-12
4	1,74E-10	1,9473E-11
…	…	…
1241	3,41E-08	6,66305E-09

Berdasarkan Tabel 17 dilakukan perbandingan pada nilai akhir data *testing* terlambat dan nilai akhir data *testing* tepat, adapun datanya dapat dilihat pada Tabel 18.

Tabel 18 Aktual dan Prediksi

No.	Aktual	Prediksi
1	Tepat	Terlambat
2	Tepat	Terlambat
3	Terlambat	Terlambat
4	Tepat	Tepat
…	…	…
1241	Tepat	Tepat

Hasil dari permodelan menggunakan algoritma Naïve Bayes mendapatkan probabilitas prediksi tepat sebesar 613 dan probabilitas prediksi terlambat sebesar 628. Tahap selanjutnya yaitu proses evaluasi, evaluasi merupakan tahapan yang digunakan untuk membantu pengukuran pada algoritma. Penulis menggunakan *Confusion Matrix* untuk melakukan pengukuran evaluasi pada model. Adapun rumus perhitungan *accuracy* dapat dilihat pada Pers. (8).

Proses evalusi *confusion matrix* dengan menggunakan 1.241 data *testing*, didapatkan hasil *confusion matrix* berupa 22 data *True Positive,* 620 data *True Negative*, 591 *False Positive*, dan 8 *False Negative.* Kemudian dilakukan proses perhitungan melalui *confusion matrix* seperti pada Tabel 19.

Hasil akurasi yang diperoleh dari perhitungan *confusion matrix* menggunakan data *testing* 1.241 memperoleh *accuracy* sebesar 51,73%.

Tabel 19 Evaluasi *Confusion Matrix*

Predict	*Actual*	
	Tepat	**Terlambat**
Tepat	22	591
Terlambat	8	620

3.5 ermodelan Algoritma Naïve Bayes *with Information Gain*

Permodelan ini akan menggunakan algoritma Naïve Bayes *with information gain* pada pemrograman Python dengan pembagian data 90:10. Data yang digunakan pada permodelan adalah data setelah melalui proses seleksi fitur *Information gain* dan dilakukan telah di transformasi. Adapun data dapat dilihat pada Tabel 20, di dalamnya terdapat 26 atribut.

Tabel 20 Data *Training*

No.	F. Ekonomi Bisnis dan Politik	F. Farmasi	…	Angkatan	Gender L	Gender P	Label
1	1	0	…	2019	1	0	Tepat
2	0	1	…	2018	0	1	Terlambat
3	0	0	…	2019	0	1	Tepat
…	…	…	…	…	…	…	…
11167	0	0	…	2018	0	1	Terlambat

Tabel 21 Data *Testing*

No.	F. Ekonomi Bisnis dan Politik	F. Farmasi	…	Angkatan	Gender L		Gender P	Label
1	0	0	…	2018	0		1	Terlambat
2	0	0	…	2017	0		1	Tepat
3	0	0	…	2020	1		0	Tepat
…	…	…	…	…	…	…	…	
1241	1	0	…	2020	0		1	Terlambat

```
from sklearn.naive_bayes import  GaussianNB
classifier = GaussianNB()
classifier.fit(X_train, y_train)
y_pred = classifier.predict(X_test)
```

Gambar 2 Permodelan Naïve Bayes *with Information Gain*

Hasil dari permodelan mendapatkan probabilitas prediksi tepat sebesar 608 dan probabilitas prediksi terlambat sebesar 633. Selanjutnya dilakukan evalusi menggunakan *confusion matrix* dengan menggunakan 1.241 data *testing*, didapatkan hasil *confusion matrix* berupa 52 data *True Positive,* 633 data *True Negative*, 556 *False Positive*, dan 0 *False Negative.* Kemudian dilakukan proses perhitungan melalui *confusion matrix* seperti pada Tabel 22.

