

ISSN : 2527-5836

e-ISSN : 2528-0074

Vol. 6 No. 3, September 2021

# *JISKa*

Jurnal Informatika Sunan Kalijaga

Jurusan Teknik Informatika  
Fakultas Sains dan Teknologi  
UIN Sunan Kalijaga Yogyakarta



## **Tim Pengelola JISKa Edisi September 2021**

### **Ketua Editor (Editor in Chief)**

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

### **Editor Bagian (Associate 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. Enny Itje Sela (Universitas Teknologi Yogyakarta, Indonesia)
4. Dr. Hamdani (Universitas Mulawarman Samarinda, Indonesia)
5. Nashrul Hakiem, Ph.D. (UIN Syarif Hidayatullah Jakarta, 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)
3. Rahmat Hidayat, M.Cs. (UIN Sunan Kalijaga Yogyakarta, Indonesia)
4. Usfita Kiftiyani, M.Sc. (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. Dr. Cahyo Crysdiان (UIN Maulana Malik Ibrahim Malang, Indonesia)
4. Dr. Eng. Ganjar Alfian (Dongguk University Seoul, Korea, Republic of)
5. Muhammad Rifqi Maarif, M.Eng. (Universitas Jenderal Achmad Yani Yogyakarta, Indonesia)
6. Mushab Al Barra, M.Kom. (Universitas Ahmad Dahlan Yogyakarta, Indonesia)
7. Dr.Eng. M. Alex Syaekhoni (Dongguk University Seoul, Korea, Republic of)
8. Norma Latif Fitriyani, M.Sc. (Dongguk University Seoul, Korea, Republic of)
9. Oman Somantri, M.Kom. (Politeknik Negeri Cilacap, Indonesia)
10. Puji Winar Cahyo, M.Cs. (Universitas Jenderal Achmad Yani Yogyakarta, Indonesia)
11. Rischian Mafrur, M.Eng. (The University of Queensland Brisbane, Australia)
12. Suhirman, Ph.D. (Universitas Teknologi Yogyakarta, Indonesia)
13. Yunita Ardilla, M.Sc (Institut Teknologi Sepuluh November Surabaya, Indonesia)

ISSN : 2527-5836

e-ISSN: 2528-0074

**JISKa**

Vol. 6, No. 3, SEPTEMBER 2021

## DAFTAR ISI

<b>Implementasi Deep Learning untuk Entity Matching pada Dataset Obat (Studi Kasus K24 dan Farmaku)</b>	<b>130-138</b>
Rivanda Putra Pratama, Rahmat Hidayat, Nisrina Fadhilah Fano, Adam Akbar, Nur Aini Rakhmawati	
<b>Optimasi Keamanan Web Server terhadap Serangan Broken Authentication Menggunakan Teknologi Blockchain</b>	<b>139-148</b>
Imam Riadi, Herman, Aulyah Zakilah Ifani	
<b>Peramalan Pelayanan Service Mobil (After-Sale) Menggunakan Backpropagation Neural Network (BPNN)</b>	<b>149-160</b>
Novianti Puspitasari, Haviluddin, Arinda Mulawardani Kustiawan, Hario Jati Setyadi, Gubtha Mahendra Putra	
<b>Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naïve Bayes</b>	<b>161-170</b>
Ami Natuzzuhriyyah, Nisa Nafisah, Rini Mayasari	
<b>Implementasi Algoritma RC4 pada Sistem Pengamanan Dokumen Digital Soal Ujian</b>	<b>171-177</b>
Fauziah Suwarsita Febriyani, Arief Arfriandi	
<b>Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up</b>	<b>178-188</b>
Adhitya Prayoga Permana, Kurniyatul Ainiyah, Khadijah Fahmi Hayati Holle	
<b>Perbandingan Faktor-Faktor Yang Mempengaruhi Penggunaan Electronic Know Your Customer (e-KYC)</b>	<b>189-200</b>
Mr. Fitree Tahe, Maria Ulfah Siregar, Mohammad Farhan Quadratullah	

## Implementasi Deep Learning untuk Entity Matching pada *Dataset* Obat (Studi Kasus K24 dan Farmaku)

Rivanda Putra Pratama <sup>(1)</sup>, Rahmat Hidayat <sup>(2)\*</sup>, Nisrina Fadhilah Fano <sup>(3)</sup>, Adam Akbar <sup>(4)</sup>,  
Nur Aini Rakhmawati <sup>(5)</sup>

Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi  
Sepuluh Nopember, Surabaya

e-mail : {rivanda.19052,rahmat.19052}@mhs.its.ac.id, {adamakbar.id,nisrina.fano}@gmail.com,  
nur.aini@is.its.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 18 Januari 2021, direvisi 6 Maret 2021, diterima 13 Maret 2021, dan  
dipublikasikan 22 September 2021.

### Abstract

*Data processing speed in companies is important to speed up their analysis. Entity matching is a computational process that companies can perform in data processing. In conducting data processing, entity matching plays a role in determining two different data but referring to the same entity. Entity matching problems arise when the dataset used in the comparison is large. The deep learning concept is one of the solutions in dealing with entity matching problems. DeepMatcher is a python package based on a deep learning model architecture that can solve entity matching problems. The purpose of this study was to determine the matching between the two datasets with the application of DeepMatcher in entity matching using drug data from farmaku.com and k24klik.com. The comparison model used is the Hybrid model. Based on the test results, the Hybrid model produces accurate numbers, so that the entity matching used in this study runs well. The best accuracy value of the 10th training with an F1 value of 30.30, a precision value of 17.86, and a recall value of 100.*

**Keywords:** *Entity Matching, Deep Learning, DeepMatcher, Dataset, Hybrid*

### Abstrak

Kecepatan dalam pengolahan data di perusahaan penting dilakukan untuk mempercepat analisis perusahaan. *Entity matching* menjadi sebuah proses komputasi yang dapat dilakukan oleh perusahaan dalam pengolahan data. Dalam melakukan pemrosesan data, *entity matching* berperan dalam menentukan dua data berbeda namun merujuk ke entitas yang sama. Permasalahan dalam *entity matching* muncul ketika *dataset* yang digunakan dalam perbandingan berukuran besar. Konsep *deep learning* menjadi salah satu solusi dalam menangani permasalahan *entity matching*. *DeepMatcher* adalah sebuah paket Python berdasarkan arsitektur model *deep learning* yang dapat menjadi solusi dalam permasalahan *entity matching*. Tujuan dari penelitian ini adalah untuk mengetahui pencocokan antara dua buah *dataset* dengan penerapan *DeepMatcher* dalam *entity matching* dengan menggunakan data obat yang berasal dari farmaku.com dan k24klik.com. Model perbandingan yang digunakan adalah model *Hybrid*. Berdasarkan hasil pengujian dari model *Hybrid* menghasilkan angka yang akurat, sehingga *entity matching* yang digunakan pada penelitian ini berjalan dengan baik. Dari hasil *training*, diperoleh nilai akurasi terbaik pada *training* ke-10 dengan nilai *F1* sebesar 30,30, *precision* sebesar 17,86, dan *recall* sebesar 100.

**Kata Kunci:** *Entity Matching, Deep Learning, DeepMatcher, Dataset, Hybrid*

## 1. PENDAHULUAN

Perkembangan dan kecepatan pengolahan data di suatu perusahaan saat ini sangat penting untuk dilakukan, langkah-langkah tersebut dilakukan untuk mempercepat analisis mengenai kemampuan perusahaan dalam membaca kebutuhan konsumen. *Entity matching* menjadi sebuah proses komputasi yang dapat dilakukan oleh perusahaan dalam pengolahan data. *Entity matching* merupakan permasalahan untuk menentukan apakah dua data merujuk ke entitas yang sama di dunia nyata (Mudgal et al., 2018). Contoh kasus dalam *entity matching* yang biasanya terjadi adalah menentukan kecocokan data dari dua *dataset* berbeda, padahal merujuk ke suatu



entitas yang sama (Christophides et al., 2015). Permasalahan muncul ketika *dataset* yang digunakan berukuran besar. Jika *dataset* yang digunakan besar, akan cukup sulit untuk menentukan pasangan entri yang cocok (Li et al., 2020). Proses yang dilakukan untuk menyelesaikan permasalahan *entity matching* menjadi kompleks. Beberapa peneliti telah menemukan metode untuk menyelesaikan permasalahan *entity matching* pada *dataset* yang berukuran besar. Salah satunya adalah menggunakan konsep *deep learning*.

*Deep learning* adalah sebuah paradigma yang sukses dalam bidang *machine learning* yang telah mencapai kesuksesan signifikan di berbagai bidang, seperti *computer vision*, *natural language processing*, *speech recognition*, *genomics*, dan lain-lain (Thirumuruganathan et al., 2020). Selain itu, *deep learning* telah berhasil diterapkan diberbagai bidang seperti *signal processing*, *artificial intelligence*, dan *emotion detection* (Abdullah et al., 2021). *Deep learning* menjadi sebuah *tool* yang secara mendalam dan berukuran besar yang berfokus pada jaringan saraf tiruan atau *artificial neural networks* (Yuan et al., 2020). *DeepMatcher* merupakan sebuah kerangka *entity matching* mendalam yang diusulkan oleh (Mudgal et al., 2018), kerangka ini memiliki tiga modul yaitu penyematan atribut, representasi tujuan atribut, dan pengklasifikasi (Fu et al., 2019). Dalam penelitian (Chen et al., 2018), mengungkapkan bahwa *DeepMatcher* adalah sebuah paket python berdasarkan arsitektur model *deep learning* yang dapat dengan baik menampilkan *entity matching*, *question answering*, dan *textual entailment*. Kemudian Mudgal et al. (2018) berpendapat bahwa *DeepMatcher Python* dapat menjadi sebuah solusi dalam permasalahan *entity matching*.

Penelitian yang dilakukan oleh Li et al. (2020) telah melakukan *deep entity matching* dengan model bahasa yang sudah dilakukan *pre-training* sebelumnya (*Pre-Trained Models*). Melalui sistem yang bernama DITTO, para peneliti melakukan *entity matching* dalam tiga tahap, yaitu tahapan *blocking*, tahap pelatihan sistem, dan tahap *matching*. Pada tahap *blocking*, dilakukan pemangkasan dari pasangan entitas-entitas. Pada tahap kedua, proses pelatihan dilakukan dengan menambahkan *domain knowledge* untuk meningkatkan performa DITTO. Pada tahap ketiga, proses *matching* dilakukan untuk menemukan pasangan entitas yang tepat (Li et al., 2020). Penelitian lain yang menggunakan *Deep Learning* untuk *Entity Matching* adalah penelitian yang dilakukan oleh Fu et al. (2019). Pada penelitian tersebut, para peneliti mengembangkan model *neural multi-perspective matching* (MPM). Terdapat empat *layer* pada model MPM. *Layer* pertama adalah *attribute representation layer* yang berfungsi untuk menyeragamkan bentuk atribut yang ada pada data. *Layer* kedua adalah *comparison layer* yang berfungsi untuk membandingkan nilai dari atribut menggunakan nilai *similarity* antara dua nilai atribut. *Layer* ketiga adalah *selection layer* yang berfungsi untuk memilih data yang mempunyai keserupaan antar atribut yang paling tinggi. Kemudian data tersebut digabungkan pada *layer* keempat, yaitu *aggregation layer* untuk mendapatkan hasil dari *entity*.

Pengembangan sebuah sistem untuk *entity matching* yang terdiri dari enam tahap telah dilakukan oleh Kasai et al. (2019). Tahap pertama adalah *input representation*. Pada tahap tersebut, dilakukan tokenisasi untuk nilai atribut dan melakukan vektorisasi kata untuk mendapatkan representasi masukan. Tahap kedua adalah *attribute representations* yang menggunakan *universal bidirectional RNN* untuk mendapatkan vektor atribut menggunakan representasi masukan dari setiap nilai atribut. Selanjutnya adalah *attribute similarity* yang berfungsi untuk menghitung keserupaan setiap pasang entitas pada data. Hasil perhitungan keserupaan atribut kemudian digunakan pada tahap *record similarity* untuk menghitung keserupaan antar entitas. Tahap kelima adalah *matching classification* yang berfungsi untuk memisahkan pasangan entitas yang *match* dengan pasangan entitas yang tidak *match*. Tahap terakhir pada sistem ini adalah *training objectives* yang bertujuan untuk melatih sistem.

