

ISSN : 2527-5836

e-ISSN : 2528-0074

Vol. 8 No. 2, Mei 2023

JISKa

Jurnal Informatika Sunan Kalijaga

Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta



Tim Pengelola JISKa Edisi Mei 2023

Ketua Editor (*Editor in Chief*)

Muhammad Taufiq Nuruzzaman, Ph.D. (UIN Sunan Kalijaga Yogyakarta, Indonesia)

Editor Bagian (*Section Editor*)

1. Dr. Ir. Agung Fatwanto (UIN Sunan Kalijaga Yogyakarta, Indonesia)
2. Dr. Ir. Bambang Sugiantoro (UIN Sunan Kalijaga Yogyakarta, Indonesia)
3. Dr. Shofwatul Uyun (UIN Sunan Kalijaga Yogyakarta, Indonesia)

Dewan Editor (*Editorial Board*)

1. Dr. Aang Subiyakto (UIN Syarif Hidayatullah Jakarta, Indonesia)
2. Andang Sunarto, Ph.D. (IAIN Bengkulu, Indonesia)
3. Dr. Hamdani (Universitas Mulawarman Samarinda, Indonesia)
4. Nashrul Hakiem, Ph.D. (UIN Syarif Hidayatullah Jakarta, Indonesia)
5. Noor Akhmad Setiawan, Ph.D. (Universitas Gadjah Mada, Indonesia)

Editor Bahasa dan Layout (*Assistant Editor*)

Sekar Minati, S.Kom. (UIN Sunan Kalijaga Yogyakarta, Indonesia)

Tim Teknologi Informasi (*Journal Manager*)

1. Eko Hadi Gunawan, M.Eng. (UIN Sunan Kalijaga Yogyakarta, Indonesia)
2. Muhammad Galih Wonoseto, M.T. (UIN Sunan Kalijaga Yogyakarta, Indonesia)

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. Alfian 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 Crysdiyan (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. 8, No. 2, MEI 2023

DAFTAR ISI

Analisis Klasifikasi <i>Broken Home</i> pada Anak Menggunakan Metode Naïve Bayes Classifier	90-101
Supiyandi Supiyandi, Almanna Hussein, Irwan Gunawan, William Lutfi Rahman Harjo	
Regresi Logistik Multinomial untuk Prediksi Kategori Kelulusan Mahasiswa	102-111
Rafika Syahranita, Suhartono Suhartono, Syahiduz Zaman	
Klasifikasi Ulasan Fasilitas Publik Menggunakan Metode Naïve Bayes dengan Seleksi Fitur <i>Chi-Square</i>	112-124
Adhitya Prayoga Permana, Totok Chamidy, Cahyo Crysdian	
Naive Bayes untuk Pengukuran Kualitas Media pada Larva BSF (<i>Black Soldier Fly</i>) Berbasis <i>Internet of Things</i>	125-139
Mohammad Faisal Fajar Fadilah, Ajib Hanani, Totok Chamidy	
Evaluasi Penerimaan Masyarakat Terhadap Aplikasi <i>Telemedicine</i> pada Masa Pandemi COVID-19	140-153
Muhammad Reza Velayani, Muhammad Taufiq Nuruzzaman, Agung Fatwanto, Bambang Sugiantoro	
Penerapan Naïve Bayes pada Potensi Akademik Siswa SD Negeri 5 Singakerta	154-163
Ni Kadek Winda Patrianingsih, I Kadek Arya Sugianta	
Pengembangan Aplikasi Audit Satuan Pengawasan Internal UIN Sunan Kalijaga Yogyakarta	164-174
Muhammad Galih Wonoseto	

Pengembangan Aplikasi Audit Satuan Pengawasan Internal UIN Sunan Kalijaga Yogyakarta

Muhammad Galih Wonoseto

Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga, Yogyakarta
e-mail : muhammad.wonoseto@uin-suka.ac.id.

Artikel ini diajukan 8 April 2023, direvisi 9 Mei 2023, diterima 10 Mei 2023, dan dipublikasikan 26 Mei 2023.

Abstract

In carrying out internal supervision duties, SPI UIN Sunan Kalijaga Yogyakarta innovates by carrying out digital document-based supervision using the Google Drive service. SPI auditors often felt problems in using Google Drive, namely having to check the completeness of financial accountability documents manually one by one by opening a file on Google Drive. One of the goals of developing this application is to facilitate SPI auditors in checking the completeness of financial accountability report documents. The problems can be solved by creating an audit application. The audit application has been developed using the FAST method. Central admin, SPI admin, operator, superior operator, auditor, team leader, technical controller, and Head of SPI are user access rights in this application. The developed application has met the minimum requirements of SPI UIN Sunan Kalijaga Yogyakarta. Other universities can adopt this application.

Keywords: FAST Method, Information System, Application, Audit, University, Indonesia

Abstrak

Dalam melaksanakan tugas pengawasan internal, SPI UIN Sunan Kalijaga Yogyakarta berinovasi dengan melaksanakan pengawasan berbasis dokumen digital menggunakan layanan Google Drive. Masalah dalam penggunaan Google Drive sering dirasakan oleh auditor SPI yaitu harus memeriksa kelengkapan dokumen pertanggungjawaban keuangan secara manual satu per satu dengan cara membuka *file* dalam Google Drive. Salah satu tujuan dari pengembangan aplikasi ini adalah untuk memudahkan auditor SPI dalam memeriksa kelengkapan dokumen laporan pertanggungjawaban keuangan. Permasalahan tersebut dapat diatasi dengan mengembangkan aplikasi audit. Aplikasi audit dikembangkan dengan metode FAST. Terdapat tujuh macam hak akses *user* dalam aplikasi ini yaitu admin pusat, admin SPI, operator, atasan operator, auditor, ketua tim, pengendali teknis, dan Kepala SPI. Aplikasi yang dikembangkan telah memenuhi *minimum requirement* SPI UIN Sunan Kalijaga Yogyakarta. Aplikasi ini juga dapat digunakan untuk universitas lain selain UIN Sunan Kalijaga Yogyakarta.

Kata Kunci: Metode FAST, Sistem Informasi, Aplikasi, Audit, Universitas, Indonesia

1. PENDAHULUAN

Berdasarkan Peraturan Menteri Agama nomor 25 tahun 2017 tentang Satuan Pengawasan Internal (SPI) (Peraturan Menteri Agama Tentang Satuan Pengawas Internal Pada Perguruan Tinggi Keagamaan Negeri, 2017), SPI memiliki kewenangan melaksanakan tugas pengawasan internal dalam bidang non akademik di Universitas Islam Negeri (Ruhiatudin et al., 2019). Pengawasan internal yang dilaksanakan meliputi reuiu, pemantauan, dan kegiatan pengawasan lain terhadap tugas dan fungsi organisasi. Hal tersebut dilaksanakan dalam rangka memberikan keyakinan bahwa kegiatan telah dilaksanakan dengan baik, efektif, dan efisien (Peraturan Pemerintah (PP) Tentang Sistem Pengendalian Intern Pemerintah, 2008). Dalam upaya mewujudkan efisiensi dan efektifitas implementasi pengawasan internal tersebut SPI UIN Sunan Kalijaga berinovasi dengan melaksanakan pengawasan berbasis dokumen digital. Pada tahun 2019 SPI UIN Sunan Kalijaga Yogyakarta telah melaksanakan Pengawasan Internal Berbasis Dokumen Digital menggunakan layanan Google Drive (Ruhiatudin et al., 2019). Hal tersebut sejalan dengan anjuran dan arahan BPK (Badan Pemeriksa Keuangan) yang mendorong auditor untuk memanfaatkan layanan TI secara optimal dalam proses audit.



Pada tahun 2022, telah dilaksanakan *preliminary research* (Wonoseto et al., 2022) untuk mengevaluasi program pengawasan berbasis dokumen digital menggunakan layanan Google Drive. Dari penelitian tersebut, disimpulkan bahwa program pengawasan berbasis dokumen digital sangat dirasakan manfaatnya oleh auditee dan institusi sehingga layak untuk dilanjutkan dan dikembangkan. Proses penelitian dan perancangan aplikasi ini (Wonoseto et al., 2022) sudah berjalan dengan melakukan beberapa kali FGD baik di internal pengurus SPI maupun dengan Wakil Dekan Bidang II, operator digitalisasi dokumen setiap unit dan fakultas dan pihak pengadaan barang dan jasa UIN Sunan Kalijaga Yogyakarta.

Masalah dalam pemanfaatan Google Drive sering dirasakan oleh internal auditor SPI UIN Sunan Kalijaga Yogyakarta. Auditor harus memeriksa kelengkapan dokumen pertanggungjawaban keuangan secara manual satu per satu dengan cara membuka *file* dalam Google Drive. Salah satu tujuan dari pengembangan aplikasi ini adalah untuk memudahkan auditor SPI dalam memeriksa kelengkapan dokumen laporan pertanggungjawaban keuangan.

Teknik Audit Berbantuan Komputer (TABK) (Akmalia & Ariani, 2022) atau dalam bahasa Inggris dikenal dengan *Computer Assisted Audit Technique Tools* (CAATs) adalah istilah untuk penggunaan teknologi informasi sebagai sarana/alat bantu kegiatan audit. Menurut Omunuk (JB & AA, 2015), CAATs atau TABK merupakan instrumen yang penting dan bermanfaat bagi auditor dalam melaksanakan tugas sehingga membuat pekerjaan auditor menjadi lebih efektif dan efisien. TABK atau CAATs tidak hanya memudahkan dalam analisa tetapi juga dapat meningkatkan efektivitas dan efisiensi waktu, biaya dan juga sumber daya manusia (Januraga & Budiarta, 2015). Salah satu strategi yang menjadi fokus BPK (Badan Pemeriksa Keuangan) saat ini adalah dengan mendorong auditor untuk memanfaatkan layanan TI dalam proses audit (Dilson & Suhery, 2016; Wicaksono et al., 2018). Namun berdasarkan laporan pertanggungjawaban kinerja pemanfaatan TI hanya mencapai 80,84% dengan target semula 97,87% (Dilson & Suhery, 2016; Wicaksono et al., 2018). Dalam penelitian berjudul *The Barriers of IT Utilization: A Case Study of Indonesian Audit Organization* (Damanik et al., 2020), Aslon Damanik menganalisis faktor-faktor yang menjadi kendala dalam penggunaan TI dalam proses audit di Indonesia. Faktor tersebut dikategorikan dalam empat dimensi yaitu: teknologi, organisasi, lingkungan, dan manusia. Hasil penelitian menunjukkan bahwa dimensi manusia merupakan faktor yang paling penting untuk diperhatikan. Namun berdasarkan kondisi saat ini faktor teknologi perlu segera ditingkatkan karena merupakan hambatan yang paling banyak dikemukakan oleh responden. Berdasarkan hasil penelitian tersebut, selain diperlukannya peningkatan SDM melalui pelatihan pemanfaatan layanan teknologi, juga diperlukan penelitian dan pengembangan TI yang mudah digunakan oleh pengguna.

Sebelumnya, peneliti telah membuat rancangan aplikasi audit internal (Wonoseto et al., 2022) dan telah diuji dengan *System Usability Scale*. Berdasarkan pengujian tersebut, rancangan aplikasi yang dibuat oleh peneliti masuk dalam kategori OK sehingga layak untuk dilanjutkan ke tahap pengembangan.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah FAST (*Framework for the Application System Thinking*) (Halim, 2020). Metode FAST adalah sebuah kerangka bisnis yang cukup fleksibel untuk membangun berbagai macam proyek sehingga dapat membantu memudahkan pelaku bisnis untuk mengidentifikasi serta mengevaluasi kebutuhan dan masalah-masalah yang timbul dalam proses informasi akuntansi (Azmi et al., 2018). Berikut delapan tahapan dalam metode FAST (Arif Nurrahman et al., 2020; Whitten & Bentley, 2007).

2.1 Definisi Ruang Lingkup

Fase definisi ruang lingkup ini adalah langkah awal untuk mengumpulkan seluruh informasi secara lengkap yang akan dikerjakan. Tahap ini dilakukan dengan wawancara atau komunikasi dengan pihak Satuan Pengawas Internal UIN Sunan Kalijaga Yogyakarta.



2.2 Analisis Permasalahan

Fase analisis masalah ini lebih meneliti masalah-masalah apa saja yang terjadi pada sistem yang sebelumnya. Dalam hal ini perlu adanya dokumentasi sistem yang berjalan sehingga hasil tahapan ini yakni perbaikan sekaligus peningkatan sistem yang akan memberikan keuntungan bagi suatu perusahaan. Tahap ini dilakukan dengan wawancara atau komunikasi dengan pihak Satuan Pengawas Internal UIN Sunan Kalijaga Yogyakarta. Selain itu, dilakukan juga proses evaluasi terhadap sistem sebelumnya dengan menggunakan *form* kuesioner. Kuesioner tersebut disebar dan diisi oleh anggota SPI, operator unit/ fakultas, ketua unit/ lembaga, dan Wakil Dekan Bidang II.

2.3 Analisis Kebutuhan

Fase analisis kebutuhan adalah tahap yang sangat krusial karena pengembangan sistem menjadi sebuah kebutuhan utama sehingga apabila terjadi kesalahan atau kekeliruan dalam analisis kebutuhan akan berakibat sangat fatal yakni ketidakpuasan pada sistem yang baru. Selanjutnya untuk mendapatkan informasi secara lengkap sesuai yang dibutuhkan maka harus melakukan identifikasi data, wawancara, dan proses interface sesuai kebutuhan yang diinginkan oleh user system actual.

2.4 Perancangan Logis

Fase perancangan logis yakni mentransformasi segala informasi usaha yang dibutuhkan sejak fase *requirements analysis* ke *model system* baru untuk dibuat. Dalam pengertian berbeda pada tahap ini memberikan jawaban sekitar bagaimana cara berteknologi (*data, process, dan interface*) sehingga menyajikan *usability, reliability, comprehensiveness, showing, dan standard* guna dikembangkan melalui *system*. Pada penelitian ini, digunakan desain Figma untuk merancang *interface* aplikasi yang akan dibuat.

2.5 Telaah Kesimpulan

Tahap telaah kesimpulan adalah tindakan untuk mengidentifikasi solusi teknis mana yang akan digunakan dalam implementasi sistem baru yang sesuai dengan *problem* dan *requirements* yang telah dijelaskan pada tahap-tahap sebelumnya. Pada fase ini, peneliti menggunakan metode *System Usability Scale* (Brooke, 2013; Prabowo & Suprpto, 2021) untuk menentukan apakah desain yang dibuat pada tahap perancangan logis layak untuk dilanjutkan ke tahap selanjutnya atau tidak.

2.6 Desain Logis

Tahap desain logis ini untuk mentransformasikan kebutuhan kebutuhan dalam suatu bisnis yang mewakili sebagai *logical design* kemudian menjadi *physical design* sehingga dapat dikembangkan menjadi lebih baik dan lebih spesifik dalam mengatasi permasalahan bisnis. Oleh sebab itu *physical design* mentransformasi persyaratan bisnis ke dalam spesifikasi desain fisik untuk menangani rincian dalam skala besar terkait bagaimana penggunaan teknologi dalam sistem informasi.

2.7 Konstruksi dan Pengujian

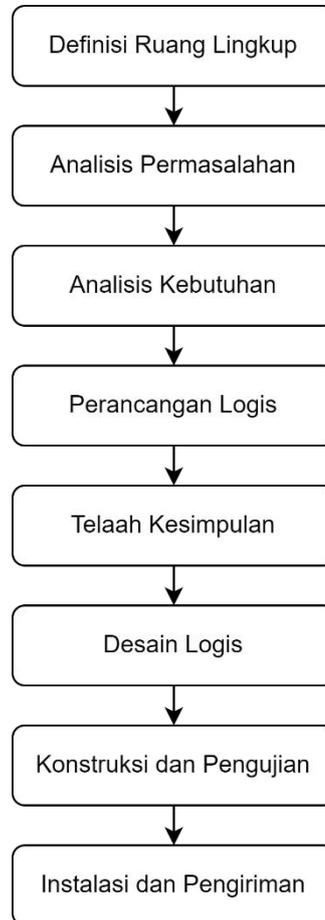
Tahap *construction* dan *testing* bertujuan untuk mengkonstruksi dan melakukan tahap pengujian pada sistem baru yang telah memenuhi kebutuhan bisnis dan spesifikasi desainnya. Dalam tahap ini harus dilakukan pengujian secara ketat dan mendetail mulai dari program aplikasi dan mengaplikasikan *interface* antara sistem baru dengan sistem yang sudah ada.

2.8 Instalasi dan Pengiriman

Tahap instalasi dan pengiriman ini adalah proses penyesuaian dari sistem lama ke sistem baru, oleh karena itu perlu melakukan beberapa kegiatan yakni, instalasi sistem, pelatihan, dan



membantu para pengguna dalam menjalankan sistem baru (*training user*) yang telah dibangun, manual sistem, mengkonversi file-file dan *database* yang ada pada sistem lama ke dalam *database* sistem baru, dan yang terakhir adalah *testing*. Jika digambarkan dalam *flowchart*, maka tahapan metode FAST akan menjadi seperti dalam Gambar 1.



Gambar 1 *Flowchart* Metode FAST

3. HASIL DAN PEMBAHASAN

Implementasi dari metode FAST menghasilkan hal-hal berikut ini dalam setiap tahapnya.

3.1 Definisi Ruang Lingkup

Aplikasi pengawasan internal berbasis dokumen digital ini adalah aplikasi yang digunakan untuk memeriksa penggunaan dana anggaran yang bersumber dari APBN (RM) maupun BLU di lingkungan UIN Sunan Kalijaga Yogyakarta. Terdapat tiga jenis mata anggaran yaitu:

- MAK 51: Untuk belanja pegawai.
- MAK 52: Untuk belanja barang dan jasa.
- MAK 53: Untuk belanja modal.

Aplikasi pengawasan internal berbasis dokumen digital digunakan untuk seluruh unit/fakultas yang ada di UIN Sunan Kalijaga Yogyakarta. Berikut ini daftar fakultas di UIN Sunan Kalijaga Yogyakarta:

- Fakultas Adab dan Ilmu Budaya
- Fakultas Dakwah dan Komunikasi



- Fakultas Ilmu Tarbiyah dan Keguruan
- Fakultas Syariah dan Hukum
- Fakultas Ushuluddin dan Pemikiran Islam
- Fakultas Sains dan Teknologi
- Fakultas Ilmu Sosial dan Humaniora
- Fakultas Ekonomi dan Bisnis Islam
- Pascasarjana

Unit dan bagian lain di UIN Sunan Kalijaga Yogyakarta di antaranya adalah:

- LPPM
- LPM
- SPI
- Perpustakaan
- PTIPD
- Pusat Pengembangan Bahasa
- Pusat Pengembangan Bisnis
- Admisi
- Biro Administrasi Umum dan Keuangan
- Biro Administrasi Akademik, Kemahasiswaan dan Kerjasama

3.2 Analisis Permasalahan

Analisis permasalahan dilakukan dengan melakukan evaluasi terhadap proses yang selama ini berjalan. Selain itu, dilakukan pula diskusi dan wawancara dengan berbagai pihak yang terlibat diantaranya:

- Wawancara dengan anggota SPI
- Wawancara dengan Operator Unit/ Fakultas
- Wawancara dengan Bagian Keuangan
- Wawancara dengan Bagian Rumah Tangga
- Wawancara dengan PTIPD

Berdasarkan wawancara dengan anggota SPI, ditemukan permasalahan di antaranya sulitnya manajemen *folder* ketika menggunakan layanan Google Drive. Selain itu, sulit untuk mengecek satu per satu kelengkapan dokumen setiap kegiatan. Sehingga dengan adanya aplikasi baru, harapannya proses mencari *file* yang dibutuhkan dapat dilakukan dengan lebih mudah dan proses mengecek kelengkapan dokumen kegiatan menjadi lebih cepat.

3.3 Analisis Kebutuhan

Aplikasi pengawasan berbasis dokumen digital SPI UIN Sunan Kalijaga Yogyakarta adalah aplikasi yang digunakan untuk melaksanakan proses audit dokumen digital keuangan di UIN Sunan Kalijaga Yogyakarta. Proses analisis kebutuhan, dilakukan dengan cara membaca dokumen SPI UIN Sunan Kalijaga Yogyakarta dan dokumen atau peraturan dari Kementerian Keuangan Republik Indonesia.

Terdapat tiga jenis dokumen keuangan:

- Belanja Pegawai (MAK 51)
- Belanja Barang dan Jasa (MAK 52)
- Belanja Modal (MAK 53)

Belanja Pegawai (MAK 51) adalah belanja pegawai PNS yang meliputi:

- Gaji Pokok
- Tunjangan
- Uang Makan
- Lembur



Belanja Barang dan Jasa (MAK 52) adalah belanja Barang dan Jasa Termasuk di antaranya belanja Jasa Pegawai Non PNS. Belanja Barang dan Jasa (MAK 52) meliputi:

- Kegiatan
- Operasional Non Personil
 - Pengadaan dan Pemeliharaan
 - Perjalanan Dinas
- Operasional Personil
 - Gaji BLU
 - Tunjangan BLU
 - Uang Makan BLU
 - Lembur BLU

Belanja Modal (MAK 53) adalah belanja yang digunakan untuk pembelian barang modal seperti tanah, peralatan mesin, gedung, dan bangunan. Dalam melaksanakan proses audit MAK 53, SPI UIN Sunan Kalijaga Yogyakarta mengklasifikasikannya berdasarkan metode/cara pembelian dan jumlah/nominal uang yang dibelanjakan dalam setiap pembelian. Sehingga, Belanja Modal (MAK 53) terdiri dari:

- E-Purchasing
- Paket < 10 juta
- Paket 10-50 juta
- Paket 50-200 juta
- Paket > 200 juta

Masing-masing nominal memiliki kebutuhan dokumen yang berbeda-beda. Semakin besar nominal belanja modal, semakin banyak pula dokumen yang dibutuhkan.

Selain dilakukan pendalaman dokumen dan peraturan yang ada, proses analisis kebutuhan juga dilaksanakan dengan wawancara dan FGD (*Focus Group Discussion*) dengan pihak-pihak yang terlibat di antaranya:

- Kepala SPI
- Sekretaris SPI (Pengendali Teknis)
- Anggota SPI (Auditor)
- Operator Unit/ Fakultas
- Wakil Dekan Bidang II

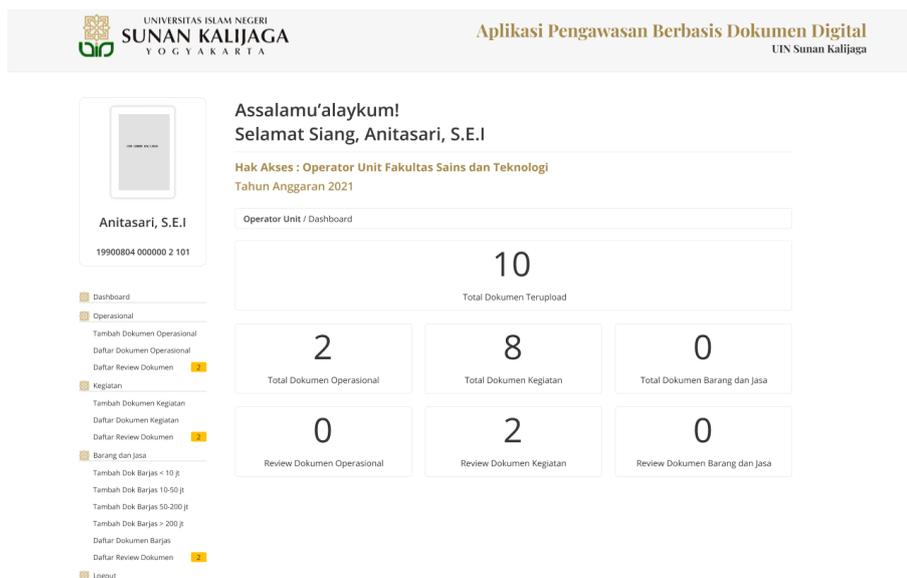
Berdasarkan hasil wawancara, ditemukan adanya kebutuhan aplikasi audit di antaranya:

- Operator unit/fakultas dapat membuat kegiatan baru.
- Operator unit/fakultas dapat melakukan *upload* dokumen untuk setiap jenis kegiatan.
- Masing-masing kegiatan memiliki karakteristik/jenis dokumen yang berbeda-beda.
- Anggota SPI (auditor) dapat melihat dokumen yang diunggah operator.
- Anggota SPI (auditor) dapat mengisi kertas kerja kegiatan dan kertas kerja rekomendasi sebagai *output* dari proses audit.
- Anggota SPI dapat melihat link dokumen kegiatan di samping form kertas kerja kegiatan.
- Kertas kerja kegiatan dan kertas kerja rekomendasi dapat ditanda-tangani oleh anggota SPI, ketua tim, pengendali teknis, dan Kepala SPI.
- Kertas kerja kegiatan dan kertas kerja rekomendasi dapat dicetak dalam bentuk pdf.
- Kertas kerja yang telah ditanda-tangani oleh Kepala SPI, dapat dilihat dan dibaca oleh pihak unit/fakultas.

3.4 Perancangan Logis

Pada tahap perancangan logis, peneliti memilih menterjemahkan keinginan *stakeholder* ke dalam tampilan antar muka yang didesain menggunakan layanan Figma. Hal ini dimaksudkan agar semua *stakeholder* lebih mudah membayangkan dan memahami seperti apa aplikasi baru yang nantinya akan dikembangkan. Gambar 2 adalah rancangan awal aplikasi pengawasan internal berbasis dokumen digital menggunakan aplikasi Figma.





Gambar 2 Rancangan *Dashboard* Aplikasi



Gambar 3 Rancangan Aplikasi

Peneliti bermaksud membuat desain antar muka aplikasi yang bertema menyerupai tampilan aplikasi lain di UIN Sunan Kalijaga Yogyakarta dengan tujuan agar pengguna merasa terbiasa dengan tampilan aplikasi seperti ini. Namun, ketika dipresentasikan kepada SPI UIN Sunan Kalijaga Yogyakarta, SPI UIN Sunan Kalijaga Yogyakarta menghendaki desain antar muka aplikasi yang lebih modern.

3.5 Telaah Kesimpulan

Tahap ini dilakukan dengan menyebarkan kuesioner *System Usability Scale (SUS)* (Brooke, 2013; Prabowo & Suprpto, 2021) kepada anggota SPI, operator unit/ fakultas, kepala unit/ lembaga, dan Wakil Dekan Bidang II Fakultas. Kuesioner tersebut digunakan untuk mengukur apakah desain aplikasi yang dibuat menggunakan Figma pada tahap sebelumnya layak atau tidak layak untuk dilanjutkan ke tahap selanjutnya. Hasil yang didapat dari tahap ini adalah desain tampilan aplikasi di atas termasuk dalam kategori OK sehingga layak untuk dikembangkan.

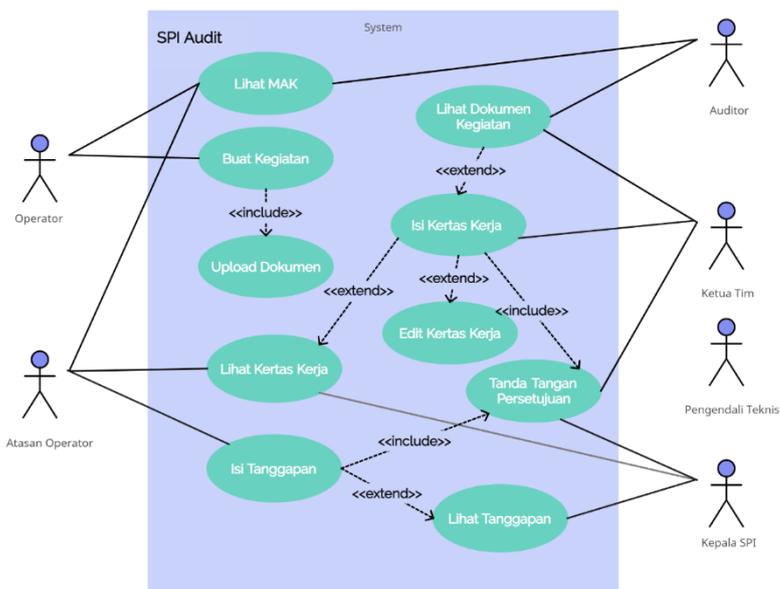
Tahap Telaah Kesimpulan juga dilakukan dengan cara menjalin komunikasi dengan PTIPD dan tim ASIKUSUKA. ASIKUSUKA adalah Aplikasi Sistem Keuangan UIN Sunan Kalijaga Yogyakarta yang dapat diakses di asik.uin-suka.ac.id. Berdasarkan hasil komunikasi yang dilakukan,



dimungkinkan adanya integrasi data antara ASIKUSUKA dengan Aplikasi Audit SPI dengan menggunakan teknologi API. Beberapa data yang mungkin dapat diambil/dimanfaatkan dari ASIKUSUKA adalah data MAK dan data serapan anggaran untuk setiap MAK.

3.6 Desain Logis

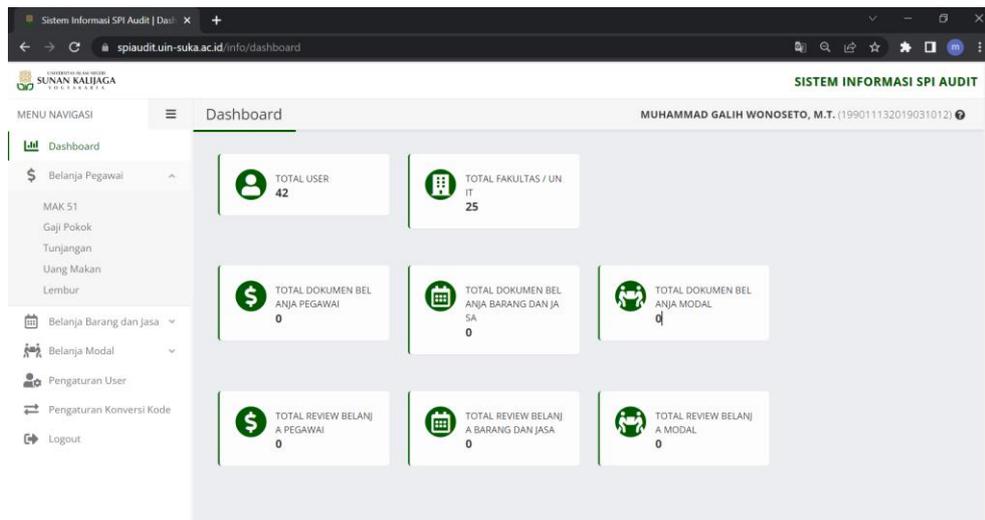
Tahap desain logis dilakukan dengan menggambarkan aplikasi yang akan dibuat dalam desain UML. Gambar 3 merupakan *use case diagram* aplikasi yang akan dibuat.



Gambar 4 Use Case Diagram Aplikasi Audit

3.7 Konstruksi dan Pengujian

Pada tahap ini dilakukan proses *coding* dan *testing*. Proses *coding* dilakukan dengan *framework* CodeIgniter. Proses *testing* dilakukan dengan tujuan untuk memastikan seluruh fungsionalitas aplikasi berjalan dengan lancar, tidak ada *bug*, dan memenuhi *requirement* dari SPI UIN Sunan Kalijaga Yogyakarta. Gambar 4 merupakan hasil *coding* aplikasi.



Gambar 5 Dashboard Aplikasi Audit SPI



Setelah aplikasi berhasil dibangun, selanjutnya dilakukan proses pengujian. Pengujian yang dilakukan di antaranya API *testing* dan *blackbox testing*. Berdasarkan hasil pengujian, aplikasi ini dinyatakan sudah memenuhi *minimum requirement* SPI UIN Sunan Kalijaga Yogyakarta dan siap untuk digunakan. Hasil API *testing* dan *blackbox testing* terdapat pada Tabel 1 dan 2.

Tabel 1 Hasil API *Testing*

No.	Deskripsi Pengujian	Hasil	Kesimpulan
1	Aplikasi dapat terhubung dengan data pegawai UIN Sunan Kalijaga Yogyakarta	Sesuai	Valid
2	Aplikasi dapat terhubung dengan data Mata Anggaran Kegiatan UIN Sunan Kalijaga Yogyakarta	Sesuai	Valid

Tabel 2 Hasil *Blackbox Testing*

No.	Deskripsi Pengujian	Hasil	Kesimpulan
1	Pengguna dapat melakukan login menggunakan akun Single Sing On universitas	Sesuai	Valid
2	Operator dapat melihat menu Belanja Pengawa, Belanja Barang dan Jasa dan Belanja Modal	Sesuai	Valid
3	Operator dapat membuat kegiatan	Sesuai	Valid
4	Operator dapat upload dokumen kegiatan	Sesuai	Valid
5	Auditor dapat melihat dokumen kegiatan yang diupload operator	Sesuai	Valid
6	Auditor dapat mengisi Kertas Kerja Kegiatan	Sesuai	Valid
7	Auditor dapat mengisi Kertas Kerja Rekomendasi	Sesuai	Valid
8	SPI dapat menandatangani Kertas Kerja Kegiatan	Sesuai	Valid
9	SPI dapat menandatangani Kertas Kerja Rekomendasi	Sesuai	Valid
10	Kertas Kerja yang telah ditandatangani Ketua SPI dapat dilihat oleh unit/ fakultas	Sesuai	Valid
11	Unit/ fakultas dapat membuat tanggapan atas hasil audit SPI	Sesuai	Valid
12	SPI dapat menerima tanggapan dari unit/ fakultas	Sesuai	Valid

3.8 Instalasi dan Pengiriman

Instalasi dilaksanakan selama pengembangan di server PTIPD UIN Sunan Kalijaga Yogyakarta di alamat spiaudit.uin-suka.ac.id. Aplikasi ini dikirimkan atau diserahkan-terimakan dengan SPI UIN Sunan Kalijaga Yogyakarta melalui mekanisme rapat/pertemuan internal dan presentasi penggunaan aplikasi. Pada pertemuan ini juga sekaligus dilakukan pelatihan penggunaan aplikasi kepada anggota SPI UIN Sunan Kalijaga Yogyakarta. Dalam pertemuan ini, aplikasi yang dibuat dinyatakan telah memenuhi kebutuhan SPI UIN Sunan Kalijaga Yogyakarta dalam melaksanakan proses audit internal berbasis dokumen digital.

4. KESIMPULAN

Dari penelitian ini dapat disimpulkan bahwa aplikasi audit kegiatan berbasis dokumen digital berhasil dibuat sesuai dengan *requirement* yang diminta oleh SPI UIN Sunan Kalijaga Yogyakarta, di antaranya:

- Operator dapat membuat kegiatan pada setiap jenis MAK.
- Operator unit/fakultas dapat melakukan *upload* dokumen untuk setiap jenis kegiatan (MAK) yang berbeda-beda.
- Anggota SPI (auditor) dapat melihat dokumen yang diunggah oleh operator.
- Anggota SPI (auditor) dapat mengisi kertas kerja kegiatan dan kertas kerja rekomendasi sebagai *output* dari proses audit.
- Anggota SPI dapat melihat link dokumen kegiatan di samping form kertas kerja kegiatan
- Kertas kerja kegiatan dan kertas kerja rekomendasi dapat ditanda-tangani oleh anggota SPI, ketua tim, pengendali teknis dan Kepala SPI



- Kertas kerja kegiatan dan kertas kerja rekomendasi dapat dicetak dalam bentuk pdf.
- Kertas kerja yang telah ditanda-tangani oleh Kepala SPI, dapat dilihat dan dibaca oleh pihak unit/fakultas.
- Unit/fakultas dapat merespon/memberi tanggapan atas hasil kertas kerja dari SPI.

Aplikasi ini juga dapat diduplikasi untuk digunakan di universitas lain atau instansi pemerintah lain karena dirancang berdasarkan peraturan Kementerian Keuangan tentang kelengkapan dokumen laporan keuangan untuk setiap jenis kegiatan.

UCAPAN TERIMA KASIH

Terima kasih LPPM UIN Sunan Kalijaga Yogyakarta yang telah mendanai penelitian ini. Terima kasih kepada PTIPD UIN Sunan Kalijaga Yogyakarta yang telah mendukung dan memfasilitasi penelitian ini.