Tabel 22 Evaluasi *Confusion Matrix*

Predict	*Actual*	
	Tepat	**Terlambat**
Tepat	52	556
Terlambat	0	633

Hasil uji coba menggunakan algoritma Naïve Bayes *with information gain* menggunakan pemrograman Python mendapatkan *accuracy* sebesar 55,19%

3.6 Perhitungan Algoritma K-Nearest Neighbor

Perhitungan Algoritma K-Nearest Neighbor dapat dilakukan dengan beberapa tahapan yaitu menentukan nilai k, menghitung jarak dengan Pers. (7), selanjutnya melakukan *sorting* dari jarak terkecil hingga terbesar, tahap terakhir yaitu klasifikasi *test* data mayoritas. Berikut contoh perhitungan manual menggunakan 10 data sampel yang diambil secara acak, di dalamnya terdapat 9 data *training* dan 1 data *testing*.

Tabel 23 Sampel Data *Training*

No.	Fakultas Ekonomi Bisnis dan Politik	Fakultas Farmasi	…	Pendidikan Ayah	Pendidikan Ibu	Label
1	1	0	…	8	4	Terlambat
2	0	0	…	4	3	Terlambat
3	0	0	…	4	2	Terlambat
4	0	0	…	4	3	Terlambat
5	0	0	…	4	0	Terlambat
6	0	1	…	4	5	Tepat
7	1	0	…	8	9	Tepat
8	0	0	…	4	4	Tepat
9	0	0	…	4	3	Tepat

Tabel 24 Sampel Data *Testing*

No.	Fakultas Ekonomi Bisnis dan Politik	Fakultas Farmasi	…	Pendidikan Ayah	Pendidikan Ibu	Label
1	1	0	…	8	4	Terlambat

Perhitungan algoritma K-Nearest Neighbor menggunakan rumus *Euclidien Distance* seperti pada Pers. (7). Setelah dilakukan perhitungan pada *data training*, selanjutnya melakukan pengurutan jarak dari yang terkecil hingga terbesar untuk menentukan nilai *k* seperti pada Tabel 25. Selanjutnya melakukan *voting* untuk memprediksi seperti yang tertera pada Tabel 26.

Tabel 25 Hasil Perhitungan Jarak

Ranking	*Euclidien Distance*	Label
1	3,162278	Terlambat
3	5,291503	Terlambat
6	5,91608	Terlambat
2	4,795832	Terlambat
9	6,557439	Terlambat
5	5,656854	Tepat
7	6,082763	Tepat
8	6,403124	Tepat
4	5,385165	Tepat

Tabel 26 Hasil Prediksi

Nilai K	*Ranking*	*Euclidien Distance*	Label
K=5	1	3,162278	Terlambat
	2	4,795832	Terlambat
	3	5,291503	Terlambat
	4	5,385165	Tepat
	5	5,656854	Tepat

Data *testing* pada percobaan ini memiliki label “Tepat” akan tetapi pada saat diprediksi dengan K=5 mendapatkan hasil label “Terlambat”. Sehingga prediksi dinyatakan salah karena data *testing* tepat saat diprediksi hasilnya terlambat.

Pada penelitian ini juga akan melakukan perhitungan algoritma K-Nearest Neighbor menggunakan K=5 dengan pembagian data 90:10. Data yang digunakan berjumlah 12.408 di dalamnya terdapat 11.167 data *training* dan 1.241 data *testing*. Adapun data yang digunakan dapat dilihat pada Tabel 11 dan 12. Proses perhitungan menggunakan pemrograman Python seperti Gambar 3.

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, accuracy_score
knn = KNeighborsClassifier (n_neighbors=5, metric='euclidean')
knn.fit(x_train, y_train)
y_pred = knn.predict(x_test)
```

Gambar 3 Permodelan Algoritma K-Nearest Neighbor

Hasil dari permodelan pada K=5 mendapatkan 316 *True Positive*, 276 *False Positif*, 304 *False Negative*, dan 345 *True Negative*. Selanjutnya dilakukan proses perhitungan menggunakan *confusion matrix* adapun rumusnya dapat dilihat pada Pers. (8).

Tabel 27 Evaluasi Confusion Matrix

Predict	***Actual***	
	Tepat	**Terlambat**
Tepat	316	276
Terlambat	304	345

Hasil uji coba menggunakan algoritma K-Nearest Neighbor menggunakan pemrograman Python mendapatkan *accuracy* sebesar 53,26%.

3.7 Permodelan Algoritma K-Nearest Neighbor *with Information Gain*

Permodelan algoritma K-Nearest Neighbor *with information gain* akan dilakukan pada pemrograman Python menggunakan K=5 dengan pembagian data 90:10. Data yang digunakan pada permodelan adalah data setelah melalui proses seleksi fitur *Information gain* dan Transformasi. Adapun data dapat dilihat pada Tabel 20 dan 21, di dalamnya terdapat 26 atribut.

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, accuracy_score
knn = KNeighborsClassifier (n_neighbors=5, metric='euclidean')
knn.fit(x_train, y_train)
y_pred = knn.predict(x_test)
```