Penelitian yang dilakukan oleh Zhao & He (2019) dalam mengembangkan sebuah sistem *end-to-end auto EM* yang terdiri dari tiga komponen. Komponen pertama adalah *attribute-type detection*, yakni pada komponen ini dilakukan identifikasi tipe atribut. Sebagai contoh apabila salah satu atribut berisi tentang nama orang, maka atribut tersebut akan ditandai sebagai tipe 'Person'. Model identifikasi tipe atribut telah dilatih sebelumnya menggunakan data *Knowledge Base* dari mesin pencarian. Komponen kedua adalah *attribute-level EM model*, di mana pada



komponen ini dilakukan penghitungan keserupaan antara nilai dua atribut. Model ini dibagi menjadi dua bagian, yaitu *type-specific models*, yaitu untuk atribut dengan tipe yang sudah terdeteksi dan *unified model* untuk atribut yang tipenya masih belum terdeteksi. Komponen ketiga adalah *table-level EM* yang menghasilkan hasil akhir dari solusi untuk permasalahan *entity matching* (Zhao & He, 2019). Penelitian dalam menganalisa *entity matching* dengan menggunakan model SIF, RNN, Attention, dan Hybrid telah dilakukan oleh (Hidayat et al., 2021). Dalam penelitian tersebut melakukan pencocokan pada *dataset smartphone* dengan menggunakan *DeepMatcher*. Hasilnya adalah pengujian dengan semua model rata-rata akurat. Model *Attention* dan Hybrid cocok untuk proses pelatihan model pada *dataset smartphone* karena masing-masing memiliki nilai F1 terbesar yaitu 82,93.

Pengujian dalam *deep learning* dapat menggunakan 4 jenis model yaitu SIF, RNN, *Attention*, dan *Hybrid*. Model ini dikemukakan oleh Mudgal et al. (2018) yang selanjutnya diringkas oleh Chen et al. (2018) sebagai berikut.

- *SIF*: Model ini berperan dalam mengumpulkan informasi tentang kata-kata yang ada di setiap contoh data, dan kemudian membandingkannya.
- *RNN*: Model ini melakukan peringkasan dari urutan setiap *instance data*, kemudian dilakukan perbandingan. Secara intuitif, contoh data yang berisi urutan kata yang mirip dianggap cocok.
- *Attention*: Model ini melakukan penyesuaian antar kata di setiap contoh data, lalu melakukan perbandingan kata demi kata berdasarkan penyesuaian, dan terakhir menggabungkan informasi ini untuk melakukan klasifikasi. Secara intuitif, contoh data yang berisi kata-kata yang selaras dianggap cocok.
- *Hybrid*: Model ini melakukan penyesuaian antara urutan kata di setiap contoh data, kemudian membandingkan urutan kata yang menyesuaikan, dan terakhir menggabungkan informasi ini untuk melakukan klasifikasi. Secara intuitif, contoh data yang berisi perataan urutan kata dianggap cocok.

Penelitian ini bertujuan untuk mengetahui pencocokan antara dua buah *dataset* dengan penerapan *DeepMatcher* dalam *entity matching*. *Dataset* yang digunakan adalah data obat yang berasal dari farmaku.com dan k24klik.com. Dalam penerapan *DeepMatcher* menggunakan Jupyter Notebook, yaitu sebuah *notebook* komputasi *open source* populer yang memungkinkan pengguna dapat menggabungkan kode, visualisasi, dan teks dalam satu dokumen (file.ipynb) yang memiliki struktur dasarnya berupa JSON (Rule et al., 2018). Pada penelitian ini akan menggunakan salah satu model perbandingan, yaitu *Hybrid*. Pengujian menggunakan model ini akan menghasilkan seberapa akurat model ini dalam penerapan *DeepMatcher* terhadap *entity matching* di data obat dari farmaku.com dan k24klik.com.

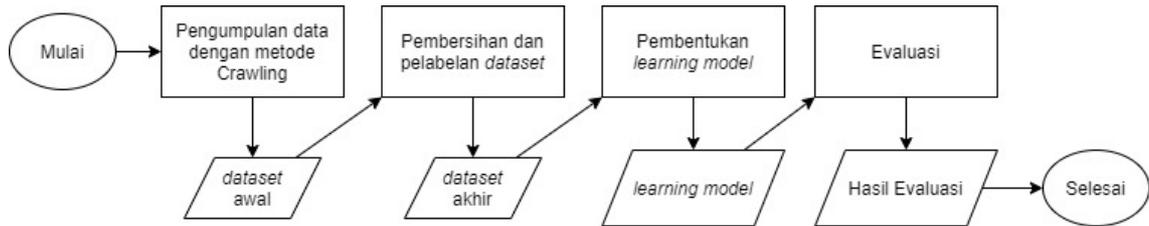
## 2. METODE PENELITIAN

Penelitian ini memiliki 4 proses utama yang dilakukan secara berurutan yaitu *data crawling*, pembersihan dan pelabelan data (*data cleansing and labeling*), pembentukan *learning model*, dan evaluasi. Gambar 1 menunjukkan proses metode penelitian dalam penelitian ini, pada masing-masing proses memiliki *output* yang diperlukan untuk menjalankan proses selanjutnya. Penjelasan yang lebih lengkap mengenai proses-proses tersebut akan dipaparkan pada bab ini.

### 2.1. Crawling

Dalam melakukan pengumpulan data, pada penelitian ini menggunakan teknik *crawling*. *Crawling* merupakan sebuah metode dalam melakukan pengumpulan data pada sebuah situs web dengan memasukkan URL (*Uniform Resource Locator*) yang menjadi acuan dalam melakukan pencarian (Arsi et al., 2021). *Dataset* yang digunakan pada penelitian ini dihimpun dengan teknik *web crawling* dari dua situs yang menjual obat secara *online* di Indonesia yaitu: farmaku.com sebagai S1, dan k24klik.com sebagai S2 (Akbar et al., 2021). Data obat yang dikumpulkan dibatasi hanya pada obat dengan kategori yang sama pada kedua situs yaitu “Batuk dan Pilek” dan “Obat Generik”. Proses *crawling* dilakukan pada bulan Januari, 2021. Hasil *crawling* tersebut disimpan ke dalam *file* berekstensi *comma-separated values* (CSV) untuk masing-masing sumber.





Gambar 1. Diagram Alir Metode Penelitian.

## 2.2. Pembersihan dan Pelabelan *Dataset*

*Dataset* menggunakan ekstensi CSV yang berupa data obat yang berhasil dihimpun kemudian dilakukan pembersihan dengan menghapus tanda koma (,) yang terdapat pada setiap baris agar tidak terjadi ambigu untuk nilai kolom. Kemudian, dilakukan juga penghapusan pada kolom-kolom yang tidak berisikan antar kedua sumber untuk dapat membandingkan kedua entitas secara setara. Sebagai contoh, seperti yang dapat dilihat pada Gambar 2, terdapat dua tabel yang memuat data obat dengan kolom yang berbeda, jika dilakukan penghapusan pada kolom yang tidak berisikan maka kolom "Dikirim Dari" pada Gambar 2A dan kolom "Diproduksi Oleh" pada Gambar 2B akan dihapus.

Nama	Dikirim Dari	Komposisi	Khasiat
Obat A	Jakarta	Parasetamol	Meredakan batuk
Obat B	Jakarta	Parasetamol	Meredakan demam
Obat C	Jakarta	Vitamin C	Meningkatkan imun

A

Nama	Komposisi	Manfaat	Diproduksi Oleh
Obat A	Parasetamol	Meredakan batuk	PT. XYZ
Obat B	Parasetamol	Meredakan demam	PT. XYZ
Obat C	Vitamin C	Meningkatkan imun	PT. XYZ

B

Gambar 2. Contoh Data Obat.

Setelah dilakukan penghapusan kolom, langkah selanjutnya adalah menyatukan data dari kedua sumber dengan cara memasang setiap baris data pada S1 dengan setiap baris data pada S2 untuk mencapai semua probabilitas yang ada. Sebagai contoh, jika kita memasang data pada Gambar 2A dan Gambar 2B maka hasil akan terlihat seperti Tabel 1.

Tabel 1. Contoh Hasil Kombinasi Antar Sumber.

No	Data A				Data B			
	Nama	Dikirim Dari	Komposisi	Khasiat	Nama	Komposisi	Manfaat	Diproduksi Oleh
1	Obat A	Jakarta	Parasetamol	Meredakan batuk	Obat A	Parasetamol	Meredakan batuk	PT. XYZ
2	Obat B	Jakarta	Parasetamol	Meredakan demam	Obat A	Parasetamol	Meredakan batuk	PT. XYZ
3	Obat C	Jakarta	Vitamin C	Meningkatkan imun	Obat A	Parasetamol	Meredakan batuk	PT. XYZ
4	Obat A	Jakarta	Parasetamol	Meredakan batuk	Obat B	Parasetamol	Meredakan demam	PT. XYZ
5	Obat B	Jakarta	Parasetamol	Meredakan demam	Obat B	Parasetamol	Meredakan demam	PT. XYZ
6	Obat C	Jakarta	Vitamin C	Meningkatkan imun	Obat B	Parasetamol	Meredakan demam	PT. XYZ
7	Obat A	Jakarta	Parasetamol	Meredakan batuk	Obat C	Vitamin C	Meningkatkan imun	PT. XYZ
8	Obat B	Jakarta	Parasetamol	Meredakan demam	Obat C	Vitamin C	Meningkatkan imun	PT. XYZ
9	Obat C	Jakarta	Vitamin C	Meningkatkan imun	Obat C	Vitamin C	Meningkatkan imun	PT. XYZ

Selanjutnya, pasangan data yang benar-benar berbeda akan dieliminasi (*blocking*) untuk mengurangi *learning* yang tidak diperlukan oleh *DeepMatcher*. Eliminasi baris pasangan data



pada penelitian ini dilakukan dengan menggunakan *package* Python dari AnHai's Group juga yang bernama *py\_entitymatching*. Pada *py\_entitymatching* terdapat pengaturan *overleap* yang bernilai desimal, yang dimaksudkan untuk jumlah kata yang sama pada kedua atribut untuk kriteria data yang dipertahankan. Sebagai contoh jika nilai *overleap* adalah 2 dan atribut yang dibandingkan adalah "Khasiat" dan "Manfaat, maka data nomor 1, 5, dan 9 pada Tabel 1 akan dipertahankan dan data nomor 2, 3, 4, 6, 7, dan 8 akan dihapus karena jumlah kata yang sama tidak mencapai dua. Nilai *overleap* bersifat minimal dalam artian jika jumlah kata yang sama melebihi nilai *overleap* maka baris tersebut tidak dieliminasi.

Baris-baris pasangan data yang tidak tereliminasi pada langkah sebelumnya akan menjadi *dataset* kandidat yang kemudian akan dipecah menjadi tiga *dataset* yang diperlukan oleh *DeepMatcher* yaitu *training* dan *validation* untuk pembentukan model dan *testing* untuk pengukuran akurasi model, pemecahan *dataset* kandidat dilakukan secara acak dengan rasio 3:1:1 sesuai dengan rekomendasi AnHai Doan (Mudgal, et al., 2018) di mana 3 untuk *dataset training* dan 1 untuk *dataset validation* dan *testing*.

Setiap baris pasangan data pada tiga *dataset* tersebut akan dilabeli secara manual dengan nilai 0 untuk pasangan data yang tidak serupa atau 1 untuk pasangan data yang serupa untuk digunakan oleh *DeepMatcher* sebagai acuan dalam pembentukan *learning model*.