DAFTAR PUSTAKA

- Akmalia, I., & Ariani, N. E. (2022). Pengaruh Teknik Audit Berbantuan Komputer (TABK), Integritas, dan Kompetensi Auditor Terhadap Kualitas Audit BPK RI Perwakilan Aceh. *Jurnal Ilmiah Mahasiswa Ekonomi Akuntansi*, 7(1), 34–44. <https://doi.org/10.24815/jimeka.v7i1.20241>
- Arif Nurrahman, A., Pizary Husen, N., & Rukmana, O. (2020). Designing Information System for Student Practicum Assessment in the Laboratory. *IOP Conference Series: Materials Science and Engineering*, 847(1), 012047. <https://doi.org/10.1088/1757-899X/847/1/012047>
- Azmi, M., Sonatha, Y., Asri, E., Rasyidah, & Putra, D. S. (2018). Implementing FAST Method on the Development of Object-Oriented Cooperative Information Systems. *JOIV : International Journal on Informatics Visualization*, 2(4–2), 366. <https://doi.org/10.30630/joiv.2.4-2.189>
- Brooke, J. (2013). SUS: a retrospective. *Journal of Usability Studies*, 8(2), 29–40. <https://doi.org/10.5555/2817912.2817913>
- Damanik, A., Handayani, P. W., & Pinem, A. A. (2020). The Barriers of IT Utilization: A Case Study of Indonesian Audit Organisation. *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)*, 280–285. <https://doi.org/10.1109/IC2IE50715.2020.9274664>
- Dilson, & Suhery, L. (2016). Sistem Informasi Pengawasan Internal Inspektorat Berbasis Client Server. *Elkawanie: Journal of Islamic Science and Technology*, 2(1), 89–108. <https://jurnal.ar-raniry.ac.id/index.php/elkawanie/article/view/2520>
- Halim, R. M. N. (2020). Sistem Informasi Penjualan Pada TB Harmonis Menggunakan Metode FAST. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(2), 203–207. <https://doi.org/10.32736/sisfokom.v9i2.868>
- Januraga, I. K., & Budiarta, I. K. (2015). Pengaruh Teknik Audit Berbantuan Komputer, Kompetensi Auditor, dan Kecerdasan Spiritual pada Kualitas Audit BPK Bali. *E-Jurnal Akuntansi*, 13(3), 1137–1163. <https://ojs.unud.ac.id/index.php/akuntansi/article/view/14469>
- JB, O., & AA, O. (2015). Computer Assisted Audit Techniques and Audit Quality in Developing Countries: Evidence from Nigeria. *The Journal of Internet Banking and Commerce*, 20(3). <https://doi.org/10.4172/1204-5357.1000127>
- Peraturan Menteri Agama tentang Satuan Pengawas Internal Pada Perguruan Tinggi Keagamaan Negeri, Pub. L. No. 25, Pemerintah Indonesia (2017). <https://peraturan.bpk.go.id/Home/Details/131267/peraturan-menag-no-25-tahun-2017>
- Peraturan Pemerintah (PP) tentang Sistem Pengendalian Intern Pemerintah, Pub. L. No. 60, Pemerintah Indonesia (2008). <https://peraturan.bpk.go.id/Home/Details/4876>
- Prabowo, M., & Suprpto, A. (2021). Usability Testing pada Sistem Informasi Akademik IAIN Salatiga Menggunakan Metode System Usability Scale. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(1), 38–49. <https://doi.org/10.14421/jiska.2021.61-05>
- Ruhatudin, B., Sriharini, Ridha, M. A., Nurochman, Rohinah, Kusumowardhani, R. P. A., Shaleh, Suswini, & Anitasari. (2019). *Pengawasan Internal Berbasis Dokumen Digital Pedoman Teknis*. Suka Press.



- Whitten, J. L., & Bentley, L. D. (2007). *System Analysis and Design Methods* (7th ed.). McGraw-Hill Irwin.
https://www.academia.edu/8787830/Whitten_and_Bentley_2007_System_Analysis_and_Design_Methods_7th_Edition
- Wicaksono, A., Laurens, S., & Novianti, E. (2018). Impact Analysis of Computer Assisted Audit Techniques Utilization on Internal Auditor Performance. *2018 International Conference on Information Management and Technology (ICIMTech)*, 267–271.
<https://doi.org/10.1109/ICIMTech.2018.8528198>
- Wonoseto, M. G., Pangestuti, D. F. R., Asyari, M., & Anggraeni, N. (2022). Evaluasi Pengawasan Berbasis Dokumen Digital SPI UIN Sunan Kalijaga Yogyakarta. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 8(2), 130. <https://doi.org/10.24014/rmsi.v8i2.16871>



Analisis Klasifikasi *Broken Home* pada Anak Menggunakan Metode Naïve Bayes Classifier

Supiyandi ⁽¹⁾, Almann Hussein ⁽²⁾, Irwan Gunawan ⁽³⁾, William Lutfi Rahman Harjo ^{(4)*}

¹ Sistem Komputer, Fakultas Ilmu Komputer, Universitas Pembangunan Panca Budi, Medan

^{2,3,4} Ilmu Komputer, Fakultas Sains dan Teknologi, UIN Sumatera Utara, Medan

e-mail : supiyandi@dosen.pancabudi.ac.id,

{almanna0701201212,irwan0701201002,william0701202183}@uinsu.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 21 Desember 2022, direvisi 24 Maret 2023, diterima 25 Maret 2023, dan dipublikasikan 26 Mei 2023.

Abstract

Broken home is a term that defines a situation in a family where most people handle no harmony, happiness, or peace. The impact of a broken home on a depressed family on children who can experience mental, emotional, and behavioral changes that are uncontrolled and undirected. Therefore, a classification is needed to categorize a child in a family as a broken home or not. The classification process will apply the Naïve Bayes Classifier classification method by taking into account the factors that refer to the statement that a child is called a broken home. With this classification, it is hoped that it can help know what and how a broken home child can be called a broken home and with this classification, it is expected that parents can minimize broken homes in children in the future by paying attention to the determining factors.

Keywords: *Broken Home, Classification, Naïve Bayes, Children, Families, Youth*

Abstrak

Broken home merupakan istilah yang mendefinisikan pada suatu keadaan dalam keluarga yang tidak adanya keharmonisan, kebahagiaan, serta ketentraman yang diharapkan dari sebagian besar anak di masyarakat, pada dasarnya dampak dari *broken home* pada keluarga tertuju pada anak yang dapat mengalami perubahan mental, emosi, serta perilaku yang tidak terkendali dan tidak terarah dikarenakan keluarga yang dimilikinya sendiri juga tidak harmonis dan kurang berkomunikasi dengan sesama anggota keluarganya, maka dari itu diperlukan klasifikasi untuk mengkategorikan suatu anak pada keluarga dikatakan *broken home* atau tidak, proses klasifikasi akan menerapkan metode klasifikasi *supervised learning* berupa Naïve Bayes *classifier* dengan memperhatikan faktor-faktor yang memicu terjadinya *broken home*, adapun manfaat dari penelitian ini adalah dapat membantu masyarakat untuk mengetahui bagaimana seorang anak dikatakan *broken home*, serta tujuan pada penelitian ini untuk menciptakan suatu model klasifikasi pada permasalahan yang diamati dan berdasarkan metode yang digunakan.

Kata Kunci: *Broken Home, Klasifikasi, Naïve Bayes, Anak, Keluarga, Remaja*

1. PENDAHULUAN

Keluarga merupakan rumah pertama atau tempat pertama bagi seorang anak dilahirkan dan berkembang, keluarga merupakan tempat untuk memperoleh kasih sayang dan kebahagiaan yang menjadi hak yang wajib dimilikinya, seorang anak seharusnya mendapatkan kasih sayang dari orang tuanya dengan penuh ketulusan tanpa adanya keterpaksaan, tetapi beberapa kasus di masyarakat yang terjadi banyak keluarga yang yang tidak harmonis dan sering terjadi keributan dalam waktu ke waktu, hal inilah yang dapat memicu mental dan emosi, serta perilaku pada anak yang tiap hari semakin tidak terkendali. Gunarsa (2008) mengemukakan “orang tua yang paling bertanggung jawab dalam perkembangan seluruh eksistensi anak”.

Sebagian besar masyarakat di Indonesia menyebutkan bahwa situasi anak yang tidak bahagia dengan sebutan *Broken Home*, istilah ini sering kali digunakan oleh masyarakat di mana “*broken home* merupakan situasi serta kondisi keluarga yang tidak terdapat keharmonisan sebagaimana banyak diharapkan orang lain.” (Wardani, 2021). John dalam Mistiani (2018) menyebutkan bahwa



“*broken home* berasal dari dua kata yaitu *broken* dan *home*, *broken* berasal dari kata *break* yang berarti keretakan, sedangkan *home* mempunyai arti rumah atau rumah tangga.”

Pada dasarnya istilah *broken home* mengacu pada kekacauan dalam sebuah keluarga (Massa et al., 2020) bisa juga didefinisikan sebagai situasi serta kondisi suatu keluarga yang tidak adanya ketentraman, dan kedekatan antara anak dan orang tua, komunikasi yang baik serta keharmonisan di dalamnya yang sangat berpengaruh pada seorang. Menurut Aziz (2015) “kondisi rumah tangga yang *broken* sering anak-anak mengalami depresi mental (tekanan mental),” sehingga kondisi *broken home* sangat mempengaruhi mental anak. Menurut Mistiani (2018) mengemukakan bahwa “dampak *broken home* bukan hanya kepada anak-anak saja, namun juga memiliki pengaruh yang besar terhadap remaja apalagi terhadap psikologinya. Hurlock dalam Ahyani & Astuti (2018) mendefinisikan bahwa “remaja berasal dari kata latin *adolescence* yang berarti tumbuh atau tumbuh menjadi dewasa. Istilah *adolescence* mempunyai arti yang lebih luas lagi yang mencakup kematangan mental, emosional sosial dan fisik.”

Keluarga merupakan lingkungan pertama bagi perkembangan emosional anak (Aisyah et al., 2022), disebabkan emosional yang berubah-ubah pada anak yang didasarkan pada keluarga yang tidak harmonis justru membuat anak semakin tidak terkontrol dan tidak terkendali. Orang tua seharusnya memberikan kasih sayang, ketulusan, serta perlindungan kepada sang anak yang memang menjadi kewajibannya sebagai orang tua. Sebagian besar anak yang berada pada keluarga yang bercerai atau tidak harmonis banyak keburukan dan ketidakadilan yang dirasakan serta tidak merasakan suatu kebahagiaan. Seligman dalam Faizah (2022) mendefinisikan bahwa:

“Kebahagiaan merupakan perasaan positif yang akan mendorong seseorang untuk melakukan berbagai tindakan yang positif. Kebahagiaan sebagai konsep yang mengacu pada emosi positif yang dirasakan individu serta aktivitas positif yang tidak memiliki komponen perasaan negatif, misalnya ketika individu terlibat dalam kegiatan yang sangat disukai.”

Seorang anak wajib mendapatkan kebahagiaan di mana “kebahagiaan adalah kesenangan dan ketenteraman hidup lahir batin, keberuntungan, dan kemujuran yang bersifat lahir batin” (Rahayu, 2016) untuk kelangsungan hidupnya pada masa kanak-kanak atau bahkan masa remaja. Hal tersebut disebabkan “kebahagiaan merupakan harapan bagi setiap orang termasuk remaja atau seorang anak, akan tetapi perceraian antara kedua orang tua juga dapat berdampak pada kebahagiaan seorang remaja atau anak.” (Muttaqin et al., 2019).

Kebahagiaan pada anak *broken home* termasuk relatif atau berbeda-beda per individu tergantung dengan kondisi keluarga serta sikap diri ketika merasakan dan menyikapinya, kebahagiaan identik dengan hal yang menyenangkan. Menurut Seligman dalam Faizah (2022) terdapat dua faktor yang mempengaruhi kebahagiaan di antaranya “faktor internal kebahagiaan, meliputi, kepuasan akan masa lalu, optimis akan masa depan, kebahagiaan pada masa sekarang, faktor eksternal, meliputi uang, perkawinan, kehidupan sosial, emosi negatif, usia, kesehatan, pendidikan, iklim, ras, dan jenis kelamin, serta agama.” Umami (2019) mengungkapkan “bahwa tingkah laku negatif pada diri remaja disebabkan adanya perlakuan lingkungan yang kurang sesuai dengan tuntutan atau kebutuhan perkembangan remaja.”

Broken home sangat berperan penting dalam perkembangan perilaku seorang anak dan sangat memiliki efek yang tidak baik atau negatif bagi seorang anak di dalam keluarga, di mana dampak inilah yang didasarkan pada perceraian kedua orang tua atau keluarga yang tidak harmonis serta komunikasi dalam keluarga yang tidak terjalin dengan baik, dampak perceraian anak sangat beragam di mana Maryanti dalam Hafiza & Mawarpury (2018) menyebutkan bahwa:

“Dampak perceraian bagi anak di antaranya anak menjadi mudah marah, frustrasi, dan ingin melampiaskannya dengan melakukan hal-hal yang berlawanan dengan peraturan-peraturan seperti memberontak dan lain sebagainya. Selain itu, bila anak tinggal dengan ibu, anak akan kehilangan figur otoritas ayah. Ketika figur otoritas itu menghilang, anak

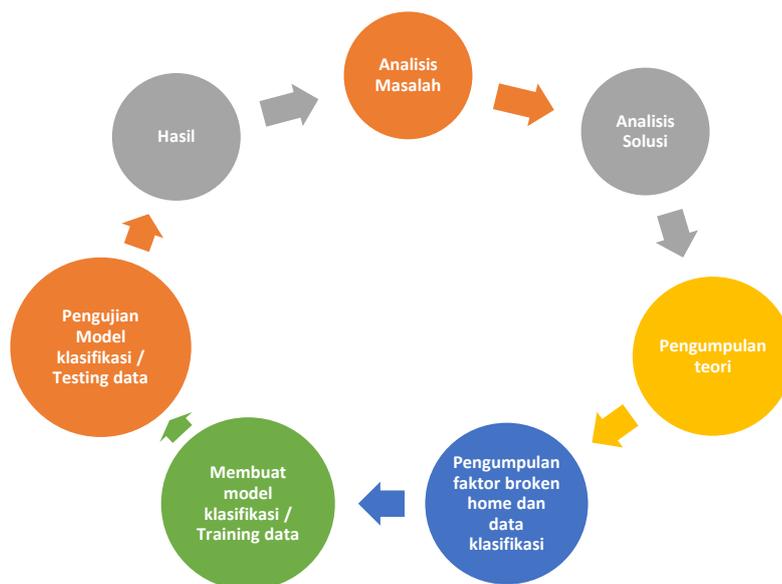


seringkali tidak begitu takut dengan ibunya. Dampak lain adalah anak menjadi kehilangan jati diri sosialnya atau identitas sosial, mendapatkan status sebagai anak cerai memberikan suatu perasaan berbeda dari anak-anak lain.”

Untuk mengurangi tingkat kecenderungan pada *broken home* anak serta dapat mengetahui seorang anak dikategorikan ke dalam *broken home* atau tidak, maka dari itu diperlukan suatu cara pemilihan keputusan atau klasifikasi untuk mengkasifikasikan seorang anak dikatakan *broken home*. Proses klasifikasi yang akan dibuat akan menggunakan sebuah metode yang di sebut dengan Naïve Bayes *classifier* dengan mengacu dan memperhatikan beberapa faktor-faktor penentu yang terdapat pada istilah *broken home*, klasifikasi ini nantinya akan menggunakan perhitungan Teorema Bayes untuk menghitung dari tiap-tiap faktor penentu pada *broken home*. Adapun tujuan dari klasifikasi ini adalah untuk mengetahui apakah sistem dapat mengklasifikasikan atau mengkategorikan *broken home* pada seorang anak di dalam keluarga. Manfaat yang didapatkan dari penelitian ini adalah dapat menganalisis dan memperhatikan *broken home* pada seorang anak dalam keluarga serta dapat menjaga dan mencegah *broken home* lainnya di luar sana dengan mengacu dan memperhatikan pada hasil klasifikasi ini. Harapannya tingkat *broken home* pada anak dapat berkurang dan segera diatasi lebih lanjut oleh ahli terkait.

2. METODE PENELITIAN

Tahapan penelitian merupakan alur dari serangkaian penelitian yang dilakukan, guna untuk menentukan tahap-tahap yang akan dikerjakan dalam sebuah penelitian. Dalam hal ini tahapan penelitian yang dimaksud dapat dilihat pada Gambar 1.



Gambar 1 Tahapan Penelitian

Untuk memperjelas pemahaman mengenai tahapan penelitian pada Gambar 1, maka dengan ini penulis menguraikan rincian dari tahapan penelitian sebagai berikut:

1) Analisis Masalah

Mencari, mengamati, dan memperhatikan permasalahan yang sedang terjadi dengan mengidentifikasi hal-hal yang belum diketahui solusinya dari permasalahan tersebut.

2) Analisis Solusi

Menentukan solusi yang dapat digunakan untuk menyelesaikan permasalahan yang telah terjadi berdasarkan pengamatan dari permasalahan.



3) Pengumpulan Teori

Mengumpulkan beberapa teori pendukung serta referensi yang terkait dengan permasalahan yang sedang diamati dengan cara studi literatur (mengumpulkan dari beberapa jurnal).

4) Pengumpulan Faktor *Broken Home* dan Data Klasifikasi

Memproleh faktor-faktor terjadinya *broken home* yang didapatkan pada beberapa referensi jurnal, serta data yang digunakan untuk klasifikasi dengan teknik pengumpulan data berupa angket atau kuesioner ke beberapa responden di masyarakat.

5) Membuat Model Klasifikasi/*Training Data*

Membuat model klasifikasi dengan menggunakan metode Naïve Bayes *Classifier* pada data latih (*training data*).

6) Pengujian Model Klasifikasi/*Testing Data*

Melakukan pengujian model klasifikasi yang telah dibuat dengan *testing data*.

7) Hasil

Mendapatkan hasil klasifikasi berupa prediksi serta *performance* atau akurasi dari model klasifikasi yang digunakan.

2.1 Klasifikasi

Klasifikasi merupakan cara pengelompokkan benda berdasarkan ciri-ciri yang dimiliki oleh objek klasifikasi. Dalam prosesnya, klasifikasi dapat dilakukan dengan banyak cara baik secara manual ataupun dengan bantuan teknologi. Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma cerdas komputer. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, di antaranya Naïve Bayes, Support Vector Machine, Decision Tree, Fuzzy, dan Jaringan Saraf Tiruan (Wibawa et al., 2018). Alasan peneliti memilih metode klasifikasi Naive Bayes disebabkan pada penelitian sebelumnya yang berjudul “*Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes*” mendapatkan tingkat akurasi sebesar 73% (Annur, 2018), pada penelitian berjudul “*Analysis of Family Health with a Combination of Naive Bayes KNN Methods*” dengan tingkat akurasi sebesar 91,24% (Maliki & Nanja, 2023), dan pada penelitian berjudul “*Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan*” dengan tingkat akurasi sebesar 92% (Putro et al., 2020). Berdasarkan ketiga penelitian sebelumnya peneliti yakin untuk memilih metode tersebut yang akan digunakan untuk klasifikasi *broken home* pada anak dalam penelitian ini.

2.2 Naïve Bayes Classifier

Bayesian *classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Bayesian *classification* didasarkan pada Teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* (Annur, 2018). Dengan menggunakan metode ini penulis dapat mengklasifikasikan *broken home* pada anak dengan melihat faktor-faktor yang mengacu pada *broken home*, mengacu pada faktor-faktor yang telah diketahui, dan akan dilanjutkan untuk membuat sebuah model berupa hasil perhitungan probabilitas pada data yang telah ditentukan dengan menggunakan persamaan yang mengacu pada Pers. (1).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Di mana X merupakan data dengan *class* yang belum diketahui dan H merupakan hipotesis data X merupakan suatu *lass* spesifik. Sehingga P(H|X) adalah probabilitas hipotesis H berdasarkan kondisi X, P(H) adalah probabilitas hipotesis H, P(X|H) adalah probabilitas X berdasarkan kondisi tertentu, dan P(X) adalah probabilitas dari X.



3. HASIL DAN PEMBAHASAN

3.1 Analisis Masalah

Permasalahan yang terjadi adalah sulitnya mengidentifikasi seorang anak terutama remaja dikategorikan sebagai *broken home* atau tidak. Mengingat perilaku seorang anak berbeda-beda serta tidak seluruh anak dan remaja dapat dengan mudah bercerita kepada keluarganya atau orang lain perihal apa yang dialaminya. Ketika seorang anak tidak dapat diketahui *broken home* atau tidak, maka dikemudian hari hal ini dapat berdampak pada perilaku, pola pikir, serta emosional pada anak itu sendiri akibat tidak dapatnya perlakuan khusus, pemahaman, dan psikologi yang dapat membantunya dalam mengatasi *broken home* tersebut.

3.2 Analisis Solusi

Klasifikasi diperlukan untuk mengkategorikan suatu data, dalam hal ini mengacu pada sumber masalah yang terjadi diperlukan suatu klasifikasi yang dapat mengklasifikasikan atau mengkategorikan suatu anak atau remaja tergolong dengan *broken home* atau tidak. Dengan mengetahui dan memperhatikan sebagian besar anak-anak atau remaja tergolong *broken home* nantinya akan diberikan penanganan lebih lanjut oleh ahli yang terkait guna mendidik emosional, karakter, perilaku, dan sikap serta pola pikir anak-anak *broken home* tersebut. Hal ini diharapkan agar dikemudian hari tidak terjadi hal-hal yang menyimpang yang terjadi dan dialami oleh anak-anak *broken home* di luar sana.

3.3 Pengumpulan Teori

Teori diperlukan untuk mengkaji suatu permasalahan dan menambah literatur guna mendapatkan hipotesis yang kuat dan wawasan yang akurat. Dalam hal ini penulis mengumpulkan beberapa teori berdasarkan beberapa jurnal yang telah dicari melalui sumber internet seperti Google Cendekia.

3.4 Pengumpulan Faktor *Broken Home* dan Data Klasifikasi

Sebelum membuat klasifikasi, penulis mengumpulkan beberapa faktor-faktor penentu yang memicu terjadinya *broken home*. Faktor ini didapat berdasarkan beberapa jurnal yang penulis kumpulkan. Adapun faktor yang dimaksud adalah faktor uang atau ekonomi (Faizah, 2022), kekerasan dalam rumah tangga atau KDRT (Aisyah et al., 2022), orang tua selingkuh (Mistiani, 2018), dan orang tua bercerai (Hafiza & Mawarpury, 2018). Faktor-faktor *broken home* yang dimaksud dapat dilihat pada Tabel 1.

Tabel 1 Data Faktor-Faktor yang Mengacu Terjadinya *Broken Home*

Faktor Broken Home	Kategori pilihan
Ekonomi	Rendah, menengah, atas
Kekerasan dalam rumah tangga (KDRT)	Tidak pernah, jarang, sering
Orang tua selingkuh	Pernah selingkuh, tidak pernah selingkuh
Orang tua bercerai	Bercerai, tidak bercerai

Faktor-faktor penentu *broken home* bisa saja lebih dari ini namun, keempat faktor ini yang umumnya terjadi pada seorang anak yang mengalami *broken home*. Masing-masing faktor penentunya memiliki parameter yang dirasakan tergantung dari apa yang dirasakan dan didapatkan pada seorang anak dirinya sendiri. Lalu data ini nantinya akan digunakan untuk membuat sebuah model klasifikasi guna mendapatkan hasil klasifikasi yang diinginkan sesuai dengan data pada Tabel 1.

3.5 Membuat Model Klasifikasi atau *Training Data*

Data faktor yang telah diketahui akan digunakan untuk membuat model klasifikasi yang telah direncanakan. Data yang dimaksud dapat dilihat pada Tabel 1. Adapun kriteria klasifikasi *broken*



home akan menetapkan 2 kategori kelas klasifikasi yaitu kelas Broken Home dan kelas Tidak Broken Home. Adapun data yang digunakan untuk klasifikasi didapat dari kuesioner yang telah penulis sebarakan ke beberapa responden di masyarakat. Totalnya adalah terdapat 36 data (36 responden) yang mana data ini akan dibagi menjadi 2 jenis yaitu 90% (32 data) untuk *training data* dan 10% (4 data) untuk *testing data*. Adapun *training data* yang didapat berdasarkan responden dapat dilihat pada Tabel 2.

Tabel 2 Data Latih atau *Training Data*

Data Ke-	Ekonomi	KDRT	Orang Tua Selingkuh	Orang Tua Bercerai	Label
1	Rendah	Sering KDRT	Pernah Selingkuh	Bercerai	Broken Home
2	Rendah	Jarang KDRT	Pernah Selingkuh	Bercerai	Broken Home
3	Rendah	Tidak Pernah KDRT	Pernah Selingkuh	Bercerai	Broken Home
4	Rendah	Sering KDRT	Tidak Selingkuh	Bercerai	Broken Home
5	Rendah	Jarang KDRT	Tidak Selingkuh	Bercerai	Broken Home
6	Rendah	Tidak Pernah KDRT	Tidak Selingkuh	Bercerai	Broken Home
7	Rendah	Sering KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home
8	Rendah	Jarang KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home
9	Rendah	Tidak Pernah KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home
10	Rendah	Sering KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home
11	Menengah	Sering KDRT	Pernah Selingkuh	Bercerai	Broken Home
12	Menengah	Jarang KDRT	Pernah Selingkuh	Bercerai	Broken Home
13	Menengah	Tidak Pernah KDRT	Pernah Selingkuh	Bercerai	Broken Home
14	Menengah	Sering KDRT	Tidak Selingkuh	Bercerai	Broken Home
15	Menengah	Tidak Pernah KDRT	Tidak Selingkuh	Bercerai	Broken Home
16	Menengah	Sering KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home
17	Menengah	Jarang KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home
18	Menengah	Tidak Pernah KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home
19	Menengah	Sering KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home
20	Menengah	Jarang KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home
21	Menengah	Tidak Pernah KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home
22	Atas	Sering KDRT	Pernah Selingkuh	Bercerai	Broken Home
23	Atas	Jarang KDRT	Pernah Selingkuh	Bercerai	Broken Home
24	Atas	Tidak Pernah KDRT	Pernah Selingkuh	Bercerai	Broken Home
25	Atas	Sering KDRT	Tidak Selingkuh	Bercerai	Broken Home



26	Atas	Jarang KDRT	Tidak Selingkuh	Bercerai	Broken Home
27	Atas	Tidak Pernah KDRT	Tidak Selingkuh	Bercerai	Broken Home
28	Atas	Sering KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home
29	Atas	Tidak Pernah KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home
30	Atas	Sering KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home
31	Atas	Jarang KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home
32	Atas	Tidak Pernah KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home

Selanjutnya, dengan mengikuti Teorema Naïve Bayes pada Pers. (1), didapatkan model klasifikasi yang mengacu pada 32 data kategori klasifikasi pada Tabel 2. Adapun model klasifikasi secara manual dapat dilihat pada Pers. (2) sebagai berikut.

Klasifikasi:

$$P(\text{BrokenHome}) = \frac{17}{32}$$

$$P(\text{TidakBrokenHome}) = \frac{15}{32}$$

Ekonomi:

$$P(\text{Rendah}|\text{BrokenHome}) = \frac{6}{17}$$

$$P(\text{Rendah}|\text{TidakBrokenHome}) = \frac{4}{15}$$

$$P(\text{Menengah}|\text{BrokenHome}) = \frac{5}{17}$$

$$P(\text{Menengah}|\text{TidakBrokenHome}) = \frac{6}{15}$$

$$P(\text{Atas}|\text{BrokenHome}) = \frac{6}{17}$$

$$P(\text{Atas}|\text{TidakBrokenHome}) = \frac{5}{15}$$

KDRT:

$$P(\text{TidakpernahKDRT}|\text{BrokenHome}) = \frac{6}{17}$$

$$P(\text{TidakpernahKDRT}|\text{TidakBrokenHome}) = \frac{5}{15}$$

$$P(\text{JarangKDRT}|\text{BrokenHome}) = \frac{5}{17}$$

$$P(\text{JarangKDRT}|\text{TidakBrokenHome}) = \frac{4}{15}$$

$$P(\text{SeringKDRT}|\text{BrokenHome}) = \frac{6}{17}$$

$$P(\text{SeringKDRT}|\text{TidakBrokenHome}) = \frac{6}{15}$$

Orang Tua Selingkuh:

$$P(\text{PernahSelingkuh}|\text{brokenHome}) = \frac{9}{17}$$

$$P(\text{PernahSelingkuh}|\text{tidakBrokenHome}) = \frac{8}{15}$$

$$P(\text{TidakSelingkuh}|\text{brokenHome}) = \frac{8}{17}$$

$$P(\text{TidakSelingkuh}|\text{TidakBrokenHome}) = \frac{7}{15}$$

Orang Tua Bercerai:

$$P(\text{Bercerai}|\text{BrokenHome}) = \frac{17}{17}$$

$$P(\text{Bercerai}|\text{TidakbrokenHome}) = \frac{0}{15}$$



$$P(\text{TidakBercerai}|\text{BrokenHome}) = \frac{0}{17}$$

$$P(\text{TidakBercerai}|\text{TidakBrokenHome}) = \frac{15}{15} \quad (2)$$

Menurut Mustofa & Mahfudh (2019) “Naïve Bayes Classifier atau NBC merupakan suatu proses pengklasifikasian probabilitas sederhana yang mengacu pada *theory* Bayes.” Maka dengan menggunakan teorema Bayes, di ketahui seperti yang telah diuraikan di atas bahwasanya masing-masing faktor penyebab *broken home* dihitung berdasarkan masing-masing kelas, baik itu kelas Broken Home atau kelas Tidak Broken Home, didapat beragam perhitungan yang akan dipakai sebagai model klasifikasi yang akan digunakan.

3.6 Pengujian Model Klasifikasi atau *Testing Data*

Untuk mengetahui apakah model klasifikasi yang telah dibuat dapat digunakan, diperlukan pengujian terlebih dahulu. Adapun pengujian model klasifikasi ini akan menggunakan paling sedikit 10% dari 36 data yang telah diketahui. Artinya terdapat 4 data responden yang akan diuji. Adapun *testing data* yang dimaksud dapat dilihat pada Tabel 3.

Tabel 3 Data Uji atau *Testing Data*

Data Ke-	Ekonomi	KDRT	Orang Tua Selingkuh	Orang Tua Bercerai	Label Mutlak	Label Prediksi
1	Rendah	Jarang KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home	Tidak Broken Home
2	Rendah	Tidak Pernah KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home	Tidak Broken Home
3	Menengah	Jarang KDRT	Tidak Selingkuh	Bercerai	Broken Home	Broken Home
4	Atas	Jarang KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home	Tidak Broken Home

Selanjutnya, perhitungan manual teorema Bayes dilakukan untuk mengetahui klasifikasi yang akan didapatkan. Adapun hasil klasifikasi data Tabel 3 pada baris 1 dapat dilihat pada persamaan Pers. (3), data baris 2 pada Pers. (4), data baris 3 pada Pers. (5), dan data baris 4 pada Pers. (6).

$$P(\text{Data1}|\text{BrokenHome})$$

$$= P(\text{Rendah}|\text{BrokenHome}) \times P(\text{JarangKDRT}|\text{BrokenHome}) \times P(\text{TidakSelingkuh}|\text{BrokenHome}) \times P(\text{TidakBercerai}|\text{BrokenHome})$$

$$= \frac{6}{17} \times \frac{5}{17} \times \frac{8}{17} \times \frac{0}{17}$$

$$= 0,352 \times 0,294 \times 0,470 \times 0$$

$$= 0$$

$$P(\text{Data1}|\text{TidakBrokenHome})$$

$$= P(\text{Rendah}|\text{TidakBrokenHome}) \times P(\text{JarangKDRT}|\text{TidakBrokenHome}) \times P(\text{TidakSelingkuh}|\text{TidakBrokenHome}) \times P(\text{TidakBercerai}|\text{TidakBrokenHome})$$

$$= \frac{4}{15} \times \frac{4}{15} \times \frac{7}{15} \times \frac{15}{15}$$

$$= 0,266 \times 0,266 \times 0,466 \times 1$$

$$= 0,032 \quad (3)$$

Mengacu pada Pers. (3), untuk data ke-1, kelas Broken Home bernilai 0 dan kelas Tidak Broken Home bernilai 0,032, maka hasil klasifikasinya adalah **Tidak Broken Home**.



$$\begin{aligned}
 &P(\text{Data2}|\text{BrokenHome}) \\
 &= P(\text{Rendah}|\text{BrokenHome}) \times P(\text{TidakPernahKDRT}|\text{BrokenHome}) \times \\
 &P(\text{TidakSelingkuh}|\text{BrokenHome}) \times P(\text{TidakBercerai}|\text{BrokenHome}) \\
 &= \frac{6}{17} \times \frac{6}{17} \times \frac{8}{17} \times \frac{0}{17} \\
 &= 0,352 \times 0,352 \times 0,470 \times 0 \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 &P(\text{Data2}|\text{TidakBrokenHome}) \\
 &= P(\text{Rendah}|\text{TidakBrokenHome}) \times P(\text{TidakPernahKDRT}|\text{TidakBrokenHome}) \times \\
 &P(\text{TidakSelingkuh}|\text{TidakBrokenHome}) \times P(\text{TidakBercerai}|\text{TidakBrokenHome}) \\
 &= \frac{4}{15} \times \frac{5}{15} \times \frac{7}{15} \times \frac{15}{15} \\
 &= 0,266 \times 0,333 \times 0,466 \times 1 \\
 &= 0,041
 \end{aligned} \tag{4}$$

Lalu, mengacu pada Pers. (4), untuk data ke-2, kelas Broken Home bernilai 0 dan kelas Tidak Broken Home bernilai 0,041, maka hasil klasifikasinya adalah **Tidak Broken Home**.

$$\begin{aligned}
 &P(\text{Data3}|\text{BrokenHome}) \\
 &= P(\text{Menengah}|\text{BrokenHome}) \times P(\text{JarangKDRT}|\text{BrokenHome}) \times \\
 &P(\text{TidakSelingkuh}|\text{BrokenHome}) \times P(\text{Bercerai}|\text{BrokenHome}) \\
 &= \frac{5}{17} \times \frac{5}{17} \times \frac{8}{17} \times \frac{17}{17} \\
 &= 0,294 \times 0,294 \times 0,470 \times 1 \\
 &= 0,040
 \end{aligned}$$

$$\begin{aligned}
 &P(\text{Data3}|\text{TidakBrokenHome}) \\
 &= P(\text{Menengah}|\text{TidakBrokenHome}) \times P(\text{JarangKDRT}|\text{TidakBrokenHome}) \times \\
 &P(\text{TidakSelingkuh}|\text{TidakBrokenHome}) \times P(\text{Bercerai}|\text{TidakBrokenHome}) \\
 &= \frac{6}{15} \times \frac{4}{15} \times \frac{7}{15} \times \frac{0}{15} \\
 &= 0,4 \times 0,25 \times 0,466 \times 0 \\
 &= 0
 \end{aligned} \tag{5}$$

Lalu, mengacu pada Pers. (5), untuk data ke-3, kelas Broken Home bernilai 0,040 dan kelas Tidak Broken Home bernilai 0, maka hasil klasifikasinya adalah **Broken Home**.

$$\begin{aligned}
 &P(\text{Data4}|\text{BrokenHome}) \\
 &= P(\text{Atas}|\text{BrokenHome}) \times P(\text{JarangKDRT}|\text{BrokenHome}) \times \\
 &P(\text{PernahSelingkuh}|\text{BrokenHome}) \times P(\text{TidakBercerai}|\text{BrokenHome}) \\
 &= \frac{6}{17} \times \frac{5}{17} \times \frac{9}{17} \times \frac{0}{17} \\
 &= 0,352 \times 0,294 \times 0,529 \times 0 \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 &P(\text{Data4}|\text{TidakBrokenHome}) \\
 &= P(\text{Atas}|\text{TidakBrokenHome}) \times P(\text{JarangKDRT}|\text{TidakBrokenHome}) \times \\
 &P(\text{PernahSelingkuh}|\text{TidakBrokenHome}) \times P(\text{TidakBercerai}|\text{TidakBrokenHome}) \\
 &= \frac{5}{15} \times \frac{4}{15} \times \frac{8}{15} \times \frac{15}{15} \\
 &= 0,333 \times 0,266 \times 0,533 \times 1 \\
 &= 0,047
 \end{aligned} \tag{6}$$



Lalu, mengacu pada Pers. (6), untuk data ke-4, kelas Broken Home bernilai 0 dan kelas Tidak Broken Home bernilai 0,047, maka hasil klasifikasinya adalah **Tidak Broken Home**.

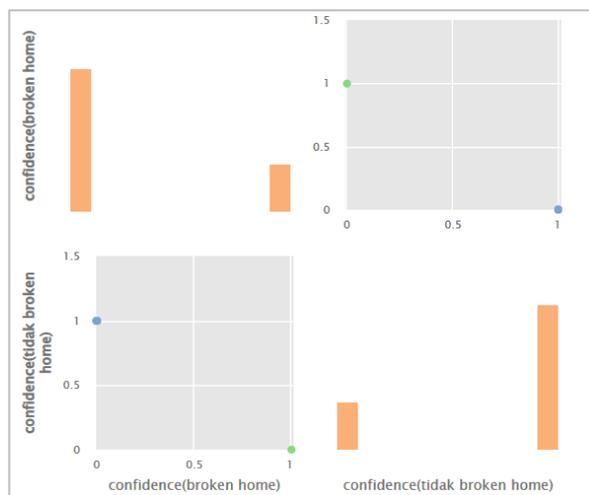
3.7 Hasil Klasifikasi

Mengacu pada Tabel 3 serta perhitungan manual yang dilakukan pada Pers. (3) sampai (6), didapatkan hasil klasifikasi yang dapat dilihat pada Tabel 4.