Gambar 4 Permodelan Algoritma K-Nearest Neighbor *with Information Gain*

Hasil dari permodelan pada K=5 mendapatkan 325 *True Positive*, 292 *False Positif*, 319 *False Negative*, dan 305 *True Negative*. Selanjutnya dilakukan proses perhitungan menggunakan confusion matrix adapun rumusnya dapat dilihat pada Pers. (8).

Hasil uji coba menggunakan algoritma K-Nearest Neighbor menggunakan pemrograman Python mendapatkan *accuracy* sebesar 50,76%.

Tabel 28 Evaluasi *Confusion Matrix*

Predict	*Actual*	
	Tepat	**Terlambat**
Tepat	325	292
Terlambat	319	305

3.8 Komparasi

Tabel 29 Hasil Komparasi

Pengujian	Akurasi
Naïve Bayes	51,73%
Naïve Bayes *with Information Gain*	55,19%
K-Nearest Neighbor	53,26%
K-Nearest Neighbor *Information Gain*	50,76%

Berdasarkan pengujian yang telah dilakukan menggunakan 12.408 data seperti pada tabel 29 tentang prediksi keterlambatan biaya kuliah menggunakan fitur *selection information gain* dapat menaikkan hasil akurasi algoritma Naïve Bayes. Hasil akurasi yang diperoleh algoritma Naïve Bayes sebesar 51,73% namun ketika menggunakan fitur seleksi *information gain* akurasi yang diperoleh meningkat sebesar 55,19% lebih tinggi dari pada algoritma K-Nearest Neighbor yang mendapatkan akurasi sebesar 53,26% sedangkan dengan menambahkan fitur seleksi *information gain* akurasi yang diperoleh menurun menjadi 50,76%. Akurasi yang diperoleh cukup rendah, penemuan ini perlu dikaji lebih jauh, dikarenakan secara teori penggunaan dataset yang besar dapat meningkatkan hasil akurasi yang diperoleh dan performa komputasi juga akan lebih baik dalam hal klasifikasi (Widystuti & Darmawan, 2018).

Beberapa penelitian yang berhasil meningkatkan akurasi menggunakan fitur seleksi *information gain* seperti pada penelitian (Setiyorini & Asmono, 2019) tentang penerapan metode KNN pada klasifikasi kinerja siswa yang berhasil meningkatkan akurasi dari 74,068% menjadi 76,553%, data yang digunakan pada penelitian bertipe numerik. Penelitian (Sari, 2017) tentang penerapan algoritma klasifikasi *machine learning* untuk memprediksi performa akademik siswa pada J48 mendapatkan akurasi 90.48%, *random forest* memperoleh 90.05%, MLP mendapatkan akurasi 88.96%, SVM memperoleh 88.1% dan Naïve Bayes memperoleh 86.68%. Data yang digunakan pada penelitian bertipe numerik. Pada penelitian yang dilakukan oleh peneliti fitur seleksi *information gain* mampu menaikkan kinerja algoritma Naïve Bayes akan tetapi tidak memberi pengaruh yang signifikan terhadap kinerja algoritma Naïve Bayes, berbeda halnya dengan penambahan seleksi fitur information gain terhadap algoritma K-Nearest Neighbor yang menurunkan tingkat akurasi yang diperoleh menjadi 50,76, sebelum menggunakan fitur optimasi memperoleh akurasi sebesar 53,26%.