### 2.3. Learning Model

*DeepMatcher* menyediakan empat jenis *deep learning* (Mudgal et al., 2018) yaitu Smooth Inverse Frequency (*sif*) yang mempertimbangkan kecocokan antar entitas berdasarkan nilai atribut tanpa memperhitungkan urutannya, Bidirectional RNN (*rnn*) yang mempertimbangkan urutan nilai atribut antar entitas, Decomposable Attention Model (*att*) yang mempertimbangkan keselarasan nilai setiap atribut antar entitas, dan Hybrid Model (*hybrid*) yang mempertimbangkan keselarasan urutan nilai atribut antar entitas.

Selain *dataset*, ada beberapa nilai yang perlu ditetapkan pada *DeepMatcher* untuk melakukan pemodelan, nilai tersebut adalah *epochs*, *batch\_size*, dan *pos\_neg\_ratio* yang semuanya ditetapkan dalam bentuk bilangan desimal. *Epoch* dimaksudkan untuk jumlah iterasi latihan yang dilakukan oleh *DeepMatcher*, *batch\_size* mengacu pada jumlah baris data yang akan dipelajari oleh *DeepMatcher* untuk setiap langkah pada satu latihan (*training step*), dan *pos\_neg\_ratio* adalah nilai bobot untuk *dataset training* yang ditentukan berdasarkan rasio jumlah baris berlabel 0 dan 1 pada *dataset* tersebut sebagai penyeimbang dalam pembentukan *learning model*. Semua parameter yang sudah ditentukan tadi akan mempengaruhi kinerja dan akurasi dari *learning model*.

Pemodelan pada tahap latihan didasarkan pada *dataset training* dengan kata lain *DeepMatcher* 'mempelajari' pelabelan yang dilakukan oleh penulis berdasarkan atribut yang terdapat pada *dataset* tersebut. *Dataset validation* digunakan untuk mengevaluasi hasil latihan sebagai acuan untuk menentukan *learning model* terbaik berdasarkan skor F1 dari *dataset validation* tersebut. *DeepMatcher* akan menyimpan *learning model* yang terbaik dari semua pelatihan yang telah dilakukan untuk nanti digunakan dalam melakukan prediksi kecocokan.

### 2.4. Evaluasi

$$prec = \frac{TP}{(TP+FP)} \quad (1)$$

$$rec = \frac{TP}{(TP+FN)} \quad (2)$$

$$F1 = \frac{2 \times prec \times recall}{(prec+recall)} \quad (3)$$

Akurasi dari *learning model* dievaluasi dengan pengukuran *precision* (*prec*), *recall* (*rec*), dan *F1-score* (*f-measure*) yang merupakan pengukuran populer dalam *machine learning* (Garreta &



Moncecchi, 2013). *Precision* atau ketepatan berkaitan dengan kemampuan sistem untuk tidak memanggil dokumen yang tidak relevan (Hardi, 2006). Secara sederhana *prec* dapat diartikan sebagai perbandingan antara prediksi positif yang tepat (TP)(*true positive*) dan semua prediksi positif yang dihasilkan mesin (TP + FP)(*false positive*). Persamaan untuk nilai *prec* dapat dilihat pada persamaan (1). Sementara itu, *recall* merupakan sebuah cara evaluasi yang sering dilakukan dalam mengukur hasil eksperimen dalam *machine learning* (Powers, 2020). Dalam persamaan *rec* dapat diartikan sebagai perbandingan antara prediksi positif yang tepat (TP) dan semua baris yang sebenarnya positif (TP + FN) (*false negative*). Persamaan untuk nilai *rec* dapat dilihat pada persamaan (2). Skor F1 dapat diartikan sebagai nilai *mean* dari *prec* dan *rec* [2], dengan persamaan yang dapat dilihat pada persamaan (3). Berdasarkan pengukuran tersebut maka dapat disimpulkan bahwa *learning model* yang paling akurat dalam artian dapat memprediksi kecocokan tanpa kesalahan adalah *learning model* yang memiliki nilai 1 untuk skor F1-nya atau 100 jika disajikan dalam bentuk persen.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Dataset

*Crawling* pada S1 didapatkan 198 baris untuk total dengan perincian 114 baris untuk kategori “Batuk Pilek” dan 84 baris untuk kategori “Obat Generik”, dan atribut yang didapatkan pada S1 adalah nama produk, URL produk, harga produk, satuan produk, deskripsi produk, indikasi penggunaan, kategori produk, komposisi produk, dan dosis penggunaan produk. *Crawling* pada S2 didapatkan 243 baris dengan perincian 97 baris untuk kategori “Batuk Pilek” dan 146 baris untuk kategori “Obat Generik”, dan atribut yang didapatkan pada S2 adalah nama produk, URL produk, harga produk, satuan produk, deskripsi produk, indikasi penggunaan, kategori produk, *tag*, komposisi produk, dan dosis penggunaan produk. Untuk rangkuman hasil *crawling* dapat dilihat pada Tabel 2.

Tabel 2. Rangkuman Hasil *Crawling*.

	S1	S2
Jumlah baris kategori Batuk Pilek	114	97
Jumlah baris kategori Obat Generik	84	146
Total baris	198	243
Atribut	nama produk	nama produk
	URL produk	URL produk
	harga produk	harga produk
	satuan produk	satuan produk
	deskripsi produk	deskripsi produk
	indikasi penggunaan	indikasi penggunaan
	kategori produk	kategori produk
	komposisi produk	<i>tag</i>
	dosis penggunaan produk	komposisi produk
		dosis penggunaan produk

Setelah itu, semua data yang berhasil dihimpun dilakukan pembersihan dengan menghapus semua tanda koma (,) pada semua nilai seperti yang sudah disebutkan pada sub-bab **Error! Reference source not found.** Kemudian penghapusan kolom yang tidak berisikan, pada hasil *crawling* atribut *tag* pada S2 dihapus karena tidak berisikan dengan S1. Setelah itu dilakukan kombinasi data antara kedua sumber yang menghasilkan  $198 \times 243 = 48.114$  baris yang kemudian dilakukan eliminasi pada baris yang benar-benar tidak mirip. Eliminasi dilakukan dalam dua tahap, tahap pertama, eliminasi dilakukan dengan membandingkan nilai atribut nama produk pada hasil kombinasi sebelumnya dengan nilai *overlap* adalah 2 yang menyisakan 566 baris pasangan data. Tahap kedua dilakukan dengan membandingkan nilai atribut satuan produk pada



hasil eliminasi tahap pertama dengan nilai *overleap* adalah 1 yang menyisakan 436 baris pasangan data. Hasil eliminasi tersebut menjadi *dataset* kandidat yang kemudian dipecah menjadi 3 *dataset* dengan rasio 3:1:1 sehingga hasil akhir pembentukan *dataset* menjadi 262 baris untuk *dataset training*, dan 87 baris untuk masing-masing *dataset validation* dan *testing*.

### 3.2. Inisialisasi DeepMatcher

Dalam menjalankan algoritma *DeepMatcher*, perlu adanya inisialisasi awal paket *DeepMatcher*. Inisialisasi awal dilakukan dengan mengunduh paket *DeepMatcher* ke modul Python. Setelah paket berhasil diunduh dan terinstal pada modul Python, maka selanjutnya adalah inisialisasi *library DeepMatcher* dengan memanggil *DeepMatcher*, *Torch*, dan OS.

### 3.3. Inisialisasi Data Training, Validation, Testing

Pada penelitian ini, *dataset* obat yang berasal dari situs K24klik dan Farmaku telah digabung menjadi satu dokumen berformat CSV. Proses penggabungan *dataset* mengikuti standar yang telah dibuat oleh algoritma *DeepMatcher*, yaitu menggabungkan kedua *dataset* menjadi *left table* (*ltable*) dan *right table* (*rtable*). Setelah dilakukan penggabungan *dataset*, maka selanjutnya adalah proses pelabelan. Proses pelabelan dilakukan secara manual ke seluruh *dataset*. Label pada algoritma *DeepMatcher* terbagi menjadi 2, yaitu *match* yang bernilai 1 dan *non-match* yang bernilai 0. Proses berikutnya adalah membagi *dataset* menjadi 3 data, yaitu *training*, *validation*, dan *testing*. Perbandingan antara data *training*, *validation*, dan *testing* adalah 3:1:1, sehingga didapatkan data *training* sejumlah 262, data *validation* sejumlah 87, serta data *testing* sejumlah 87. Data *training*, *validation*, dan *testing* yang digunakan pada penelitian ini dapat dilihat pada (Akbar et al., 2021). Proses yang terakhir adalah inisialisasi data *training*, *validation*, dan *testing* menggunakan fungsi *data.process* yang dimiliki oleh *library DeepMatcher*. Proses inisialisasi ini berfungsi untuk mengenali *dataset* yang digunakan, serta membersihkan *dataset* sesuai kebutuhan sistem.

### 3.4. Proses Training Dataset

Data yang digunakan dalam proses *training* adalah data *training* dan *validation*. Proses *training* pada penelitian ini menggunakan model *Hybrid* yang terdapat pada algoritma *DeepMatcher*. Konfigurasi proses *training* pada penelitian ini menggunakan nilai *epoch* sebesar 10 dan *batch\_size* sebesar 16. Konfigurasi ini dipilih karena mengingat keterbatasan sumber daya dan spesifikasi *hardware* yang digunakan. Serta, nilai *pos\_neg\_ratio* sebesar 17, yang didapatkan dengan membagi banyaknya label *non-match* dengan label *match* pada data *training* dan *validation*.

Proses *training dataset* pada penelitian ini dilakukan sebanyak 10 kali sesuai dengan nilai *epoch* yang diberikan. Proses *training data* menghasilkan nilai akurasi terbaik yang ditunjukkan dengan nilai *F1* sebesar 30,30, *precision* sebesar 17,86, dan *recall* sebesar 100. Dari hasil akurasi yang diperoleh, terlihat bahwa nilai *Precision* rendah. Hal ini dimungkinkan karena perbandingan label *match* dan *non-match* yang tidak seimbang, dimana data berlabel *match* jauh lebih sedikit dibandingkan dengan data berlabel *non-match*.

### 3.5. Prediksi Kecocokan Data

Dalam mendapatkan prediksi kecocokan data, perlu membuat sebuah *dataset* yang berisi data obat yang berasal dari kedua situs. *Dataset* ini juga dibuat dengan format yang disesuaikan dengan standar algoritma *DeepMatcher* tetapi tanpa menggunakan kolom Label. *Dataset* ini biasa disebut dengan data *unlabeled*. Data *unlabeled* yang digunakan pada penelitian ini berjumlah 10 dan dapat dilihat pada (Akbar et al., 2021). Data *unlabeled* ini juga perlu dilakukan proses inisialisasi dengan fungsi *data.process\_unlabeled* yang dimiliki oleh *library DeepMatcher*.

Setelah data *unlabeled* berhasil di inisialisasi, maka proses selanjutnya adalah melakukan proses prediksi kecocokan data menggunakan fungsi *model.run\_prediction*. Proses prediksi kecocokan data ini akan menghasilkan *match\_score* berdasarkan urutan pada data *unlabeled*. *Match\_score* ini yang dapat digunakan sebagai acuan apakah kedua data cocok atau tidak. Pada penelitian



ini, kedua data akan dikatakan cocok apabila *match\_score* lebih dari 0,9 karena dianggap paling mendekati nilai 1. Proses yang terakhir adalah menyimpan hasil prediksi kecocokan data ke dalam file dokumen berformat CSV dengan fungsi `predictions.to_csv`. Hasil prediksi kecocokan data secara lengkap dapat dilihat pada (Akbar et al., 2021).