Tabel 4 Hasil klasifikasi Data Uji atau *Testing*

Data Ke-	Ekonomi	KDRT	Orang Tua Selingkuh	Orang Tua Bercerai	Label	Label Prediksi
1	Rendah	Jarang KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home	Tidak Broken Home
2	Rendah	Tidak Pernah KDRT	Tidak Selingkuh	Tidak Bercerai	Tidak Broken Home	Tidak Broken Home
3	Menengah	Jarang KDRT	Tidak Selingkuh	Bercerai	Broken Home	Broken Home
4	Atas	Jarang KDRT	Pernah Selingkuh	Tidak Bercerai	Tidak Broken Home	Tidak Broken Home

Adapun hasil pengujian berupa visualisasi sebaran data berdasarkan label yang dilakukan pada *testing data* terhadap *training data* dalam penelitian ini dapat dilihat pada Gambar 2.



Gambar 2 Hasil Pengujian Berupa Sebaran Data Berdasarkan Label Klasifikasi

Terlihat jelas bahwasanya 4 *testing data* tersebar ke masing-masing label pada klasifikasi yang telah ditetapkan berdasarkan model yang dihasilkan.

Tabel 5 Tabel *Confusion* Matriks

4 Record Testing Data	Broken Home	Tidak Broken Home	
Broken Home	1	0	1
Tidak Broken Home	0	3	3
	1	3	



Berdasarkan Tabel 5, maka dapat dibuat kinerja atau *performance* dari penggunaan metode klasifikasi Naïve Bayes yang dapat dilihat pada Pers. (7).

$$\text{Akurasi} = \frac{(TP/TN)}{\text{Total}} = \frac{(3/1)}{4} = 100\% \quad (7)$$

4. KESIMPULAN

Pada permasalahan yang diamati dalam penelitian ini, peneliti menyimpulkan bahwasanya Naïve Bayes Classifier dapat cukup baik digunakan untuk mengklasifikasikan permasalahan *broken home* pada anak, dengan mengacu pada 4 faktor pemicu terjadinya *broken home*. Data yang diperoleh berdasarkan angket atau kuesioner berjumlah 36 data dengan membagi data menjadi 2 bagian yaitu *training data* sebanyak 90% atau berjumlah 32 data dan *testing data* sebanyak 10% atau berjumlah 4 data. Model yang didapatkan berupa *numerical* atau angka dari perhitungan pada Teorema Bayes, dan pada pengujian ini mendapatkan akurasi (*performance*) sebesar 100%.

Saran penulis terhadap penelitian selanjutnya adalah agar *dataset* yang digunakan bisa lebih banyak dan atributnya juga beragam atau bervariasi, serta dapat menggunakan atau mengkombinasikan metode-metode *data mining* lainnya, agar model yang diciptakan juga bervariasi dan akurat.

DAFTAR PUSTAKA

- Ahyani, L., & Astuti, Rr. D. (2018). *Buku Ajar Psikologi Perkembangan Anak dan Remaja*. https://www.researchgate.net/publication/330577631_Buku_Ajar_Psikologi_Perkembangan_Anak_dan_Remaja
- Aisyah, S. H., Bahiyah, K., Prasetya, B., & Kusumawati, D. (2022). Dampak Keluarga Broken Home Terhadap Psikologis Anak. *AI-ATHFAL: Jurnal Pendidikan Anak*, 3(2), 75–81. <https://doi.org/10.46773/ALATHFAL.V3I2.485>
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160–165. <https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Aziz, M. (2015). Perilaku Sosial Anak Remaja Korban Broken Home dalam Berbagai Perspektif (Suatu Penelitian di SMPN 18 Kota Banda Aceh). *JURNAL AL-IJTIMAIYYAH*, 1(1). <https://doi.org/10.22373/al-ijtimaiyyah.v1i1.252>
- Faizah, S. K. (2022). Pemahaman Kebahagiaan oleh Remaja Broken Home. *Jurnal Bimbingan Konseling Dan Dakwah*, 3(1), 28–39. <https://jurnal.iairmngabar.com/index.php/taqorrub/article/download/188/124>
- Putro, H. F., Vlandari, R. T., & Saptomo, W. L. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*, 8(2). <https://doi.org/10.30646/TIKOMSIN.V8I2.500>
- Gunarsa, S. D. (2008). *Psikologi Perkembangan Anak dan Remaja*. BPK Gunung Mulia.
- Hafiza, S., & Mawarpury, M. (2018). Pemaknaan Kebahagiaan oleh Remaja Broken Home. *Psymphatic : Jurnal Ilmiah Psikologi*, 5(1), 59–66. <https://doi.org/10.15575/psy.v5i1.1956>
- Maliki, O., & Nanja, M. (2023). Analysis of Family Health with a Combination of Naive Bayes KNN Methods. *JTKSI (Jurnal Teknologi Komputer Dan Sistem Informasi)*, 6(1), 52. <https://doi.org/10.56327/jtksi.v6i1.1350>
- Massa, N., Rahman, M., & Napu, Y. (2020). Dampak Keluarga Broken Home Terhadap Perilaku Sosial Anak. *Jambura Journal Community Empowerment*, 1(1), 1–10. <https://doi.org/10.37411/jjce.v1i1.92>
- Mistian, W. (2018). Dampak Psikologi Terhadap Kehidupan Anak Korban Broken Home. *Musawa: Journal for Gender Studies*, 10(2), 322–354. <https://doi.org/10.24239/MSW.V10I2.528>
- Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes. *Walisongo Journal of Information Technology*, 1(1), 1. <https://doi.org/10.21580/wjit.2019.1.1.3915>
- Muttaqin, I., Muttaqin, I., & Sulisty, B. (2019). Analisis Faktor Penyebab dan Dampak Keluarga



- Broken Home. *Raheema*, 6(2), 245–256. <https://doi.org/10.24260/raheema.v6i2.1492>
- Rahayu, T. P. (2016). Determinan Kebahagiaan Di Indonesia. *Jurnal Ekonomi Dan Bisnis*, 19(1), 149–170. <https://doi.org/10.24914/JEB.V19I1.485>
- Umami, I. (2019). *Psikologi Remaja*. Idea Press Yogyakarta. <https://repository.metrouniv.ac.id/id/eprint/1447/>
- Wardani, A. K. (2021). Kebahagiaan Anak Broken Home. *Jurnal Pendidikan Tambusai*, 5(3), 6718–6727. <https://jptam.org/index.php/jptam/article/view/2038>
- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding SAKTI (Seminar Ilmu Komputer Dan Teknologi Informasi)*, 3(1), 134–138. <https://e-journals.unmul.ac.id/index.php/SAKTI/article/view/2101>



Evaluasi Penerimaan Masyarakat Terhadap Aplikasi *Telemedicine* pada Masa Pandemi COVID-19

Muhammad Reza Velayani ^{(1)*}, Muhammad Taufiq Nuruzzaman ⁽²⁾, Agung Fatwanto ⁽³⁾, Bambang Sugiantoro ⁽⁴⁾

Magister Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga, Yogyakarta
e-mail : velayanireza@gmail.com, {m.taufiq,agung.fatwanto,bambang.sugiantoro}@uin-suka.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 18 Januari 2023, direvisi 16 Mei 2023, diterima 17 Mei 2023, dan dipublikasikan 26 Mei 2023.

Abstract

Telemedicine is a technology that provides benefits during the COVID-19 pandemic, which has been going on for more than two years. However, we have never conducted an evaluation or assessment of Telemedicine applications. In this study, we tried to look at people's acceptance of using the application, and 104 participants became respondents. This study uses the TAM (Technology Acceptance Model) method, which this method measures the influence between variables. TAM has five indicator variables: Perceived Ease of Use, Perceived Usefulness, Attitude Toward Using, Behavioral Intention, and Actual Usage. In this study, four models were made with three comparison models; then, for the results of testing the relationship between variables, three hypotheses are always accepted in each model, namely Perceived Ease of Use with Perceived Usefulness, Perception Ease of Use with Attitude Toward Using, and Behavioral Intention with Actual Usage. Then the two hypotheses always rejected in each model are Perceived Usefulness with Attitude Toward Using, Attitude Toward Using with Behavioral Intention. Then there is one hypothesis for each model that is refused, namely Perceived Usefulness with Behavioral Intention, Perceived Usefulness with Actual Usage, Perceived Ease of Use with Behavioral Intention, and Perceived Ease of Use with Actual Usage.

Keywords: *Telemedicine, COVID-19, Technology Acceptance Model (TAM), SPSS, SEM*

Abstrak

Telemedicine menjadi teknologi yang memberikan manfaat di masa pandemi COVID-19 yang sudah berlangsung lebih dari dua tahun. Namun, kami belum pernah melakukan evaluasi atau penilaian aplikasi *telemedicine*. Dalam penelitian ini kami mencoba melihat penerimaan masyarakat terhadap penggunaan aplikasi, dan 104 partisipan menjadi responden. Penelitian ini menggunakan metode TAM (*Technology Acceptance Model*), di mana metode ini mengukur pengaruh antar variabel. TAM memiliki lima variabel indikator, yaitu *Perceived Ease of Use*, *Perceived Usefulness*, *Attitude Toward Using*, *Behavioral Intention*, dan *Actual Usage*. Pada penelitian ini dibuat empat model dengan tiga model pembandingan, kemudian untuk hasil pengujian hubungan antar variabel selalu diterima tiga hipotesis pada masing-masing model yaitu *Perceived Ease of Use* dengan *Perceived Usefulness*, *Perception Ease of Use* dengan *Attitude Toward Using*, dan *Behavioral Intention* dengan *Actual Usage*. Kemudian dua hipotesis yang selalu ditolak pada masing-masing model adalah *Perceived Usefulness* dengan *Attitude Toward Using*, *Attitude Toward Using* dengan *Behavioral Intention*. Kemudian ada satu hipotesis untuk setiap model yang ditolak, yaitu *Perceived Usefulness* dengan *Behavioral Intention*, *Perceived Usefulness* dengan *Actual Usage*, *Perceived Ease of Use* dengan *Behavioral Intention*, dan *Perceived Ease of Use* dengan *Actual Usage*.

Kata Kunci: *Telemedicine, COVID-19, Technology Acceptance Model (TAM), SPSS, SEM*

1. PENDAHULUAN

Kasus terkonfirmasi COVID-19 pertama kali ditemukan di Wuhan, Provinsi Hubei, China pada Desember 2019, dan kasus terkonfirmasi COVID-19 pertama di Indonesia ditemukan pada 2 Maret 2020. Terlebih lagi pandemi COVID-19 yang sudah berlangsung selama kurang lebih 2



tahun yang memberikan dampak sangat signifikan. Oleh karena itu pemerintah banyak mengeluarkan peraturan yang bertujuan untuk mengurangi penyebaran COVID-19. Pemberlakuan pembatasan kegiatan masyarakat (PPKM) merupakan salah satu metode atau cara yang diterapkan oleh pemerintah untuk mengurangi penyebaran tersebut, dengan adanya metode PPKM ini dapat mengurangi atau bahkan memutuskan rantai penyebaran COVID-19, karena pandemi ini banyak masyarakat tidak dapat melakukan kegiatan keseharian mereka di luar rumah (Direktorat Jenderal Pencegahan dan Pengendalian Penyakit, 2020).

Rumah sakit yang saat ini merupakan garda terdepan yang bertugas menangani virus tersebut tidak bisa melakukan kegiatan kesehatan seperti sebelum pandemi, sehingga di masa ini banyak teknologi pendukung yang bertujuan membantu seseorang untuk melakukan konsultasi kesehatan secara online tanpa harus bertemu langsung dengan dokter, dengan harapan untuk mengurangi mobilitas seseorang di luar rumah. *Telemedicine* merupakan salah satu teknologi pendukung tersebut, *telemedicine* telah ada sejak abad ke 19 dan terpublikasi pada abad ke 20, mengutip dari WHO. Pengembangan pelayanan berbasis *telemedicine* mampu mengatasi masalah kesenjangan dan ketidakmerataan fasilitas dan pelayanan kesehatan di Indonesia, jadi pemerintah Indonesia hanya perlu mengalokasikan anggaran yang cukup dalam menyiapkan pelayanan berbasis *telemedicine* (Pasaribu et al., 2018).

Technology Acceptance Model (TAM) secara garis besar berfungsi untuk mengukur tingkat persepsi dan sikap dari responden terhadap aplikasi *telemedicine*, dalam TAM terdapat dua faktor yang berpengaruh yaitu *perceived usefulness* (manfaat yang dirasakan) dan *perceived ease of use* (kemudahan penggunaan yang dirasakan) dan yang menjadi rumusan masalah adalah bagaimana setiap relasi antar indikator pada TAM saling memberikan pengaruh, sehingga nantinya relasi tersebut menjadi pengukur untuk menguji efektivitas dari sebuah aplikasi. Penelitian bertujuan untuk mengevaluasi penerimaan masyarakat terhadap aplikasi ini menggunakan *Technology Acceptance Model* (TAM) sebagai metode untuk mengukur seberapa besar penerimaan masyarakat terhadap aplikasi *telemedicine* yang telah banyak beredar, karena setiap masyarakat mempunyai penilaian tersendiri terkait aplikasi tersebut maka dari itu dengan penelitian ini peneliti dapat memberikan informasi tentang seberapa berperannya aplikasi *telemedicine* dalam kehidupan sehari-hari terlebih hingga saat ini kita masih dilanda oleh pandemi dan pada penelitian ini kita mengambil pasien yang terlibat langsung dalam penggunaan aplikasi tersebut (Adnan & Pramaningtyas, 2020; Ganiem, 2021; Jamil et al., 2015; Lubis, 2021; Riyanto, 2021; Sari & Wirman, 2021).

2. METODE PENELITIAN

2.1 Jenis Penelitian

Penelitian Kuantitatif adalah jenis penelitian yang kami pilih pada artikel ini. Penelitian kuantitatif adalah penelitian yang memanfaatkan pengukuran yang bersifat objektif dan juga memanfaatkan analisis yang bersifat *numerical* dengan tujuan untuk menjelaskan, meramalkan, dan/atau mengontrol atas suatu fenomena yang sedang terjadi di masyarakat (Moleong, 2018). Data tentang variabel-variabel yang dapat mempengaruhi penerimaan pengguna atas aplikasi *telemedicine*. Setelah itu, kami melakukan pengukuran terkait aplikasi *telemedicine* menggunakan *Technology Acceptance Model* (TAM).

2.2 Jenis dan Sumber Data

Penelitian ini menggunakan data primer sebagai jenis dan sumber data, data primer adalah data yang didapatkan langsung dari subjek penelitian dengan menggunakan alat pengukuran data langsung pada subjek sebagai sumber informasi yang dicari. Data yang dicari pada penelitian ini adalah data mengenai tanggapan pengguna aplikasi *telemedicine* terhadap kegunaan, kemudahan, sikap, minat, serta penggunaan aplikasi *telemedicine* sesungguhnya. Sumber data pada penelitian ini yaitu pengguna aplikasi *telemedicine*.



2.3 Teknik Pengumpulan Data

Data pada penelitian ini dikumpulkan melalui kuesioner online yang dibagikan kepada pengguna aplikasi *telemedicine*. Kuesioner online merupakan metode pengumpulan data di mana peneliti mengajukan pertanyaan dengan formulir online sebagai wadah untuk subjek dapat menjawab pertanyaan yang diajukan oleh peneliti. Peneliti menggunakan kuesioner online untuk mendapatkan data-data bagaimana variabel-variabel pada *Technology Acceptance Model* (TAM) berpengaruh terhadap penerimaan pengguna aplikasi *telemedicine*.

Penentuan skala pengukuran jawaban pada kuesioner ini menggunakan skala Likert yang biasa digunakan dalam perhitungan untuk mengukur sikap, pendapat, dan persepsi seseorang. Jawaban pada kuesioner disusun dengan gradasi positif hingga negatif. Pemberian skor berdasarkan pada ketentuan, sebagaimana terlihat pada Tabel 1.

Tabel 1 Skala Likert

Pilihan Jawaban	Nilai Jawaban
Sangat Setuju	5
Setuju	4
Biasa	3
Tidak Setuju	2
Sangat Tidak Setuju	1

Tabel 2 Variabel Indikator Konstruksi

	Variabel Indikator
<i>Perceived Ease of Use (Perceived Ease of Use)</i>	PEU1 = Kemudahan Penggunaan PEU2 = Jelas dan mudah dipahami PEU3 = Mudah untuk diingat PEU4 = Cekatan dalam menggunakan
<i>Perceived Usefulness (Perceived Usefulness)</i>	PU1 = Kemanfaatan penggunaan PU2 = Efisiensi penggunaan PU3 = Kecepatan proses PU4 = Kemudahan memesan obat
<i>Attitude Toward Using (Attitude Toward Using)</i>	ATU1 = Senang untuk menggunakan ATU2 = Nyaman untuk menggunakan
<i>Behavioral Intention (Behavioral Intention)</i>	BI1 = Bisa digunakan kapan saja BI2 = Bisa digunakan dalam kondisi apapun BI3 = Bisa digunakan terus menerus
<i>Actual Usage (Actual Usage)</i>	AU1 = Penggunaan secara rutin AU2 = Intensitas penggunaan

Variabel endogen dan juga variabel eksogen merupakan variabel-variabel yang digunakan pada penelitian ini. Variabel eksogen adalah variabel independen yang mempengaruhi variabel dependen dan variabel endogen adalah variabel dependen yang dipengaruhi oleh variabel independen (eksogen). Variabel PEU (*Perceived Ease of Use*) merupakan variabel eksogen yang terdiri atas tiga indikator, di antaranya: fleksibilitas penggunaan, kemudahan penggunaan, dan kemudahan belajar dan dipahami. Variabel berikut ini merupakan variabel endogen adalah PU (*Perceived Usefulness*) yang terdiri dari tiga indikator, yaitu: efisiensi penggunaan, kemanfaatan penggunaan, dan kecepatan proses, itu merupakan indikator yang dipengaruhi oleh indikator pada variabel eksogen. ATU (*Attitude Toward Using*) juga merupakan faktor endogen dan didalamnya terdapat dua indikator yaitu kesenangan serta kenyamanan pengguna. Variabel endogen selanjutnya adalah BI (*Behavioral Intention*) yang didalamnya terdiri berdasarkan indikator yang mampu mendukung dalam segala kondisi baik waktu maupun tempat. Variabel endogen yang terakhir adalah AU (*Actual Usage*) yaitu membuat pengguna ingin terus



menggunakan karena kebermanfaatan dan kemudahan yang diberikan. Variabel indikator konstruk disajikan pada Tabel 2.

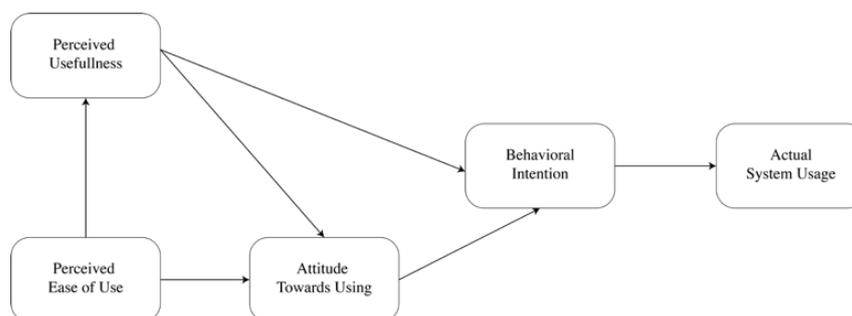
Penelitian ini menggunakan statistik deskriptif untuk pengolahan datanya yang meliputi kegiatan mengumpulkan data, mengolah data, dan menyajikan data. Statistik deskriptif ditunjukkan dengan adanya data frekuensi, ukuran tendensi sentral (*mean, median, modus*), dan dispersi (kisaran, varian, standar deviasi). Dalam penelitian ini, statistik deskriptif digunakan untuk mendeskripsikan profil responden serta variabel-variabel penelitian (Suryani & Hendryadi, 2015). Selama penelitian, metode penelitian kuantitatif juga dapat digunakan untuk dapat melakukan penilaian statistik tambahan. Pada penelitian ini digunakan alat analisis SPSS (*Statistical Product and Service Solution*) dan AMOS (*Analysis of Moment Structures*) digunakan dalam penelitian ini.

Menurut (Sujarweni, 2014) uji validitas dilakukan untuk mengetahui kelayakan butir-butir dalam suatu daftar pertanyaan untuk mendeskripsikan suatu variabel dan pada umumnya untuk mendukung suatu kelompok variabel tertentu. Suatu butir pertanyaan dinyatakan valid jika hasil nilai r hitung lebih besar dari r tabel di mana $df=n-2$ dengan sig 5%. Jika r tabel < dari r hitung maka hasil tidak valid.

Dalam penelitian ini SEM (*Structural Equation Modeling*) digunakan sebagai metode untuk menguji hipotesis. Model SEM dapat terdiri atas *measurement model* dan *structural model* dan tujuan utama analisis SEM adalah menguji kesesuaian antara model tersebut fit dengan data yang ada. Sebelum melakukan SEM terlebih dahulu melakukan *Test of Model Fit* yang terdiri dalam beberapa bagian, di antaranya:

- 1) Pengujian *chi-square*, pengujian ini bertujuan untuk mengetahui apakah matriks kovarians sampel berbeda secara signifikan dengan matriks kovarians estimasi. Dalam model *chi-square*, jika nilai yang dihasilkan rendah maka dianggap tepat. Semakin rendah nilai *chi-square*, semakin baik peringkatnya
- 2) *Goodness of Fit Index*, nilai berkisar antara 0-1, jika nilai lebih dari 1 maka lebih baik. $GFI > 0,90$ adalah *good fit*, sedangkan $0,80 < GFI < 0,90$ adalah nilai yang bisa diterima.
- 3) *The Minimum sampel discrepancy function* (CMIN) dibagi dengan derajat bebas akan menghasilkan indeks CMIN/DF, yaitu salah satu indikator untuk mengukur tingkat kesesuaian sebuah model. Dalam hal ini CMIN/DF tidak lain adalah statistics Chisquare dibagi derajat bebasnya (χ^2 relatif). Nilai (χ^2 relatif) yang diharapkan adalah kurang dari atau sama dengan 2,00 (Sahadi & Wibowo, 2014).
- 4) RMSEA (*Root Mean Square Error of Approximation*) menunjukkan seberapa baik model dengan estimasi parameter yang tidak diketahui tetapi dipilih secara optimal dan cocok dengan matriks kovarians populasi. Dalam beberapa tahun terakhir, indeks ini telah dianggap sebagai "salah satu indeks kecocokan yang paling informatif" karena sensitivitasnya terhadap jumlah parameter yang diestimasi dalam model. Salah satu keuntungan terbesar RMSEA adalah memiliki interval kepercayaan. Secara umum, batas bawah RMSEA dari model yang dipasang mendekati nol, sedangkan batas atas harus kurang dari 0,08.

Model penelitian yang akan diuji dalam penelitian digambarkan pada Gambar 1.



Gambar 1 Model Penelitian



Hipotesis yang akan diujikan dalam penelitian ini adalah sebagai berikut:

HP1: *Perceived Ease of Use* (PEU) memiliki pengaruh positif terhadap *Perceived Usefulness* (PU).

HP2: *Perceived Ease of Use* (PEU) memiliki pengaruh positif terhadap *Attitude Toward Using* (ATU).

HP3: *Perceived Usefulness* (PU) memiliki pengaruh positif terhadap *Attitude Toward Using* (ATU).

HP4: *Perceived Usefulness* (PU) memiliki pengaruh positif terhadap *Behavioral Intention* (BI).

HP5: *Attitude Toward Using* (ATU) memiliki pengaruh positif terhadap *Behavioral Intention* (BI).

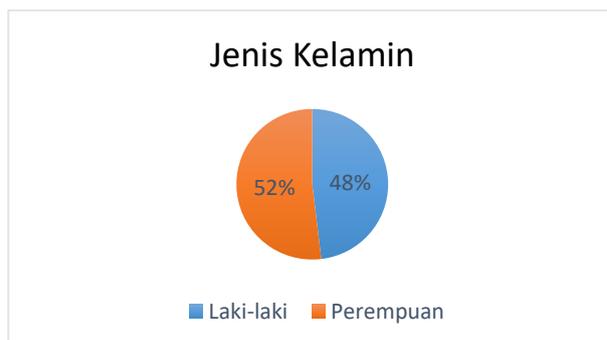
HP6: *Behavioral Intention* (BI) memiliki pengaruh positif terhadap *Actual Usage* (AU).

3. HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

Penelitian ini menggunakan pengambilan data secara kuantitatif melalui kuesioner dengan skala likert yang melibatkan 104 *user* (pengguna) aktif yang menggunakan aplikasi *telemedicine* di masa pandemi, untuk pendistribusian kuesioner ini dilakukan kepada seluruh masyarakat tanpa ada batasan daerah. Di bawah akan ditampilkan 3 bagan yang menampilkan 3 kategori responden di antaranya jenis kelamin, jenis penggunaan, dan rentang umur.

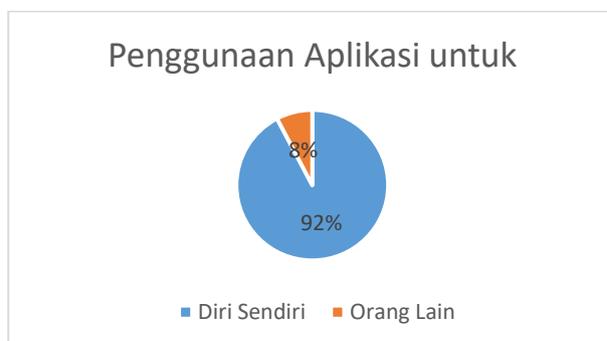
3.1.1 Jenis Kelamin



Gambar 2 Persentase Responden Berdasarkan Jenis Kelamin

Gambar 2 menunjukkan data responden berdasarkan jenis kelamin, di mana untuk responden dengan jenis kelamin laki-laki sebanyak 48% dan untuk perempuan sebanyak 52%. Melihat data tersebut kita bisa mengetahui bahwa penggunaan aplikasi *telemedicine* tidak dominan pada suatu jenis kelamin.

3.1.2 Jenis Penggunaan

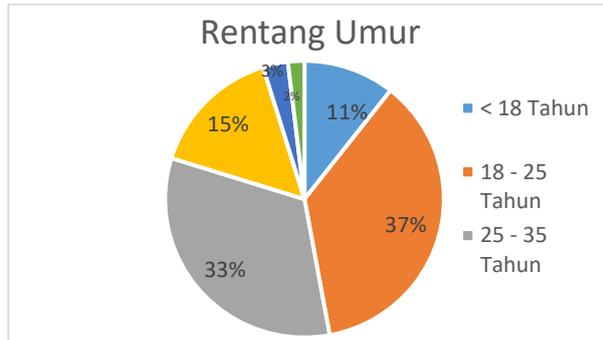


Gambar 3 Persentase Responden Berdasarkan Jenis Penggunaan



Berdasarkan bagan atau grafik pada Gambar 3, kita bisa melihat bahwa untuk penggunaan aplikasi untuk diri sendiri lebih dominan sebanyak 92% sedangkan penggunaan aplikasi untuk orang lain sebanyak 8%, yang artinya banyak orang peduli akan kesehatan secara personal di masa pandemi.

3.1.3 Rentang Umur



Gambar 4 Persentase Responden Berdasarkan Rentang Umur

Berdasarkan data pada Gambar 4, pada penelitian ini membagi 6 kelompok umur dalam penggunaan aplikasi *telemedicine*, di antaranya adalah usia <18 tahun, usia 18-25 tahun, usia 25-35 tahun, usia 35-45 tahun, usia 45-55 tahun, dan usia > 65 tahun. Melihat dari signifikansi kelompok di atas bahwa dengan rentang usia 18-25 tahun sebanyak 37% kemudian usia direntang 25-35 tahun sebanyak 33%. Berarti kita bisa melihat tingkat penggunaan aplikasi ini cukup tenar dikalangan usia 18-35 tahun dan pada usia ini mereka memiliki pemahaman teknologi yang lebih baik dibandingkan dengan rentang usia di atasnya.

Tabel 3 Persepsi Kemudahan Penggunaan

No.	Item Pertanyaan	Skor		Mean
		Min	Maks	
1	PEU1	1	5	3,54
2	PEU2	1	5	3,24
3	PEU3	1	5	2,70
4	PEU4	1	5	3,30
Rata-rata Persepsi Kemudahan Penggunaan				3,19

Hasil dari Tabel 3, menunjukkan bahwa nilai rata-rata dari persepsi kemudahan penggunaan aplikasi *telemedicine* cukup tinggi, yaitu sebesar 3,19 yang berarti secara umum partisipan atau pengguna memiliki persepsi bahwa aplikasi *telemedicine* mudah untuk digunakan.

Tabel 4 Persepsi Kegunaan

No.	Item Pertanyaan	Skor		Mean
		Min	Maks	
1	PU1	1	5	3,54
2	PU2	1	5	3,24
3	PU3	1	5	2,70
4	PU4	1	5	3,30
Rata-rata Persepsi Kegunaan				3,41

Hasil dari Tabel 4, menunjukkan bahwa nilai rata-rata dari persepsi kegunaan aplikasi *telemedicine* cukup tinggi, yaitu sebesar 3,41 yang berarti secara umum partisipan atau pengguna memiliki persepsi bahwa aplikasi *telemedicine* memiliki manfaat dan berguna.



Tabel 5 Sikap Terhadap Penggunaan

No.	Item Pertanyaan	Skor		Mean
		Min	Maks	
1	ATU1	1	5	3,54
2	ATU2	1	5	3,24
Rata-rata Sikap Terhadap Penggunaan				3,50

Hasil dari Tabel 5, menunjukkan bahwa nilai rata-rata dari sikap terhadap penggunaan aplikasi *telemedicine* tergolong tinggi, yaitu sebesar 3,50 yang berarti secara umum partisipan atau pengguna memiliki penilaian yang positif terhadap penerepan aplikasi *telemedicine*.

Tabel 6 Minat Berperilaku

No.	Item Pertanyaan	Skor		Mean
		Min	Maks	
1	BI1	1	5	3,54
2	BI2	1	5	3,24
3	BI3	1	5	
Rata-rata Minat Berperilaku				3,81

Hasil dari Tabel 6, menunjukkan bahwa nilai rata-rata dari minat berperilaku menggunakan aplikasi *telemedicine* tergolong tinggi, yaitu sebesar 3,81 yang berarti secara umum partisipan atau pengguna memiliki minat yang positif untuk menggunakan aplikasi *telemedicine*.

Tabel 7 Penggunaan Nyata

No.	Item Pertanyaan	Skor		Mean
		Min	Maks	
1	AU1	1	5	3,54
2	AU2	1	5	3,24
Rata-rata Penggunaan Nyata				3,84

Hasil dari Tabel 7, menunjukkan bahwa nilai rata-rata dari penggunaan nyata menggunakan aplikasi *telemedicine* tergolong tinggi, yaitu sebesar 3,84 yang berarti secara umum partisipan atau pengguna memiliki intensitas penggunaan aplikasi yang tinggi dalam menggunakan aplikasi ini.

3.2 Analisis Data

Pada tahapan ini dilakukan dua proses pengujian, yaitu uji validitas dan uji reliabilitas. Uji validitas digunakan untuk mengetahui ada atau tidaknya korelasi antara variabel dengan variabel lainnya atau suatu konstruk dengan konstruk lainnya. Uji validitas bertujuan untuk membuktikan keabsahan data yang dibagikan kepada para responden, uji validitas bisa dikerjakan dengan bantuan SPSS menggunakan korelasi *sig-Pearson* dan menilai setiap variabel indikator yang akan diujikan. Suatu kuesioner dianggap valid apabila sebuah pertanyaan pada kuesioner mampu merepresentasikan sesuatu yang akan diukur oleh kuesioner tersebut.

Dasar pengambilan keputusan dalam uji validitas adalah:

- 1) Jika nilai $r_{hitung} > r_{tabel}$, maka *item* angket atau pertanyaan memiliki korelasi yang signifikan dengan skor total (valid)
- 2) Jika nilai $r_{hitung} < r_{tabel}$, maka *item* angket atau pertanyaan maka tidak berkorelasi signifikan terhadap skor total (tidak valid)



Uji validitas ini dilakukan dengan melihat setiap variabel indikator dan membandingkannya dengan nilai R-tabel pada taraf signifikansi 0,05 (5%), diketahui jumlah responden yang mengisi kuesioner (N) adalah sebanyak 104 orang. Di mana $df = N-2$ adalah $df = 104-2 = 102$ yang berarti untuk nilai signifikasinya sebesar 0,192. Hasil pengujian seluruh konstruk dinyatakan valid karena *r-score* lebih besar dari 0,192 maka angket dinyatakan valid. Hasil uji validitas ditunjukkan pada Tabel 8.

Tabel 8 Uji Validitas

Item	r_{hitung}	$r_{tabel\ 5\% (102)}$	Sig.	Kriteria
PEU1	0,674	0,192	.000	Valid
PEU2	0,780	0,192	.000	Valid
PEU3	0,654	0,192	.000	Valid
PEU4	0,487	0,192	.000	Valid
PU1	0,758	0,192	.000	Valid
PU2	0,702	0,192	.000	Valid
PU3	0,746	0,192	.000	Valid
PU4	0,634	0,192	.000	Valid
ATU1	0,802	0,192	.000	Valid
ATU2	0,591	0,192	.000	Valid
BI1	0,850	0,192	.000	Valid
BI2	0,886	0,192	.000	Valid
BI3	0,873	0,192	.000	Valid
AU1	0,914	0,192	.000	Valid
AU2	0,991	0,192	.000	Valid

Di sisi lain, uji realibilitas atau tingkat kepercayaan, dalam analisis statistik, dipergunakan sebagai pengukur tingkat konsistensi dari suatu angket yang telah disebar dan digunakan oleh peneliti untuk mengambil data dari responden. Oleh karena itu, angket dapat diukur kehandalannya untuk mengukur variabel penelitian. (Sujarweni, 2014) juga menjelaskan bahwa uji realibilitas dapat dilakukan secara bersama-sama untuk setiap butir atau *item* pertanyaan dalam angket (kuesioner) penelitian yang telah dibuat. Berikut dasar pengambilan keputusan dalam uji realibilitas: Jika nilai *Cronbach's Alpha* > 0,6 maka kuesioner atau angket dinyatakan reliabel atau konsisten. Jika nilai *Cronbach's Alpha* < 0,6 maka kuesioner atau angket dinyatakan tidak reliabel atau tidak konsisten. Setelah dilakukan uji realibilitas untuk 15 *item* pertanyaan dalam angket (kuesioner) mendapatkan nilai *Cronbach's Alpha* sebesar 0,675 yang berarti reliabel karena lebih besar dari 0,6. Hasil uji realibilitas ditunjukkan pada Tabel 9.

Tabel 9 Uji Realibilitas

<i>Cronbach's Alpha</i>	N of Items
0,675	15

3.3 Analisis Structural Equation Model (SEM)

3.3.1 Uji Asumsi Kecukupan Sampel

Sampel adalah bagian dari populasi yang mewakilinya. Sampel yang digunakan dalam penelitian SEM (*Structural Equation Modeling*) adalah minimal 100 sampel (Ferdinand, 2014). Menurut (Ghozali, 2009) dalam metode SEM besarnya sampel adalah antara 100-200. Pedoman penentuan besarnya *sample size* (ukuran sampel) untuk SEM. Menurut (Solimun, 2002) adalah:

- 1) Bila pendugaan parameter menggunakan metode kemungkinan maksimum (*maximum likelihood estimation*), besar sampel yang disarankan antara 100 hingga 200, dengan minimum sampel adalah 50.
- 2) Sebanyak 5–10 kali jumlah parameter yang ada di dalam model.
- 3) Sama dengan 5–10 kali jumlah indikator dari keseluruhan variabel laten.



Sebanyak 15 indikator digunakan dalam penelitian ini, merujuk pada poin ketiga maka ukuran sampel minimal 5 x 15 atau sebesar 100 sampel, sehingga sampel penelitian ini adalah 75 responden.

3.3.2 Evaluasi *Outlier* Model

Kasus *outlier* adalah data yang memiliki karakteristik yang sangat unik dan yang terlihat sangat jauh berbeda dibandingkan dengan data-data dari observasi-observasi lainnya. Kasus *outlier* bisa muncul dalam bentuk nilai yang ekstrim baik untuk sebuah variabel tunggal atau pun variabel kombinasi (Ghozali, 2013). Data outliers atau biasa disebut data pencilan. *Outliers* bisa diartikan sebagai data observasi yang muncul dengan nilai-nilai ekstrim. Nilai ekstrim disini adalah nilai yang jauh berbeda dengan sebagian besar nilai yang muncul, baik itu secara univariate ataupun multivariate, data *outlier* bisa menyebabkan taksiran data menjadi bias sehingga data menjadi tidak normal dan mampu menurunkan derajat validitas dan reliabilitas. Dalam penelitian ini tidak ditemukannya *outliers*, kita memiliki data dengan nilai *mahalanobis distance* (MD) yang tertinggi yaitu 31,395 masih di bawah nilai *chi-square* yaitu 37,697. Berikut cara menentukan data *outliers*:

Chi-square:

- Prob= 0,001
- Jumlah indikator = 15

Kriteria:

Nilai MD > *chi-square* (prob, jumlah indikator) maka **Outlier**.

3.3.3 Uji Normalitas

Sebelum melakukan penelitian, penting untuk memastikan bahwa normalitas dari sebuah indikator yang diamati harus diuji normalitasnya untuk memastikan data terdistribusi normal. Uji normalitas univariat dengan mengamati nilai *skewness* dan kurva dari data yang digunakan. Jika nilai *CR skewness* data antara -2,58 dan 2,58 dan nilai kurtosis < 7 maka data yang digunakan kemungkinan normal. Hasil uji normalitas data dilihat pada Tabel 10.