Menurut (Id, 2021) untuk memudahkan algoritma *machine learning* dalam melakukan prediksi salah satunya pendekatan *one-hot encoding* dengan membuat suatu kolom baru dari variabel kategorikal, di mana setiap kategori menjadi kolom baru dengan nilai 0 atau 1 yang di mana nilai 0 menandakan tidak mewakili ada dan 1 mewakili ada. Akan tetapi setelah dilakukan penelitian terdapat beberapa macam gap atau beberapa hasil yang kurang memuaskan seperti pada penelitian (Setiyorini & Asmono, 2019) tentang implementasi kinerja algoritma Naïve Bayes untuk prediksi masa studi mahasiswa memperoleh akurasi yang cukup rendah 68%, termasuk penelitian yang peneliti lakukan mendapatkan tingkat akurasi yang kurang maksimal sebesar 55,19 pada algoritma Naïve Bayes dan 50,76% pada algoritma KNN.

Perbedaan penelitian ini dengan penelitian sebelumnya adalah tipe data yang digunakan, pada penelitian sebelumnya menggunakan tipe data numerik sehingga hasil akurasi yang diperoleh cukup tinggi. Hasil akurasi yang diperoleh dari penelitian ini cukup rendah dikarenakan pada penelitian ini data yang digunakan bersifat kategorikal dan banyaknya atribut menyulitkan algoritma untuk mengambil keputusan, ketika terlalu banyak decision maka mempengaruhi

kinerja algoritma tersebut. pada algoritma Naïve Bayes keputusan yang diambil berdasarkan probabilitas namun jika jenis data banyak bertipe kategori maka tingkat kesulitan dalam mengambil keputusan meningkat.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik beberapa kesimpulan yaitu Atribut keterlambatan biaya kuliah yang terbaik dipilih berdasarkan perhitungan seleksi fitur *informain gain* seperti fakultas, prodi, angkatan dan gender. Hasil pengujian dengan pembagian data 90:10 menggunakan algoritma Naïve Bayes dengan menambahkan fitur seleksi *information gain* mendapatkan akurasi 55,19%, sedangkan pada algoritma K-Nearest Neighbor dengan *information gain* memperoleh akurasi sebesar 50,76%. Dapat disimpulkan bahwa algoritma yang memiliki kinerja terbaik yaitu algoritma Naïve Bayes dan penambahan fitur seleksi *information gain* mampu menaikkan kinerja algoritma Naïve Bayes akan tetapi tidak memberi pengaruh yang signifikan terhadap kinerja algoritma Naïve Bayes, berbeda halnya dengan penambahan seleksi fitur *information gain* terhadap algoritma K-Nearest Neighbor yang menurunkan tingkat akurasi yang diperoleh.

DAFTAR PUSTAKA

Akhmad, M. R., & Siswa, T. A. Y. (2022). Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi. *Progresif: Jurnal Ilmiah Komputer*, *18*(2), 185. https://doi.org/10.35889/progresif.v18i2.921

Ali, H., Mohd Salleh, M. N., Saedudin, R., Hussain, K., & Mushtaq, M. F. (2019). Imbalance class problems in data mining: a review. *Indonesian Journal of Electrical Engineering and Computer Science*, *14*(3), 1552. https://doi.org/10.11591/ijeecs.v14.i3.pp1552-1563

Amelia, M. winny, Lumenta, A. S. ., & Jacobus, A. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Informatika*, *11*(1). https://doi.org/10.35793/jti.11.1.2017.17652

Id, I. D. (2021). *Machine Learning : Teori, Studi Kasus dan Implementasi Menggunakan Python*. UR PRESS. https://doi.org/10.5281/zenodo.5113507

Kinoto, J., Damanik, J. L., Situmorang, E. T. S., Siregar, J., & Harahap, M. (2020). Prediksi Employee Churn Dengan Uplift Modeling Menggunakan Algoritma Logistic Regression. *Jurnal Teknologi Dan Ilmu Komputer Prima (JUTIKOMP)*, *3*(2), 503–508. https://doi.org/10.34012/jutikomp.v3i2.1645

Kurniawan, D. (2020). *Pengenalan Machine Learning dengan Python*. PT Elex Media Komputindo.