Dari 10 data yang digunakan pada proses prediksi kecocokan data, terdapat 4 data yang dapat dikatakan cocok karena *match\_score* yang diperoleh lebih dari 0,9. Selanjutnya, keempat data ini digunakan sebagai data perbandingan harga obat yang dapat dilihat pada **Error! Reference source not found.**

No.	Farmaku		K24klik		Match Score
	Nama Obat	Harga	Nama Obat	Harga	
1	Decadryl Exp 120 ml	Rp 17.700	Decadryl Exp Syr 120ml	Rp 21.357	0,937635779
2	Mucera Drop	Rp 48.800	Mucera Drop 15mg/1ml 15ml	Rp 54.078	0,907262385
3	OBH Combi Flu Jahe 60 ml New	Rp 13.500	Obh Combi Dewasa Batuk Flu Jahe 60ml	Rp 13.867	0,977404475
4	Ciprofloxacin 500 mg Tab Hexp	Rp 600	Ciprofloxacin Hexpharm 500mg Tab 50s	Rp 1.326	0,92949605

Berdasarkan data perbandingan obat, dapat diketahui bahwa untuk produk obat pertama yang bernama Decadryl Exp 120 ml pada situs Farmaku memiliki harga lebih murah dibandingkan K24klik dengan harga Rp 17.700. Produk obat kedua bernama Mucera Drop pada situs Farmaku lebih murah dibandingkan K24klik dengan harga Rp 48.800. Produk obat ketiga bernama OBH Combi Flu Jahe 60ml pada situs Farmaku lebih murah dibandingkan K24klik dengan harga Rp 13.500. Serta, produk obat keempat bernama Ciprofloxacin Hexpharm 500 mg pada situs Farmaku lebih murah dibandingkan K24klik dengan harga Rp 600.

#### 4. KESIMPULAN

Penelitian ini mengimplementasikan *deep learning* untuk *entity matching* pada *dataset* obat yang diperoleh dari situs K24Klik dan Farmaku. Penelitian ini menggunakan *library DeepMatcher* untuk mengolah *dataset*. Proses *training* pada penelitian ini dilakukan sebanyak 10 kali menggunakan model *Hybrid*. Dari hasil *training*, diperoleh nilai akurasi terbaik yang ditunjukkan dengan nilai *F1* sebesar 30,30, *precision* sebesar 17,86, dan *recall* sebesar 100. Nilai *Precision* yang rendah dimungkinkan karena data berlabel *match* jauh lebih sedikit dibandingkan dengan data berlabel *non-match*. Namun, percobaan prediksi terhadap 10 data menunjukkan hasil yang sesuai. Sehingga, metode *entity matching* yang digunakan pada penelitian ini dapat dikatakan cukup baik dan dapat digunakan untuk kebutuhan lanjutan seperti membandingkan harga pada produk obat yang sama.

#### DAFTAR PUSTAKA

- Abdullah, S. M. S. A., Ameen, S. Y. A., M. Sadeeq, M. A., & Zeebaree, S. (2021). Multimodal Emotion Recognition using Deep Learning. *Journal of Applied Science and Technology Trends*, 2(02), 52–58. <https://doi.org/10.38094/jastt20291>
- Akbar, A., Fano, N. F., Pratama, R. P., Hidayat, R., & Rakhmawati, N. A. (2021). *Dataset Obat Untuk Penelitian Entity Matching*. <https://doi.org/10.5281/ZENODO.4445466>
- Arsi, P., Wahyudi, R., & Waluyo, R. (2021). Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 231–237. <https://doi.org/10.29207/resti.v5i2.2698>
- Chen, C., Golshan, B., Halevy, A., Tan, W., & Doan, A. (2018). BigGorilla: An Open-Source Ecosystem for Data Preparation and Integration. *IEEE Data Eng. Bull.*, 41, 10–22.
- Christophides, V., Efthymiou, V., & Stefanidis, K. (2015). Entity Resolution in the Web of Data. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 5(3), 1–122. <https://doi.org/10.2200/S00655ED1V01Y201507WBE013>



- Fu, C., Han, X., Sun, L., Chen, B., Zhang, W., Wu, S., & Kong, H. (2019). End-to-End Multi-Perspective Matching for Entity Resolution. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 4961–4967. <https://doi.org/10.24963/ijcai.2019/689>
- Garreta, R., & Moncecchi, G. (2013). *Learning Scikit-Learn: Machine Learning in Python*. Packt Publishing Ltd.
- Hardi, W. (2006). Mengukur kinerja search engine : sebuah eksperimentasi penilaian precision and recall untuk informasi ilmiah bidang ilmu perpustakaan dan informasi [Search Engines performance evaluation: an experimental the value of precision and recall for scientific information in LIS field.]. In *Visi Pustaka [National Library of Indonesia]*. Perpustakaan Nasional RI [National Library of Indonesia].
- Hidayat, R., Pratama, R. P., & Rakhmawati, N. A. (2021). ANALISIS ENTITY MATCHING PADA DATASET SMARTPHONE MENGGUNAKAN METODE SIF, RNN, ATTENTION, DAN HYBRID. *TEKNOSAINS: MEDIA INFORMASI SAINS DAN TEKNOLOGI*, 15(1), 67–77. <https://doi.org/10.24252/teknosains.v15i1.17583>
- Kasai, J., Qian, K., Gurajada, S., Li, Y., & Popa, L. (2019). Low-resource Deep Entity Resolution with Transfer and Active Learning. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5851–5861.
- Li, Y., Li, J., Suhara, Y., Doan, A., & Tan, W.-C. (2020). Deep entity matching with pre-trained language models. *Proceedings of the VLDB Endowment*, 14(1), 50–60. <https://doi.org/10.14778/3421424.3421431>
- Mudgal, S., Li, H., Rekatsinas, T., Doan, A., Park, Y., Krishnan, G., Deep, R., Arcaute, E., & Raghavendra, V. (2018). Deep Learning for Entity Matching. *Proceedings of the 2018 International Conference on Management of Data*, 19–34. <https://doi.org/10.1145/3183713.3196926>
- Powers, D. M. W. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *International Journal of Machine Learning Technology*, 2(1), 37–63.
- Rule, A., Tabard, A., & Hollan, J. D. (2018). Exploration and Explanation in Computational Notebooks. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3173574.3173606>
- Thirumuruganathan, S., Tang, N., Ouzzani, M., & Doan, A. (2020). Data Curation with Deep Learning. *Proceedings of the 23rd International Conference on Extending Database Technology (EDBT)*, 277–286. <https://doi.org/https://dx.doi.org/10.5441/002/edbt.2020.25>
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241, 111716. <https://doi.org/10.1016/j.rse.2020.111716>
- Zhao, C., & He, Y. (2019). Auto-EM: End-to-end Fuzzy Entity-Matching using Pre-trained Deep Models and Transfer Learning. *The World Wide Web Conference on - WWW '19*, 2413–2424. <https://doi.org/10.1145/3308558.3313578>



## Optimasi Keamanan Web Server terhadap Serangan Broken Authentication Menggunakan Teknologi Blockchain

Imam Riadi <sup>(1)</sup>, Herman <sup>(2)</sup>, Aulyah Zakilah Ifani <sup>(3)\*</sup>

<sup>1</sup> Sistem Informasi, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan, Yogyakarta

<sup>2,3</sup> Teknik Informatika, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan, Yogyakarta

e-mail : imam.riadi@is.uad.ac.id, hermankaha@mti.uad.ac.id,  
aulyah1908048022@webmail.uad.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 29 April 2021, direvisi 25 Juni 2021, diterima 11 Juli 2021, dan dipublikasikan 22 September 2021.

### Abstract

*The aspect of the internet that needs to be considered a security is the login system. The login system usually uses a username and password as an authentication method because it is easy to implement. However, data in the form of usernames and passwords are very vulnerable to theft, so it is necessary to increase the security of the login system. The purpose of this research is to investigate the security of the system. Whether the system is good at protecting user data or not, minimizing execution errors from the system and minimizing risk errors on the system so that the login system can be used safely. This research is conducted to test the system security with Burp Suite on the login system that has been built. Testing the security of this system by experimenting with POST data which is secured using blockchain technology makes the data sent in the form of hash blocks safer and more confidential so that the system is safer than before. Blockchain technology has successfully secured usernames and passwords from broken authentication attacks. By using the Burp Suite testing system, login is more specific in conducting security testing.*

**Keywords:** Authentication, Broken Authentication, Blockchain, Burp Suite, Login System

### Abstrak

Salah satu aspek di internet yang perlu diperhatikan keamanannya adalah sistem *login*. Sistem *login* biasanya menggunakan *username* dan *password* sebagai metode autentikasi karena mudah dalam mengimplementasikannya. Data *username* dan *password* sangat rentan diretas sehingga perlu dilakukan peningkatan keamanan pada sistem *login*. Penelitian bertujuan untuk mengetahui keamanan dari sistem dalam melindungi data pengguna, meminimalkan kesalahan eksekusi dari sistem serta mengurangi risiko *error* pada sistem, sehingga sistem *login* bisa digunakan secara aman. Penelitian ini dilakukan untuk menguji keamanan sistem dengan Burp Suite pada sistem *login* yang dibangun. Pengujian keamanan sistem ini dengan percobaan data POST yang diamankan menggunakan teknologi *blockchain* membuat data yang dikirimkan dalam bentuk blok *hash* menjadi lebih aman dan rahasia sehingga sistem lebih aman daripada sebelumnya. Teknologi *blockchain* berhasil mengamankan *username* dan *password* dari serangan *broken authentication*. Pengujian menggunakan Burp Suite pada sistem *login* lebih spesifik dalam melakukan pengujian keamanan.

**Kata Kunci:** Autentikasi, Broken Authentication, Burp Suite, Sistem Login, Teknologi Blockchain

## 1. PENDAHULUAN

Keamanan terhadap data, informasi, dan sistem secara keseluruhan semakin penting seiring berkembangnya teknologi informasi. Adanya sistem *login* merupakan salah satu aspek di internet yang perlu diperhatikan keamanannya (Ramadhan & Ariyani, 2018). Kebutuhan informasi dalam internet yang luas memberikan kemudahan dalam mengaksesnya. Data atau informasi menjadi sangat rentan terhadap pencurian sehingga perlu menjaga integritas data ataupun informasi. Dalam mengimplementasikan suatu web diperlukan proses *login*. Sistem keamanan dan proses



*login* biasanya menggunakan *username* dan *password* sebagai metode autentikasi. Hal ini digunakan karena kemudahan dalam mengimplementasikannya (Sudiarto Raharjo et al., 2017). Akan tetapi autentikasi menggunakan *username* dan *password* rentan terhadap peretasan. Terutama ketika *username* dan *password* disimpan dalam sebuah *database*. Hal ini terbukti ketika dilakukan *vulnerability assesment* menggunakan beberapa *tools* seperti *openvas*, Burp Suite dan *wireshark*. Kelemahan lainnya, ketika sistem autentikasi diretas, sulit bagi pengguna untuk mengetahuinya (A. W. P. Putra et al., 2018).

Autentikasi adalah suatu pembuktian identitas terhadap suatu entitas seperti pada mesin, kartu kredit, dan orang (Rusdan & Sabar, 2020). Autentikasi dibagi menjadi tiga kategori di antaranya yaitu: *What the entity knows* contoh berupa kata sandi, kedua *What the entity owns* seperti kartu pintar, kunci privasi atau tiket kerberos, dan ketiga *What the entity is* yang mencakup teknik autentikasi berdasarkan fitur *biometric* pengguna (sidik jari, bentuk wajah, bentuk tangan, dll.) (Sudiarto Raharjo et al., 2017). Salah satu teknologi inovasi yang mampu menyelesaikan permasalahan tersebut adalah teknologi *Blockchain*.

*Blockchain* merupakan sebuah teknologi dalam pertukaran informasi tanpa melibatkan pihak ketiga. Informasi berupa informasi dalam bentuk digital, entri data transaksi, aset (Dilley et al., 2016). *Blockchain* adalah teknologi dengan *database* terdistribusi disimpan dan dibagikan ke pengguna yang berwenang (Parizi et al., 2018) (Bouscaren, 1989). Paling tidak *blockchain* melibatkan tiga unsur teknologi yang sebetulnya sudah ada sejak lama, yaitu internet, protokol dari perangkat lunak, dan kriptografi (Fadlil et al., 2020). Teknologi *blockchain* membuat peretas akan sulit mengubah dan memodifikasi data yang sama di semua komputer di saat yang sama karena membutuhkan waktu yang sangat lama untuk memecahkan kode enkripsi pada setiap blok data di seluruh jaringan komputer (Riadi, Umar, & Busthomi, 2020).