Tabel 10 Uji Normalitas

Variable	min	max	skew	c.r.	kurtosis	c.r.
au_1	2,000	5,000	-0,094	-0,392	-0,261	-0,543
au_2	2,000	5,000	-0,193	-0,803	-0,065	-0,135
bi_1	2,000	5,000	-0,122	-0,507	-0,202	-0,420
bi_2	2,000	5,000	-0,404	-1,681	0,233	0,486
bi_3	2,000	5,000	-0,616	-2,565	0,393	0,818
atu_2	2,000	5,000	0,322	1,342	-0,261	-0,543
atu_1	1,000	5,000	-0,272	-1,132	0,260	0,541
peu_1	1,000	5,000	-0,447	-1,860	-0,003	-0,006
peu_2	1,000	5,000	-0,367	-1,529	-0,669	-1,393
peu_3	1,000	5,000	0,192	0,798	-0,887	-1,846
peu_4	1,000	5,000	0,263	1,096	0,793	1,650
pu_1	1,000	5,000	-0,537	-2,236	-0,049	-0,101
pu_2	1,000	5,000	-0,591	-2,460	0,379	0,788
pu_3	1,000	5,000	0,012	0,051	0,512	1,066
pu_4	1,000	5,000	-0,060	-0,251	-0,209	-0,435
Multivariate					13,698	3,093

3.3.4 Analisa *Goodness-of-Fit*

Tujuan pengujian ini untuk melihat seberapa tepat antara frekuensi yang diamati dengan yang diharapkan, apabila tingkat dari *goodness of fit* baik maka model ini akan digunakan dalam



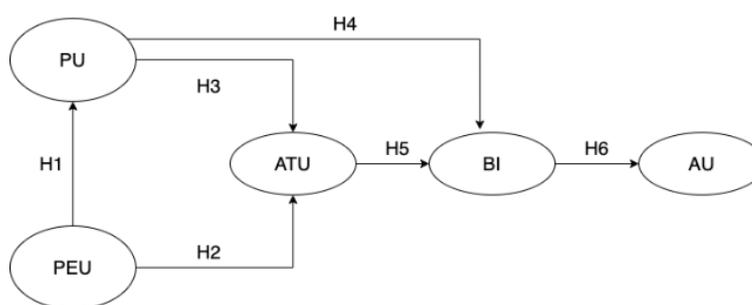
tahapan *Structural Equation Modelling*. Tahapan pertama yaitu menguji *goodness of fit model* penelitian ini.

Tabel 11 Goodness of Fit Index Model

<i>Goodness of Index</i>	Kriteria	<i>Cut of Value</i>	Keterangan
<i>Chi-square</i>	<Chi-square\ DF\ (128.803)	102.364	Baik
Prob	>0,05	.084	Baik
RMSEA	\le0,08	.046	Baik
GFI	\geq0,80	.882	Fit
AGFI	\geq0,80	.831	Baik
CMIN/DF	\le2.00	1.219	Baik

Melihat Tabel 11 hasil pengujian *Goodness of fit index* menunjukkan hasil yang baik untuk setiap kriteria, dari hasil tersebut kita dapat menyimpulkan bahwa model tersebut telah *fit* dan sesuai. Model yang diujikan pada penelitian ini didukung oleh data yang dikumpulkan di lapangan dan ini membuktikan bahwa matriks varians-kovarians populasi sesuai dengan varians-kovarians sampel atau data pengamatan. Kemudian hasil uji hipotesis ditunjukkan pada tabel *Regression Weight Analisis*. Hasil dari *Regression Weight* ditunjukkan pada Tabel 12 dengan gambar model digambarkan pada Gambar 5.

a) Model 1



Gambar 5 Model TAM (Model Pertama)

Tabel 12 Regression Weights Analisis SEM

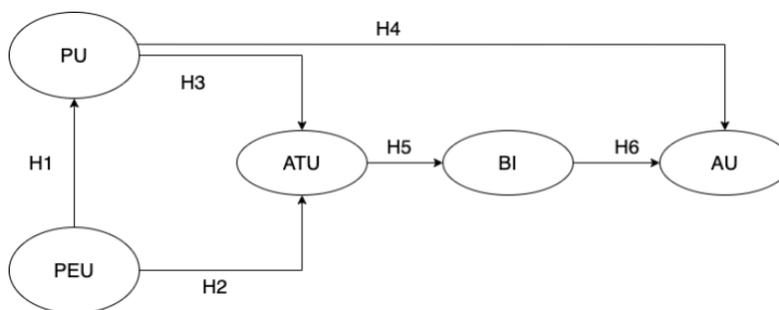
Hubungan variabel	<i>Estimate</i>	<i>Square Error</i>	<i>Critical Ratio (>1.96)</i>	<i>Probability (<0,05)</i>	Keterangan
<i>Perceived Ease of Use</i> → <i>Perceived Usefulness</i>	1.217	.555	2.193	.028	Diterima
<i>Perceived Ease of Use</i> → <i>Attitude Toward Using</i>	1.997	.994	2.008	.045	Diterima
<i>Perceived Usefulness</i> → <i>Attitude Toward Using</i>	-.728	.345	-2.057	.040	Ditolak
<i>Perceived Usefulness</i> → <i>Behavioral Intention</i>	-.021	.187	-.112	.911	Ditolak
<i>Attitude Toward Using</i> → <i>Behavioral Intention</i>	.565	.294	1.923	.054	Ditolak
<i>Behavioral Intention</i> → <i>Actual Usage</i>	.579	.118	4.887	***	Diterima

Ket: *** = Signifikan < 0,001



Hasil pengujian hipotesis dilakukan dengan menganalisa nilai *Critical Ratio* (CR) dan nilai *probability* (P) hasil olah data, dibandingkan dengan Batasan statistik yang telah ditentukan, yaitu jika nilai CR lebih besar dari 1.96 dan nilai *probability* di bawah 0,05. Apabila data yang muncul menunjukkan nilai yang telah sesuai dengan syarat yang ada maka data tersebut diterima.

b) Model 2



Gambar 6 Model 2 (Pembanding)

Dalam model 2 pembanding, peneliti mencoba merubah jalur hubungan dari PU (*Perceived Usefulness*) dengan BI (*Behavioral Intention*) menjadi PU (*Perceived Usefulness*) dengan AU (*Actual Usage*), sehingga kita bisa melihat apakah hubungan tersebut akan menghasilkan hipotesis yang diterima atau ditolak seperti pada model pertama. Hasil hipotesis ditunjukkan pada Tabel 13 Hasil Hipotesis Model 2 (Pembanding) dengan gambar model digambarkan pada Gambar 6.

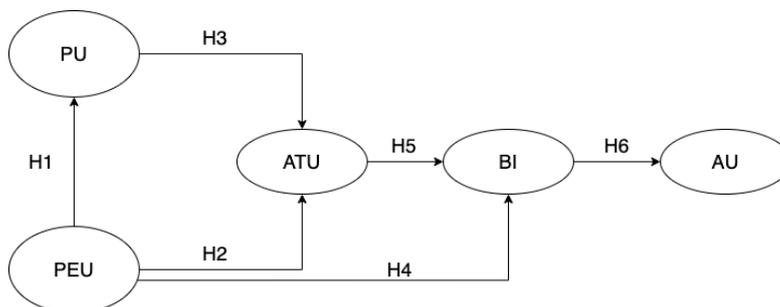
Tabel 13 Hasil Hipotesis Model 2 (Pembanding)

Hubungan variabel	Estimate	Square Error	Critical Ratio (>1.96)	Probability (<0,05)	Keterangan
<i>Perceived Ease of Use</i> → <i>Perceived Usefulness</i>	1.219	.555	2.193	.028	Diterima
<i>Perceived Ease of Use</i> → <i>Attitude Toward Using</i>	1.993	.992	2.009	.045	Diterima
<i>Perceived Usefulness</i> → <i>Attitude Toward Using</i>	-.741	.340	-2.179	.029	Ditolak
<i>Perceived Usefulness</i> → <i>Actual Usage</i>	-.002	.134	-.015	.988	Ditolak
<i>Attitude Toward Using</i> → <i>Behavioral Intention</i>	.569	.292	1.945	.052	Ditolak
<i>Behavioral Intention</i> → <i>Actual Usage</i>	.579	.119	4.886	***	Diterima

Ket: *** = Signifikan < 0,001



c) Model 3 (Pembanding)



Gambar 7 Model 3 (Pembanding)

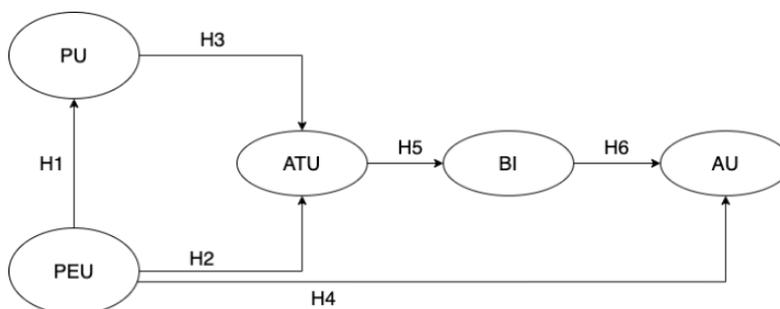
Dalam model 3 pembanding, peneliti mencoba merubah jalur hubungan dari PU (*Perceived Usefulness*) dengan BI (*Behavioral Intention*), menjadi PEU (*Perceived Ease of Use*) dengan BI (*Behavioral Intention*), sehingga kita bisa melihat apakah hubungan tersebut akan menghasilkan hipotesis yang diterima atau ditolak seperti pada model pertama. Hasil hipotesis ditunjukkan pada Tabel 14 Hasil Hipotesis Model 3 (Pembanding) dengan gambar model digambarkan pada Gambar 7.

Tabel 14 Hasil Hipotesis Model 3 (Pembanding)

Hubungan variabel	Estimate	Square Error	Critical Ratio (>1.96)	Probability (<0,05)	Keterangan
<i>Perceived Ease of Use</i> → <i>Perceived Usefulness</i>	1.217	.555	2.194	.028	Diterima
<i>Perceived Ease of Use</i> → <i>Attitude Toward Using</i>	1.995	.994	2.008	.045	Diterima
<i>Perceived Usefulness</i> → <i>Attitude Toward Using</i>	-0.730	.354	-2.062	.039	Ditolak
<i>Perceived Ease of Use</i> → <i>Behavioral Intention</i>	-0.048	.519	-0.092	.927	Ditolak
<i>Attitude Toward Using</i> → <i>Behavioral Intention</i>	.590	.373	1.582	.114	Ditolak
<i>Behavioral Intention</i> → <i>Actual Usage</i>	.579	.118	4.887	***	Diterima

Ket: *** = Signifikan < 0,001

d) Model 4 (Pembanding)



Gambar 8 Model 4 (Pembanding)



Dalam model 4 perbandingan, peneliti mencoba merubah jalur hubungan dari PU (*Perceived Usefulness*) dengan BI (*Behavioral Intention*) menjadi PEU (*Perceived Ease of Use*) dengan AU (*Actual Usage*), sehingga kita bisa melihat apakah hubungan tersebut akan menghasilkan hipotesis yang diterima atau ditolak seperti pada model pertama. Hasil hipotesis ditunjukkan pada Tabel 15 Hasil Hipotesis Model 4 (Pembanding) dengan gambar model digambarkan pada Gambar 8.

Tabel 15 Hasil Hipotesis Model 4 (Pembanding)

Hubungan variabel	Estimate	Square Error	Critical Ratio (>1.96)	Probability (<0,05)	Keterangan
<i>Perceived Ease of Use</i> → <i>Perceived Usefulness</i>	1.215	.551	2.206	.027	Diterima
<i>Perceived Ease of Use</i> → <i>Attitude Toward Using</i>	2.005	.993	2.019	.043	Diterima
<i>Perceived Usefulness</i> → <i>Attitude Toward Using</i>	-.754	.344	-2.193	.039	Ditolak
<i>Perceived Ease of Use</i> → <i>Behavioral Intention</i>	.125	.296	.421	.674	Ditolak
<i>Attitude Toward Using</i> → <i>Behavioral Intention</i>	.559	.289	1.931	.054	Ditolak
<i>Behavioral Intention</i> → <i>Actual Usage</i>	.562	.121	4.644	***	Diterima

Ket: *** = Signifikan < 0,001

Setelah melihat hasil dari 3 model yang dibuat sebagai perbandingan ternyata terdapat 3 hipotesis yang ditolak, sehingga kita bisa melihat indikator yang memiliki hubungan atau relasi yang hipotesisnya ditolak di setiap modelnya. Pada Tabel 16 menampilkan hasil hipotesis dari setiap model yang telah diujicobakan.

Tabel 16 Hasil Hipotesis dari Semua Model

Model 1	Model 2	Model 3	Model 4
H1: PEU dengan PU			
H2: PEU dengan ATU			
H3: PU dengan ATU			
H4: PU dengan BI	H4: PU dengan AU	H4: PEU dengan BI	H4: PEU dengan AU
H5: ATU dengan BI			
H6: BI dengan AU			

Keterangan:

- Hijau = diterima
- Kuning = ditolak

4. KESIMPULAN

Dari hasil pengujian terhadap enam hipotesis pada model pertama dan tiga model perbandingan untuk melihat hasil hipotesis dari setiap model. Setelah melihat hasil dari empat model tersebut, ditemukan tiga hipotesis yang ditolak pada semua model, yaitu hubungan antara *Perceived Usefulness* dengan *Attitude Toward Using*, dan hubungan antara *Attitude Toward Using* dengan *Behavioral Intention*. Hipotesis-hipotesis yang ditolak menunjukkan bahwa ekspektasi dan minat pengguna dalam menggunakan aplikasi tidak selalu sesuai dengan kenyataan penggunaan aplikasi tersebut. Namun, terdapat tiga hipotesis yang selalu diterima dalam setiap model, yaitu relasi antara *Perceived Ease of Use* dengan *Perceived Usefulness*, *Perceived Ease of Use* dengan *Attitude Toward Using*, dan *Behavioral Intention* dengan *Actual Usage*. Hal ini menunjukkan



bahwa pengguna memiliki pendapat bahwa aplikasi *telemedicine* berguna dan bisa memberikan kemudahan dalam penggunaan, sehingga menciptakan rasa senang dan nyaman dalam menggunakan aplikasi tersebut, dan minat pengguna terhadap aplikasi *telemedicine* mempengaruhi motivasi mereka dalam menggunakan aplikasi secara nyata.

4.1 Keterbatasan dan Saran Penelitian

- 1) Pada penelitian ini, pertanyaan pada kuesioner tidak menjadi acuan pasti namun bisa dijadikan bahan referensi dalam melakukan penelitian sejenis lainnya.
- 2) Metode analisis yang digunakan pada penelitian ini adalah SEM-AMOS. Oleh karena itu untuk penelitian selanjutnya dapat menggunakan metode analisis lainnya, seperti PLS, Lisrel ataupun *software* lainnya.
- 3) Jumlah partisipan pada penelitian ini terbatas, karena jika mengambil data dari seluruh masyarakat Indonesia akan memakan waktu lama, untuk penelitian selanjutnya bisa mengumpulkan data pada suatu populasi tertentu atau pada daerah tertentu.
- 4) Penelitian ini menggunakan 2 (dua) jenis variabel, yaitu variabel dependen dan variabel independent. Maka dari itu diharapkan untuk penelitian selanjutnya dapat melakukan pengembangan dengan menambahkan jenis variabel lain, seperti variabel *moderating* maupun variabel *intervening*.

DAFTAR PUSTAKA

- Adnan, M. L., & Pramaningtyas, M. D. (2020). Penggunaan Telemedicine pada Masa Pandemi COVID-19: Prospek dan Tantangan. *JIMKI: Jurnal Ilmiah Mahasiswa Kedokteran Indonesia*, 8(3), 225–233.
- Direktorat Jenderal Pencegahan dan Pengendalian Penyakit. (2020). *Pedoman Pencegahan dan Pengendalian Coronavirus Disese (COVID-19)*. Kementerian Kesehatan RI. https://infeksiemerging.kemkes.go.id/download/REV-04_Pedoman_P2_COVID-19_27_Maret2020_TTD1.pdf
- Ferdinand, A. (2014). Metode Penelitian Manajemen. In *Badan Penerbit Universitas Diponegoro*. (5th ed.). UNDIP Press.
- Ganiem, L. M. (2021). Efek Telemedicine pada Masyarakat (Kajian Hukum Media McLuhan: Tetrad). *Interaksi: Jurnal Ilmu Komunikasi*, 9(2), 87–97. <https://doi.org/10.14710/interaksi.9.2.87-97>
- Ghozali, I. (2009). *Aplikasi analisis multivariate dengan program SPSS* (4th ed.). Badan Penerbit Universitas Diponegoro. <https://onesearch.id/Record/IOS3107.47186/Description>
- Ghozali, I. (2013). Aplikasi Analisis Multivariate Dengan Program IBM SPSS 21 Update PLS Regresi. In *International Journal of Physiology* (7th ed.). Badan Penerbit Undip.
- Jamil, M., Khairan, A., & Fuad, A. (2015). Implementasi Aplikasi Telemedicine Berbasis Jejaring Sosial dengan Pemanfaatan Teknologi Cloud Computing. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 1(1). <https://doi.org/10.26418/jp.v1i1.9930>
- Lubis, Z. I. (2021). Analisis Kualitatif Penggunaan Telemedicine sebagai Solusi Pelayanan Kesehatan di Indonesia pada Masa Pandemi COVID-19. *Physiotherapy Health Science (PhysioHS)*, 2(2), 76–82. <https://doi.org/10.22219/physiohs.v2i2.15148>
- Moleong, L. J. (2018). *Metodologi penelitian kualitatif*. PT Remaja Rosdakarya. <https://opac.perpusnas.go.id/DetailOpac.aspx?id=1133305>
- Pasaribu, K. F., Arisjulyanto, D., & Hikmatushaliha, B. T. (2018). Pengembangan telemedicine dalam mengatasi konektivitas dan aksesibilitas pelayanan kesehatan. *Berita Kedokteran Masyarakat*, 34(11). <https://jurnal.ugm.ac.id/bkm/article/view/40576>
- Riyanto, A. (2021). Faktor-Faktor yang Mempengaruhi Pelaksanaan Telemedicine (Systematic Review). *Jurnal Manajemen Informasi Kesehatan Indonesia*, 9(2), 174. <https://doi.org/10.33560/jmiki.v9i2.337>
- Sahadi, S., & Wibowo, M. A. (2014). Pengaruh Faktor Pengembangan Sumber Daya Manusia terhadap Komitmen dan Kinerja pada Manajer Proyek Konstruksi. *Media Komunikasi Teknik Sipil*, 19(1), 67–76. <https://doi.org/10.14710/MKTS.V19I1.7836>



- Sari, G. G., & Wirman, W. (2021). Telemedicine sebagai Media Konsultasi Kesehatan di Masa Pandemi COVID 19 di Indonesia. *Jurnal Komunikasi*, 15(1), 43-54. <https://doi.org/10.21107/ilkom.v15i1.10181>
- Solimun. (2002). *Multivariate Analysis Structural Equation Modelling (SEM) Lisrel dan Amos: Aplikasi Manajemen, Ekonomi Pembangunan, Psikologi Sosial, Kedokteran, dan Agrokompleks*. UM PRESS.
- Sujarweni, V. W. (2014). *SPSS untuk Penelitian* (1st ed.). Pustaka Baru Press.
- Suryani, & Hendryadi. (2015). *Metode riset kuantitatif : teori dan aplikasi pada penelitian bidang manajemen dan ekonomi Islam*. Prenadamedia Group. <https://opac.perpusnas.go.id/DetailOpac.aspx?id=924005>



Naive Bayes untuk Pengukuran Kualitas Media pada Larva BSF (*Black Soldier Fly*) Berbasis *Internet of Things*

Mohammad Faisal Fajar Fadilah ^{(1)*}, Ajib Hanani ⁽²⁾, Totok Chamidy ⁽³⁾

Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim, Malang
e-mail : 18650095@student.uin-malang.ac.id, ajib@uin-malang.ac.id, to2k2013@ti.uin-malang.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 27 Desember 2022, direvisi 16 Mei 2023, diterima 16 Mei 2023, dan dipublikasikan 26 Mei 2023.

Abstract

Piles of waste increase in line with population growth and consumption patterns. The concept of bioconversion using black soldier fly larvae can solve the problem of organic waste management. From these problems, an application of Internet of Things technology is needed. The system implemented aims to allow the system to find out how much accuracy, precision, and recall are in making decisions on media quality values using the Naive Bayes method. The main feature of this Naive Bayes Classifier is the very strong assumption of the independence of each condition or event. From the research results, the system has been successfully built according to the research design, as well as the goals that have been fulfilled in completing the development of the smart maggot. Several sensors used in this study were tested so that sensor performance could be determined by finding the average error value. Three parameters are measured; namely, the temperature obtained an average error of 1.6%, air humidity obtained an average error of 2.03%, and soil moisture obtained an average error of 2.7%. By measuring using Python, the Confusion Matrix is obtained so that the test results from the calculation of the Naive Bayes method can find the data in the form of accuracy, precision, and recall. Accuracy percentage results obtained 92%, precision percentage average results obtained 93%, and recall percentage average results obtained 92%. The conclusion shows the results of the system's accuracy obtained have worked well.

Keywords: Media Measurement, BSF Larvae, Internet of Things, NodeMCU, Naive Bayes

Abstrak

Timbunan sampah bertambah seiring dengan pertumbuhan jumlah penduduk serta pola konsumsi masyarakat. Konsep biokonversi menggunakan larva *Black Soldier Fly* bisa menjadi solusi dalam mengatasi persoalan pengelolaan sampah organik. Dari permasalahan tersebut maka dibutuhkan suatu penerapan teknologi *Internet of Things*. Sistem yang diterapkan bertujuan agar sistem dapat mengetahui seberapa besar tingkat akurasi, presisi, dan *recall* dalam pengambilan keputusan nilai kualitas media menggunakan metode Naive Bayes. Ciri utama dari Naive Bayes Classifier ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian. Dari hasil penelitian yang dilakukan, sistem sudah berhasil dibangun sesuai rancangan penelitian, serta tujuan yang telah terpenuhi dalam menyelesaikan pembangunan *smart maggot*. Beberapa sensor yang digunakan pada penelitian diuji sehingga dapat diketahui kinerja sensor dengan mencari nilai eror rata-rata. Ada 3 parameter yang diukur yaitu suhu diperoleh eror rata-rata yaitu 1,6%, kelembaban udara diperoleh eror rata-rata yaitu 2,03%, dan kelembaban tanah diperoleh eror rata-rata yaitu 2,7%. Dengan mengukur menggunakan Python diperoleh Confusion Matrix, Sehingga hasil pengujian dari perhitungan metode Naive Bayes dapat mengetahui data berupa *accuracy*, *precision*, dan *recall*. Hasil persentase *accuracy* diperoleh 92%, hasil rata-rata persentase *precision* diperoleh 93%, hasil rata-rata persentase *recall* diperoleh 92%. Kesimpulan menunjukkan hasil akurasi dari sistem yang diperoleh berhasil berjalan dengan baik.

Kata Kunci: Pengukuran Media, Larva BSF, *Internet of Things*, NodeMCU, Naive Bayes



1. PENDAHULUAN

Timbunan sampah bertambah seiring dengan pertumbuhan jumlah penduduk serta pola konsumsi masyarakat. Sampah merupakan bentuk konsekuensi dari adanya aktivitas alam juga manusia yang belum mempunyai nilai ekonomis. Tidak dapat dipungkiri, sampah terus selalu ada selama aktivitas kehidupan masih berjalan. Menganalisis permasalahan yang dihadapi dalam pengelolaan sampah di Indonesia, di antaranya kurangnya dasar hukum yang tegas, tempat pembuangan sampah yang tidak memadai, kurangnya usaha dalam melakukan pengomposan, dan kurangnya pengelolaan TPA dengan sistem yang tepat (Chaerul et al., 2007).

Selain sampah organik diolah menjadi kompos dan biogas, ada alternatif lain yang lebih menguntungkan dan cepat yaitu daur ulang sampah organik yang dilakukan menggunakan metode biokonversi. Adapun biokonversi merupakan suatu proses yang melibatkan mikroorganisme seperti bakteri, jamur, ragi, serta larva yang bertujuan untuk mengubah sampah organik menjadi produk yang memiliki nilai tinggi. Konsep biokonversi bisa menjadi solusi dalam mengatasi persoalan pengelolaan sampah organik. Pemanfaatan larva BSF menjadi agen bioreduksi sampah organik sudah dibuktikan pada skala besar menggunakan kemampuan mereduksi sampah organik mencapai 200 ton per hari. Larva BSF sangat rentan dengan perubahan suhu dan hal itu dapat membuat larva BSF mengalami penurunan daya tahan tubuh sebagai akibatnya bisa menyebabkan kematian. Suhu yang optimal bagi larva BSF pada kondisi suhu 30°-38° C pada kandang (Y. Putra & Ariesmayana, 2020), dengan kelembaban udara optimum sekitar 60%-70% (Holmes et al., 2012), dan kondisi kelembaban tanah 70%-80% (Wahyuni et al., 2020), suhu tersebut sudah sesuai dengan standar yang dibutuhkan larva BSF.

Dibutuhkan pengawasan dan pengontrolan terhadap media larva BSF agar menghasilkan kualitas media yang sesuai dengan standar kualitas suhu, kelembaban udara, dan kelembaban tanah agar media layak digunakan oleh larva BSF. Dari permasalahan tersebut maka dibutuhkan suatu penerapan teknologi *Internet of Things* untuk sistem pengontrolan media larva BSF. Tahapan proses kerja dari *Internet of Things* dengan memanfaatkan pemrograman di setiap perintah untuk sebuah instruksi kepada mesin tanpa bantuan manusia. Dengan menggunakan sambungan atau koneksi internet. Seperti bagaimana mengolah data yang diperoleh dari peralatan elektronik melalui sebuah *interface* antara pengguna dan peralatan itu. Penggunaan sensor secara *real time* mengkonversikan ke dalam mesin format yang dimengerti sehingga akan mudah dipertukarkan antara berbagai bentuk format data (Suresh et al., 2014).

Untuk menentukan klasifikasi kualitas media larva BSF yang optimal terdapat banyak metode klasifikasi yang dapat digunakan, salah satunya dengan menggunakan algoritma Naïve Bayes. Penelitian yang dilakukan oleh (Listiana et al., 2015) dalam kasus identifikasi tumbuh kembang anak balita dengan perbandingan algoritma *decision tree* (C4.5) dan Naïve Bayes, diperoleh hasil pengujian rata-rata nilai akurasi metode Naïve Bayes sebesar 96,89% dan algoritma *decision tree* (C4.5) sebesar 89,78%. Penelitian lain yang dilakukan oleh (Hastuti, 2012) pada prediksi data mahasiswa nonaktif dengan melakukan perbandingan metode klasifikasi *logistic regression*, *decision tree*, *neural network*, dan Naïve Bayes. Penelitian tersebut menghasilkan tingkat akurasi metode *logistic regression* sebesar 81,64%, metode *decision tree* sebesar 95,29%, metode *neural network* sebesar 94,56%, dan metode Naïve Bayes sebesar 93,47%. Berdasarkan dari beberapa penelitian sebelumnya, Naïve Bayes merupakan metode yang memiliki akurasi cukup tinggi. Untuk itu penelitian ini, peneliti akan menerapkan algoritma *naïve bayes* untuk mengklasifikasikan kualitas media larva BSF. Berdasarkan uraian latar belakang, sistem yang dibangun berguna untuk menganalisis seberapa besar tingkat akurasi, presisi, dan *recall* dalam pengambilan keputusan nilai kualitas media pada sistem pengukuran kualitas media larva BSF menggunakan metode Naive Bayes.



1.1 Landasan Teori

1.1.1 Penelitian Terkait

Dalam penelitian yang telah dilakukan Suciati & Faruq (2017) ditulis bahwa sampah organik seringkali menyebabkan bau dan penyakit. Belakangan ini ditemukan kegiatan untuk mendaur ulang sampah organik dengan metode biokonversi. Umumnya, organisme yang berperan dalam proses biokonversi adalah bakteri, jamur, dan larva serangga. Larva serangga *Hermetia illucens/Black Soldier Fly* (BSF) ini sering dipergunakan menjadi agen biokonversi, serta lebih dikenal dengan istilah “maggot”. Penelitian ini bertujuan untuk mengetahui pengaruh berbagai jenis media pertumbuhan larva, dengan parameter biomassa larva, wadah, dan kondisi media tumbuh larva.

Penelitian Abidin & Saragih (2020) telah menghasilkan sebuah sistem pada kandang burung puyuh dengan menerapkan *Internet of Things* dan kecerdasan Naive Bayes di dalamnya. Membuat suatu alat untuk mengontrol suhu kandang burung puyuh agar tetap dalam kondisi yang stabil. Serta mengurangi penggunaan listrik yang berlebih daripada lampu yang digunakan. Pada penelitian ini secara khusus dilakukan pengembangan sistem pada *layer* aplikasi dan *layer* servis disistem berbasis *Internet of Things* dengan melakukan penggunaan platform Node-Red dan pembuatan basis data *local* menampung data dari platform. Data dianalisis menggunakan salah satu metode *data mining* berupa metode Naive Bayes.

Penelitian yang telah dilakukan R. A. D. Putra (2021) merancang sistem kandang *maggot* BSF *Internet of Things* menggunakan sensor otomatis. Pada penelitian ini belum memanfaatkan kecerdasan di dalamnya. Adapun penelitian ini membuat sistem *monitoring* serta kontrol suhu lampu untuk budidaya *maggot* BSF berbasis *Internet of Things* sebagai pengontrol sistem. Dalam penelitian ini komponen utama yang digunakan yaitu Arduino UNO, *relay*, Bluetooth HC-05, I2C LCD 16x2, lampu, dan modul DHT22. Cara kerjanya alat ini nantinya jika ingin memonitor suhu atau mengontrol lampu, menggunakan koneksi via Bluetooth dengan membuka aplikasi di *smartphone*. Hasil dari model sistem ini adalah mampu mempertahankan keseimbangan pada kondisi Suhu 30°-38° C, suhu tersebut sudah sesuai standar suhu yang dibutuhkan oleh larva BSF.

1.1.2 Larva Black Soldier Fly (BSF)



Gambar 1 Larva Black Soldier Fly

Larva *Hermetia Illucens* adalah larva yang dihasilkan serangga *hermetia illucens* atau *black soldier fly* yang akhir-akhir ini banyak dikembangkan untuk menjadi pakan ikan, unggas, dan hewan peliharaan. Larva BSF dikenal juga dengan *maggot*. Larva BSF mengandung protein tinggi yang berpotensi dijadikan pakan alternatif pengganti tepung ikan, bungkil kedelai, dan *meat bone meal*. Penggunaan larva sebagai bahan baku pakan mempunyai banyak keunggulan yaitu keberadaannya dapat ditemukan hampir diseluruh dunia, bisa mereduksi bahan organik, bisa hidup pada toleransi pH yang cukup luas, tidak membawa penyakit, masa hidup yang cukup lama (Indariyanti & Barades, 2018). Larva BSF dapat dilihat pada Gambar 1.



Dalam siklus hidupnya lalat *hermetia illucens* mempunyai beberapa stadia yaitu fase dewasa, fase telur, fase larva, fase prepupa, dan fase pupa (Diener & Tov, 2007). Siklus hidup lalat BSF membutuhkan total daur hidup selama 40 hari di mana fase telur akan terjadi selama 3 hari dilanjutkan 18 hari fase larva. Larva menuju tahap prepupa 14 hari, lalu tiga hari setelahnya akan menjadi pupa kemudian bertransformasi menjadi lalat dewasa yang bertahan selama 3 hari dan akan mati jika telah kawin. Untuk sekali bertelur mampu menghasilkan 500 sampai 900 telur dan kekhawatiran akan over populasi tidak akan terjadi karena predator sangat banyak (Fatmasari, 2018).

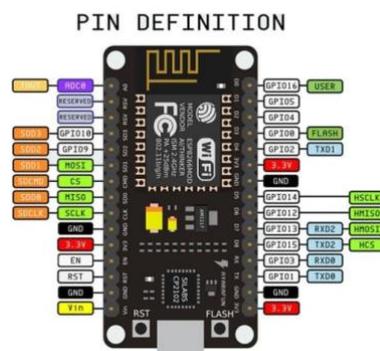
1.1.3 Naïve Bayes

Menurut Olson & Delen (2008) menjelaskan Naive Bayes untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari tabel keputusan. Naïve Bayes Classifier bekerja dengan sangat baik dibandingkan model *classifier* lainnya.

1.1.4 Internet of Things

Ekosistem *Internet of Things* atau disingkat dengan IoT terdiri dari perangkat pintar berkemampuan web yang menggunakan sistem tertanam (*embedded system*), seperti prosesor, sensor, dan perangkat keras komunikasi untuk mengumpulkan, mengirim, dan bertindak sesuai data yang mereka peroleh dari lingkungan mereka. Perangkat IoT berbagi data sensor yang mereka kumpulkan dengan menghubungkan ke *gateway* IoT atau *edge device* lainnya, di mana data dikirim ke komputer *cloud* untuk dianalisis secara lokal. Terkadang, perangkat ini berkomunikasi dengan perangkat terkait lainnya dan bertindak berdasarkan informasi yang mereka dapatkan satu dengan yang lain. Sebagian besar perangkat melakukan pekerjaan tanpa campur tangan manusia, meskipun orang masih dapat berinteraksi dengan perangkat tersebut, misalnya ketika saat mengkonfigurasinya, memberi mereka instruksi, atau mengakses data. Konektivitas jaringan dan protokol komunikasi yang digunakan pada perangkat yang mendukung web ini sangat tergantung pada spesifikasi aplikasi IoT yang digunakan. IoT juga dapat menggunakan kecerdasan buatan (*artificial intelligence*) dan pembelajaran mesin (*machine learning*) untuk membantu membuat proses pengumpulan data lebih mudah dan lebih dinamis.

1.1.5 Mikrokontroler NodeMCU ESP8266



Gambar 2 NodeMCU ESP8266

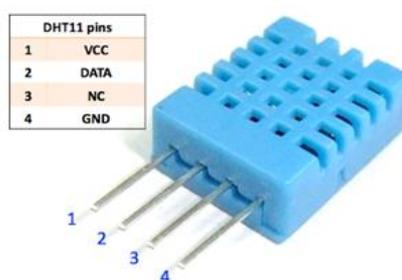
NodeMCU adalah sebuah *open-source* platform *Internet of Things* serta pengembangan kit menggunakan bahasa pemrograman untuk membantu dalam membuat *prototype* produk *Internet of Things* dan juga bisa menggunakan *sketch* dengan Arduino IDE. Pengembangan kit didasarkan pada modul ESP8266 yang mengintegrasikan GPIO, PWM (*Pulse Width Modulation*), IIC, 1-Wire, serta ADC (*Analog to Digital Converter*) semuanya dalam sebuah *board*. NodeMCU



memiliki ukuran panjang 4,83 cm, lebar 2,54 cm, serta berat 7 gram. *Board* ini sudah dilengkapi dengan fitur *WiFi* dan *Firmware* yang bersifat *open source*. NodeMCU dapat dilihat pada Gambar 2.

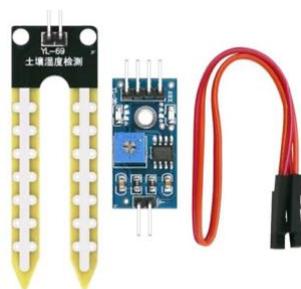
1.1.6 Sensor Suhu DHT11

Sensor DHT11 adalah rangkaian komponen sensor dan IC kontroler yang dikemas dalam satu paket. Pada sensor ini terdapat 4 pin dan juga ada yang memiliki 3 pin. Akan tetapi tidak menjadi masalah karena dalam penerapannya tidak ada yang berbeda. Pada *body* sensor yang berwarna biru atau putih terdapat sebuah Resistor tipe NTC (*Negative Temperature Coefficient*). Resistor jenis ini memiliki karakteristik nilai resistansinya berbanding terbalik dengan kenaikan suhu. Artinya, semakin tinggi suhu ruangan maka nilai resistansi NTC semakin kecil. Sebaliknya nilai resistansi meningkat ketika suhu disekitar sensor menurun. Selain itu di dalamnya terdapat sebuah sensor kelembaban dengan karakteristik resistif terhadap kadar air yang berubah di udara. Kedua data sensor diolah di dalam IC kontroler. IC kontroler mengeluarkan *output* data dalam bentuk *single wire bi-directional*. Sensor suhu DHT11 dapat dilihat pada Gambar 3.



Gambar 3 Sensor DHT11

1.1.7 Sensor Soil Moisture YL-69



Gambar 4 Sensor Soil Moisture YL-69

Soil moisture sensor merupakan sensor yang dapat mendeteksi kelembaban pada tanah. Sensor ini terdiri dua *probe* untuk melewati arus melalui tanah, kemudian membaca resistansinya untuk mendapatkan nilai tingkat kelembaban. Semakin banyak air membuat tanah lebih mudah menghantarkan listrik (resistansi kecil), sedangkan tanah yang kering sulit menghantarkan listrik (resistansi besar). Sensor ini membantu untuk mengingatkan atau memantau tingkat kelembaban tanah. Sensor kelembaban tanah *Soil Moisture YL-69* dapat dilihat pada Gambar 4.