Muqorobin, M., Kusrini, K., & Luthfi, E. T. (2019). Optimasi Metode Naive Bayes dengan Feature Selection Information Gain untuk Prediksi Keterlambatan Pembayaran SPP Sekolah. *Jurnal Ilmiah SINUS*, *17*(1), 1. https://doi.org/10.30646/sinus.v17i1.378

Muqorobin, M., Kusrini, K., Rokhmah, S., & Muslihah, I. (2020). Estimation System For Late Payment Of School Tuition Fees. *International Journal of Computer and Information System (IJCIS)*, *1*(1), 1–6. https://doi.org/10.29040/ijcis.v1i1.5

Mustakim, M., & Oktaviani, G. (2016). Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa. *Jurnal Sains, Teknologi, Dan Industri*, *13*(2), 195–202. https://doi.org/10.24014/sitekin.v13i2.1688

Primartha, R. (2021). *Algoritma Machine Learning*. Informatika.

Rahmatullah, S. (2019). Prediksi Tingkat Kelulusan Tepat Waktu dengan Metode Naïve Bayes dan K-Nearest Neighbor. *Jurnal Informasi Dan Komputer*, *7*(1), 7–16. https://doi.org/10.35959/jik.v7i1.118

Rajaraman, A., & Ullman, J. D. (2011). Data Mining. In *Mining of Massive Datasets* (Vol. 2, Issue January 2013, pp. 1–17). Cambridge University Press. https://doi.org/10.1017/CBO9781139058452.002

Rifai, M. F., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS). *PETIR*, *12*(2), 131–144. https://doi.org/10.33322/petir.v12i2.471

Rohmayani, D. (2020). Analysis Of Student Tuition Fee Pay Delay Prediction Using Naive Bayes

Algorithm With Particle Swarm Optimation Optimazation (Case Study : Politeknik TEDC Bandung). *Jurnal Teknologi Informasi Dan Pendidikan*, *13*(2), 1–8. https://doi.org/10.24036/tip.v13i2.317

Salmu, S., & Solichin, A. (2017). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta. *Seminar Nasional Multidisiplin Ilmu (SENMI)*, 701–709.

Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, *6*(1), 1. https://doi.org/10.24076/citec.2019v6i1.178

Sari, B. N. (2016). Implementasi Teknik Seleksi Fitur Information Gain Pada Algoritma Klasifikasi Machine Learning Untuk Prediksi Performa Akademik Siswa. *Seminar Nasional Teknologi Informasi Dan Multimedia 2016*, *March*, 55–60.

Setiyorini, T., & Asmono, R. T. (2019). Penerapan Metode K-Nearest Neighbor dan Information Gain pada Klasifikasi Kinerja Siswa. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *5*(1), 7–14. https://doi.org/10.33480/jitk.v5i1.613

Suardika, I. G. I. (2019). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus Fakultas Ekonomi dan Bisnis Universitas Pendidikan Nasional. *Jurnal Ilmu Komputer Indonesia*, *4*(2), 37–44. https://doi.org/10.23887/jik.v4i2.2775

Suntoro, J. (2019). *Data Mining Algoritma dan Implementasi dengan Pemrograman PHP*. PT Elex Media Komputindo.

Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Informatika.

Wanto, A., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitupulu, D., Negara, E. S., Lubis, M. R., Dewi, S. V., & Prianto, C. (2020). *Data Mining : Algoritma Klasifikasi*. Yayasan Kita Menulis.

Widaningsih, S. (2019). Perbandingan Metode Data Mining untuk Prediksi Nilai dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika dengan Algoritma C4.5, Naïve Bayes, KNN Dan SVM. *Jurnal Tekno Insentif*, *13*(1), 16–25. https://doi.org/10.36787/jti.v13i1.78

Widystuti, W., & Darmawan, J. B. B. (2018). Pengaruh jumlah data set terhadap akurasi pengenalan dalam deep convolutional network. *Konferensi Nasional Sistem Informasi (KNSI)*, 634–636.

9 772527 583007