Di balik teknologi *blockchain* terdapat 6 karakteristik utama yaitu *blockchain* adalah kriptografi yang di dalamnya tercatat proses enkripsi yang tingkat keamanan setiap transaksi tinggi. *Blockchain* adalah akuntansi yang di dalamnya tercatat tentang sebuah transaksi. *Blockchain* adalah rantai, ini dikarenakan *blockchain* terdiri dari kumpulan blok, di mana blok sebelumnya harus sama sehingga dapat menyambung seperti sebuah rantai. *Blockchain* merupakan catatan terdistribusi di mana data transaksi dalam *blockchain* tersimpan pada suatu buku besar terdistribusi (distribusi *ledger*) di seluruh *nodes* dan sulit dimanipulasi oleh *adversaries*. *Blockchain* adalah *mining* karna setiap seseorang berhasil melakukan validasi kebenaran transaksi maka akan mendapatkan sejumlah imbalan dalam bentuk *native coin* dari *blockchain* tersebut. *Blockchain* adalah *smart contract* karna selain menyimpan data dan transaksi, *blockchain* juga bisa mengeksekusi kontrak perjanjian yang telah disimpan sebelumnya (Rahardja et al., 2019).

Jenis *record* yang terdapat pada sistem *blockchain* yaitu blok dan transaksi. Setiap transaksi *blockchain* tersimpan dalam satu blok secara bersama. Setiap blok membentuk jaringan yang berisi algoritma kriptografi. Algoritma kriptografi digunakan untuk mengambil data dari blok sebelumnya dan diubah ke *Compact String* yang dapat mendeteksi sabotase (Fauzan, 2018). Setiap blok memiliki nilai *hash* yang didapatkan dari blok sebelumnya (Hu et al., 2019). Cara kerja fungsi *hash* yaitu panjang variabel diubah ke dalam bentuk biner. Setiap biner memiliki panjang yang sama. Fungsi *hash* digunakan dalam sistem keamanan, salah satu contohnya autentikasi pesan, penyimpanan *password*, dan tanda tangan digital (H. F. Putra et al., 2019). Untuk mengetahui seberapa besar tingkat keamanan dari suatu sistem penting dilakukan *penetration testing*, berupa simulasi terhadap serangan nyata yang mungkin menyerang sistem tersebut.

*Penetration testing* merupakan mengamati serangan dan menganalisis risiko terkait pelanggaran dari keamanan. Penguji tidak hanya dapat mengetahui keberadaan celah bagi *hacker*, tetapi juga dapat mengeksploitasi lebih jauh untuk mengevaluasi tingkat kerentanan sebuah sistem (Azis & Fattah, 2019). *Penetration testing* memerlukan analisa intensif pada kerentanan sistem yang di dapatkan dari kelemahan sistem. Seluruh data analisa yang telah dilakukan akan didokumentasikan dan serahkan ke *user* yang pemilik sistem juga dampak beserta solusi yang didapatkan penguji dari celah keamanan yang ada (Pangalila et al., 2015). Penelitian ini



menggunakan Burp Suite untuk menguji keamanan sistem. Burp Suite merupakan *tool* untuk melakukan keamanan *open source* yang digunakan untuk menjalankan dan menguji fitur keamanan pada sebuah aplikasi *website* (Sai Kiran et al., 2020). Burp Suite digunakan untuk menangkap aliran data dengan mengatur sebagai pendengar *proxy* yang bertindak sebagai *server proxy HTTP* lokal (Joshi & Kumar, 2016). Burp Suite secara keseluruhan merupakan kerangka pengujian untuk penetrasi web. Burp Suite terbagi menjadi dua yaitu ada *community edition* dan *professional edition*. Selain itu Burp Suite sebagai platform terintegrasi berbasis Java untuk melakukan pengujian keamanan suatu aplikasi web. Burp Suite awalnya hanya merupakan aplikasi *proxy server* untuk melakukan *intercept* baik terhadap *http-request* ataupun *http-response* ke server dan web *application* (T & Sasikala, 2019). Terdapat beberapa kerentanan yang menjadi risiko pada sebuah aplikasi web yaitu injeksi, *insecure direct object references*, *broken authentication and session management (XSS)*, *security misconfiguration*, *sensitive data exposure*, *cross site request forgery (CSRF)*, *cross site scripting (XSS)*, *unvalidated redirects and forwards*, *using components with known vulnerabilities*, *missing function level access control* (Guntoro et al., 2020).

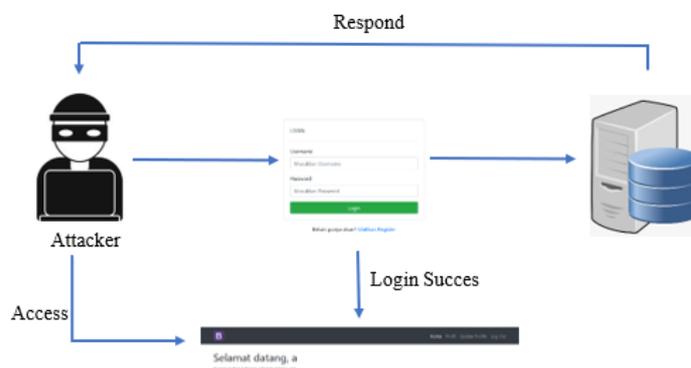
Kebanyakan dari pengguna menggunakan pengujian *vulnerability* untuk meningkatkan kesadaran tentang pentingnya keamanan informasi. Kerentanan (*vulnerability*) dari sebuah sistem disebabkan oleh faktor eksternal dan faktor internal (Wibowo et al., 2019). Dalam menguji *vulnerability* sistem dapat dilakukan dengan 2 tipe yaitu *external testing* dan *internal testing*. *External Testing* merupakan analisa terhadap informasi *public* yang tersedia. Untuk menampilkan jumlah *network access point* merupakan *internal testing* yang mewakili beberapa *logical* dan *physical segment* (Harjowinoto et al., 2016). metode dalam *vulnerability* yaitu pertama, *passive vulnerability testing* dengan melakukan pengujian terhadap kontrol *login*, konfigurasinya, dan kontrol *web application* sehingga dapat memetakan target sistem. kedua, *active vulnerability testing*, di mana pengujian dilakukan dengan memanipulasi input terhadap kerentanan yang ada, pengambilan hak akses. Ketiga, *aggressive vulnerability testing*, dilakukan *reverse engineering* terhadap *software* dan *system* (Harjowinoto et al., 2016).

Kerentanan (*vulnerability*) banyak diterapkan dan dilakukan contohnya pengujian terhadap web OJS versi 3.0, hasil pengujian didapatkan celah untuk melakukan serangan *Cross-site Scripting (XSS)* (Riadi, Yudhana, et al., 2020). Penelitian juga dilakukan pada sistem administrasi rumah sakit X (Harjowinoto et al., 2016), *security auditing* pada *vulnerable machine* (Sitinjak et al., 2020), dan juga pada pengujian celah keamanan pada CMS (Kunang et al., 2013). Selain menggunakan *OpenVAS* penelitian juga dilakukan pada aplikasi *smart payment* menggunakan *OWASP*, pengujian menggunakan *framework OWASP* menguji kerentanan terhadap serangan *XSS (Cross-site Scripting)* (Riadi, Umar, & Lestari, 2020). Ada beberapa jenis *vulnerability* di antaranya pemindaian berbasis jaringan, pemindaian berbasis *host*, pemindaian jaringan nirkabel, pemindaian aplikasi. Pemindaian berbasis jaringan untuk mengidentifikasi keamanan jaringan. Pemindaian berbasis *host* untuk mengidentifikasi kerentanan server, *host* jaringan lainnya. Pemindaian jaringan nirkabel lebih berfokus pada titik serangan dalam nirkabel. Pemindaian aplikasi untuk mendeteksi kerentanan perangkat lunak. Pemindaian basis data mengidentifikasi titik lemah dalam basis data (Laksmiati, 2020). Dalam penelitian ini menggunakan *broken authentication* sebagai serangan akan digunakan.

*Broken authentication* merupakan kerentanan web yang terjadi karna kesalahan konfigurasi manajemen *session*. Hal yang harus diperhatikan dari *broken authentication* yaitu pertama, *password strength*, di mana pada aplikasi yang kita bangun harus memiliki level minimal dari keamanan *password*, yang dapat dilihat pada panjang *password* dan kompleksitasnya. Kedua, *password use*, di mana aplikasi yang kita buat harus mempunyai batasan *user* mengaksesnya dalam tenggang waktu tertentu. Ketiga, *password storage*, di mana *password* yang kita miliki tidak boleh disimpan dalam aplikasi, dalam hal ini *password* harus ada dalam keadaan terenkripsi. Keempat, *issue* lainnya yang berhubungan contohnya *password* di dalam *source code* tidak boleh dalam bentuk *hard-coded*. Kelima, *Session ID Protection* hal ini digunakan biasanya untuk server mengidentifikasi *user* yang akan masuk ke dalam *session* menggunakan *session ID* (Hassan et al., 2018). *Broken authentication* memiliki fungsi untuk autentikasi dan manajemen *session* yang tidak dapat diterapkan dengan baik, memungkinkan



penyerang menyusup untuk mendapatkan *username* dan *password* dan mengeksploitasi kelemahan implementasi untuk mengasumsikan identitas pengguna lain secara permanen atau sementara (OWASP, 2017).



**Gambar 1. Skenario Broken Authentication.**

Gambar 1 menunjukkan *attacker* mengirimkan permintaan kredensial pengguna yang dihasilkan sampai sistem menemukan itu benar. Server memverifikasi kredensial pengguna dan membuat sesi yang kemudian disimpan dalam *database*. Setelah kredensial yang didapat ditebak cocok dengan *database*, sistem mengirimkan respons ke *attacker* atau penyerang dengan akses di akun.

Teknologi *blockchain* digunakan dalam penelitian ini untuk mengamankan data *username* dan *password* pengguna. Sebagaimana diketahui *blockchain* menggunakan fitur tanda tangan digital untuk melakukan transaksi sehingga data dari *user* tidak dapat diubah atau dirusak, hal ini menjadi kelebihan dari *blockchain*. Selain itu, *blockchain* menggunakan sistem terdesentralisasi yang dapat membuat transaksi lebih aman, cepat, dan lancar (Parizi et al., 2018). Sehingga penelitian ini menggunakan *blockchain* sebagai tujuan untuk meningkatkan keamanan *username* dan *password* dari suatu sistem *login*. Pengujian sistem menggunakan *tool* Burp Suite untuk memastikan sistem *login* aman digunakan.

## 2. METODE PENELITIAN

### 2.1. Objek Penelitian

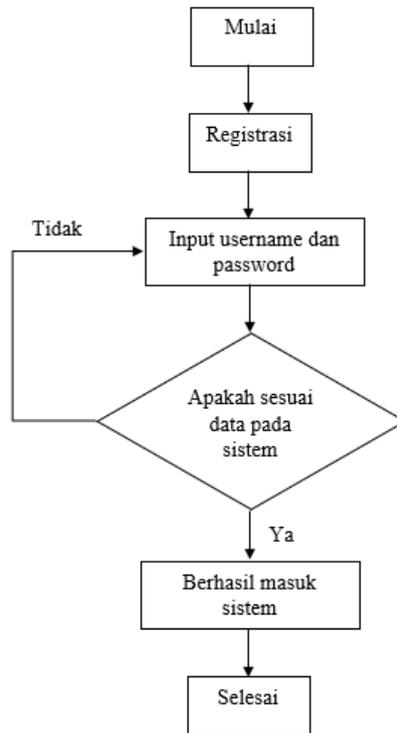
Sistem *login* adalah aplikasi yang digunakan sebagai objek pada penelitian ini. Aplikasi ini merupakan aplikasi atau *website* yang akan digunakan di berbagai aplikasi lainnya. Sistem *login* terdapat *username* dan *password* yang akan diisikan terlebih dahulu sebelum masuk ke sistemnya. Memasukkan *username* dan *password* menjadi sangat rentan terhadap peretasan. Bahnyaknya *tools* yang tersedia membuat data pengguna menjadi beragam. Hal ini tentunya akan sangat berbahaya apabila dibiarkan terus menerus. Oleh karena itu pada penelitian ini akan dilakukan optimasi keamanan. Dengan menggunakan teknologi *blockchain* diharapkan dapat digunakan sebagai pengamanan sistem *login*. Pengujian dalam penelitian ini menggunakan serangan *broken authentication*.