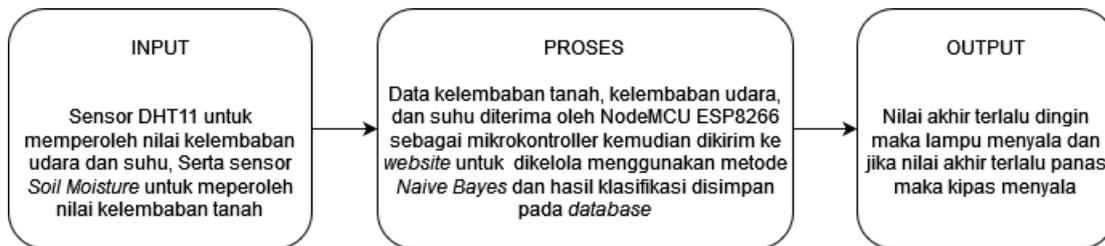
2. METODE PENELITIAN

2.1 Sistem

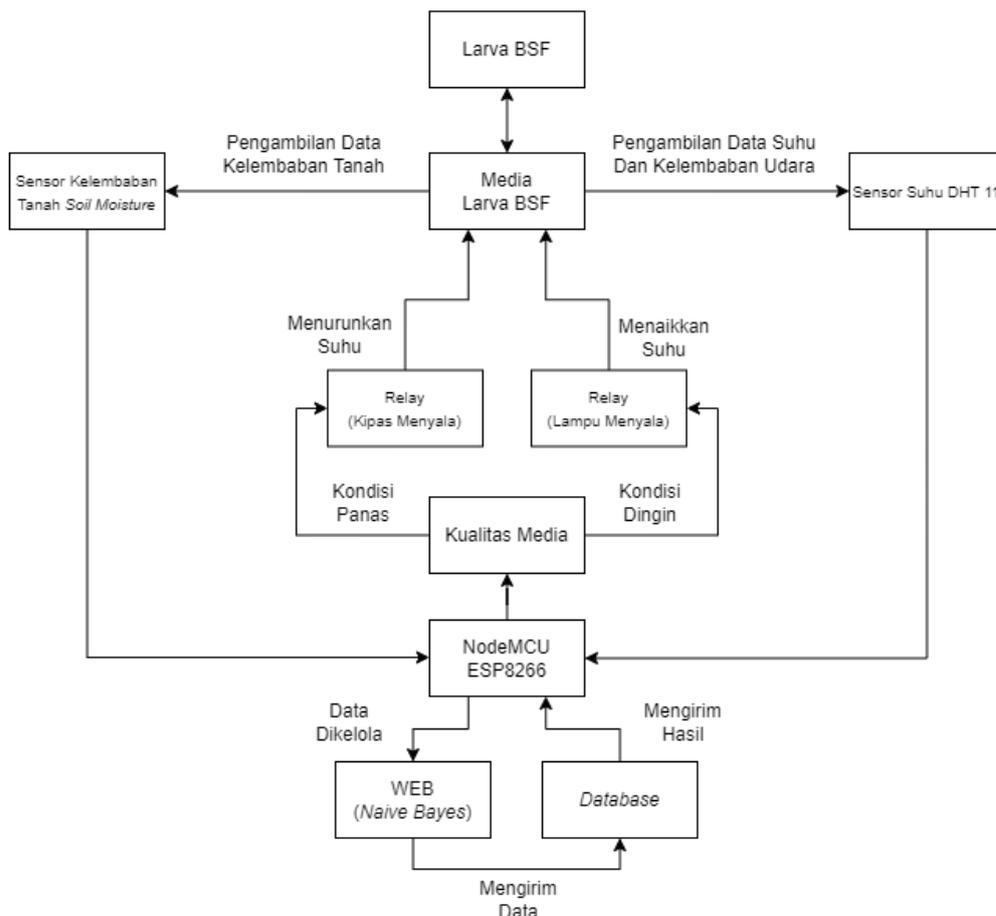
Alur sistem yang dibuat memiliki tujuan untuk memudahkan dalam mengetahui proses berjalannya suatu sistem yang dirancang. Alur sistem dijelaskan pada Gambar 5 dan desain sistem ditunjukkan pada Gambar 6.



Pada desain sistem terdapat alur berjalannya sistem. Untuk dapat menjalankan aksi *output* sesuai hasil klasifikasi, sistem akan melewati beberapa proses. Dengan menggunakan NodeMCU ESP8266 sebagai mikrokontroler yang terhubung sensor suhu DHT11 dan sensor *soil moisture* YL-69 untuk mendapatkan data masukan. NodeMCU ESP8266 menerima data nilai suhu, nilai kelembaban udara, dan kelembaban tanah dari sensor, kemudian data tersebut dikelola pada web menggunakan metode Naive Bayes untuk mendapatkan hasil klasifikasi berupa label dan menyimpannya ke *database*. Setelah mendapatkan hasil klasifikasi berupa label maka NodeMCU ESP8266 memberikan aksi berupa *output* pendinginan atau penghangatan atau tidak keduanya. *Output* yang dihasilkan pada penelitian ini yaitu label dingin, optimal, dan panas. Kemudian sistem merespon dengan melakukan pengontrolan suhu pada media larva BSF dan melakukan aksi tidak mengaktifkan relay atau mengaktifkan relay (lampu menyala atau kipas angin menyala) sesuai hasil klasifikasi media sehingga media tetap dalam keadaan optimal.



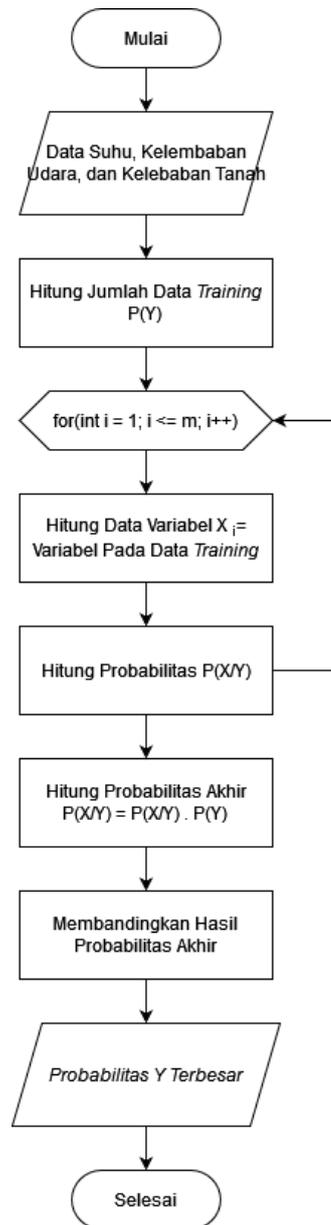
Gambar 5 Alur Sistem



Gambar 6 Desain Sistem



2.2 Metode Naïve Bayes



Gambar 7 Flowchart Klasifikasi Metode Naïve Bayes

Pada *flowchart* klasifikasi metode Naïve Bayes yang terdapat pada Gambar 7, terdapat beberapa tahap untuk memperoleh hasil klasifikasi. Tahap pertama yaitu nilai kelembaban tanah, kelembaban udara, dan suhu sebagai data input. Tahap berikutnya yaitu menghitung jumlah probabilitas Y (dingin, optimal, dan panas) atau label pada data *training*. Kemudian menghitung variabel X ke-*i*. Tahap selanjutnya yaitu menghitung jumlah probabilitas X terhadap kondisi Y. Proses tersebut dilakukan berulang pada setiap variabel X. Jika semua variabel X sudah dihitung maka langkah selanjutnya menghitung probabilitas akhir. Perhitungan probabilitas akhir yaitu probabilitas X berdasarkan kondisi Y dikalikan dengan probabilitas X1, X2, dan X3. Dari hasil proses tersebut maka memperoleh nilai masing-masing probabilitas Y akhir. Kemudian membandingkan hasil probabilitas akhir untuk memperoleh probabilitas Y terbesar. Tabel 1 merupakan tabel kategori tiap variabel dengan contoh data *training* ditunjukkan pada Tabel 2 dan contoh data *testing* ditunjukkan pada Tabel 3.



Tabel 1 Kategori Tiap Variabel

Variabel	Nilai	Keterangan
Suhu	<30°	Terlalu Dingin
	30°– 38°	Normal
	>38°	Terlalu Panas
Kelembaban Udara	<60%	Kurang
	60% - 70%	Cukup
	>70%	Lebih
Kelembaban Tanah	<70%	Kering
	70% - 80%	Cukup
	>80%	Basah

Tabel 2 Contoh Data Training

No	Suhu	Kelembaban Udara	Kelembaban Tanah	Label
1	39°	75%	60%	Panas
2	37°	78%	77%	Optimal
3	29°	70%	78%	Dingin
4	40°	59%	50%	Panas
5	31°	69%	72%	Optimal
6	27°	75%	87%	Dingin
7	29°	70%	65%	Optimal
8	27°	78%	90%	Dingin
9	28°	79%	79%	Dingin
10	26°	72%	86%	Dingin

Tabel 3 Contoh Data Testing

No	Kelembaban Tanah	Kelembaban Udara	Suhu	Label
1	79%	75%	28°	?

Hasil perhitungan dari tabel data *training* dapat digunakan untuk mengklasifikasi kualitas media tabel *testing* yang dijabarkan menggunakan Pers. (1) sampai (3).

$$P(Y_i) = \frac{Y_i}{\sum_{j=1}^n Y_j} \quad (1)$$

Di mana Y_i merupakan label Y dengan entitas ke-i dan n adalah jumlah label dengan i merupakan bilangan 1, ..., m.

a) Menghitung jumlah probabilitas variabel Y

$$P(Y = Dingin) = \frac{5}{10} = 0,5$$

$$P(Y = Optimal) = \frac{3}{10} = 0,3$$

$$P(Y = Panas) = \frac{2}{10} = 0,2$$

$$P(X_i|Y_j) = \frac{X_i}{Y_j} \quad (2)$$

Di mana X_i adalah label X ke-I dengan i menunjukkan bilangan 1, ..., m.

b) Menghitung probabilitas variabel X terhadap variabel Y

$$P(X_1 = Cukup|Y = Dingin) = \frac{2}{5}$$



$$P(X_1 = Cukup|Y = Optimal) = \frac{2}{3}$$

$$P(X_1 = Cukup|Y = Panas) = \frac{0}{2}$$

$$P(X_2 = Lebih|Y = Dingin) = \frac{4}{5}$$

$$P(X_2 = Lebih|Y = Optimal) = \frac{1}{3}$$

$$P(X_2 = Lebih|Y = Panas) = \frac{1}{2}$$

$$P(X_3 = TerlaluDingin|Y = Dingin) = \frac{5}{5}$$

$$P(X_3 = TerlaluDingin|Y = Optimal) = \frac{1}{3}$$

$$P(X_3 = TerlaluDingin|Y = Panas) = \frac{0}{2}$$

$$P(X_1 \dots X_m|Y_i) = \left(\prod_{j=1}^m P(X_j|Y_i)\right)P(Y_i) \quad (3)$$

Di mana Y_i yaitu label Y ke-i dan X_i adalah label X ke-i dengan i merupakan 1, ..., m.

c) Membandingkan hasil probabilitas tiap label

$$P(X_1 = Cukup, X_2 = Lebih, X_3 = TerlaluDingin|Y = Dingin) = \frac{2}{5} \times \frac{4}{5} \times \frac{5}{5} \times \frac{5}{10} = 0,16$$

$$P(X_1 = Cukup, X_2 = Lebih, X_3 = TerlaluDingin|Y = Optimal) = \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{3}{10} = 0,02$$

$$P(X_1 = Cukup, X_2 = Lebih, X_3 = TerlaluDingin|Y = Panas) = \frac{0}{2} \times \frac{1}{2} \times \frac{0}{2} \times \frac{2}{10} = 0$$

Dari perhitungan di atas diperoleh bahwa probabilitas terbesar adalah pada $P(Y = Dingin)$ maka dapat disimpulkan data pada contoh data *testing* yaitu label berada pada kelas Dingin.

Untuk membandingkan akurasi nilai suhu dan nilai kelembaban udara pada sensor DHT11 dengan alat pengukur termometer. Sementara itu, untuk membandingkan akurasi nilai kelembaban tanah pada sensor *soil moisture* YL-69 dengan akurasi alat pengukur *moisture meter*. Setelah itu, untuk memperoleh hasil persentase *error* dapat menggunakan perhitungan dengan rumus pada Pers. (4).

$$\% \text{ error} = \frac{|X-Y|}{Y} \times 100 \quad (4)$$

Di mana X merupakan nilai dari sensor dan Y merupakan nilai dari alat ukur termometer dan *moisture meter*.

Pada pengujian metode Naïve Bayes dalam memprediksi kualitas media larva BSF memerlukan *confusion matrix* untuk mengevaluasi performa suatu algoritma. *Confusion matrix* merepresentasikan prediksi dan kondisi sebenarnya dari data yang dihasilkan oleh algoritma. Dengan adanya *confusion matrix*, maka *accuracy*, *precision*, dan *recall* dapat ditentukan. Tabel 4 merupakan tabel *confusion matrix* variabel.

Tabel 4 Confusion Matrix

	TRUE	FALSE
Nilai Prediksi TRUE	(TP)	(FP)
FALSE	(FN)	(TN)

Pada tabel *confusion matrix* terdapat *True Positive* (TP), *False Negative* (FN), *False Positive* (FP), dan *True Negative* (TN). Dari keempat prediksi tersebut memiliki pengertian yang berbeda, di mana *True Positive* (TP) merupakan jumlah data dengan nilai sebenarnya positif dan diprediksi benar oleh algoritmal. *False Negative* (FN) merupakan jumlah data dengan nilai sebenarnya



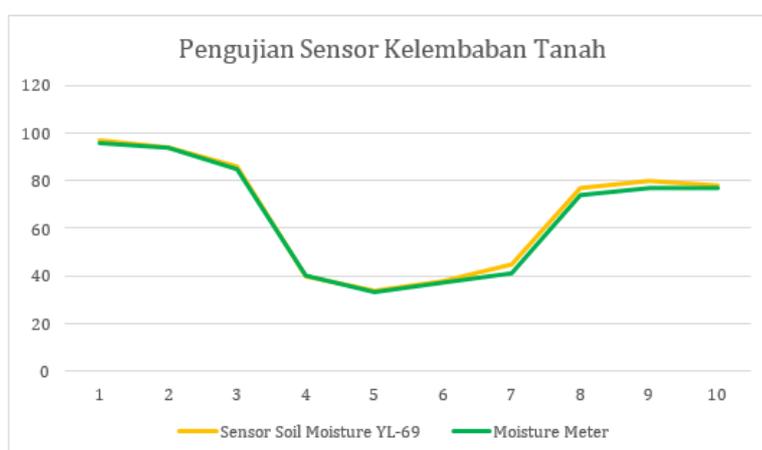
positif dan diprediksi salah oleh algoritma. *False Positive* (FP) merupakan jumlah data dengan nilai sebenarnya negatif dan diprediksi benar oleh algoritma. Sedangkan *True Negative* (TN) merupakan jumlah data dengan nilai sebenarnya negatif dan dinilai salah oleh algoritma (Fibrianda & Bhawiyuga, 2018).

3. HASIL DAN PEMBAHASAN

3.1 Pengujian Sensor

3.1.1 Sensor *Soil Moisture* YL-69

Perbandingan sensor kelembaban tanah *soil moisture* YL-69 dengan alat ukur *moisture* meter dilakukan sebagai pengujian terhadap sensor agar mengetahui perbedaan nilai kelembaban tanah. Grafik hasil pengujian sensor *soil moisture* YL-69 ditunjukkan pada Gambar 8.



Gambar 8 Grafik Hasil Pengujian Sensor *Soil Moisture* YL-69

Grafik perbedaan hasil pengujian kelembaban tanah menggunakan sensor *soil moisture* YL-69 dengan alat ukur *moisture meter*. Dari data yang diperoleh menggunakan sensor maupun alat ukur, selanjutnya adalah mencari nilai persentase *error* dan mencari nilai rata-rata *error*.

$$Error = \frac{Sensor\ Soil\ Moisture\ YL - 69 - Moisture\ Meter}{Moisture\ Meter} \times 100\% = \frac{97 - 96}{96} \times 100\% = 1,04\%$$

Diperoleh hasil persentase dari perhitungan *error* antara data sensor *soil moisture* YL-69 dengan alat ukur *moisture* meter yaitu 1,04%

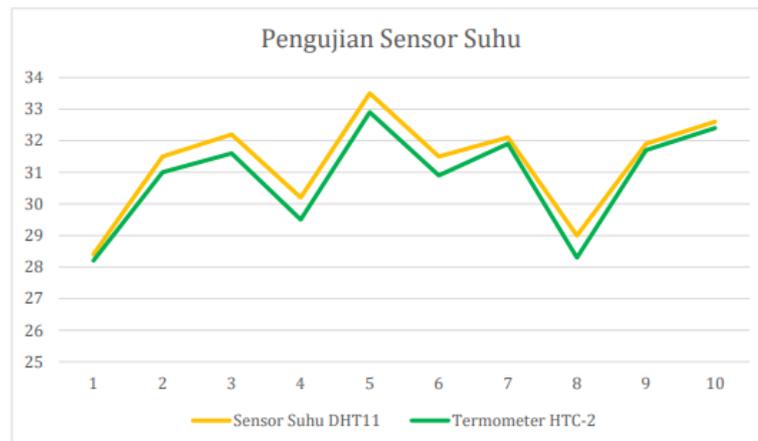
$$Rata - rata\ error = \frac{\sum 27,0}{\sum 10} = \frac{27,0}{10} = 2,70$$

Perbandingan antara sensor *soil moisture* YL-69 dengan alat ukur *moisture* meter diperoleh nilai rata-rata *error* yaitu 2,70%.

3.1.2 Sensor Suhu DHT11

Perbandingan sensor suhu DHT11 dengan alat ukur termometer dilakukan sebagai pengujian terhadap sensor agar mengetahui perbedaan nilai suhu. Grafik hasil pengujian sensor suhu DHT11 ditunjukkan pada Gambar 9.





Gambar 9 Grafik Hasil Pengujian Sensor Suhu DHT11

Grafik perbedaan hasil pengujian suhu menggunakan sensor suhu DHT11 dengan alat ukur termometer. Dari data yang diperoleh menggunakan sensor maupun alat ukur, selanjutnya adalah mencari nilai persentase *error* dan mencari nilai rata-rata *error*.

$$Error = \frac{Sensor\ Suhu\ DHT11 - Termometer\ HTC-2}{Termometer\ HTC-2} \times 100\% = \frac{28,4 - 28,2}{28,2} \times 100\% = 0,71\%$$

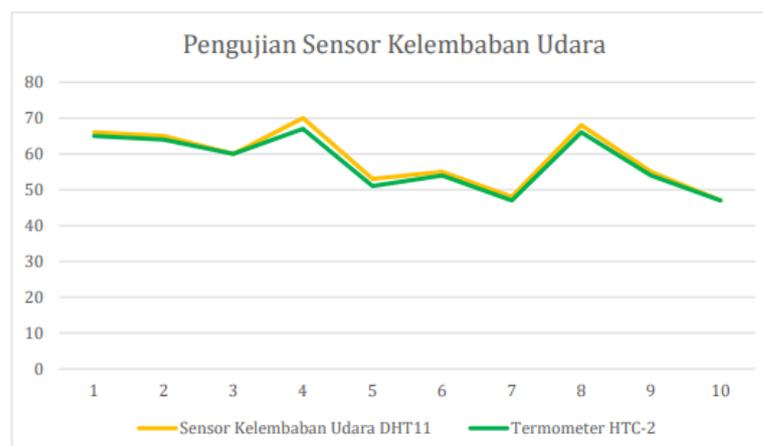
Diperoleh hasil persentase dari perhitungan *error* antara data sensor suhu DHT11 dengan alat ukur termometer yaitu 1,04%.

$$Rata - rata\ error = \frac{\sum 16,0}{\sum 10} = \frac{16,0}{10} = 1,60$$

Perbandingan antara sensor suhu DHT11 dengan alat ukur termometer diperoleh nilai rata-rata *error* yaitu 1,60%.

3.1.3 Sensor Kelembaban Udara DHT11

Perbandingan sensor kelembaban udara DHT11 dengan alat ukur termometer dilakukan sebagai pengujian terhadap sensor agar mengetahui perbedaan nilai kelembaban udara. Grafik hasil pengujian sensor kelembaban udara DHT11 ditunjukkan pada Gambar 10.



Gambar 10 Grafik Hasil Pengujian Sensor Kelembaban Udara DHT11



Grafik perbedaan hasil pengujian kelembaban udara menggunakan sensor kelembaban udara DHT11 dengan alat ukur termometer. Dari data yang diperoleh menggunakan sensor maupun alat ukur, selanjutnya adalah mencari nilai persentase *error* dan mencari nilai rata-rata *error* seperti pada.

$$Error = \frac{Sensor\ Kelembaban\ Udara\ DHT11 - Termometer\ HTC-2}{Termometer\ HTC-2} \times 100\% = \frac{66-65}{65} \times 100\% = 1,54$$

Diperoleh hasil persentase dari perhitungan *error* antara data sensor kelembaban udara DHT11 dengan alat ukur termometer yaitu 1,54%.

$$Rata - rata\ error = \frac{\sum 20,3}{\sum 10} = \frac{20,3}{10} = 2,03$$

Perbandingan antara sensor kelembaban udara DHT11 dengan alat ukur termometer diperoleh nilai rata-rata *error* yaitu 2,03%.

3.2 Confusion Matrix

Hasil *confusion matrix* ditunjukkan pada Gambar 11.

		PREDIKTED		
		Dingin	Optimal	Panas
AKTUAL	Dingin	6	1	0
	Optimal	0	11	0
	Panas	0	1	6

Gambar 11 Hasil Confusion Matrix

Perhitungan pada Pers. (5) merupakan perhitungan untuk mendapatkan nilai akurasi.

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (5)$$

$$Akurasi = \frac{6 + 17}{6 + 17 + 1 + 1} \times 100\% = 92\%$$

Untuk melakukan perhitungan *precision* dengan rumus pada Pers. (6).

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

$$Precision_{dingin} = \frac{6}{6 + 0} \times 100\% = 100\%$$

$$Precision_{optimal} = \frac{11}{11 + 1} \times 100\% = 85\%$$

$$Precision_{panas} = \frac{6}{6 + 0} \times 100\% = 100\%$$

Nilai *recall* digunakan untuk menentukan seberapa lengkap sebuah algoritma. Perhitungan *recall* dengan rumus pada Pers. (7).



$$Recall = \frac{TP}{TP+FN} \quad (7)$$

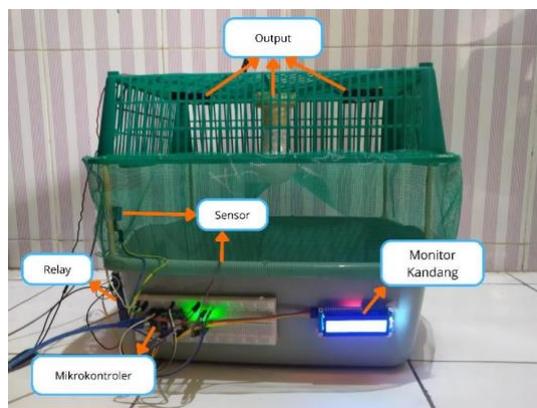
$$Recall_{dingin} = \frac{6}{6+1} \times 100\% = 86\%$$

$$Recall_{optimal} = \frac{11}{11+0} \times 100\% = 100\%$$

$$Recall_{panas} = \frac{6}{6+1} \times 100\% = 86\%$$

3.3 Aplikasi

3.3.1 Kandang *Smart Maggot*



Gambar 12 Kandang *Smart Maggot*

Gambar 12 menunjukkan kandang *smart maggot*. Kandang larva BSF dibangun dengan boks ukuran lebar 30 cm, panjang 38 cm, dan tinggi 36 cm dengan kerangka menggunakan boks plastik dan kayu sebagai penyangga. Atap yang digunakan yaitu plastik berongga. Pemasangan instalasi pendinginan dan pemanasan dipasang pada kandang di atas media larva BSF dengan pendinginan menggunakan 1 kipas ukuran 5 cm sebanyak 2 buah dan pemanasan menggunakan lampu sebanyak 1 buah.

3.3.2 Aplikasi *Smart Maggot*



Gambar 13 Aplikasi Web *Smart Maggot*



Aplikasi smart maggot dibangun dengan gabungan teknologi Internet of Things dan aplikasi berbasis web dengan bahasa PHP seperti pada Gambar 13. Aplikasi ini dapat melakukan pemantauan dengan jarak jauh serta melakukan pengontrolan secara otomatis. Aplikasi dibangun dengan memperhatikan *User Interface* (UI) dan *User Experience* (UX) dengan tujuan agar pengguna mudah dan nyaman saat menggunakan aplikasi.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan oleh penulis dari hasil pengujian dan analisis dapat disimpulkan bahwa algoritma Naïve Bayes dapat diaplikasikan untuk sistem kandang *smart maggot*. Hasil dari pengujian yang dilakukan, penelitian dengan menggunakan *dataset* yang diambil pada penelitian ini memiliki akurasi yang cukup tinggi dibandingkan pada penelitian sebelumnya yang belum memanfaatkan algoritma Naïve Bayes Classifier di dalamnya. Penelitian ini sudah berhasil dibangun sesuai rancangan penelitian, serta tujuan yang telah terpenuhi dalam menyelesaikan pembangunan *smart maggot*. Beberapa sensor yang digunakan pada penelitian diuji sehingga dapat diketahui kinerja dari sensor-sensor tersebut dengan cara mencari nilai *error* rata-rata. Ada 3 parameter yang diukur pada penelitian yaitu suhu diperoleh *error* rata-rata 1,6%, kelembaban udara diperoleh *error* rata-rata 2.03%, dan kelembaban tanah diperoleh *error* rata-rata 2,7%. Dengan mengukur menggunakan Python diperoleh *Confusion Matrix*, Sehingga hasil pengujian dari perhitungan metode Naive Bayes dapat diketahui dengan data yang didapatkan berupa *accuracy*, *precision*, dan *recall*. Hasil persentase *accuracy* yang diperoleh adalah 92%, hasil rata-rata persentase *precision* diperoleh 93%, dan hasil rata-rata persentase *recall* diperoleh 92%. Kesimpulan menunjukkan hasil akurasi dari sistem yang diperoleh berhasil berjalan dengan baik.

DAFTAR PUSTAKA

- Abidin, A. Z. Z., & Saragih, N. A. A. (2020). Sistem Monitoring Kandang Burung Puyuh Berbasis Internet of Things pada Platform Node-RED Menggunakan Metode Naive Bayes. *Jurnal Teknologi Dan Komunikasi STMIK Subang*, 13(1), 16–25. <https://doi.org/10.47561/a.v13i1.164>
- Chaerul, M., Tanaka, M., & Shekdar, A. V. (2007). Municipal Solid Waste Management in Indonesia: Status and The Strategic Actions. *Journal of the Faculty of Environmental Science and Technology*, 2(1), 41–49.
- Diener, E., & Tov, W. (2007). Subjective Well-Being and Peace. *Journal of Social Issues*, 63(2), 421–440. <https://doi.org/10.1111/j.1540-4560.2007.00517.x>
- Fatmasari, L. (2018). *Tingkat Densitas Populasi, Bobot, dan Panjang Maggot (Hermetia Illucens) pada Media yang Berbeda* [UIN Raden Intan Lampung]. <http://repository.radenintan.ac.id/3265/>
- Fibrianda, M. F., & Bhawiyuga, A. (2018). Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 3112–3123. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2559>
- Hastuti, K. (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif. *Seminar Nasional Teknologi Informasi & Komunikasi Terapan*, 241–249. <https://publikasi.dinus.ac.id/index.php/semantik/article/view/132/>
- Holmes, T. R. H., Jackson, T. J., Reichle, R. H., & Basara, J. B. (2012). An assessment of surface soil temperature products from numerical weather prediction models using ground-based measurements. *Water Resources Research*, 48(2). <https://doi.org/10.1029/2011WR010538>
- Indariyanti, N., & Barades, E. (2018). Evaluasi Biomassa dan Kandungan Nutrisi Magot (*Hermetia illucens*) pada Media Budidaya yang Berbeda. *Prosiding Seminar Nasional Pengembangan Teknologi Pertanian*, 137–141. <https://doi.org/10.25181/PROSEMNAS.V2018I0.1151>
- Listiana, M., Sujalwo, & Gunawan, D. (2015). *Perbandingan Algoritma Decision Tree (C4.5) Dan Naïve Bayes Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita (Studi Kasus Puskesmas Kartasura)* [Universitas Muhammadiyah Surakarta]. <https://eprints.ums.ac.id/36124/>
- Olson, D. L., & Delen, D. (2008). Advanced data mining techniques. In *Advanced Data Mining*



- Techniques*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-76917-0/COVER>
- Putra, R. A. D. (2021). Monitoring dan Kontrol Suhu Lampu untuk Budidaya Maggot BSF Berbasis IoT (Lamp Temperature Monitoring and Control for IoT-Based Maggot BSF Cultivation). *JURNAL TRANSIT*, 9. <https://repository.usm.ac.id/detail-jurnalmahasiswa-979.html>
- Putra, Y., & Ariesmayana, A. (2020). Efektifitas Penguraian Sampah Organik Menggunakan Maggot (BSF) di Pasar Rau Trade Center. *Jurnal Lingkungan Dan Sumberdaya Alam (JURNALIS)*, 3(1), 11–24. <https://ejournal.lppm-unbaja.ac.id/index.php/jls/article/view/888>
- Suciati, R., & Faruq, H. (2017). Efektifitas Media Pertumbuhan Maggots *Hermetia Illucens* (Lalat Tentara Hitam) Sebagai Solusi Pemanfaatan Sampah Organik. *BIOSFER: Jurnal Biologi Dan Pendidikan Biologi*. <https://doi.org/10.23969/biosfer.v2i1.356>
- Suresh, G., Gunasekar, P. H., Kokila, D., Prabhu, D., Dinesh, D., Ravichandran, N., Ramesh, B., Koodalingam, A., & Vijaiyan Siva, G. (2014). Green synthesis of silver nanoparticles using *Delphinium denudatum* root extract exhibits antibacterial and mosquito larvicidal activities. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 127, 61–66. <https://doi.org/10.1016/j.saa.2014.02.030>
- Wahyuni, L., Ramdani, M. R., Imama, N. O., Larasati, V. E., Fahmi, A. R., & Hermana, W. (2020). Suplementasi Sari Belimbing Wuluh (*Averrhoa bilimbi* L.) dalam Air Minum terhadap Produktivitas Telur Puyuh. *Jurnal Ilmu Nutrisi Dan Teknologi Pakan*, 18(2), 54–61. <https://doi.org/10.29244/jintp.18.2.54-61>



Regresi Logistik Multinomial untuk Prediksi Kategori Kelulusan Mahasiswa

Rafika Syahrani^{(1)*}, Suhartono⁽²⁾, Syahiduz Zaman⁽³⁾

Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim, Malang
e-mail : 17650003@student.uin-malang.ac.id, {suhartono,syahid}@ti.uin-malang.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 22 Desember 2022, direvisi 8 April 2023, diterima 10 April 2023, dan dipublikasikan 26 Mei 2023.

Abstract

Students must meet certain goals to earn a degree but can extend their time at university or drop out (DO). The problem of dropping out of students has become an important issue for tertiary institutions to ensure the success or graduation of students and reduce dropouts. DO can affect the accreditation of the tertiary institution. The quality of higher education institutions in Indonesia is measured based on accreditation from the National Accreditation Board for Higher Education or BAN-PT. One of the main standards measured is the Quality of Students and Graduates. The quality of educational accreditation is measured by the percentage of student graduation and the university's strategy to retain students. To predict student graduation based on graduation time categories, researchers collected academic data from students in 2012-2018 at the Informatics Engineering Study Program, State Islamic University of Maulana Malik Ibrahim Malang. The variables used as predictors are gender, type of entry pathway, and grade point average from semesters one to six. The resulting model was evaluated to obtain an accuracy value of 85.5%, a precision of 78.5%, a recall of 93.9%, and a micro f1-score of 89.8%. An accuracy value of 85.5% indicates that the system can classify properly using the logistic regression model.

Keywords: Categories, Graduation, Prediction, Logistic Regression, Machine Learning

Abstrak

Mahasiswa harus memenuhi tujuan tertentu untuk mendapatkan gelar, tetapi dapat memperpanjang waktu di universitas atau *dropout* (DO). Masalah DO mahasiswa telah menjadi isu penting bagi pengelola perguruan tinggi untuk menjamin keberhasilan atau kelulusan mahasiswa dan mengurangi DO. DO dapat mempengaruhi akreditasi perguruan tinggi tersebut. Kualitas perguruan tinggi di Indonesia diukur berdasarkan akreditasi yang dilaksanakan oleh Badan Akreditasi Nasional Perguruan Tinggi atau BAN PT. Salah satu standar utama yang diukur adalah Kualitas Mahasiswa dan Lulusan. Kualitas akreditasi pendidikan diukur dari persentase kelulusan mahasiswa dan strategi perguruan tinggi untuk mempertahankan mahasiswa. Untuk memprediksi kelulusan mahasiswa berdasarkan kategori waktu kelulusan, peneliti mengumpulkan data akademik mahasiswa tahun 2012-2018 pada Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang. Variabel yang digunakan sebagai prediktor adalah jenis kelamin, jenis jalur masuk, dan indeks prestasi semester satu sampai enam. Model yang dihasilkan diuji dengan evaluasi, sehingga didapatkan nilai *accuracy* sebesar 85,5%, *precision* sebesar 78,5%, *recall* sebesar 93,9%, dan *micro f1-score* sebesar 89,8%. Nilai *accuracy* sebesar 85,5% menyatakan sistem dapat melakukan klasifikasi dengan baik menggunakan model regresi logistik.

Kata Kunci: Kategori, Kelulusan, Prediksi, Regresi Logistik, Pembelajaran Mesin

1. PENDAHULUAN

Mahasiswa harus memenuhi tujuan tertentu untuk mendapatkan gelar, dan dalam banyak kasus dapat memperpanjang waktu mereka di perguruan tinggi atau mereka *dropout* (Dalipi et al., 2018). *Dropout* adalah tindakan yang dilakukan oleh perguruan tinggi untuk mengeluarkan mahasiswa karena perilaku yang tidak baik dan masa studi kuliah yang sudah melewati batas (Lu et al., 2018). Umumnya mahasiswa S1 menyelesaikan studinya dalam waktu yang ditentukan, yaitu delapan semester atau 4 tahun. Masalah yang timbul pada mahasiswa *dropout*



adalah batas masa studi mahasiswa yang ditempuh mahasiswa maksimal sampai 14 semester atau 7 tahun (Agwil et al., 2020).

Dropout merupakan masalah yang memengaruhi semua perguruan tinggi di seluruh dunia, seperti, penurunan pendaftaran, pendapatan berkurang untuk perguruan tinggi, kemalangan moneter untuk negara yang mendukung penyelidikan, dan menetapkan masalah sosial untuk studi sarjana, keluarga mereka, dan masyarakat pada umumnya (Asha et al., 2020). Mahasiswa *dropout* terdiri jenis mahasiswa yang dikeluarkan, putus kuliah, dan mengundurkan diri.

Masalah mahasiswa *dropout* telah mendapat perhatian yang luas dan telah menjadi isu penting bagi pengelola perguruan tinggi untuk menjamin keberhasilan atau kelulusan mahasiswa dan mengurangi angka mahasiswa *dropout* (Yaacob et al., 2020). Pada tahun 2019 angka mahasiswa *dropout* mencapai sebesar 602.208 mahasiswa dari jumlah mahasiswa yang terdaftar, data mahasiswa yang *dropout* didapatkan dari statisik pendidikan tinggi tahun 2020 (Handini et al., 2020).

Mencegah meningkatnya angka mahasiswa *dropout* sangat penting di berbagai lembaga pendidikan. Karakteristik *dropout* menjadi perhatian utama. Konsekuensinya termasuk kerugian finansial, tingkat kelulusan yang lebih rendah, dan reputasi perguruan tinggi yang buruk (Alturki et al., 2022). Kualitas perguruan tinggi di Indonesia diukur berdasarkan akreditasi yang dilaksanakan oleh Badan Akreditasi Nasional Perguruan Tinggi atau BAN PT. Kualitas tersebut diukur berdasarkan 9 standar utama, salah satunya adalah mahasiswa dan lulusan. Akreditasi pada perguruan tinggi sering dilihat dari keberhasilan mahasiswa menyelesaikan masa studinya. Meningkatnya jumlah mahasiswa *dropout* dapat mempengaruhi akreditasi perguruan tinggi tersebut.

Kualitas pendidikan diukur dari persentase kelulusan mahasiswa dan strategi perguruan tinggi untuk mempertahankan mahasiswa. Jika sebuah institusi kehilangan siswa karena alasan apapun, tingkat retensinya menurun. Identifikasi awal mahasiswa yang berisiko gagal sangat penting untuk keberhasilan strategi retensi apapun. Penting untuk mengidentifikasi mahasiswa ini sejak dini, dengan memberikan perhatian untuk mencapai retensi siswa. Kunci untuk mengurangi angka *dropout* adalah identifikasi dini mahasiswa berisiko, selain mempertahankan intervensi yang intensif dan berkelanjutan (Perez et al., 2018).

Penelitian ini membahas tipe pertama, yaitu memprediksi keberhasilan studi mahasiswa berdasarkan Kategori kelulusan. Sistem dibangun menggunakan teknik klasifikasi *machine learning* regresi logistik multinomial untuk memprediksi potensi lama kategori kelulusan mahasiswa pada tahap awal. Metode ini memiliki dua kegunaan utama, pertama untuk melakukan prediksi keanggotan kelompok di mana variabel dependen (respon) memiliki sifat politomi (lebih dari dua kategori). Kedua, metode regresi logistik dapat mengidentifikasi faktor kesuksesan yang paling berpengaruh (Hoffait & Schyns, 2017).

Al-Balushi & Islam (2020) mengungkapkan bahwa usia mahasiswa, jenis kelamin, perguruan tinggi, dan wilayah tempat tinggal merupakan prediktor penting dari nilai mahasiswa. Pengaruh variasi perguruan tinggi untuk usia mahasiswa, jenis kelamin mahasiswa, dan wilayah tempat tinggal lebih lanjut menyiratkan bahwa ada perbedaan yang cukup besar dalam nilai mahasiswa di antara perguruan tinggi (Urrutia-Aguilar et al., 2016). Mahasiswa yang proses penerimaannya melalui ujian memiliki motivasi diri lebih tinggi dibandingkan dengan mahasiswa yang diterima melalui metode tanpa ujian atau transfer otomatis.

Hashim et al. (2020) melakukan penelitian dengan menggunakan data mahasiswa selama enam semester menggunakan beberapa algoritma *Supervised Machine Learning*, seperti *decision tree*, Naïve Bayes, dan regresi logistik, *Support Vector Machine* (SVM), K-Nearest Neighbor, *Sequential Minimal Optimization* dan *neural network*. Hasil penelitian menunjukkan bahwa pengklasifikasi regresi logistik adalah yang paling akurat dalam memprediksi nilai akhir aktual mahasiswa dengan akurasi 68,7% untuk status lulus dan 88,8% untuk status gagal.



Umer et al. (2017) melakukan prediksi prestasi akademik mahasiswa dalam *Massive Open Online Courses* (MOOCs). Studi tersebut mengevaluasi empat teknik klasifikasi *machine learning*, yaitu regresi logistik, Naïve Bayes, *random forest*, dan K-Nearest Neighbor (KNN) di mana hasil studi menunjukkan bahwa metode regresi logistik dan Naïve Bayes memiliki tingkat akurasi yang tinggi serta mengungguli *random forest* dan KNN. Ashraf et al. (2018) melakukan penelitian yang menyimpulkan bahwa setiap pendekatan *data mining* menggunakan suatu algoritma memberikan hasil yang bervariasi sesuai dengan *dataset* dan variabel yang digunakan saat memprediksi. Berdasarkan hasil pengamatan, *decision tree*, regresi logistik, dan pendekatan *neural network* mampu memberikan hasil akurasi yang sangat baik mengenai prediksi masa mendatang serta membantu peningkatan sistem pendidikan.

Meskipun banyak penelitian tentang penggunaan metode regresi logistik untuk memprediksi keberhasilan akademik mahasiswa. Namun, masih terdapat kekurangan yang diperlukan untuk meningkatkan *performance model* prediksi meliputi identifikasi dan analisis variabel *input-output* yang mempengaruhi keberhasilan akademik mahasiswa dalam konteks yang berbeda yang telah dikembangkan oleh peneliti. Oleh karena itu, peneliti lebih lanjut diperlukan dalam membangun model menggunakan metode regresi logistik untuk memprediksi kategori kelulusan mahasiswa berdasarkan data akademik mahasiswa Program Studi Teknik Informatika UIN Malang Program S1 dari tahun 2012-2018 yang telah dilakukan pra-pemrosesan yang spesifik sehingga menghasilkan klasifikasi yang baik dengan nilai akurasi 85,5%.

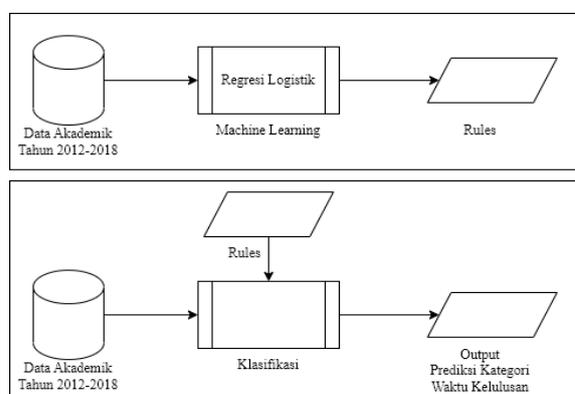
2. METODE PENELITIAN

2.1 Pengumpulan Data

Sumber data yang dibutuhkan dalam penelitian adalah data akademik mahasiswa Program Studi Teknik Informatika UIN Malang program S1 dari tahun 2012 - 2018. Data tersebut diperoleh melalui Bagian Administrasi Akademik (BAK) UIN Malang. Data induk mahasiswa merupakan data mahasiswa yang didata ketika mahasiswa pertama kali masuk perguruan tinggi setelah melakukan registrasi ulang. Sedangkan, data kelulusan merupakan data mahasiswa yang sudah menyelesaikan studinya di jurusan Teknik Informatika. *Dataset* berisi kumpulan data dari tujuh tahun akademik pertama yang diperlakukan untuk membangun model (2012-2018).

2.2 Desain Sistem

Desain sistem menjelaskan alur sistem dalam mengimplementasikan metode regresi logistik multinomial untuk melakukan prediksi Kategori kelulusan mahasiswa. Seperti diilustrasikan dalam Gambar 1, alur desain sistem terdiri dari masukan, *preprocessing*, proses *learning* dengan menggunakan metode regresi logistik multinomial, proses prediksi, dan evaluasi hasil prediksi, Dalam proses *learning*, pembagian data *training* dan data *testing* menggunakan perbandingan 70:30 di mana 70% untuk data *training* sedangkan 30% untuk data *testing*. Sistem akan dibangun menggunakan RStudio.



Gambar 1 Desain Sistem



2.3 Persiapan Data

Persiapan data diawali dengan tahapan *preprocessing* data, yaitu langkah yang dilakukan untuk membuat data mentah menjadi data berkualitas dengan input yang dibutuhkan untuk *data mining*. Tidak semua variabel data induk mahasiswa dimasukkan ke dalam *dataset*. Berdasarkan kebutuhan penelitian, variabel yang digunakan dalam *dataset* diterangkan dalam Tabel 1.

Tabel 1 Karakteristik *Dataset*

Variabel	Definisi	Indikator
Dependen		
IP	Penilaian yang diterapkan dikampus dengan hasil penjumlahan dari semua nilai mata kuliah dalam satu semester	Data IP selama semester 1 sampai semester 6
Jenis Kelamin	Pembeda antara laki-laki dan perempuan dilihat dari sudut pandang	L dan P
Jalur Masuk	Bagian penerimaan yang dilalui mahasiswa saat masuk ke perguruan tinggi (S1)	Prestasi dan Non Prestasi
Independen	Definisi	Indikator
Kategori waktu kelulusan	Lama mahasiswa menyelesaikan masa studinya	Cepat Tepat Lambat

Kategori cepat, tepat, dan lambat dilakukan secara manual dengan melihat data mahasiswa yang didapat dari Bagian Administrasi Akademik (BAK) UIN Maulana Malik Ibrahim Malang. Kategori cepat apabila pada data mahasiswa mahasiswa tersebut lulus pada semester 7 atau sebelum semester 8 dan telah memenuhi jumlah minimal SKS yang telah ditentukan yaitu sejumlah 151 SKS. Kategori tepat apabila pada data mahasiswa mahasiswa tersebut lulus tepat pada semester 8 dan telah memenuhi jumlah minimal SKS yang telah ditentukan yaitu sejumlah 151 SKS. Kategori lambat apabila pada data mahasiswa mahasiswa tersebut lulus lebih dari semester 8 dan telah memenuhi jumlah minimal SKS yang telah ditentukan yaitu sejumlah 151 SKS (Satsangkaryon, 2018).

2.4 Modeling

Sub bab ini menjelaskan implementasi metode klasifikasi regresi logistik multinomial untuk memprediksi kategori kelulusan mahasiswa. Untuk mencapai ini, model dikembangkan dari data *training* di mana nilai *input* maupun label *output* telah diketahui. Model menggeneralisasi hubungan antara *input* dan *output* kemudian menggunakannya untuk melakukan prediksi kategori kelulusan mahasiswa pada kumpulan data lain di mana hanya nilai *input* saja diketahui. Langkah-langkah dari metode yang diterapkan sebagai berikut:

- 1) Melakukan *input dataset* penelitian ke dalam sistem. Kemudian, pisahkan *dataset* menjadi dua *subset*: 70% untuk pelatihan dan 30% untuk validasi.
- 2) Komputasi regresi logistik pada data latih di mana, setiap paramater (x) menghasilkan nilai koefisien (β) dan *intercept* ditambahkan ke dalam sistem. Nilai koefisien dari masing-masing parameter dikalikan dengan nilai paramater bersangkutan. Hasil setiap perkalian lalu dijumlahkan dengan menambahkan nilai *intercept*, seperti Pers. (1).

$$g(x) = \beta_0 + \beta_1x_1 + \dots + \beta_px_p \quad (1)$$

- 3) Nilai penjumlahan diterapkan ke dalam fungsi aktivasi—fungsi sigmoid atau fungsi logistik (Scott et al., 1991), yang mana mengubah data yang diberikan dalam kisaran (0, 1) ditunjukkan pada Pers. (2). Nilai *output* yang diukur dengan regresi logistik menyesuaikan dengan probabilitas, artinya nilai 1 menunjukkan peluang lebih baik untuk berada dalam kelas kategori kelulusan tertentu daripada nilai, contoh 0,7.



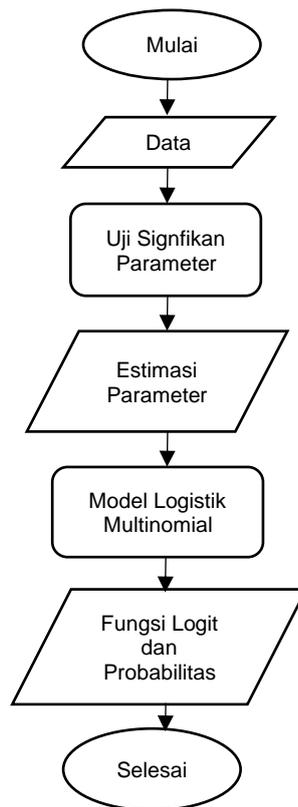
$$f(z) = \frac{1}{1+e^{-z}} \quad -\infty < z < +\infty \quad (2)$$

Di mana $z = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p$ sehingga probabilitas bersyarat yang dimodelkan pada penelitian ini dapat dilambangkan pula dengan Pers. (3).

$$P(Y_i | x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4)}} \quad (3)$$

4) Evaluasi model menggunakan *multi-class confusion matrix*.

Diagram alir dari regresi logistik multinomial ditunjukkan pada Gambar 2.



Gambar 2 Flowchart Regresi Logistik Multinomial

2.5 Uji Signifikansi Parameter

Variabel-variabel yang diamati dalam penelitian ini dapat mempengaruhi keberhasilan akademik mahasiswa, menurut statistik deskriptif pada bab sebelumnya. Dengan nilai alpha 0,05 (Sari et al., 2017), peneliti selanjutnya meneliti statistik Wald dan perubahan statistik *Likelihood Ratio* untuk mengetahui signifikansi variabel regresi logistik. Hasil pengujian signifikansi parameter secara serentak ditunjukkan pada Tabel 2.

Tabel 2 Estimasi Parameter Secara Serentak

Pearson's chi-squared test
Data: Training\$Lama_Studi and Predict(mutinomModel)
X-squared = 136,83, df= 4, p-value <2.2e-16

Berdasarkan keterangan dari Tabel 2, diketahui bahwa hasil pengujian *Likelihood Ratio* memiliki nilai *p-value* sebesar 2.2e-16, di mana nilai ini lebih kecil dari 0,05. Hal ini mengartikan bahwa



terdapat paling sedikit satu variabel independen yang berpengaruh signifikan terhadap kategori kelulusan mahasiswa.

Tabel 3 Estimasi Parameter Secara Parsial

Coefficients:									
(Intercept)	JK	JM	IP1	IP2	IP3	IP4	IP5	IP6	
2.	-17.07503	0.2906574	0.1301306	-6.649710	2.659724	4.350708	3.3276037	-2.133633	3.407288
3.	19.61444	0.2878672	0.7440973	-4.489338	1.600311	0.223948	0.1354936	-5.101165	2.082269

Tabel 3 di atas menampilkan estimasi parameter secara parsial untuk kategori kelulusan mahasiswa di mana tepat diberi keterangan 2, lambat diberi keterangan 3. Berdasarkan pengujian tersebut, kategori kelulusan yang dijadikan pembanding adalah cepat dengan keterangan 1.

2.6 Model Logistik

Proses *training* menghasilkan dua buah fungsi logit, yaitu fungsi logit untuk kategori kelulusan tepat, dan lambat sementara kategori yang digunakan sebagai pembanding adalah kategori kelulusan cepat. Merujuk pada Pers. (1) maka fungsi regresi logistik multinomial untuk prediksi kategori kelulusan seorang mahasiswa apabila berjenis kelamin laki-laki, jalur masuk prestasi, IP semester 1 sebesar 3.55, IP semester 2 sebesar 3.15, IP semester 3 sebesar 3.38, IP semester 4 sebesar 3, IP semester 5 sebesar 3.46, IP semester 6 sebesar 3.7 adalah sebagai berikut:

- 1) $g_2(x)$

$$= -17.075 + 0.290x_1 + 0.130x_2 - 6.649x_3 + 2.659x_4 + 4.350x_5 + 3.327x_6 - 2.133x_7 + 3.407x_8$$

$$= -17.075 + 0.290(1) + 0.130(1) - 6.649(3.55) + 2.659(3.15) + 4.350(3.38) + 3.327(3) - 2.133(3.46) + 3.407(3.7)$$

$$= -1.974$$
- 2) $g_3(x)$

$$= 19.614 + 0.287x_1 + 0.744x_2 - 4.489x_3 + 1.600x_4 + 0.223x_5 + 0.135x_6 - 5.101x_7 + 2.082x_8$$

$$= 19.614 + 0.287(1) + 0.744(1) - 4.489(3.55) + 1.600(3.15) + 0.223(3.38) + 0.135(3) - 5.101(3.46) + 2.082(3.7)$$

$$= 0.962$$

Setelah mendapatkan nilai dari kedua model logit maka langkah berikutnya adalah mengukur nilai probabilitas masing-masing kategori dalam kategori kelulusan mahasiswa menggunakan Pers. (3) sehingga didapatkan nilai probabilitas sebagaimana berikut:

- 1) $\pi_1(x) = \frac{1}{1 + \exp(-1.974) + \exp(0.962)} = 0.266$
- 2) $\pi_2(x) = \frac{\exp(-1.974)}{1 + \exp(-1.974) + \exp(0.962)} = 0.036$
- 3) $\pi_3(x) = \frac{\exp(0.962)}{1 + \exp(-1.974) + \exp(0.962)} = 0.696$

Dari fungsi probabilitas di atas, menunjukkan bahwa peluang mahasiswa dengan deskripsi berjenis kelamin laki-laki, jalur masuk prestasi, IP semester 1 sebesar 3,71, IP semester 2 sebesar 3,6, IP semester 3 sebesar 3,43, IP semester 4 sebesar 3,55, IP semester 5 sebesar 3,1, IP semester 6 sebesar 3,38, IP semester 7 sebesar 3,3, IP semester 8 sebesar 2,71 berpeluang paling besar mendapatkan kategori waktu kelulusan lambat di masa studinya dengan nilai peluang paling sebesar 0,696 (EI-Habil, 2012).



2.7 Evaluasi Model

Langkah selanjutnya adalah evaluasi terhadap model yang telah didapatkan dalam proses *learning* metode regresi logistik multinomial untuk prediksi kategori kelulusan mahasiswa. Evaluasi model diukur menggunakan *multiclass confusion matrix* dikarenakan penelitian ini termasuk ke dalam *multi-class classification*. Evaluasi model meliputi pengukuran akurasi, presisi, *recall*, dan *micro f1-score*. Akurasi klasifikasi adalah persentase data *testing* yang diklasifikasi secara benar oleh model. Apabila akurasi klasifikasi dianggap dapat diterima maka model dapat diterapkan untuk mengklasifikasi set data masa mendatang yang label kelas tidak diketahui (Agarwal, 2013). Presisi atau *precision* menyatakan proporsi unit yang diprediksi positif oleh model yang juga positif pada data sebenarnya. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut (Mayadewi & Rosely, 2015). *Recall* merupakan hasil perhitungan terhadap semua data uji yang positif yang telah diprediksi benar sebagai positif pada klasifikasi. *Recall* juga dikenal sebagai *True Positive Rate* (TPR), sensitivitas, dan probabilitas deteksi (Grandini et al., 2020). Seluruh kelas observasi digabung untuk mengukur *micro-averaged* presisi dan *micro-averaged recall* sehingga diperoleh rata-rata harmonik *micro-F1* (Zhang et al., 2015). Pers. (4) sampai (7) merupakan rumus dari akurasi, presisi, *recall*, dan *micro-F1*.

$$accuracy = \frac{TP}{Total\ data\ testing} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$MicroF1 = 2 \times \left(\frac{Recall \times Precision}{Recall + Precision} \right) \quad (7)$$

3. HASIL DAN PEMBAHASAN

Metode regresi logistik multinomial diterapkan pada data *training* untuk proses *learning* sehingga menghasilkan dua buah model logit kategori kelulusan mahasiswa dengan satu kategori lainnya dijadikan pembanding. Peneliti membagi data menjadi dua bagian yang digunakan untuk proses *training* dan *testing* model dengan menggunakan perbandingan 70:30 di mana artinya 70% atau sebanyak 210 data digunakan sebagai data *training*, sedangkan 30% atau sebanyak 90 data digunakan sebagai data *testing*. Tabel 4 menunjukkan lima sampel dari data *training* sedangkan Tabel 5 menunjukkan lima sampel dari data *testing*.

Setelah mendapatkan model, maka pengujian dilanjutkan dengan membandingkan hasil prediksi model terhadap kategori kelulusan dalam 90 data *testing*. Hal ini bertujuan untuk mengukur kemampuan sistem prediksi Kategori kelulusan mahasiswa yang telah dibangun dalam hal presisi, *recall*, akurasi, dan *f1-score*. Hasil prediksi Kategori kelulusan mahasiswa tahun 2012-2018 menggunakan metode regresi logistik multinomial ditunjukkan pada Tabel 6.

Tabel 4 Sampel Data *Training*

No.	X_train							Y_train	
	Jenis kelamin	Jalur masuk	IP 1	IP 2	IP3	IP 4	IP5	IP 6	Kategori
1	2	2	3,4	3,26	3,59	3,57	3,61	3,61	2
2	2	2	3,38	3,31	3,26	3,17	3,23	3,5	3
3	1	2	2,88	2,29	2,81	2,73	3,16	2,23	3
4	2	2	3,57	3,59	3,72	3,69	3,46	3,73	2
5	2	2	3,45	3,07	3,83	3,85	3,76	3,79	2



Tabel 5 Sampel Data *Testing*

No.	X_train								Y_train
	Jenis kelamin	Jalur masuk	IP 1	IP 2	IP3	IP 4	IP5	IP 6	Kategori
1	2	1	3,55	3,31	3,42	3,59	3,05	3,34	3
2	1	2	3,45	3,4	3,23	3,02	3,46	2,95	1
3	1	2	3,62	3,56	3,48	3,56	3,15	3,31	3
4	2	1	3,76	3,61	3,74	3,79	3,52	3,86	2
5	2	2	3,38	2,98	3,1	3,36	3	3,52	3

Tabel 6 Hasil Prediksi Mahasiswa Tahun 2012-2018

No.	X_pred	Y_pred	Keterangan
1.	3	3	Sesuai
2.	1	3	Tidak Sesuai
3.	3	3	Sesuai
4.	2	2	Sesuai
5.	3	3	Sesuai

Visualisasi dari *multiclass confusion matrix* berdasarkan hasil prediksi model pada Tabel 6 ditunjukkan oleh Tabel 7.

Tabel 7 *Multiclass Confusion Matrix* Kategori Kelulusan

		Prediksi		
		1	2	3
Aktual	1	1	2	3
	2	1	16	2
	3	4	6	60

Analisis perhitungan nilai *True Postive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) pada *multiclass confusion matrix* dilakukan dengan menghitung per kelas kategori kelulusan. Kalkulasi perolehan nilai TP, TN, FP, dan FN dalam semua kelas kategori kelulusan tertera dalam Tabel 8.

Tabel 8 Kalkulasi *Multiclass Confusion Matrix*

Kategori Kelulusan	Confusion Matrix			
	TP	TN	FP	FN
Cepat	1	84	5	0
Tepat	16	65	6	3
Lambat	60	18	10	2
Total	77	167	21	5

Berdasarkan Tabel 8 maka nilai presisi, *recall*, akurasi, dan *micro f1-score* dapat diukur menggunakan Pers. (4) sampai Pers. (7). Hasil pengukuran sebagai berikut.

$$Accuracy = \frac{77}{90} = 0.855 = 85.5\%$$

$$Precision = \frac{77}{77 + 21} = 0.785 = 78.5\%$$

$$Recall = \frac{77}{77 + 5} = 0.939 = 93.9\%$$



$$Micro F1 = 2 \frac{(0.785 \times 0.939)}{(0.785 + 0.855)} = 0.898 = 89.8\%$$

Berdasarkan hasil evaluasi sistem dalam melakukan prediksi kategori kelulusan mahasiswa menggunakan metode regresi logistik multinomial menunjukkan nilai *accuracy* sebesar 85,5%, *precision* sebesar 78,5%, nilai *recall* sebesar 93,9%, dan nilai *micro F1* sebesar 89,8%. Nilai *accuracy* 85,5% menunjukkan sistem dapat melakukan klasifikasi dengan baik.

4. KESIMPULAN

Peneliti telah membangun model prediksi dengan menggunakan regresi logistik multinomial berdasarkan *dataset* studi prediksi kategori kelulusan mahasiswa terkumpul sebanyak 300 data dengan variabel berupa jenis kelamin, jenis jalur masuk, indeks prestasi semester satu sampai enam sedangkan tiga tingkat klasifikasi kategori kelulusan program Sarjana (S1) dijadikan hasil. Model prediksi telah dilakukan uji dan evaluasi sehingga didapatkan nilai *accuracy* sebesar 85,5%, *precision* sebesar 78,5%, *recall* sebesar 93,9%, dan *micro f1-score* sebesar 89,8%. Kesimpulan berdasarkan nilai *accuracy* sebesar 85,5% menyatakan model prediksi dapat melakukan klasifikasi dengan baik sehingga penggunaan metode regresi logistik multinomial dapat diterima dan dapat digunakan untuk memprediksi kategori kelulusan mahasiswa. Saran untuk penelitian selanjutnya dapat menggunakan tambahan metode optimasi yang lain untuk menghasilkan nilai akurasi yang lebih optimal dan membuat model prediksi tersedia bagi administrator untuk membiarkan mereka membuat keputusan tentang penerapan dan kegunaannya secara dinamik sesuai dengan dunia nyata.

DAFTAR PUSTAKA

- Agarwal, S. (2013). Data Mining: Data Mining Concepts and Techniques. *2013 International Conference on Machine Intelligence and Research Advancement*, 203–207. <https://doi.org/10.1109/ICMIRA.2013.45>
- Agwil, W., Fransiska, H., & Hidayati, N. (2020). Analisis Ketepatan Waktu Lulus Mahasiswa dengan Menggunakan Bagging CART. *FIBONACCI: Jurnal Pendidikan Matematika Dan Matematika*, 6(2), 155. <https://doi.org/10.24853/fbc.6.2.155-166>
- Al-Balushi, M. S., & Islam, M. M. (2020). Predicting Academic Performance of Students of Sultan Qaboos University, Oman, Using Multilevel Modeling Approach. *Far East Journal of Theoretical Statistics*, 58(1), 59–76. <https://doi.org/10.17654/TS058010059>
- Alturki, S., Hulpuş, I., & Stuckenschmidt, H. (2022). Predicting Academic Outcomes: A Survey from 2007 Till 2018. *Technology, Knowledge and Learning*, 27(1), 275–307. <https://doi.org/10.1007/s10758-020-09476-0>
- Asha, P., Vandana, E., Bhavana, E., & Shankar, K. R. (2020). Predicting University Dropout through Data Analysis. *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, 852–856. <https://doi.org/10.1109/ICOEI48184.2020.9142882>
- Ashraf, A., Anwer, S., & Khan, M. G. (2018). A Comparative Study of Predicting Student's Performance by use of Data Mining Techniques. *American Scientific Research Journal for Engineering, Technology, and Sciences*, 44(1), 122–136. https://www.asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/4170
- Dalipi, F., Imran, A. S., & Kastrati, Z. (2018). MOOC dropout prediction using machine learning techniques: Review and research challenges. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 1007–1014. <https://doi.org/10.1109/EDUCON.2018.8363340>
- El-Habil, A. M. (2012). An Application on Multinomial Logistic Regression Model. *Pakistan Journal of Statistics and Operation Research*, 8(2), 271. <https://doi.org/10.18187/pjsor.v8i2.234>
- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: an Overview*. <http://arxiv.org/abs/2008.05756>
- Handini, D., Hidayat, F., Putri, D. A. V., Rouf, M. R., Anjani, N. R., & Attamimi, A. N. R. (2020). *Statistik pendidikan tinggi tahun 2020 (higher education statistics 2020)*. Sekretariat Direktorat Jenderal Pendidikan Tinggi. <https://repositori.kemdikbud.go.id/22653/>
- Hashim, A. S., Awadh, W. A., & Hamoud, A. K. (2020). Student Performance Prediction Model based on Supervised Machine Learning Algorithms. *IOP Conference Series: Materials*



- Science and Engineering*, 928(3), 032019. <https://doi.org/10.1088/1757-899X/928/3/032019>
- Hoffait, A.-S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems*, 101, 1–11. <https://doi.org/10.1016/j.dss.2017.05.003>
- Lu, O. H. T., Huang, J. C. H., Huang, A. Y. Q., & Yang, S. J. H. (2018). Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. In O. H. T. Lu, J. C. H. Huang, A. Y. Q. Huang, & S. J. H. Yang (Eds.), *Learning Analytics* (pp. 78–92). Routledge. <https://doi.org/10.4324/9780429428500-7>
- Mayadewi, P., & Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. *Seminar Nasional Sistem Informasi Indonesia (SESINDO) 2015*, 2015. <https://is.its.ac.id/pubs/oajis/index.php/home/detail/1582/PREDIKSI-NILAI-PROYEK-AKHIR-MAHASISWA-MENGGUNAKAN-ALGORITMA-KLASIFIKASI-DATA-MINING>
- Perez, B., Castellanos, C., & Correal, D. (2018). Applying Data Mining Techniques to Predict Student Dropout: A Case Study. *2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (CoCACI)*, 1–6. <https://doi.org/10.1109/CoCACI.2018.8484847>
- Sari, A. Q., Sukestiyarno, Y., & Agoestanto, A. (2017). Batasan Prasyarat Uji Normalitas Dan Uji Homogenitas pada Model Regresi Linear. *Unnes Journal of Mathematics*, 6(2), 168–177. <https://doi.org/10.15294/UJM.V6I2.11887>
- Satsangkaryon, S. (2018). Analisis Korelasi Pemanfaatan Hasil Perkembangan Teknologi Informatika Terhadap Tingkat Kelulusan Mahasiswa pada Fakultas Ekonomi Universitas Pakuan Bogor. *JIMFE (Jurnal Ilmiah Manajemen Fakultas Ekonomi)*, 2(1), 73–87. <https://doi.org/10.34203/jimfe.v2i1.722>
- Scott, A. J., Hosmer, D. W., & Lemeshow, S. (1991). Applied Logistic Regression. *Biometrics*, 47(4), 1632. <https://doi.org/10.2307/2532419>
- Umer, R., Susnjak, T., Mathrani, A., & Suriadi, S. (2017). On predicting academic performance with process mining in learning analytics. *Journal of Research in Innovative Teaching & Learning*, 10(2), 160–176. <https://doi.org/10.1108/JRIT-09-2017-0022>
- Urrutia-Aguilar, M. E., Fuentes-García, R., Martínez, V. D. M., Beck, E., León, S. O., & Guevara-Guzmán, R. (2016). Logistic Regression Model for the Academic Performance of First-Year Medical Students in the Biomedical Area. *Creative Education*, 07(15), 2202–2211. <https://doi.org/10.4236/ce.2016.715217>
- Yaacob, W. F. W., Sobri, N. M., Nasir, S. A. M., Yaacob, W. F. W., Norshahidi, N. D., & Husin, W. Z. W. (2020). Predicting Student Drop-Out in Higher Institution Using Data Mining Techniques. *Journal of Physics: Conference Series*, 1496, 012005. <https://doi.org/10.1088/1742-6596/1496/1/012005>
- Zhang, D., Wang, J., & Zhao, X. (2015). Estimating the Uncertainty of Average F1 Scores. *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, 317–320. <https://doi.org/10.1145/2808194.2809488>



Klasifikasi Ulasan Fasilitas Publik Menggunakan Metode Naïve Bayes dengan Seleksi Fitur *Chi-Square*

Adhitya Prayoga Permana ⁽¹⁾, Totok Chamidy ^{(2)*}, Cahyo Crysdian ⁽³⁾
Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim, Malang
e-mail : {adhityaprayoga48,totokchamidy}@gmail.com, cahyo@ti.uin-malang.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 26 Desember 2022, direvisi 4 April 2023, diterima 5 April 2023, dan dipublikasikan 26 Mei 2023.

Abstract

Government builds public facilities to support the needs of the community. The use of these public facilities needs to be re-evaluated, and one way to do it is through community response. Google Maps is one platform that receives the most responses from the community about location. Google Maps Reviews allow us to see how the public reacts to a location. Naïve Bayes method is used for classification in this study because it is one of the simple methods in machine learning that can be easily applied to several experiments conducted by the author. In the classification process, reviews produce many features that will be calculated based on their class. More features generated, more features processed too in the system. Chi-Square feature selection will be used to reduce features that have low dependence on the system. In this study, performance values will be calculated based on the experimental use of feature ratios of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%. The results show that the use of 10% Chi-Square features produces the best performance, with an accuracy rate of 86.94%, precision of 80.42%, recall of 80.42%, and f-measure of 80.42%.

Keywords: Sentiment Analysis, Public Facility, Google Maps Reviews, Naïve Bayes, Chi-Square Feature Selection

Abstrak

Pemerintah membangun fasilitas publik untuk mendukung kebutuhan masyarakat. Penggunaan fasilitas publik ini perlu ditinjau kembali di mana salah satu caranya adalah melalui respon masyarakat. Google Maps merupakan salah satu platform yang paling banyak menerima respon dari masyarakat tentang suatu tempat atau lokasi. Ulasan Google Maps memungkinkan untuk melihat bagaimana reaksi masyarakat terhadap suatu lokasi. Metode Naïve Bayes digunakan untuk melakukan klasifikasi dalam penelitian karena termasuk salah satu metode dalam *machine learning* yang sederhana sehingga mudah diterapkan pada beberapa percobaan yang dilakukan penulis. Dalam proses klasifikasi, ulasan-ulasan tersebut menghasilkan banyak fitur dan akan dihitung berdasarkan kelasnya. Semakin banyak fitur yang dihasilkan maka semakin banyak pula fitur yang diproses dalam sistem. Seleksi fitur *Chi-Square* akan digunakan untuk mereduksi fitur-fitur yang memiliki ketergantungan rendah dalam sistem. Dalam penelitian ini akan menghitung nilai performa berdasarkan uji coba penggunaan rasio fitur sebanyak 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, dan 100%. Hasilnya penggunaan 10% fitur *chi-square* menghasilkan performa terbaik yaitu dengan nilai akurasi 86,94%, presisi 80,42%, *recall* 80,42% dan *f-measure* 80,42%.

Kata Kunci: Analisis Sentimen, Fasilitas Publik, Ulasan Google Maps, Naïve Bayes, Seleksi Fitur *Chi-Square*

1. PENDAHULUAN

Fasilitas publik atau fasilitas umum adalah sarana dan prasarana yang disediakan oleh pemerintah yang digunakan untuk kepentingan bersama dalam melaksanakan kegiatan sehari-hari. Dengan adanya fasilitas-fasilitas publik yang telah dibangun oleh pemerintah daerah, diharapkan masyarakat pada suatu daerah tersebut dapat memanfaatkan fasilitas tersebut dengan semestinya.



Fasilitas publik ini perlu ditinjau kembali penggunaannya karena fasilitas publik yang pembangunannya tidak sesuai dengan kebutuhan masyarakat tidak akan membantu produktivitas. Untuk menghindari adanya fasilitas-fasilitas yang tidak sesuai dengan tujuan pembuatan ini, maka perlu adanya evaluasi dari pemerintah. Evaluasi penggunaan fasilitas publik dapat dilakukan melalui respon masyarakat. Masyarakat biasa memberikan responnya melalui media sosial, di mana salah satu platform yang banyak mendapat respon masyarakat apabila berkaitan dengan suatu tempat atau lokasi adalah Google Maps. Google Maps sebenarnya merupakan aplikasi peta dunia milik perusahaan Google yang memiliki fitur ulasan sehingga tempat-tempat yang tertera pada peta yang ditampilkan dapat diberikan ulasan berdasarkan pengalaman orang-orang yang pernah mengunjungi tempat tersebut. Dari fitur ulasan Google Maps tersebut maka akan diketahui bagaimana respon masyarakat terhadap suatu tempat tersebut.

Ulasan-ulasan yang diberikan masyarakat melalui platform Google Maps tersebut terdiri dari ulasan yang bersifat mendukung adanya fasilitas dan menentang adanya fasilitas. Jumlah ulasan tersebut semakin lama akan meningkat secara signifikan sehingga untuk mengetahui sebuah ulasan tergolong ulasan positif atau negatif akan membutuhkan banyak waktu apabila dilakukan secara manual. Untuk mengatasi hal tersebut akan dikembangkan sistem otomatis untuk mengklasifikasikan ulasan tersebut menggunakan metode dalam *machine learning*.

Analisis sentimen, juga disebut klasifikasi opini, adalah bidang studi yang menganalisis pendapat orang, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, topik, dan atributnya (B. Liu, 2012).

Linawati et al. (2020) melakukan penelitian tentang perbandingan penggunaan algoritma Naïve Bayes dan Support Vector Machine untuk penentuan pemberian beasiswa PPA. Data yang digunakan berasal dari data pemberian beasiswa PPA berjumlah 122 data dengan 5 variabel untuk menentukan kelas diterima atau ditolak. Variabelnya terdiri dari semester, pekerjaan, orang tua, penghasilan, dan IPK. Pengujian dilakukan menggunakan aplikasi WEKA 3.8 dan hasilnya algoritma Naïve Bayes memiliki tingkat akurasi lebih tinggi dengan nilai 90,90% sedangkan metode SVM memiliki nilai akurasi 89,25%.

Dulhare (2018) dalam penelitiannya membuat sistem prediksi penyakit jantung menggunakan Naïve Bayes dan Particle Swarm Optimization (PSO). Data yang digunakan merupakan data *stat log (heart)* dari UCI *dataset* yang berjumlah 270 data dengan 14 fitur dan dibagi menjadi dua kelas yaitu *presence* dan *absence*. Pengujian 7 dilakukan dengan sepuluh kali iterasi dengan 100 persen sampai dengan 90 persen seleksi fitur PSO dengan akurasi tertinggi 89,71 persen pada iterasi dengan fitur 70 persen dan 100 persen. Harahap et al. (2018) juga melakukan penelitian mengenai implementasi metode Naïve Bayes untuk memprediksi pembelian. Hasilnya 15 dari 20 data berhasil memprediksi dengan benar yang diujikan berdasar data latih dan nilai akurasinya sebesar 75%.

Ruz et al. (2020) dalam penelitiannya menganalisis data *tweet* dari Twitter selama terjadi peristiwa penting seperti bencana alam maupun pergerakan sosial. Data yang digunakan merupakan *dataset* dalam bahasa Spanyol pada saat peristiwa gempa di Chile tahun 2010 yang berjumlah 2187 *tweet* serta peristiwa *Catalan Independence Referendum* tahun 2017 yang berjumlah 60000 *tweet*. Data tersebut dibagi menjadi dua kelas yaitu positif dan negatif. Metode *bayesian network* digunakan untuk menyelesaikan permasalahan ini. Mereka juga menggunakan *Synthetic Minority Over-sampling Technique* pada data *training* untuk menangani masalah ketidakseimbangan kelas. Hasilnya pada *dataset* gempa di Chile didapatkan nilai akurasi sebesar $0,742 \pm 0,027$, presisi $0,895 \pm 0,004$, *recall* $0,790 \pm 0,009$, *F1-score* $0,841 \pm 0,020$, *Nº Edges* 0. Sedangkan pada *dataset* kedua didapatkan nilai akurasi sebesar $0,781 \pm 0,013$, presisi $0,885 \pm 0,005$, *recall* $0,852 \pm 0,004$, *F1-score* $0,868 \pm 0,000$, *Nº Edges* 0.