### 2.2. Desain Sistem

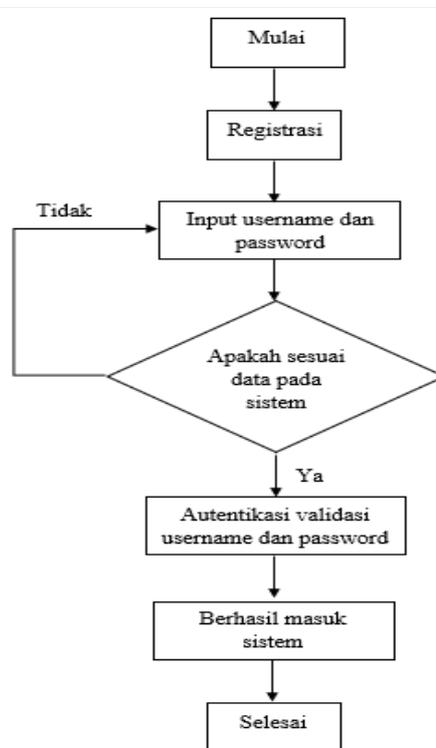
Desain *flowchart* sistem menggambarkan alur kerja sistem yang akan dirancang. Sistem mempunyai kemampuan untuk melakukan pengecekan data yang ada pada sistem *login*. Sistem mempunyai kemampuan untuk melakukan autentikasi validasi *username* dan *password*. Sebelum melakukan autentikasi sistem terlebih dahulu melakukan pengecekan data yang ada pada sistem *blockchain*. Tahapan *flowchart* sistem, pertama pengguna atau *user* akan melakukan *login* terlebih dahulu. Setelah *login* pengguna memasukan *username* dan *password*. Sistem akan mengecek apakah data ada pada sistem *login*, jika data yang dimasukkan sudah benar maka sistem akan melakukan autentikasi validasi *username* dan *password*, setelah semua



proses selesai maka pengguna akan masuk sistem. Berikut perancangan *flowchart* sistem dapat dilihat pada Gambar 2.



Gambar 2. *Flowchart* Sistem Sebelum Menggunakan *Blockchain*.



Gambar 3. *Flowchart* Sistem Setelah Menggunakan *Blockchain*.

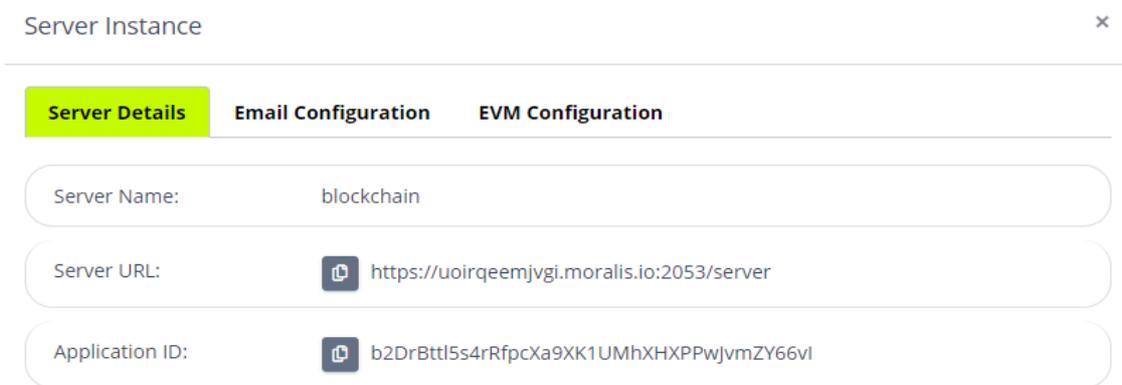


Gambar 2 dan Gambar 3 menunjukkan *flowchart* sistem di mana pengguna atau *user* akan melakukan *login* terlebih dahulu sebelum masuk ke sistem. Setelah memasukkan *username* dan *password* sistem akan melakukan autentikasi *username* dan *password*. Jika sudah sesuai maka proses *login* akan berhasil. Pada Gambar 2 setelah pengguna melakukan registrasi lalu *login* ke sistem, sistem akan mengecek dan berhasil masuk apabila data yang dimasukkan sesuai. Berbeda dengan *flowchart* yang ada pada gambar 3 yang menggunakan pengecekan autentikasi *username* dan *password* terlebih dahulu sebelum masuk ke dalam sistem, di mana Gambar 3 terdapat proses autentikasi dengan MetaMask. MetaMask digunakan sebagai jembatan antara sistem *login* dengan *blockchain* Ethereum. Sedangkan Gambar 2 tidak melalui proses tersebut sehingga sangat rentan terhadap peretasan.

### 3. HASIL DAN PEMBAHASAN

Sistem *login* pada penelitian ini memanfaatkan penggunaan *platform blockchain* Ethereum yang mengimplementasikan teknologi *blockchain* dan *smart contract*. Dengan *web3.js* sebagai *application programming interface* (API) untuk menghubungkan *browser* dengan ekstensi yang dinamakan MetaMask sebagai jembatan antara sistem *login* dengan *blockchain* Ethereum. MetaMask ini bertindak sebagai *wallet* Ethereum untuk pengelolaan informasi. Sementara itu untuk *smart contract* dibangun dengan menggunakan tools Visual Studio Code (VSCode). Pada sistem *login* juga menggunakan Moralis sebagai autentikasi ke MetaMask. Moralis mengambil informasi dari akun MetaMask. Untuk menghubungkan Moralis dengan MetaMask perlu adanya penyinkronan yang dilakukan di VSCode.

*User* harus menjadi anggota jaringan *blockchain* ketika jadi anggota maka *user* akan mempunyai *user ID* Ethereum. Sebuah aplikasi berbasis *blockchain* Ethereum harus menjalankan *smart contract*. Sehingga untuk *smart contract* akan ditanda tangani terlebih dahulu oleh orang-orang yang sudah mempunyai *ID* Ethereum untuk menjalankan sebuah aplikasi berbasis Ethereum. *User* tersebut selanjutnya jika ingin *login* maka harus sudah terdaftar sebagai penandatanganan *smart contract*. Berikut merupakan tampilan dari akun Moralis pada Gambar 4.



Gambar 4. Server URL dan *application ID* akun Moralis.

Gambar 4 merupakan server URL dan *application ID* pada akun Moralis yang akan disalin untuk menyinkronkan ke MetaMask. Link URL tersebut kemudian disalin dan dimasukkan ke dalam *source code* di VSCode. Berikut Gambar 5 merupakan URL untuk menyinkronkan akun MetaMask dan Moralis.

```
<script>
Moralis.initialize("b2DrBttl5s4rRfpcXa9XK1UMhXHPPwJvmZY66vI"); // Application id from moralis.io
Moralis.serverURL = "https://uoirqeemjvgi.moralis.io:2053/server"; //Server url from moralis.io
```

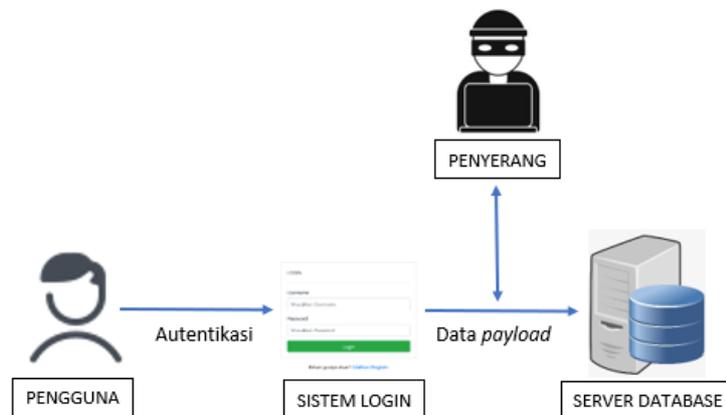
Gambar 5. *Source code* penyinkronan akun MetaMask dan Moralis.



Gambar 5 menunjukkan bagaimana menghubungkan antara akun MetaMask dengan Moralis. Menghubungkan akun tersebut terlebih dahulu masuk ke akun Moralis poc.moralis.io untuk mendapatkan *application* ID dan server URL tersebut. *Application ID from Moralis.io* digunakan untuk menghubungkan web3. Sedangkan server URL digunakan untuk menghubungkan Moralis. Ketika sistem sudah terhubung selanjutnya mengisi *username* dan *password* dan akan mendapatkan notifikasi dari MetaMask dan diminta untuk memberikan tanda tangan terlebih dahulu.

Hasil penerapan teknologi *blockchain* sebagai autentikasi pada sistem *login* mempunyai kemampuan untuk melakukan pengecekan data yang ada pada sistem *blockchain*. Komponen-komponen yang ada pada sistem meliputi menu registrasi, menu *login*, menu lihat profil, menu *update* profil dan *logout*. Halaman registrasi merupakan gambaran awal ketika pengguna belum memiliki akun untuk *login* yang digunakan untuk mendaftarkan *username* dan *password* sebelum masuk ke sistem. Menu *login* merupakan langkah pertama sebelum masuk ke sistem. Halaman ini dibuat untuk memberikan batasan kepada pihak yang tidak berkepentingan agar tidak dapat mengakses dan mengolah data tanpa melakukan *login* terlebih dahulu. Sehingga sebelum melakukan *login*, pengguna atau *user* harus melakukan registrasi terlebih dahulu untuk mendapatkan akun. *User* memasukkan *username* dan *password* dengan benar maka akan diarahkan ke menu awal dari sistem. Setelah *login* berhasil maka *user* dapat melihat profil dan melakukan *update* data dari *user*.

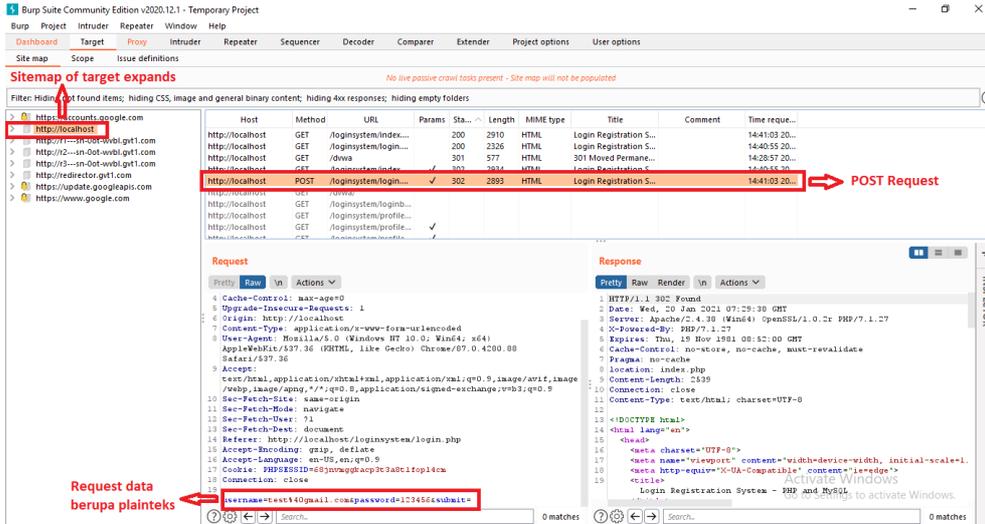
Pengujian dilakukan menggunakan aplikasi Burp Suite sebagai *software* untuk melakukan percobaan serangan *broken authentication* pada sistem *login*. Autentikasi yang diperlukan yaitu *username* dan *password* yang sudah terdaftar sebelumnya. Autentikasi dengan *username* dan *password* sangat mudah untuk mengimplementasikannya. Akan tetapi, menggunakan *username* dan *password* membuat para peretas mudah untuk melakukan serangan. Gambar 6 akan dipaparkan skenario yang akan dilakukan sebelum menggunakan teknologi *blockchain*.



**Gambar 6. Skenario keamanan sistem *login*.**

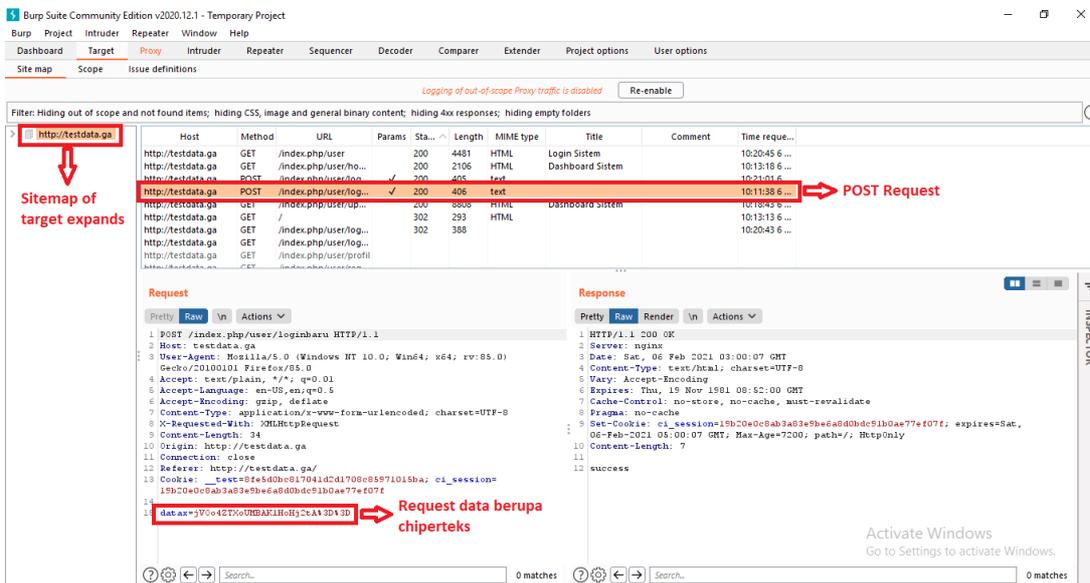
Gambar 6 menunjukkan mengenai skenario keamanan yang ada pada sistem. Pengguna akan menginputkan *username* dan *password* ke dalam sistem *login*. *Data payload* akan dikirimkan ke *database* server. Penambahan *blockchain* pada autentikasi sistem *login* membuat penyerang tidak dapat membaca isi *payload* data yang dikirimkan pengguna. Pengujian sistem menggunakan *tool* Burp Suite sebagai *software* untuk melakukan penetrasi pada halaman *login* untuk mencoba masuk ke dalam sistem. Penetrasi ini melakukan bantuan beberapa aplikasi dalam pengujiannya nanti. Pengujian ini difokuskan pada celah keamanan *login* dan autentikasi pada sistem. Percobaan data POST yang diamankan dengan teknologi *blockchain* membuat data yang dikirimkan dalam bentuk blok *hash*. Blok *hash* menjadi lebih aman dan rahasia, data POST dikirimkan ke *database* server. Pengujian sebelum menggunakan *blockchain* dapat dilihat pada Gambar 7.





Gambar 7. Aplikasi Burp Suite menampilkan *username* dan *password*.

Gambar 7 menunjukkan hasil pengujian setelah diimplementasikannya *blockchain* pada sistem *login*. Aplikasi Burp Suite menampilkan *username* dan *password* ketika masuk ke dalam sistem. Pengujian ini menggunakan *tool* Burp Suite untuk mencari kombinasi *username* dan *password* yang benar untuk masuk ke sistem. Setelah berhasil menangkap status dan proses *login* ini pada sistem bisa dilakukan penetrasi *password*. Hasil pengujian dapat dilihat pada Gambar 8. Pengujian percobaan serangan *broken authentication* pada *tool* Burp Suite menghasilkan data yang dikirimkan berupa blok *hash* atau terenkripsi, sehingga data yang ada dapat terjamin keamanannya dan terjaga rahasianya.



Gambar 8. Capture data setelah menggunakan *Blockchain*.

Gambar 8 menunjukkan hasil *capture* data setelah menggunakan *blockchain*. Data yang didapatkan berupa blok yang terenkripsi. *Username* dan *password* yang dimasukkan sebelumnya berhasil diubah. Serangan dengan menggunakan Burp Suite tidak dapat mendeteksi data dari pengguna setelah menggunakan teknologi *blockchain*.



#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, kesimpulan yang dapat diambil yaitu sistem *login* berhasil dibangun. Dengan menggunakan teknologi *blockchain*, data dari *user* terjaga kerahasiaannya. Pengujian dilakukan dengan menggunakan Burp Suite. Data yang dikirimkan berupa blok *hash* dan terenkripsi, sehingga data terjamin keamanannya. Pengujian keamanan sistem *login* ini berhasil dilakukan dengan tepat sehingga sistem lebih aman daripada sebelumnya. Burp Suite sebagai pengujian sistem *login* lebih spesifik dalam melakukan pengujian keamanan. Data *username* dan *password* diubah menjadi chiperteks sehingga penyerang tidak dapat mengetahui data dari pengguna.

#### DAFTAR PUSTAKA

- Azis, H., & Fattah, F. (2019). ANALISIS LAYANAN KEAMANAN SISTEM KARTU TRANSAKSI ELEKTRONIK MENGGUNAKAN METODE PENETRATION TESTING. *ILKOM Jurnal Ilmiah*, 11(2), 167–174. <https://doi.org/10.33096/ilkom.v11i2.447.167-174>
- Bouscaren, E. (1989). Elementary pairs of models. *Annals of Pure and Applied Logic*, 45(2), 129–137. [https://doi.org/10.1016/0168-0072\(89\)90057-2](https://doi.org/10.1016/0168-0072(89)90057-2)
- Dilley, J., Poelstra, A., Wilkins, J., Piekarska, M., Gorlick, B., & Friedenbach, M. (2016). *Strong Federations: An Interoperable Blockchain Solution to Centralized Third-Party Risks*.
- Fadlil, A., Riadi, I., & Nugrahantoro, A. (2020). Data Security for School Service Top-Up Transactions Based on AES Combination Blockchain Technology. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 11(3), 155. <https://doi.org/10.24843/LKJITI.2020.v11.i03.p04>
- Fauzan, N. I. (2018). TEKNOLOGI BLOCKCHAIN DAN PERANANNYA DALAM ERA DIGITAL. *Jurnal BJB University*, 4, 1–15.
- Guntoro, G., Costaner, L., & Musfawati, M. (2020). ANALISIS KEAMANAN WEB SERVER OPEN JOURNAL SYSTEM (OJS) MENGGUNAKAN METODE ISSAF DAN OWASP (STUDI KASUS OJS UNIVERSITAS LANCANG KUNING). *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 5(1), 45. <https://doi.org/10.29100/jupi.v5i1.1565>
- Harjowinoto, D., Noertjahyana, A., & Andjarwirawan, J. (2016). VULNERABILITY TESTING PADA SISTEM ADMINISTRASI RUMAH SAKIT X. *Jurnal Infra*, 4(1), 227–232.
- Hassan, M. M., Nipa, S. S., Akter, M., Haque, R., Deepa, F. N., Rahman, M. M., Siddiqui, M., & Sharif, M. H. (2018). Broken Authentication and Session Management Vulnerability: A Case Study of Web Application. *International Journal of Simulation: Systems, Science & Technology*, 1–11. <https://doi.org/10.5013/IJSSST.a.19.02.06>
- Hu, S. D. K., Palit, H. N., & Handojo, A. (2019). IMPLEMENTASI BLOCKCHAIN: STUDI KASUS E-VOTING. *Jurnal Infra*, 7(1), 183–189.
- Joshi, C., & Kumar, U. (2016). Security Testing and Assessment of Vulnerability Scanners in Quest of Current Information Security Landscape. *International Journal of Computer Applications*, 145(2), 1–7. <https://doi.org/10.5120/ijca2016910563>
- Kunang, Y. N., Muklis, F., & Sauda, S. (2013). PENGUJIAN CELAH KEAMANAN PADA CMS (CONTENT MANAGEMENT SYSTEM). *Prosiding Seminar Nasional Ilmu Komputer (SeNAIK 2013)*, 398–406.
- Laksmiati, D. (2020). VULNERABILITY ASSESSMENT PADA SITUS WWW.HATSEHAT.COM MENGGUNAKAN OPENVAS. *Jurnal Akrab Juara*, 5(3), 240–246.
- OWASP. (2017). *OWASP Top Ten Web Application Security Risks*. OWASP. <https://owasp.org/www-project-top-ten/>
- Pangalila, R., Noertjahyana, A., & Andjarwirawan, J. (2015). PENETRATION TESTING SERVER SISTEM INFORMASI MANAJEMEN DAN WEBSITE UNIVERSITAS KRISTEN PETRA. *Jurnal Infra*, 3(2), 271–276.
- Parizi, R. M., Dehghantaha, A., Choo, K.-K. R., & Singh, A. (2018). Empirical Vulnerability Analysis of Automated Smart Contracts Security Testing on Blockchains. *In Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering (CASCON18)*, 103–113. <https://doi.org/10.5555/3291291.3291303>
- Putra, A. W. P., Bhawiyuga, A., & Data, M. (2018). Implementasi Autentikasi JSON Web Token (JWT) Sebagai Mekanisme Autentikasi Protokol MQTT Pada Perangkat NodeMCU. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK)*, 2(2), 584–593.
- Putra, H. F., Wirawan, W., & Penangsang, O. (2019). Penerapan Blockchain dan Kriptografi untuk



- Keamanan Data pada Jaringan Smart Grid. *Jurnal Teknik ITS*, 8(1). <https://doi.org/10.12962/j23373539.v8i1.38525>
- Rahardja, U., Harahap, E. P., & Christianto, D. D. (2019). PENGARUH TEKNOLOGI BLOCKCHAIN TERHADAP TINGKAT KEASLIAN IJAZAH. *Technomedia Journal*, 4(2), 211–222. <https://doi.org/10.33050/tmj.v4i2.1107>
- Ramadhan, M. S., & Ariyani, F. (2018). PENINGKATAN KEAMANAN LOGIN WEBSITE DENGAN IMPLEMENTASI ONE TIME PASSWORD MENGGUNAKAN ALGORITMA SHA1 DAN MD5 BERBASIS MOBILE. *SKANIKA*, 1(2), 689–696.
- Riadi, I., Umar, R., & Busthomi, I. (2020). Optimasi Keamanan Autentikasi dari Man in the Middle Attack (MiTM) Menggunakan Teknologi Blockchain. *Journal of Information Engineering and Educational Technology*, 4(1), 15–19. <https://doi.org/http://dx.doi.org/10.26740/jjeet.v4n1.p15-19>
- Riadi, I., Umar, R., & Lestari, T. (2020). Analisis Kerentanan Serangan Cross Site Scripting (XSS) pada Aplikasi Smart Payment Menggunakan Framework OWASP. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 5(3), 146–152. <https://doi.org/10.14421/jiska.2020.53-02>
- Riadi, I., Yudhana, A., & W, Y. (2020). Analisis Keamanan Website Open Journal System Menggunakan Metode Vulnerability Assessment. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(4), 853. <https://doi.org/10.25126/jtiik.2020701928>
- Rusdan, M., & Sabar, M. (2020). Design and Analysis of Wireless Network with Wireless Distribution System using Multi-Factor Authentication-based User Authentication. *Journal of Information Technology*, 2(1), 17–24. <https://doi.org/10.47292/joint.v2i1.004>
- Sai Kiran, K. V. V. N. L., Devisetty, R. N. K., Kalyan, N. P., Mukundini, K., & Karthi, R. (2020). Building a Intrusion Detection System for IoT Environment using Machine Learning Techniques. *Procedia Computer Science*, 171(2019), 2372–2379. <https://doi.org/10.1016/j.procs.2020.04.257>
- Sitinjak, H. S. F., Hedyanto, U. Y. K. S., & Widjajarto, A. (2020). Security Auditing Pada Vulnerable Machine Menggunakan Open Source Ids Dan Vulnerability Scanner Berdasarkan Nist Cybersecurity Framework. *EProceedings of Engineering*, 7(2), 7638–7646.
- Sudiarto Raharjo, W., E.K. Ratri, I. D., & Susilo, H. (2017). IMPLEMENTASI TWO FACTOR AUTHENTICATION DAN PROTOKOL ZERO KNOWLEDGE PROOF PADA SISTEM LOGIN. *Jurnal Teknik Informatika Dan Sistem Informasi*, 3(1), 127–136. <https://doi.org/10.28932/jutisi.v3i1.579>
- T, G. S., & Sasikala, D. (2019). Vulnerability Assessment of Web Applications using Penetration Testing. *International Journal of Recent Technology and Engineering*, 8(4), 1552–1556. <https://doi.org/10.35940/ijrte.B2133.118419>
- Wibowo, F., Harjono, H., & Wicaksono, A. P. (2019). Uji Vulnerability pada Website Jurnal Ilmiah Universitas Muhammadiyah Purwokerto Menggunakan OpenVAS dan Acunetix WVS. *Jurnal Informatika*, 6(2), 212–217. <https://doi.org/10.31311/ji.v6i2.5925>