Banyak metode dalam *machine learning* yang dapat digunakan untuk melakukan analisis sentimen. Berdasar beberapa penelitian yang telah dilakukan sebelumnya, penulis memutuskan



untuk menggunakan metode Naïve Bayes. Metode Naïve Bayes merupakan metode statistik klasifikasi sederhana berdasarkan probabilitas. Klasifikasi Naïve Bayes mengasumsikan independensi kondisi kelas yaitu nilai atribut pada kelas tertentu tidak tergantung pada nilai atribut lainnya di mana hal ini digunakan untuk menyederhanakan perhitungan yang terlibat (Han & Kamber, 2001). Dalam Naïve Bayes, hipotesis langsung dibentuk tanpa proses pencarian, hanya dengan menghitung frekuensi kemunculan suatu kata pada data training, sedangkan pada metode pembelajaran lainnya pencarian hipotesis biasanya dilakukan dari ruang hipotesis (Aziz et al., 2021).

Analisis sentimen dalam proses klasifikasinya akan membagi dokumen ke dalam kelas dengan memisahkan fitur-fitur dari setiap dokumen untuk dihitung nilai fitur berdasarkan kelasnya. Banyaknya fitur yang dihasilkan tentu akan membebani jalannya sistem komputasi. Penggunaan seleksi fitur *chi-square* digunakan untuk mengurangi dimensi data inputan sehingga tidak membebani sistem. Seleksi fitur *chi-square* akan mengurangi fitur yang tidak memiliki ketergantungan terhadap kelas label sehingga diharapkan dapat meningkatkan performa dari analisis sentimen menggunakan metode Naïve Bayes.

Pratama et al. (2018) melakukan analisis sentimen dengan seleksi fitur *chi-square* terhadap ulasan konsumen untuk rekomendasi lokasi makanan tradisional. Metode yang digunakan dalam penelitian ini merupakan metode Naïve Bayes. Data yang diperoleh merupakan data ulasan konsumen makanan tradisional di Kota Malang dari website tripadvisor.co.id dan diklasifikasikan menjadi dua kelas yaitu positif dan negatif. Hasilnya dilakukan seleksi fitur sebesar 25%, 50%, 75%, 100%, di mana seleksi fitur 25% mendapatkan nilai akurasi tertinggi yaitu 81% serta seleksi fitur 75% mendapatkan nilai terendah dengan akurasi 77%.

Irvantoro (2019) melakukan seleksi fitur menggunakan *chi-square* dan *n-gram* untuk menganalisis ulasan produk elektronik. Metode yang digunakan menggunakan metode Naïve Bayes. Data yang digunakan didapat dari *library RStudio* sejumlah 500 data yang dibagi menjadi dua kelas positif dan negatif dengan jumlah yang sama dan diambil 100 data uji secara acak dari data tersebut. Pengujian dilakukan 4 kali dengan 25% fitur, 50% fitur, 75% fitur dan 100% fitur. Prediksi dilakukan menggunakan *N-gram tokenizer* di mana terdapat dua model yaitu *uni-gram* yang menggunakan satu per satu kata dan *bi-gram* yang menggunakan dua kata terkait. Hasilnya pada model *unigram* mendapat nilai tertinggi dengan 89% akurasi, 89% presisi, dan 100% *recall* pada seleksi fitur 75%. Sedangkan model *bi-gram* mendapatkan nilai tertinggi dengan 89% akurasi, 89% presisi, dan 100% *recall* pada seleksi fitur 25%.

Xu et al. (2019) menganalisa data komentar berbasis teks menggunakan metode BiLSTM. Data yang digunakan merupakan data komentar hotel yang didapat dari website ctrip.com yang berjumlah 15.000 teks komentar. Pada penelitian ini juga dilakukan ekstraksi fitur yang mana mengintegrasikan kontribusi informasi sentimen kedalam TF-IDF. Hasil dari integrasi ini didapatkan nilai presisi sebesar 91.50, *recall* 92.87, dan *F1-score* 92.18 yang mana secara keseluruhan lebih baik dari seluruh ekstraksi fitur yang dicoba. Sedangkan hasil akhir metode BiLSTM yang digunakan pada penelitian ini didapatkan nilai terbaik dari keseluruhan metode yang dicoba dengan nilai presisi sebesar 91.54, *recall* 92.82, dan *F1-score* 92.18

Singh et al. (2019) dalam penelitiannya membandingkan penggunaan *multinomial* Naïve Bayes dengan *multivariate* Bernoulli Naïve Bayes. Data yang digunakan berjumlah 312 data polaritas berita. Hasilnya *multinomial* Naïve Bayes memiliki nilai akurasi 73% sedangkan *multivariate* Bernoulli Naïve Bayes memiliki nilai akurasi 69%.

Melihat penelitian-penelitian yang telah dilakukan sebelumnya, penulis memutuskan untuk melakukan klasifikasi ulasan fasilitas publik menggunakan seleksi fitur *chi-square* dengan metode Naïve Bayes. Dalam penelitian ini, akan dilakukan uji coba penggunaan rasio fitur sebanyak 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, dan 100% untuk menghitung nilai performa dalam sistem. Perbedaan rasio fitur ini digunakan untuk mengetahui pengaruh penggunaan seleksi fitur



chi-square pada klasifikasi ulasan fasilitas publik serta menganalisis faktor-faktor apa saja yang memengaruhi nilai performa dalam sistem.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data yang diperoleh dari ulasan pada platform Google Maps. Penulis melakukan pengambilan data menggunakan fitur pada *website* <https://apify.com/compass/google-maps-reviews-scrapert> terhadap data ulasan tempat-tempat fasilitas publik di Kota Malang yang dengan total jumlah data sebanyak 3000 data ulasan yang didapat dari 44 entitas fasilitas publik. Data yang digunakan cukup besar dan diharapkan dapat memberikan hasil yang cukup akurat dan representatif dalam penelitian. Data ulasan yang dipilih hanya data ulasan yang berupa teks sehingga data ulasan yang hanya menampilkan *rating* dan atau foto saja tidak akan digunakan. Ulasan akan diklasifikasikan menjadi 3 kelas yaitu mendukung (D), netral (N), dan tidak mendukung (TD). Ulasan yang termasuk kelas mendukung (D) merupakan ulasan positif dan pro terhadap fasilitas publik. Sebaliknya, ulasan yang termasuk kelas tidak mendukung (TD) merupakan ulasan negatif dan kontra terhadap fasilitas publik. Sedangkan ulasan yang termasuk kelas netral (N) merupakan ulasan yang tidak termasuk kategori dua kelas sebelumnya.

2.2 Desain Sistem

Untuk tercapainya penelitian, peneliti membangun sistem menggunakan bahasa pemrograman Python. Data ulasan fasilitas publik yang telah didapatkan akan diproses melalui beberapa langkah yaitu *preprocessing*, *bag-of-words*, seleksi fitur *chi-square*, Naïve Bayes sehingga hasilnya ulasan-ulasan tersebut akan terklasifikasi berdasarkan kelas dalam label.

2.3 Preprocessing

Ulasan fasilitas publik dari pengumpulan data perlu melalui tahapan *preprocessing* sebelum digunakan dalam proses selanjutnya karena dalam tahapan *preprocessing* dapat memperbaiki struktur inputan data. Proses yang dibutuhkan dalam tahapan *preprocessing* disesuaikan berdasarkan karakteristik dari data inputan. Dalam penelitian ini, proses yang akan dilakukan adalah *cleaning*, *lowercase*, *tokenizing*, *stopword removal*, dan *stemming*.

Sentimen perlu melewati tahapan *preprocessing* di mana pada tahap ini data sentimen akan diolah terlebih dahulu sebelum data diproses dalam klasifikasi atau sebelum diproses dalam algoritma (Faisal et al., 2020).

Cleaning akan menghapus karakter selain huruf yang ada pada data inputan sehingga data inputan hanya akan tersisa data berupa teks. Setelah itu teks perlu diseragamkan untuk menghindari *case sensitive*. Sehingga untuk menyeragamkan data, seluruh teks akan diubah menjadi huruf kecil dalam proses *lowercase*.

Data ulasan yang masih berupa kalimat dipecah menjadi satuan kata atau token, dan dilanjutkan pada proses *stemming*. Token kata diubah kembali ke dalam bentuk kata dasar tanpa imbuhan apapun. Segala jenis prefiks maupun sufiks yang ada dalam token dihilangkan. Terakhir adalah proses *stopword removal*. Pada proses ini kata-kata yang cenderung tidak penting dan tidak ada hubungannya dengan dokumen seperti konjungsi dan preposisi akan dihilangkan.

2.3.1 Bag-of-Words

Bag-of-words adalah representasi kemunculan kata atau *term* terhadap sebuah dokumen. Dalam *bag-of-words* dihitung frekuensi kemunculan kata dalam setiap kategori dokumen. *Bag-of-words* adalah teknik sederhana namun efektif untuk merepresentasikan teks sebagai vektor jumlah kata. Teknik ini melibatkan dua langkah: tokenisasi dan penghitungan. Tokenisasi membagi teks



menjadi kata-kata dan menghapus tanda baca dan karakter non-kata lainnya. Penghitungan menghitung berapa kali setiap kata muncul dalam teks. Vektor yang dihasilkan dapat digunakan sebagai masukan untuk berbagai algoritma pembelajaran mesin seperti Naive Bayes dan regresi logistik (Jurafsky & Martin, 2023). Tahapan ini perlu dilakukan sebab data yang telah diperoleh akan dilakukan transformasi data untuk mentransformasikan atribut-atribut pada variable ke dalam bentuk yang dapat diolah pada tahap selanjutnya (Baskoro et al., 2022).

2.3.2 Seleksi Fitur *Chi-Square*

Proses seleksi fitur diharapkan dapat mengurangi jumlah *noise* dan mengeliminasi fitur yang kurang relevan (Putri, 2018). Fitur yang nilainya rendah dianggap tidak banyak memengaruhi kinerja sistem sehingga bisa mengurangi beban sistem. *Chi-square* dipilih karena dapat digunakan untuk mengetahui tingkat independensi sebuah *term* atau fitur terhadap kelas label. *Chi-square* adalah alat yang berguna dan andal untuk diskritisasi dan pemilihan fitur atribut numerik (Huan Liu & Setiono, 1995). Bobot *chi-square* tiap kelas dapat dihitung menggunakan Pers. (1).

$$x^2(t, c) = \frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (1)$$

Di mana x^2 merupakan nilai *chi-square* fitur t dalam kelas c , dengan N merupakan jumlah keseluruhan dokumen. Nilai A didapat dari probabilitas jumlah dokumen pada kelas c yang mengandung fitur t , nilai B merupakan probabilitas jumlah dokumen selain pada kelas c yang mengandung fitur t , nilai C didapat dari probabilitas jumlah dokumen pada kelas c yang tidak mengandung fitur t , dan nilai D merupakan probabilitas jumlah dokumen selain pada kelas c yang tidak mengandung fitur t . Kemudian nilai total *chi-square term* dihitung menggunakan Pers. (2).

$$X^2(t) = \sum_{c=1}^k x^2(t, c) \quad (2)$$

Di mana X^2 merupakan nilai *chi-square* fitur t yang didapatkan dari penjumlahan $x^2(t, c)$ *chi-square term* sebanyak c kelas pertama sampai dengan k jumlah keseluruhan kelas pada label.

Setelah nilai *chi-Square* masing-masing *term* diketahui, maka *term* tersebut akan dilakukan pemeringkatan secara *descending* dari nilai *chi-square* dari *term* yang tertinggi. *Term* yang memiliki nilai *chi-square* tinggi merupakan *term* yang memiliki ketergantungan tinggi terhadap kelas label dan berpengaruh dalam sistem,

2.3.3 Naïve Bayes

Klasifikasi Naïve Bayes yang merupakan metode statistik yang dihitung menggunakan konsep probabilitas. Klasifikasi Naïve Bayes mengasumsikan bahwa efek dari nilai atribut pada kelas tertentu tidak tergantung pada kelas lainnya. Nilai atribut tersebut kemudian dicari nilai probabilitas *prior*, probabilitas atribut tiap kelas, dan hasilnya nilai probabilitas yang paling tinggi merupakan hasil dari klasifikasi Naïve Bayes. Nilai probabilitas *prior* dihitung menggunakan Pers. (3).

$$P(c) = \frac{Nc}{N} \quad (3)$$

Probabilitas *prior* $P(c)$ dihitung dari Nc yang merupakan jumlah dokumen pada kelas c dibagi dengan N jumlah keseluruhan dokumen. Selanjutnya adalah menghitung nilai probabilitas *likelihood* kelas menggunakan Pers. (4).

$$P(t|c) = \frac{freq(t,c) + 1}{freq(c) + V} \quad (4)$$

Dengan $P(t|c)$ merupakan probabilitas *term* t pada kelas c , $freq(t,c)$ merupakan jumlah frekuensi *term* t pada kelas c , $freq(c)$ merupakan jumlah *term* yang ada pada kelas c , dan V



merupakan jumlah keseluruhan *term* pada data latih. Kemudian dihitung probabilitas *posterior term* menggunakan Pers. (5).

$$P(c|t) = \frac{P(t|c)P(c)}{P(t)} \quad (5)$$

Dengan $P(c|t)$ merupakan probabilitas kelas c untuk *term* t , $P(t|c)$ merupakan probabilitas *term* t pada kelas c , $P(c)$ merupakan probabilitas *prior* kelas dan $P(t)$ merupakan probabilitas *prior* prediktor atau probabilitas jumlah dokumen yang terdapat *term* t terhadap keseluruhan dokumen.

Tetapi, hasil perhitungan $P(t)$ berbanding terbalik terhadap $P(c|t)$ serta nilainya selalu tetap saat nilai hasil $P(c|t)$ dibandingkan untuk mencari nilai kelas tertinggi. Sehingga $P(t)$ nilainya dapat dianggap 1 dan hanya menyisakan Pers. (6).

$$P(c|t) = P(t|c).P(c) \quad (6)$$

Metode Naïve Bayes yang digunakan dalam penelitian ini merupakan multinomial Naïve Bayes. Dalam multinomial Naïve Bayes digunakan distribusi multinomial untuk menghitung likelihood, di mana hal ini bekerja dengan baik untuk data yang merepresentasikan frekuensi atau jumlah, misalnya jumlah kemunculan kata pada dokumen (Russell & Norvig, 1995). Untuk menghitung probabilitas posterior dokumen, didapatkan dengan Pers. (7).

$$P(c|u) = \prod_{i=1}^n P(t_i|c) \cdot P(c) \quad (7)$$

Berdasar pada Pers. (6), $P(c|t)$ diubah menjadi $P(c|u)$ yang merupakan probabilitas kelas c untuk dokumen u , $P(t|c)$ diubah menjadi $\prod_{i=1}^n P(t_i|c)$ yang merupakan probabilitas *term* t pertama sampai *term* ke- n pada dokumen terhadap kelas c , probabilitas *prior* tetap.

Dalam pengklasifikasian Naïve Bayes kita perlu mendapatkan nilai tertinggi dari nilai probabilitas setiap kelas, yang dinyatakan sebagai *Hypothesis Maximum A Posteriori* (HMAP) dengan Pers. (8).

$$HMAP = \arg \max (P(c|u)) \quad (8)$$

2.4 Skenario Uji Coba

Data yang telah dikumpulkan akan dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Data *training* akan diimplementasikan ke dalam sistem yang telah dibuat untuk membangun model klasifikasi dokumen Naïve Bayes. Sedangkan data *testing* digunakan untuk melakukan pengujian kinerja pada sistem. Ulasan-ulasan yang ada akan dibagi menjadi data *testing* dan data *training* dengan rasio perbandingan 8:2 berdasarkan *Pareto Principle* (Sanders, 1987). *Stratified sampling* digunakan untuk membagi data berdasar jumlah label yang mendukung (D), netral (N), dan tidak mendukung (TD) sehingga perbandingan data *training* dan data *testing* terhadap rasio label tetap sama. Uji coba dilakukan berdasarkan seleksi fitur *chi-square* dengan menggunakan skenario variasi rasio penggunaan fitur sebanyak 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, serta 100% di mana penggunaan 100% fitur juga dianggap sebagai pengujian tanpa seleksi fitur. Hasil klasifikasi akan dibandingkan apakah sama dengan label (*ground truth*). Dan dicari nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN) dan akan dijumlahkan untuk menghitung nilai performa dari sistem menggunakan Pers. (9) sampai (12).

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (9)$$

$$Presisi = \frac{TP}{TP+FP} \quad (10)$$



$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$F - measure = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (12)$$

3. HASIL DAN PEMBAHASAN

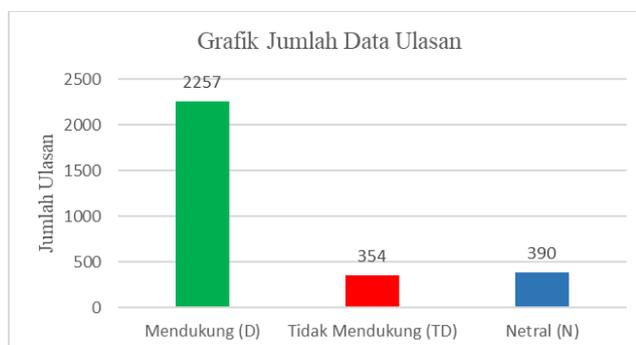
3.1 Data Understanding

Tabel 1 Entitas Penelitian

No.	Entitas	Jumlah Ulasan
1	Taman Merjosari	300
2	Rumah Sakit Saiful Anwar	97
3	Rumah Sakit Lavalette	136
4	RSUD Kota Malang	67
5	Bandara Abdurrahman Saleh	100
6	Terminal Arjosari	93
7	Stasiun Malang Kotabaru	107
8	Vihara Vajra Earth Kertanegara	4
9	Pura Luhur Dwijawarsa	22
10	Klenteng Eng An Kiong	26
11	Gereja Cathedral Ijen	86
12	Masjid Sabilillah	162
13	MAN 1 Kota Malang	13
14	MAN 2 Kota Malang	28
15	SMA N 1 Kota Malang	9
16	SMA N 3 Kota Malang	9
17	SMA N 4 Kota Malang	7
18	SMK N 2 Kota Malang	43
19	SMK N 3 Kota Malang	26
20	SMK N 4 Kota Malang	66
21	SMK Telkom	41
22	SMAK Kolese Santo Yusup	58
23	Lapangan Rampal	131
24	Stadion Gajayana	169
25	Perpustakaan Kota Malang	63
26	Perpustakaan Universitas Negeri Malang	57
27	Perpustakaan Universitas Muhammadiyah Malang	11
28	Perpustakaan UIN Maulana Malik Ibrahim Malang	21
29	Perpustakaan Universitas Brawijaya	26
30	Pasar Buku Wilis	113
31	Toko Buku Togamas	9
32	Pasar Besar Kota Malang	300
33	Mall Olympic Garden	171
34	Malang Town Square	129
35	Indigo Space	20
36	Ruang Perintis	19
37	East Java Super Corridor Malang	92
38	Ngalup Coworking Space	53
39	MOA Coworking Space	5
40	Seikopi Coworking Space	9
41	Malang Digital Core	30
42	EZO Coworking Space	3
43	Orbits Coworking	6
44	Alfath Coworking Space	63



Data ulasan fasilitas publik di Kota Malang yang berjumlah 3000 data ulasan yang berasal dari 44 entitas fasilitas publik. Berikut Tabel 1 merupakan tabel entitas dan Gambar 1 merupakan grafik *ground truth* data ulasan yang telah dilabeli.



Gambar 1 Grafik Jumlah Data Ulasan

Berdasarkan data tersebut didapatkan 2415 data *training* dan 585 data *testing* yang akan digunakan untuk dalam sistem.

3.2 Preprocessing dan Bag-of-Words

Data ulasan yang telah dilabeli dan disimpan dalam format csv diinputkan ke dalam sistem dalam bentuk *dataframe* dan akan diambil kolom yang berisi ulasan untuk masuk ke tahap *preprocessing* di antaranya *cleaning*, *lowercase*, *tokenizing*, *stopword removal*, dan *stemming*. Tabel 2 merupakan contoh hasil dari tahap *preprocessing*.

Tabel 2 Tahap *Preprocessing*

Keterangan	Hasil
Ulasan	Untuk tempatnya strategis sebenarnya, tapi sangat kurang perawatan menurut saya. Seandainya dirawat dengan baik pasti lebih oke. Skybikenya juga mangkrak.
<i>Cleaning</i>	Untuk tempatnya strategis sebenarnya tapi sangat kurang perawatan menurut saya Seandainya dirawat dengan baik pasti lebih oke Skybikenya juga mangkrak
<i>Lowercase</i>	untuk tempatnya strategis sebenarnya tapi sangat kurang perawatan menurut saya seandainya dirawat dengan baik pasti lebih oke skybikenya juga mangkrak
<i>Tokenizing</i>	"untuk", "tempatnya", "strategis", "sebenarnya", "tapi", "sangat", "kurang", "perawatan", "menurut", "saya", "seandainya", "dirawat", "dengan", "baik", "pasti", "lebih", "oke", "skybikenya", "juga", "mangkrak"
<i>Stemming</i>	untuk tempat strategis benar tapi sangat kurang awat turut saya anda rawat dengan baik pasti lebih oke skybikenya juga mangkrak
<i>Stopword Removal</i>	untuk tempat strategis benar tapi kurang awat turut anda rawat baik pasti lebih oke skybikenya mangkrak

Proses selanjutnya adalah *bag-of-word*. Data ulasan yang telah melewati tahap *preprocessing* akan dihitung frekuensi kemunculan kata unik atau *unique term* dalam tiap ulasan. Gambar 2 merupakan hasil dari tahap *bag-of-words*.



Gambar 2 Hasil Tahap *Bag-of-Words*

3.3 Seleksi Fitur *Chi-Square*

Tahapan seleksi fitur digunakan untuk menyeleksi fitur atau *term* berdasarkan bobot fitur. Fitur yang nilainya rendah dianggap tidak banyak memengaruhi kinerja sistem sehingga bisa menambah beban sistem. *Term* yang memiliki nilai *chi-square* tinggi merupakan *term* yang memiliki ketergantungan tinggi terhadap kelas label dan berpengaruh dalam sistem. Sedangkan *term* yang nilainya rendah merupakan *term* yang tidak memiliki ketergantungan pada label dan tidak berpengaruh banyak terhadap kinerja sistem sehingga nantinya akan dilakukan seleksi. Hasil perhitungan nilai *chi-square term* dapat dilihat pada Gambar 3.

	term	D	N	TD	Xtot	rank
0	sempit	5.546670	23.458788	0.814283	29.819740	1.0
1	peduli	4.250257	0.820863	20.452176	25.523296	2.0
2	bumi	2.779750	0.273167	17.257320	20.310237	3.0
3	canggih	2.779750	0.273167	17.257320	20.310237	4.0
4	customer	2.779750	0.273167	17.257320	20.310237	5.0
...
3395	butuh	0.011430	0.024750	0.000798	0.036978	3396.0
3396	diskusi	0.024728	0.006882	0.004997	0.036606	3397.0
3397	kompli	0.024728	0.006882	0.004997	0.036606	3398.0
3398	naik	0.024728	0.006882	0.004997	0.036606	3399.0
3399	samping	0.024728	0.006882	0.004997	0.036606	3400.0

Gambar 3 Hasil Perhitungan Nilai *Chi-Square Term*

3.4 Uji Coba

Percobaan selanjutnya akan dilakukan berdasarkan seleksi fitur *chi-square* dengan menggunakan skenario variasi rasio penggunaan fitur sebanyak 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, serta 100% di mana penggunaan 100% fitur juga dianggap sebagai pengujian tanpa seleksi fitur. Dari setiap penggunaan variasi rasio fitur ini, kemudian nilai performa akan dihitung sehingga dapat diketahui di mana rasio penggunaan fitur dengan performa terbaik.

Dalam penelitian ini menggunakan 3 kelas label sehingga sistem akan membuat klasifikasi untuk masing-masing ulasan yang diproses kemudian hasil klasifikasi tersebut akan dipadankan dengan label (*ground truth*). Jika hasil klasifikasi sistem dengan *ground truth* sama, nilai TP adalah 1 sebab hasil klasifikasi sistem benar pada satu kelas, dan nilai TN adalah 2 sebab hasil klasifikasi sistem terdapat 2 kelas selain *ground truth* dan diklasifikasikan secara benar, selain



kelas dari ulasan. Kemudian jika hasil klasifikasi sistem dengan *ground truth* berbeda atau salah, nilai TP adalah 0 sebab tidak ada hasil klasifikasi sistem yang sama dengan *ground truth*, nilai FP adalah 1 sebab seharusnya ada hasil klasifikasi sistem benar pada satu kelas, nilai FN adalah 1 sebab seharusnya ada hasil klasifikasi sistem salah pada satu kelas. Berdasarkan pemaparan tersebut didapatkan nilai yang ditunjukkan pada Tabel 3.

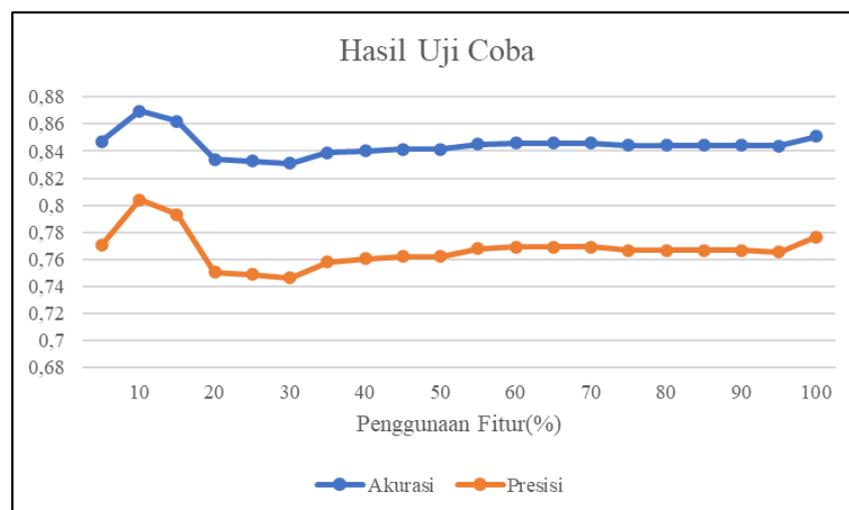
Tabel 3 Hasil Uji Coba

Features (%)	TP	TN	FP	FN
10	152	341	37	37
20	286	667	95	95
30	321	751	109	109
40	381	882	120	120
50	391	904	122	122
60	413	950	124	124
70	413	950	124	124
80	414	954	126	126
90	414	954	126	126
100	445	1018	128	128

Kemudian dihitung nilai performa dalam sistem yang ditunjukkan pada Tabel 4 serta representasinya pada Gambar 4.

Tabel 4 Nilai Performa

Penggunaan Fitur (%)	Akurasi	Presisi	Recall	F-Measure
10	0,869489	0,804233	0,804233	0,804233
20	0,833771	0,750656	0,750656	0,750656
30	0,831008	0,746512	0,746512	0,746512
40	0,840319	0,760479	0,760479	0,760479
50	0,841455	0,762183	0,762183	0,762183
60	0,846058	0,769088	0,769088	0,769088
70	0,846058	0,769088	0,769088	0,769088
80	0,844444	0,766667	0,766667	0,766667
90	0,844444	0,766667	0,766667	0,766667
100	0,851076	0,776614	0,776614	0,776614



Gambar 4 Grafik Nilai Performa



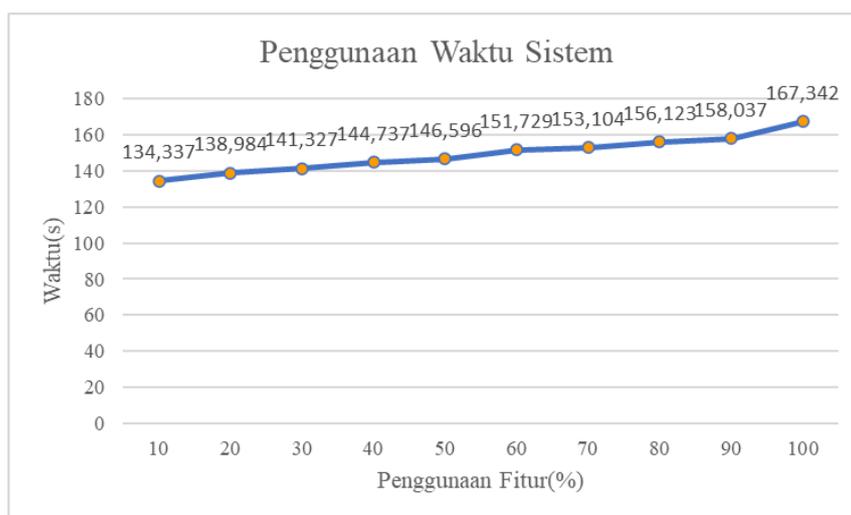
Berdasarkan percobaan yang telah dilakukan, terdapat kesalahan atau kelalaian dalam pengklasifikasian pemberitahuan fasilitas umum. Hal ini dikarenakan kesalahan ketik yang menyebabkan terdapat kekurangan pada tahap *preprocessing* dalam mengubah kata menjadi kata yang tepat seperti penggunaan huruf yang berulang-ulang “mantulll”, “hantuuu”, dan lain-lain. Selain itu, terdapat unsur kesalahan ejaan dan dua kata yang saling menempel seperti “pelajaeen”, “ruangtunggunya” dan lain-lain. Hal ini dikarenakan gramatika yang digunakan pengunjung tempat-tempat fasilitas publik dalam memberikan ulasan menggunakan kata-kata informal. Penggunaan kata pada ulasan yang bercampur dengan bahasa lain seperti “experience”, “nice” juga memengaruhi hasil klasifikasi.

Seleksi fitur *chi-square* yang digunakan dalam penelitian ini juga dapat memengaruhi hasil pengujian performa karena seleksi fitur *chi-square* akan menyeleksi kata yang tidak memiliki ketergantungan terhadap label sistem. Diketahui semakin banyak rasio penggunaan fitur yang digunakan dalam sistem, nilai performa cenderung semakin naik. Sebaliknya, semakin sedikit rasio fitur yang digunakan dalam sistem, nilai performa cenderung turun. Tetapi terdapat anomali pada hasil penggunaan fitur sebesar 10% yang menghasilkan nilai akurasi 86,94%, presisi 80,42%, *recall* 80,42% dan *f-measure* 80,42% di mana hasil ini merupakan nilai performa terbaik.

Faktor yang dapat memengaruhi hasil dan anomali pada hasil uji coba adalah fitur-fitur pada seleksi fitur *chi-square* itu sendiri. Misalnya, dalam penggunaan 10% fitur, fitur-fitur ini merupakan fitur dengan peringkat teratas yang memiliki nilai *chi-square* total tertinggi dari fitur-fitur yang digunakan dalam sistem. Meskipun tiap-tiap fitur tersebut merupakan fitur dengan peringkat teratas, dalam fitur kata tersebut masih banyak terdapat kata yang tidak baku, kata singkatan, serta kata bahasa asing.

Faktor lain yang dapat memengaruhi nilai performa adalah ulasan-ulasan yang tidak dapat terklasifikasi akibat pengurangan penggunaan fitur dalam sistem. Jika seluruh kata yang terdapat dalam sebuah ulasan tereduksi akibat seleksi fitur, maka ulasan tersebut tidak dapat diklasifikasikan dan akan dihilangkan dari data uji coba yang akan digunakan untuk menghitung nilai performa.

Durasi penggunaan waktu sistem dalam melakukan *training data* hingga data terklasifikasi keseluruhan dapat dilihat pada Gambar 5.



Gambar 5 Grafik Penggunaan Waktu Sistem

Diketahui bahwa semakin banyak rasio fitur yang digunakan, maka semakin lama pula durasi waktu yang dibutuhkan untuk melakukan klasifikasi. Di mana hal ini menunjukkan bahwa penggunaan seleksi fitur *chi-square* secara efisien dapat melakukan klasifikasi dengan durasi



waktu relatif lebih cepat. Hal ini dikarenakan fitur-fitur telah direduksi sebelum diproses dalam sistem sehingga tidak memberatkan sistem dalam melakukan perhitungan.

4. KESIMPULAN

Penggunaan 10% fitur *chi-square* menghasilkan performa terbaik yaitu nilai akurasi 86,94%, presisi 80,42%, *recall* 80,42%, dan *f-measure* 80,42%. Faktor yang dapat memengaruhi hal tersebut adalah banyaknya kata tidak baku yang masih terproses dalam sistem karena kurang optimalnya tahap *preprocessing* dalam menangani kata-kata yang memiliki huruf berulang, kesalahan penulisan, hingga kata yang seharusnya ditulis terpisah tetapi tertulis secara tergabung. Penggunaan seleksi fitur *chi-square* juga memengaruhi hasil performa di mana semakin banyak penggunaan fitur semakin tinggi pula nilai performa dalam sistem, akan tetapi terdapat anomali pada penggunaan 10% fitur pada penelitian ini yang menghasilkan nilai performa tertinggi. Tetapi, semakin banyak fitur yang digunakan juga menyebabkan waktu yang dibutuhkan untuk melakukan klasifikasi semakin lama.

Dikarenakan keterbatasan yang ada dalam penelitian ini, penulis menyarankan hal-hal berikut untuk dicoba pada penelitian selanjutnya:

- 1) Mengoptimalkan tahap *preprocessing* terutama untuk mengatasi, memperbaiki, mengeliminasi kesalahan kata, kata tidak baku, maupun kata yang tidak berpengaruh terhadap sistem.
- 2) Melakukan percobaan menggunakan algoritma *oversampling* atau *undersampling* untuk mengatasi *imbalance dataset*.
- 3) Melakukan percobaan menggunakan *N-gram* pada penggunaan fitur untuk mengetahui apakah fitur yang dipasangkan dapat memengaruhi performa sistem.

DAFTAR PUSTAKA

- Aziz, K. E., Crysdian, C., & Imamudin, M. (2021). Identification of Student Academic Performance in Computer Science Based on Naive Bayes. *MATICS*, 13(1), 33–41. <https://doi.org/10.18860/mat.v13i1.9726>
- Baskoro, S. E., Suhartono, S., Chamidy, T., & Zaman, S. (2022). Pengujian akurasi model regresi logistik multinomial untuk memprediksi keberhasilan mahasiswa di perguruan tinggi menggunakan r. *Fair Value: Jurnal Ilmiah Akuntansi Dan Keuangan*, 5(3), 1551–1565. <https://doi.org/10.32670/fairvalue.v5i3.2472>
- Dulhare, U. N. (2018). Prediction system for heart disease using Naive Bayes and particle swarm optimization. *Biomedical Research (India)*, 29(12), 2646–2649. <https://doi.org/10.4066/BIOMEDICALRESEARCH.29-18-620>
- Faisal, M., Nugroho, F., Sulthan, M. M. El, Amini, F., Hariyadi, M. A., & Sedayu, A. (2020). Plagiarism Detection Using Manber and Winnowing Algorithm. *International Journal of Advanced Science and Technology*, 29(6s), 2130–2136. <http://sersc.org/journals/index.php/IJAST/article/view/10924>
- Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Harahap, F., Harahap, A. Y. N., Ekadiansyah, E., Sari, R. N., Adawiyah, R., & Harahap, C. B. (2018). Implementation of Naive Bayes Classification Method for Predicting Purchase. *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 1–5. <https://doi.org/10.1109/CITSM.2018.8674324>
- Irvantoro, D. (2019). *Feature Selection Chi-Square N-Gram Naive Bayes Classifier Review* [Universitas Muhammadiyah Jember]. <http://repository.unmuhjember.ac.id/7128/>
- Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed.). Prentice Hall. <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>
- Linawati, S., Nurdiani, S., Handayani, K., & Latifah. (2020). *Prediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma Random Forest dan C4.5*. 8(1), 6–13. <https://doi.org/10.31294/jki.v8i1.7827>
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language*



- Technologies*, 5(1), 1–167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Huan Liu, & Setiono, R. (1995). Chi2: feature selection and discretization of numeric attributes. *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*, 388–391. <https://doi.org/10.1109/TAI.1995.479783>
- Pratama, N. D., Sari, Y. A., & Adikara, P. P. (2018). Analisis Sentimen pada Review Konsumen Menggunakan Metode Naive Bayes dengan Seleksi Fitur Chi Square untuk Rekomendasi Lokasi Makanan Tradisional. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 2982–2988. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2494>
- Putri, L. A. A. R. (2018). Seleksi Fitur dalam Klasifikasi Genre Musik. *Jurnal Ilmu Komputer*, 10(1), 19–26. <https://ojs.unud.ac.id/index.php/jik/article/view/39772>
- Russell, S., & Norvig, P. (1995). Artificial intelligence—a modern approach. In *The Knowledge Engineering Review* (Issue 1). Pearson Education. https://www.cambridge.org/core/product/identifier/S0269888900007724/type/journal_article
- Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*, 106, 92–104. <https://doi.org/10.1016/j.future.2020.01.005>
- Sanders, R. (1987). THE PARETO PRINCIPLE: ITS USE AND ABUSE. *Journal of Services Marketing*, 1(2), 37–40. <https://doi.org/10.1108/eb024706>
- Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, 593–596. <https://doi.org/10.1109/ICACTM.2019.8776800>
- Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment Analysis of Comment Texts Based on BiLSTM. *IEEE Access*, 7, 51522–51532. <https://doi.org/10.1109/ACCESS.2019.2909919>



Penerapan Naïve Bayes pada Potensi Akademik Siswa SD Negeri 5 Singakerta

Ni Kadek Winda Patrianingsih ^{(1)*}, I Kadek Arya Sugianta ⁽²⁾

Informatika, Fakultas Bisnis, Teknologi, Sosial dan Humaniora, Universitas Bali Internasional, Bali

e-mail : {windapatrianingsih,aryasugianta}@iikmpbali.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 28 Maret 2023, direvisi 12 April 2023, diterima 13 April 2023, dan dipublikasikan 26 Mei 2023.

Abstract

Student potential cannot only be measured based on the result of academic scores, and many things influence student academic determination. The purpose of this research is to prove that students' potential is influenced by many things, such as character, academic activity, socioeconomic status, and distance of residence. By using the naïve Bayes method and testing with the confusion matrix, it will give results for this research. The data is from V-grade students at SD Negeri 5 Singakerta, with 120 students assisted by the homeroom teacher. Based on the results of the tests that have been carried out using a data sample of 10 students and 1 data using the Naïve Bayes, it is obtained that students have academic potential, and the results with the confusion matrix are accuracy of 75%, precision of 81%, and recall of 89%. In this case, it can be concluded that the academic potential of students can not only be measured based on the results of the final grade, but many other factors have an effect, the application of the Naïve Bayes in students' academic potential is appropriate to use.

Keywords: *Student Potential, Student Academic, Data Mining, Naïve Bayes, Confusion Matrix*

Abstrak

Potensi siswa tidak hanya dapat diukur berdasarkan hasil nilai akademik, banyak hal yang berpengaruh dalam penentuan akademik siswa. Tujuan penelitian ini adalah untuk membuktikan bahwa potensi siswa dipengaruhi oleh banyak hal seperti, karakter, aktivitas akademik, status sosial ekonomi, dan jarak tempat tinggal. Dengan menggunakan algoritma Naïve Bayes dan pengujian dengan *confusion matrix*, akan memberikan hasil terhadap penelitian ini. Data yang digunakan yaitu data siswa kelas V di SD Negeri 5 Singakerta dengan perolehan data melalui peserta didik yang didampingi wali kelas sebanyak 120 siswa. Berdasarkan hasil perhitungan manual dengan Naïve Bayes memberikan keputusan terhadap data *testing* dengan hasil siswa tersebut memiliki potensi akademik. Pengujian dengan *confusion matrix* memberikan hasil yaitu *accuracy* sebesar 75%, *precision* sebesar 81%, dan *recall* sebesar 89%. Dalam hal ini dapat disimpulkan bahwa potensi akademik siswa tidak hanya dapat diukur berdasarkan hasil nilai akhir, namun banyak faktor lain yang berpengaruh, oleh karena itu penerapan algoritma Naïve Bayes dalam potensi akademik siswa layak digunakan.

Kata Kunci: *Potensi Siswa, Akademik Siswa, Data Mining, Naïve Bayes, Confusion Matrix*

1. PENDAHULUAN

Pendidikan memiliki peranan penting dalam kemajuan bangsa. Ilmu pengetahuan merupakan salah satu gambaran dari penggunaan dan pemanfaatan kemajuan teknologi komputer, yang terdiri dari berbagai bagian dari ilmu. Setiap individu memiliki atribut karakter yang berbeda-beda. Karakter yang disebut merupakan suatu ciri khas yang terdapat di dalam diri sendiri dalam keadaan yang berbeda (Martantoh & Yanih, 2022). Sekolah merupakan suatu jembatan bagi siswa di dalam menimba ilmu sebagai bekal siswa tersebut di masa depan. Faktor yang mempengaruhi pencapaian dari siswa adalah guru yang mampu memberikan pengajaran yang efektif dan berkualitas. Perkembangan proses pembelajaran merupakan suatu unsur dasar sebagai bentuk keberhasilan yang di mana setiap siswa memiliki tingkat kecepatan dalam



memahami materi dan nantinya akan menjadi perhatian khusus bagi sekolah untuk menentukan tingkat keberhasilan (Triandini et al., 2021). Perolehan apresiasi siswa akan ditampung dalam sebuah rapat yang di mana proses pembelajaran siswa telah dikumpulkan selama enam bulan atau satu semester. Hasil dari pembelajaran siswa tersebut akan diukur dan menghasilkan sebuah penilaian potensi dari masing-masing siswa (Susilo & Nur'aini, 2018).

SD Negeri 5 Singkerta merupakan sekolah dasar yang berada di Banjar Tunon, Kecamatan Ubud, Kabupaten Gianyar. Terdapat 120 siswa kelas lima pada semester ini. Perkembangan dunia digital memiliki berbagai manfaat dalam pemenuhan kebutuhan manusia yang salah satunya yaitu perkembangan *data science*. *Data science* memiliki fungsi signifikan dalam pengolahan data (Saleh, 2015b). Sistem penilaian dan pencatatan pada SD Negeri 5 Singkerta masih bersifat manual sehingga para wali kelas mempunyai kendala dalam proses penentuan potensi akademik siswa. Potensi siswa di SD Negeri 5 Singkerta masih dinilai dari hasil nilai akhir masing-masing siswa saja, sedangkan masih ada aspek penilaian yang dapat mendukung dari potensi siswa tersebut. Sistem penilaian tersebut tentu saja belum cukup untuk kemajuan potensi akademik siswa di SD Negeri 5 Singkerta. Salah satu alternatif yang dapat dilakukan untuk membantu pihak wali kelas dalam proses penilaian akhir yaitu dengan memanfaatkan perkembangan teknologi yang ada, khususnya *data mining*.

Dalam penelitian ini teknik *data mining* yang akan diterapkan yaitu algoritma Naïve Bayes. Pemilihan algoritma ini berdasarkan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang didefinisikan. Naïve Bayes adalah algoritma yang sederhana dan tiap-tiap atribut bersifat bebas, yang memungkinkan tiap atribut dapat berkontribusi terhadap hasil akhir (Marpaung et al., 2021). Penggunaan *data mining* khususnya Naïve Bayes diharapkan dapat memberikan informasi yang dapat digunakan untuk menentukan potensi akademik siswa di SD Negeri 5 Singkerta, sehingga kemampuan siswa dapat dilihat tidak hanya dari nilai akhir saja, namun dari segi ekonomi maupun jarak tempat tinggal dapat mempengaruhi.

Klasifikasi algoritma Naïve Bayes dalam *data mining* untuk menentukan konsentrasi siswa menghasilkan bahwa metode Naïve Bayes berhasil mengklasifikasikan 109 data siswa dari 120 data yang diuji dengan persentase keakuratan 90,8333% (Saleh, 2015a). Selanjutnya, dalam penelitian (Putri, 2020) yaitu penerapan model Naïve Bayes untuk memprediksi potensi pendaftaran siswa di SMK Taman Siswa Teluk Betung berbasis *web* memperoleh hasil akhir sebuah sistem informasi yang memiliki kemampuan untuk memprediksi potensi banyaknya siswa yang mendaftar dan tidak mendaftar dengan menggunakan aplikasi yang menghasilkan perhitungan akurasi sebesar 86%. Implementasi algoritma naïve bayes untuk klasifikasi penerima beasiswa (studi kasus universitas hamzanwadi) memberikan kesimpulan bahwa algoritma naïve bayes sangat baik digunakan dalam seleksi kelulusan pemberkasan beasiswa bidikmisi (Nurhidayati et al., 2023).

Dalam penelitian sebelumnya yaitu penentuan topik tugas akhir dengan metode Naïve Bayes, membuktikan bahwa metode ini menunjukkan hasil yang sangat layak berdasarkan kuisioner yang disebar ke dua responden yaitu dosen dan ahli sistem oleh (Ghaniy & Sihotang, 2019). Prediksi potensi mahasiswa berprestasi dengan metode Naïve Bayes di Stikom Dinamika Bangsa Jambi memberikan hasil persentase akurasi sebesar 87% oleh (Resti et al., 2021).

Selanjutnya, dalam penerapan metode Naïve Bayes Classifier untuk mengklasifikasi tingkat prestasi akademik santri Pondok Pesantren Mahasiswa (PPM) Baitul Jannah Malang (Y. M. Firdaus, 2019), penelitian tersebut dibuatkan sebuah sistem untuk mengklasifikasi tingkat prestasi akademik agar memudahkan para guru dalam proses penentuan prestasi akademik dan aplikasi diuji dengan menghasilkan keakuratan dengan nilai 76%.

Kelebihan dari penerapan Naïve Bayes yang dijelaskan dalam penelitian sebelumnya yaitu Naïve Bayes mampu memberikan hasil keakuratan dalam melakukan prediksi prestasi siswa, sehingga penelitian ini diambil berdasarkan penelitian sebelumnya bahwa penggunaan algoritma Naïve Bayes akan mampu menangani kumpulan data siswa yang terpilih yaitu siswa kelas lima di SD



Negeri 5 Singakerta untuk memberikan hasil yang akurat dan tepat. Selain itu faktor pendukung dari penggunaan naïve bayes dinilai bahwa berpotensi baik dalam melakukan klasifikasi suatu data dan memberikan hasil akurasi dan efesiensi yang optimal (Kawani, 2019).

Mengacu pada penjelasan latar belakang di atas, maka pentingnya dilakukan pengukuran hasil potensi akademik siswa di SD Negeri 5 Singakerta, yang kedepannya akan membantu pihak sekolah dalam peningkatan mutu pendidikan serta memberikan kriteria-kriteria yang tidak biasa dijumpai dalam melakukan penentuan potensi dengan hasil secara akurat dan tepat.

2. METODE PENELITIAN

2.1 Naïve Bayes

Kategorisasi probabilitas berdasarkan Teorema Bayes disebut Naïve Bayes. Untuk membuat perhitungan yang diperlukan lebih mudah, Naïve Bayes memperhitungkan dampak dari nilai atribut lainnya. Rumus probabilitas Bayesian yang luas dapat dilihat pada Pers. (1) (Sarasvananda et al., 2022).

$$\frac{P(C_i|X)=P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

Berdasarkan rumus probabilitas Naïve Bayes dijelaskan bahwa $P(C_i|X)$ merupakan probabilitas C_i jika diberi bukti X , selanjutnya $P(C_i)$ merupakan probabilitas C_i tanpa memandang bukti, $P(X|C_i)$ yaitu probabilitas X terjadi akan mempengaruhi C_i , dan $P(X)$ adalah probabilitas X tanpa memandang bukti apapun.

2.2 Confusion Matrix

Confusion matrix merupakan suatu cara untuk mengukur performa dari kinerja klasifikasi yang di mana mendapatkan keluaran dua kelas atau lebih. *Confusion matrix* digambarkan berupa tabel dengan empat kombinasi berbeda dari nilai aktual dan nilai prediksi, dari hasil tersebut dimasukkan ke dalam tabel yaitu *true positif*, *true negatif*, *false positif* dan *false negatif*. Dengan menggunakan perhitungan ini akan menghasilkan seberapa tingkat *accuracy*, *precision*, dan *recall*. Adapun persamaan dalam melakukan pengujian dengan *confusion matrix* terdapat pada Tabel 1 (A. F. Firdaus et al., 2021).

Tabel 1 *Confusion Matrix*

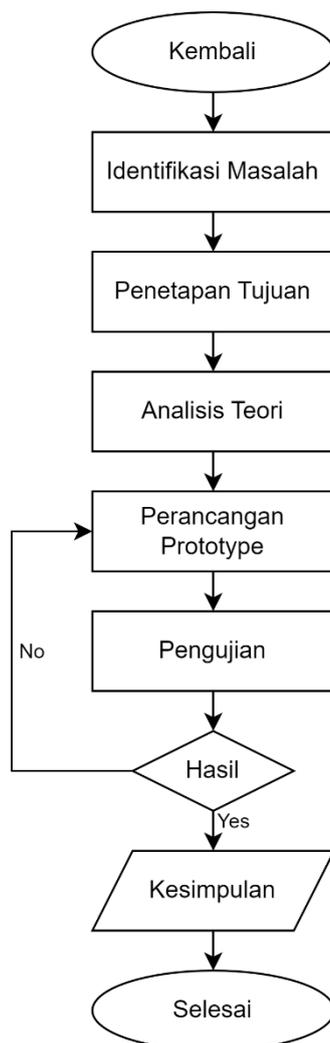
Nilai Aktual	Nilai Prediksi	
	Positif	Negatif
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

Berdasarkan Tabel 1 dijelaskan bahwa *true positif* merupakan data positif yang diprediksi benar, *true negatif* merupakan data negatif yang diprediksi benar, *false positif* merupakan data negatif namun diprediksi sebagai data positif, dan nilai prediksi *false negatif* merupakan data positif namun diprediksi sebagai data negatif.

2.3 Alur Penelitian

Dalam penelitian ini data yang ditambah adalah data siswa kelas V di SD Negeri 5 Singakerta. Penelitian ini difokuskan untuk mengetahui keakuratan dari penerapan metode Naïve Bayes dalam menentukan potensi akademik siswa di SD Negeri 5 Singakerta. Adapun rancangan alur penelitian digambarkan pada Gambar 1.





Gambar 1 Alur Penelitian

Berdasarkan rancangan alur pada Gambar 1, proses penelitian ini akan diawali dengan identifikasi masalah. Dalam penelitian ini, ditemukan permasalahan yang terjadi di SD Negeri 5 Singakerta yaitu dalam penentuan potensi akademik siswa, selanjutnya dilakukan penetapan tujuan sehingga mendapatkan hasil sesuai dengan perumusan masalah. Proses selanjutnya yaitu analisis teori, kemudian perancangan *prototype*, dan dilakukan pengujian yang menggunakan *confusion matrix*. Pengujian berhasil dan memperoleh hasil yang akurat, jika hasil yang diperoleh akurat dan baik dari proses pengujian bisa dilanjutkan ke tahap kesimpulan, jika tidak maka pengujian harus diulang kembali sampai menemukan hasil yang baik dan akurat, dan proses penelitian bisa selesai.

2.4 Pengumpulan Data

Data yang digunakan bersumber dari SD Negeri 5 Singakerta, yang di mana pengisian data tersebut diisi oleh peserta didik yang didampingi oleh wali kelas, kemudian data digunakan sebagai penunjang di data rapot siswa. Data yang diambil untuk pemilihan atribut adalah data siswa kelas V SD Negeri 5 Singakerta yang berjumlah 120 siswa dengan rincian sebagai pada Gambar 2.



Kode	Atribut				Kode	Atribut				Kode	Atribut			
	A1	A2	A3	A4		A1	A2	A3	A4		A1	A2	A3	A4
V01	A	Aktif	Sedang	Tinggi	V21	A	Tidak Aktif	Sedang	Rendah	V41	B	Aktif	Dekat	Tinggi
V02	A	Aktif	Sedang	Sedang	V22	B	Tidak Aktif	Sedang	Sedang	V42	B	Aktif	Dekat	Sedang
V03	A	Sedang	Sedang	Rendah	V23	B	Tidak Aktif	Dekat	Tinggi	V43	B	Aktif	Sedang	Sedang
V04	B	Aktif	Sedang	Sedang	V24	B	Aktif	Dekat	Sedang	V44	B	Aktif	Sedang	Tinggi
V05	B	Aktif	Dekat	Tinggi	V25	A	Aktif	Sedang	Sedang	V45	B	Aktif	Dekat	Sedang
V06	C	Aktif	Dekat	Sedang	V26	C	Tidak Aktif	Dekat	Tinggi	V46	B	Aktif	Dekat	Tinggi
V07	C	Tidak Aktif	Sedang	Sedang	V27	A	Aktif	Dekat	Sedang	V47	B	Sedang	Sedang	Tinggi
V08	A	Tidak Aktif	Dekat	Tinggi	V28	A	Tidak Aktif	Jauh	Tinggi	V48	B	Aktif	Dekat	Sedang
V09	B	Sedang	Dekat	Sedang	V29	B	Tidak Aktif	Dekat	Tinggi	V49	B	Aktif	Sedang	Tinggi
V10	C	Aktif	Jauh	Tinggi	V30	B	Aktif	Sedang	Sedang	V50	B	Aktif	Sedang	Sedang
V11	B	Sedang	Dekat	Tinggi	V31	B	Aktif	Sedang	Sedang	V51	B	Aktif	Dekat	Tinggi
V12	C	Tidak Aktif	Dekat	Tinggi	V32	B	Tidak Aktif	Sedang	Tinggi	V52	B	Aktif	Aktif	Tinggi
V13	C	Tidak Aktif	Dekat	Sedang	V33	C	Tidak Aktif	Dekat	Tinggi	V53	B	Aktif	Dekat	Sedang
V14	C	Aktif	Sedang	Sedang	V34	B	Aktif	Dekat	Sedang	V54	B	Aktif	Dekat	Rendah
V15	B	Aktif	Dekat	Tinggi	V35	B	Aktif	Jauh	Tinggi	V55	B	Tidak Aktif	Jauh	Sedang
V16	C	Tidak Aktif	Dekat	Sedang	V36	B	Aktif	Dekat	Tinggi	V56	B	Aktif	Dekat	Tinggi
V17	B	Tidak Aktif	Jauh	Tinggi	V37	B	Sedang	Sedang	Tinggi	V57	B	Aktif	Sedang	Sedang
V18	C	Tidak Aktif	Dekat	Tinggi	V38	B	Aktif	Sedang	Sedang	V58	A	Aktif	Dekat	Tinggi
V19	A	Tidak Aktif	Sedang	Tinggi	V39	B	Aktif	Sedang	Rendah	V59	C	Tidak Aktif	Dekat	Tinggi
V20	B	Aktif	Sedang	Sedang	V40	B	Aktif	Sedang	Sedang	V60	B	Aktif	Sedang	Sedang

Kode	Atribut				Kode	Atribut				Kode	Atribut			
	A1	A2	A3	A4		A1	A2	A3	A4		A1	A2	A3	A4
V61	B	Aktif	Dekat	Rendah	V81	B	Aktif	Sedang	Tinggi	V101	A	Aktif	Dekat	Sedang
V62	B	Aktif	Dekat	Sedang	V82	B	Aktif	Dekat	Sedang	V102	C	Tidak Aktif	Dekat	Tinggi
V63	B	Aktif	Jauh	Tinggi	V83	B	Aktif	Dekat	Tinggi	V103	A	Aktif	Dekat	Sedang
V64	B	Aktif	Dekat	Sedang	V84	B	Aktif	Sedang	Tinggi	V104	B	Aktif	Dekat	Tinggi
V65	C	Tidak Aktif	Sedang	Sedang	V85	B	Aktif	Dekat	Sedang	V105	B	Aktif	Dekat	Sedang
V66	B	Aktif	Sedang	Tinggi	V86	B	Aktif	Sedang	Tinggi	V106	A	Aktif	Sedang	Sedang
V67	B	Aktif	Sedang	Sedang	V87	B	Aktif	Sedang	Sedang	V107	A	Aktif	Dekat	Sedang
V68	B	Aktif	Sedang	Rendah	V88	B	Tidak Aktif	Sedang	Rendah	V108	A	Aktif	Jauh	Tinggi
V69	B	Aktif	Sedang	Sedang	V89	B	Aktif	Sedang	Sedang	V109	B	Aktif	Dekat	Tinggi
V70	C	Tidak Aktif	Dekat	Tinggi	V90	B	Aktif	Dekat	Tinggi	V110	A	Aktif	Sedang	Sedang
V71	B	Aktif	Dekat	Sedang	V91	B	Aktif	Dekat	Sedang	V111	A	Aktif	Dekat	Tinggi
V72	C	Tidak Aktif	Sedang	Sedang	V92	A	Tidak Aktif	Sedang	Sedang	V112	A	Aktif	Dekat	Sedang
V73	B	Aktif	Dekat	Tinggi	V93	B	Aktif	Dekat	Tinggi	V113	B	Aktif	Sedang	Sedang
V74	C	Aktif	Dekat	Sedang	V94	B	Aktif	Dekat	Sedang	V114	B	Aktif	Sedang	Tinggi
V75	B	Aktif	Jauh	Tinggi	V95	B	Aktif	Jauh	Tinggi	V115	B	Aktif	Dekat	Sedang
V76	B	Tidak Aktif	Dekat	Tinggi	V96	B	Aktif	Dekat	Tinggi	V116	B	Aktif	Dekat	Tinggi
V77	B	Aktif	Sedang	Sedang	V97	B	Aktif	Sedang	Sedang	V117	B	Aktif	Sedang	Tinggi
V78	B	Aktif	Dekat	Tinggi	V98	B	Tidak Aktif	Dekat	Sedang	V118	B	Aktif	Dekat	Sedang
V79	A	Aktif	Dekat	Sedang	V99	B	Aktif	Sedang	Sedang	V119	A	Aktif	Sedang	Tinggi
V80	B	Aktif	Sedang	Sedang	V100	B	Tidak Aktif	Sedang	Tinggi	V120	A	Aktif	Sedang	Sedang

Gambar 2 Atribut Data Siswa

2.5 Variabel Penelitian

Atribut penilaian mencakup antara lain yaitu karakter, aktivitas akademik, tempat tinggal, dan status sosial ekonomi. Uraian atribut dalam variabel penelitian ini adalah:

- 1) Atribut Karakter
Karakter siswa yang diukur dalam kategori sangat baik (A), baik (B), cukup (C), kurang baik (D), dan buruk (E).
- 2) Atribut Aktivitas Akademik
Aktivitas siswa dalam bidang akademik terdiri dari aktif dan tidak aktif.
- 3) Atribut Tempat Tinggal
Jarak dari rumah ke sekolah diukur berdasarkan kategori dekat, sedang, dan jauh.
- 4) Atribut Sosial Ekonomi
Atribut status sosial ekonomi diukur berdasarkan kategori tinggi, sedang, dan rendah.

3. HASIL DAN PEMBAHASAN

Metode yang akan digunakan untuk memprediksi potensi akademik siswa untuk mempermudah penilaian oleh wali kelas sehingga dapat berjalan dengan lancar dan metodis serta mencapai tujuan yang telah ditetapkan. Langkah-Langkah berikut termasuk dalam tahapan penelitian ini yaitu:

3.1 Analisis Data

Berdasarkan data yang telah terkumpulkan hasil dari implementasi Naïve Bayes dalam menentukan potensi akademik siswa akan dilakukan perhitungan secara manual dari data



training dan data *testing* yang telah ditentukan. Selanjutnya pengujian dengan *confusion matrix* akan dilakukan menggunakan keseluruhan data yang diberikan dari pihak sekolah. Hasil prediksi dalam Naïve Bayes akan membantu pihak sekolah tentunya para wali kelas dalam melakukan pengelolaan nilai dan memberikan keputusan terhadap siswa yang memiliki potensi akademik dan tidak memiliki potensi akademik berdasarkan kriteria yang sudah ditetapkan dalam penelitian ini.

3.2 Kriteria dan Sub Kriteria

Analisis potensi akademik siswa dengan menggunakan Naïve Bayes membutuhkan kriteria dan subkriteria dalam proses perhitungannya. Subkriteria yang ada juga perlu didefinisikan rentang nilai bobotnya. Kriteria dan subkriteria dapat dilihat pada Tabel 2 (Yendarman, 2016).

Tabel 2 Kriteria dan Sub Kriteria

Kriteria	Subkriteria	Nilai
Karakter	A (Sangat Baik)	5
	B (Baik)	4
	C (Cukup)	3
	D (Kurang Baik)	2
	E (Buruk)	1
Aktivitas Akademik	Aktif	5
	Sedang	3
	Tidak Aktif	1
Tempat Tinggal	Dekat	5
	Sedang	3
	Jauh	1
Status Sosial Ekonomi	Tinggi	5
	Sedang	3
	Rendah	1

3.3 Hasil Perhitungan Naïve Bayes

Berdasarkan hasil pengumpulan data yang telah dilakukan, jumlah populasi data dalam penelitian ini yaitu 120 orang. Pada proses perhitungannya data akan dibagi menjadi 2, yaitu data *training* yang menggunakan data dengan kode V01 sampai V10 dan data *testing* menggunakan data dengan kode V11. Adapun simulasi disajikan seperti pada Tabel 3.

Tabel 3 Data Training

Kode	ATRIBUT				Potensi
	A1	A2	A3	A4	
V01	A	Aktif	Sedang	Tinggi	Ya
V02	A	Aktif	Sedang	Sedang	Ya
V03	A	Sedang	Sedang	Rendah	Ya
V04	B	Aktif	Sedang	Sedang	Ya
V05	B	Aktif	Dekat	Tinggi	Ya
V06	C	Aktif	Dekat	Sedang	Tidak
V07	C	Tidak Aktif	Sedang	Sedang	Tidak
V08	A	Tidak Aktif	Dekat	Tinggi	Tidak
V09	B	Sedang	Dekat	Sedang	Ya
V10	C	Aktif	Jauh	Tinggi	Ya

Tabel 3 merupakan tabel data *training* dari algoritma Naïve Bayes. Selanjutnya diperlukan data *testing* untuk dapat menghasilkan keputusan terkait data siswa, data *testing* tersebut seperti pada Tabel 4.



Tabel 4 Data Testing

Kode	ATRIBUT				Potensi
	A1	A2	A3	A4	?
V11	B	Sedang	Dekat	Tinggi	

Berikut merupakan tahapan penyelesaian data uji menggunakan perhitungan Naïve Bayes.

1) Tahap 1, menghitung jumlah kelas/label

- $P(Y = Ya) = 7/10$
 Jumlah data "Ya" pada kolom "Potensi" dibagi jumlah data.
- $P(Y = Tidak) = 3/10$
 Jumlah data "Tidak" pada kolom "Potensi" dibagi jumlah data.

2) Tahap 2, menghitung jumlah kasus yang sama dengan kelas yang sama

- $P(Karakter = A|Y = Ya) = 3/7$
- $P(Karakter = A|Y = Tidak) = 1/3$
- $P(Karakter = B|Y = Ya) = 3/7$
- $P(Karakter = B|Y = Tidak) = 0/3$
- $P(Karakter = C|Y = Ya) = 1/7$
- $P(Karakter = C|Y = Tidak) = 2/3$

- $P(AktivitasAkademik = Aktif|Y = Ya) = 5/7$
- $P(AktivitasAkademik = Aktif|Y = Tidak) = 1/3$
- $P(AktivitasAkademik = Sedang|Y = Ya) = 2/7$
- $P(AktivitasAkademik = Sedang|Y = Tidak) = 0/3$
- $P(AktivitasAkademik = TidakAktif|Y = Ya) = 0/7$
- $P(AktivitasAkademik = TidakAktif|Y = Tidak) = 2/3$

- $P(TempatTinggal = Dekat|Y = Ya) = 2/7$
- $P(TempatTinggal = Dekat|Y = Tidak) = 2/3$
- $P(TempatTinggal = Sedang|Y = Ya) = 4/7$
- $P(TempatTinggal = Sedang|Y = Tidak) = 1/3$
- $P(TempatTinggal = Jauh|Y = Ya) = 1/7$
- $P(TempatTinggal = Jauh|Y = Tidak) = 0/3$

- $P(StatusSosialEkonomi = Rendah|Y = Ya) = 1/7$
- $P(StatusSosialEkonomi = Rendah|Y = Tidak) = 0/3$
- $P(StatusSosialEkonomi = Sedang|Y = Ya) = 3/7$
- $P(StatusSosialEkonomi = Sedang|Y = Tidak) = 2/3$
- $P(StatusSosialEkonomi = Tinggi|Y = Ya) = 3/7$
- $P(StatusSosialEkonomi = Tinggi|Y = TIDAK) = 1/3$

3) Tahap 3, kalikan semua hasil variabel Ya dan Tidak, sesuai dengan sampel data *testing*.

V11	B	Sedang	Dekat	Tinggi
-----	---	--------	-------	--------

Selanjutnya, dilakukan proses perhitungan terhadap variabel-variabel tersebut, untuk mendapatkan nilai dari sampel data *testing*.

Ya

$$\begin{aligned}
 R &= P(Karakter = B|Y = Ya) \times P(AktivitasAkademik = TidakAktif|Y = Ya) \times \\
 &P(TempatTinggal = Dekat|Y = Ya) \times P(StatusSosialEkonomi = Sedang|Y = YA) \times P(Y = YA) \\
 &= 3/7 \times 2/7 \times 2/7 \times 3/7
 \end{aligned}$$



= 0,01499375

Tidak

$$R = P(\text{Karakter} = B|Y = Ya) * P(\text{Aktivitas Akademik} = \text{Tidak Aktif}|Y = Ya) * P(\text{Tempat Tinggal} = \text{Dekat}|Y = Ya) * P(\text{Status Sosial Ekonomi} = \text{Sedang}|Y = Ya) * P(Y = Ya)$$

$$= 3/7 \times 2/7 \times 2/7 \times 3/7$$

$$= 0$$

4) Tahap 4, bandingkan hasil pada kelas Ya dan Tidak. Karena hasil (P|Ya) lebih besar dibanding (P|Tidak), maka keputusan yang dihasilkan dari data uji bahwa siswa dengan kode V11 **memiliki potensi akademik.**

3.4 Hasil Pengujian

Berdasarkan hasil perhitungan yang telah didapatkan, selanjutnya dilakukan pengujian dengan menggunakan *confusion matrix*. Pengujian ini dilakukan oleh hasil yang diberikan oleh pihak sekolah sesuai dengan penilaian per semester oleh masing-masing wali kelas dengan jumlah data sebanyak 120 siswa. Pengujian dengan *confusion matrix* ini akan menghasilkan tiga nilai yaitu *accuracy*, *precision*, dan *recall*. Tabel 5 menunjukkan hasil prediksi dari algoritma Naïve Bayes.

Tabel 5 Hasil Prediksi Naïve Bayes

Aktual	Prediksi	
	Berpotensi	Tidak Berpotensi
Berpotensi	85	10
Tidak Berpotensi	20	5

$$Accuracy(\%) = \frac{85 + 5}{120} = 75\%$$

Kedekatan nilai prediksi dengan nilai asli berada pada tingkat 75 %.

$$Precision(\%) = \frac{85}{85 + 20} = 81\%$$

Prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif berada pada tingkat 81 %.

$$Recall = \frac{85}{85 + 10} = 89\%$$

Ukuran ketepatan dari suatu kejadian yang diinginkan berada pada tingkat 89%.

Hasil dari pengujian *confusion matrix* dengan metode Naïve Bayes yaitu memiliki nilai *accuracy* sebesar 75 %, nilai *precision* sebesar 81 %, dan nilai *recall* sebesar 89%.

4. KESIMPULAN

Sistem penilaian di SD Negeri 5 Singakerta masih bersifat subjektif dengan menggunakan nilai akhir sebagai acuan dalam menentukan potensi akademik siswa. Potensi siswa tidak dapat diukur hanya dengan satu faktor, melainkan banyak faktor pendukung yang dapat berpengaruh dalam penentuan potensi akademik siswa. Penelitian ini menjelaskan bahwa pengukuran terhadap potensi akademik siswa dapat berpengaruh dengan status sosial ekonomi, karakter, aktivitas akademik, dan juga jarak tempat tinggal. Perkembangan teknologi yang semakin maju, dapat membantu pihak sekolah dalam melakukan penyelesaian permasalahan ini, yaitu dengan teknik *data mining* dengan menggunakan algoritma Naïve Bayes. Naïve Bayes merupakan suatu algoritma yang berfungsi untuk memprediksi suatu kelas dengan melakukan pengelompokan



secara efektif. Adapun data dalam penelitian ini bersumber dari data SD Negeri 5 Singakerta. Data tersebut terkumpul sebanyak 120 data siswa, yang di mana data dengan jumlah 120 tersebut sudah diberikan oleh pihak sekolah yang digunakan untuk melakukan pengujian dengan *confusion matrix*. Pengujian ini dilakukan dengan mendapatkan hasil yaitu *accuracy* sebesar 75%, *precision* sebesar 81%, dan *recall* sebesar 89%. Hal ini membuktikan bahwa penilaian potensi akademik siswa dapat dilakukan dengan algoritma Naïve Bayes dengan menghasilkan data akurasi yang cukup baik. Hal lain yang perlu dikembangkan dalam penelitian selanjutnya yaitu dapat dilakukan implementasi dengan perangkat lunak berbasis *web*, dengan menambahkan algoritma lain dalam perhitungannya sehingga menghasilkan tingkat akurasi yang lebih baik.

DAFTAR PUSTAKA

- Firdaus, A. F., Saedudin, R., & Andeswari, R. (2021). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa. *EProceedings of Engineering*, 8(5), 9274–9279. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15825/15538>
- Firdaus, Y. M. (2019). Penerapan Metode Naïve Bayes Classifier untuk Mengklasifikasi Tingkat Prestasi Akademik Santri Pondok Pesantren Mahasiswa (PPM) Baitul Jannah Malang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 3(1), 327–336. <https://doi.org/10.36040/JATI.V3i1.1398>
- Ghaniy, R., & Sihotang, K. (2019). Penerapan Metode Naïve Bayes Classifier Untuk Penentuan Topik Tugas Akhir. *Teknois: Jurnal Ilmiah Teknologi Informasi Dan Sains*, 9(1), 63–72. <https://doi.org/10.36350/jbs.v9i1.7>
- Kawani, G. P. (2019). Implementasi Naive Bayes. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 1(2), 73–81. <https://doi.org/10.20895/inista.v1i2.73>
- Marpaung, S., -, S., & -, I. (2021). Penerapan Metode Naïve Bayes Dalam Memprediksi Prestasi Siswa Di SMA Negeri 1 Panombeian Panei. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 4(2), 8–13. <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v4i2.1522>
- Martantoh, E., & Yanih, N. (2022). Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan Php Mysql. *Jurnal Teknologi Sistem Informasi*, 3(2), 166–175. <https://doi.org/10.35957/jtsi.v3i2.2896>
- Nurhidayati, N. I., Yahya, Y., Fathurrahman, F., Samsu, L. M., & Amnia, W. (2023). Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi). *Infotek: Jurnal Informatika Dan Teknologi*, 6(1), 177–188. <https://doi.org/10.29408/jit.v6i1.7529>
- Putri, S. eka Y. (2020). Penerapan Model Naive Bayes untuk Memprediksi Potensi Pendaftaran Siswa di SMK Taman Siswa Teluk Betung Berbasis Web. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 1(1), 93–99. <https://doi.org/10.33365/jatika.v1i1.228>
- Resti, R., Abidin, D. Z., & Rasywir, E. (2021). Penerapan Data Mining Klasifikasi Untuk Memprediksi Potensi Mahasiswa Berprestasi Di Stikom Dinamika Bangsa Jambi Dengan Metode Naive Bayes. *Jurnal Ilmiah Mahasiswa Teknik Informatika*, 3(2), 97–114. <https://ejournal.stikom-db.ac.id/index.php/jimti/article/view/1047>
- Saleh, A. (2015a). Klasifikasi Metode Naive Bayes dalam Data Mining untuk Menentukan Konsentrasi Siswa (Studi Kasus di MAS PAB 2 Medan). *KeTIK - Konferensi Nasional Pengembangan Teknologi Informasi Dan Komunikasi*, 200–208.
- Saleh, A. (2015b). Penerapan Data Mining dalam Menentukan Jurusan Siswa. *Seminar Nasional Informatika 2015*, 1(1), 351–355. <https://e-journal.potensi-utama.ac.id/ojs/index.php/SNI/article/view/291>
- Sarasvananda, I. B. G., Selivan, D., Radhitya, M. L., & Putra, I. N. T. A. (2022). Analisis Sentimen Pada Pembelajaran Daring Di Indonesia Melalui Twitter Menggunakan Naïve Bayes Classifier. *SINTECH (Science and Information Technology) Journal*, 5(2), 227–233. <https://doi.org/10.31598/sintechjournal.v5i2.1241>
- Susilo, G., & Nur'aini, T. A. (2018). Pengaruh Tes Potensi Akademik Siswa terhadap Prestasi



- Belajar Matematika. *JKPM (Jurnal Kajian Pendidikan Matematika)*, 4(1), 21. <https://doi.org/10.30998/jkpm.v4i1.2975>
- Triandini, M., Defit, S., & Nurcahyo, G. W. (2021). Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia. *Jurnal Informasi Dan Teknologi*, 3, 167–173. <https://doi.org/10.37034/jidt.v3i3.120>
- Yendarman, Y. (2016). Peningkatan Kemampuan Menetapkan Kkm dengan Diskusi Kelompok Kecil Bagi Guru Matematika SMKN 1, SMKN 4, Smk PGRI 2 Kota Jambi. *Jurnal Ilmiah Universitas Batanghari Jambi*, 16(3), 122–129. <https://doi.org/10.33087/JIUBJ.V16I3.25>





9 772527 583007

LABORATORIUM AGAMA
MASJID SUNAN KALIJAGA