## Peramalan Pelayanan Service Mobil (After-Sale) Menggunakan Backpropagation Neural Network (BPNN)

Novianti Puspitasari <sup>(1)\*</sup>, Haviluddin <sup>(2)</sup>, Arinda Mulawardani Kustiawan <sup>(3)</sup>, Hario Jati Setyadi <sup>(4)</sup>, Gubtha Mahendra Putra <sup>(5)</sup>

<sup>1,2,3,5</sup> Informatika, Fakultas Teknik, Universitas Mulawarman, Samarinda

<sup>4</sup> Sistem Informasi, Fakultas Teknik, Universitas Mulawarman, Samarinda

e-mail : {novia.ftik.unmul,arinda.mulawardani}@gmail.com, haviluddin@unmul.ac.id, {hario.setyadi,gubthamp}@fkti.unmul.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 4 Juni 2021, direvisi 20 Juni 2021, diterima 28 Juni 2021, dan dipublikasikan 22 September 2021.

### Abstract

*The automotive industry in Indonesia, primarily cars, is getting more and more varied. Along with increasing the number of vehicles, Brand Holder Sole Agents (ATPM) compete to provide after-sale services (mobile service). However, the company has difficulty knowing the rate of growth in the number of mobile services handled, thus causing losses that impact sources of income. Therefore, we need a standard method in determining the forecasting of the number of car services in the following year. This study implements the Backpropagation Neural Network (BPNN) method in forecasting car service services (after-sale) and Mean Square Error (MSE) for the process of testing the accuracy of the forecasting results formed. The data used in this study is car service data (after-sale) for the last five years. The results show that the best architecture for forecasting after-sales services using BPNN is the 5-10-5-1 architectural model with a learning rate of 0.2 and the learning function of trainlm and MSE of 0.00045581. This proves that the BPNN method can predict mobile service (after-sale) services with good forecasting accuracy values.*

**Keywords:** After-Sale, Backpropagation Neural Network, Cars, MSE, Prediction

### Abstrak

Industri otomotif di Indonesia terutama mobil semakin banyak dan bervariasi. Seiring dengan meningkatkan jumlah mobil maka Agen Tunggal Pemegang Merk (ATPM) berlomba-lomba untuk memberikan pelayanan *after-sale* (*service* mobil). Namun, pihak perusahaan mengalami kesulitan dalam mengetahui laju pertumbuhan jumlah *service* mobil yang ditangani, sehingga memberikan kerugian yang berdampak pada sumber pendapatan. Oleh karena itu diperlukan sebuah metode baku dalam menentukan peramalan jumlah *service* mobil di tahun berikutnya. Penelitian ini mengimplementasikan metode *Backpropagation Neural Network* (BPNN) dalam peramalan pelayanan *service* mobil (*after-sale*) dan *Mean Square Error* (MSE) untuk metode pengujian akurasi hasil peramalan yang terbentuk. Adapun data yang digunakan pada penelitian ini adalah data pelayanan *service* mobil (*after-sale*) selama lima tahun terakhir. Hasil penelitian menunjukkan bahwa arsitektur terbaik untuk peramalan pelayanan *after-sale* menggunakan BPNN adalah model arsitektur 5-10-5-1 dengan *learning rate* sebesar 0,2 dan fungsi pembelajaran yaitu *trainlm* serta MSE sebesar 0,00045581. Hal ini membuktikan bahwa metode BPNN mampu memprediksi pelayanan *service* mobil (*after-sale*) dengan nilai akurasi peramalan yang baik.

**Kata Kunci:** After-Sale, Backpropagation Neural Network, Mobil, MSE, Peramalan

## 1. PENDAHULUAN

Industri otomotif di Indonesia terutama jumlah pertumbuhan mobil semakin meningkat dan bervariasi dipasaran. Menurut laporan Gabungan Industri Kendaraan Bermotor Indonesia (Gaikindo) jumlah penjualan mobil sebanyak 1.151.413 unit di akhir tahun 2018 (GAIKINDO, 2018). Agen Tunggal Pemegang Merk (ATPM) sebagai salah satu perusahaan yang memproduksi mobil berlomba-lomba dalam memberikan pelayanan *after-sale* kepada para pemilik mobil yang jumlahnya semakin banyak dipasaran. Pelayanan *after-sale* merupakan



pelayanan *service* mobil yang disediakan oleh bengkel resmi dan pengadaan suku cadang kendaraan *original*. Layanan purna jual termasuk perbaikan kendaraan (*service*) pada bengkel resmi merupakan salah satu indikator penting dalam sebuah perusahaan (Untari et al., 2018). Jika tingkat pelayanan *service* yang dihasilkan oleh sebuah perusahaan tersebut tinggi, maka perusahaan akan menghasilkan laba yang besar pula, sehingga perusahaan akan bertahan dalam persaingan bisnis (Solihat, 2019).

Berdasarkan data pelayanan *after-sale* ATPM dalam lima tahun terakhir, dari tahun 2014 sampai 2018 mengalami ketidakstabilan. Hal ini menyebabkan perusahaan mengalami kerugian yang berdampak pada sumber pendapatan perusahaan. Oleh karena itu, pentingnya untuk mengetahui hasil pelayanan *after-sale* di masa yang akan datang agar perusahaan dapat mengetahui tingkat pelayanan *service* yang dihasilkan telah sesuai dengan target atau tidak dalam rangka mencegah kerugian. Kesulitan perusahaan dalam mengetahui laju pertumbuhan jumlah *service* mobil di masa yang akan datang disebabkan belum adanya metode baku dalam menentukan peramalan jumlah penjualan pelayanan *after-sale* untuk tahun berikutnya. Peramalan atau prediksi telah banyak dilakukan menggunakan berbagai metode kecerdasan buatan. Di antaranya adalah metode *Naïve Bayes* untuk memprediksi pengunduran diri mahasiswa dengan menghitung sekumpulan probabilitas dengan tingkat akurasi sebesar 77,78% yang menunjukkan bahwa metode *Naïve Bayes* mampu memprediksi dengan baik namun perlu menambah atribut dan menggunakan *dataset* yang lebih banyak terutama data aktual (Mahanggara & Laksito, 2019). Prediksi peramalan kedatangan turis menggunakan metode *Average Based Fuzzy Time Series* yang memiliki nilai *Mean Absolute Percentage Error* (MAPE) sebesar 0,77375% (Widians et al., 2019), metode *Adaptive Neural Network Backpropagation* (ANNBP) untuk memprediksi tinggi muka air dengan hasil akurasi MSE dan MAPE sebesar 9,7% (Mislani et al., 2018), serta masih banyak lagi penelitian tentang peramalan menggunakan berbagai metode kecerdasan buatan (Alfajriani et al., 2020; Ardianto et al., 2018; Bisht & Kumar, 2016; Gadaleta et al., 2016; Minarni & Aldyanto, 2016; Puspitasari et al., 2019). Namun dari beberapa metode tersebut masih memiliki akurasi yang lebih kecil dibandingkan metode Jaringan Syaraf Tiruan (JST).

Metode JST merupakan metode yang sangat baik untuk mendapatkan hasil peramalan yang lebih baik (Aini et al., 2019; Haviluddin & Dengen, 2016; Ma et al., 2016; Majhi et al., 2014; Purnawansyah & Haviluddin, 2014; Sakinah et al., 2018; Simanungkalit et al., 2020; Syafiq et al., 2020). Model akurasi peramalan terbaik yang digunakan dalam JST adalah MSE (Haviluddin et al., 2016; Simanungkalit et al., 2020). Hal inilah yang menjadi dasar penelitian ini menerapkan metode *Backpropagation Neural Network* (BPNN) dalam meramalkan data pelayanan *service* mobil (*after-sale*) pada tahun berikutnya. Penelitian ini akan mengeksplorasi arsitektur jaringan dan laju pembelajaran (*learning rate*) yang merupakan parameter BPNN. Hasil penelitian ini diharapkan dapat membantu pihak perusahaan dalam mengambil kebijakan terkait pelayanan *service* mobil (*after-sale*) sehingga perusahaan dapat meningkatkan keuntungan.

## 2. METODE PENELITIAN

Pada bab ini akan dijelaskan tentang metode yang digunakan untuk peramalan menggunakan pelatihan yang terdapat pada metode BPNN, metode evaluasi terhadap akurasi dari hasil peramalan yang telah dilakukan, dan data peramalan yang digunakan.

### 2.1. *Backpropagation Neural Network* (BPNN)

*Backpropagation Neural Network* (BPNN) merupakan salah satu metode pelatihan dari JST yang banyak digunakan untuk memecahkan berbagai persoalan simulasi termasuk peramalan (Aini et al., 2019). Pelatihan BPNN terdiri dari tiga tahapan, yaitu umpan maju (*feed-forward*), umpan mundur (*backpropagation*) dan perubahan bobot. Adapun pelatihan BPNN adalah sebagai berikut.

- Tahap 0 : *Inisialisasi* bobot
- Tahap 1 : Jika kondisi terpenuhi, lakukan tahap 2-8
- Tahap 2 : Untuk setiap pasangan data, lakukan tahap 3-8



**Fase 1** : **Feed Forward**

**Tahap 3** : Setiap unit menerima *signal* masukan dan dikirim ke lapisan tersembunyi

**Tahap 4** : Hitung semua nilai dalam lapisan tersembunyi  $Z_j, (j = 1, 2, \dots, p)$ , kemudian setiap *signal* masukan dijumlahkan dengan persamaan (1).

$$z_{net_j} = v_{j0} + \sum_{i=1}^n x_i v_{kj} \quad (1)$$

Lebih lanjut hitung fungsi aktivasi untuk menghitung *signal* keluaran dan dikirim ke semua unit keluaran. Fungsi aktivasi yang biasanya dipakai adalah fungsi *sigmoid*, baik *biner* maupun *bipolar*. Fungsi *sigmoid biner* memiliki *range* antara 0 sampai dengan 1, untuk fungsi *sigmoid bipolar* memiliki *range* antara -1 sampai dengan 1. Fungsi aktivasi dihitung menggunakan persamaan (2).

$$z_j = f(z_{net_j}) = \frac{1}{1 + e^{-z_{net_j}}} \quad (2)$$

**Tahap 5** : Hitung semua nilai keluaran dalam semua jaringan *output*  $y_k, (k = 1, 2, \dots, m)$ , kemudian setiap *signal* masukan dijumlahkan dengan persamaan (3).

$$y_{net_k} = w_{k0} + \sum_{j=1}^p z_j w_{kj} \quad (3)$$

Hitung fungsi aktivasi untuk menghitung *signal* keluaran dan dikirimkan ke semua unit keluaran dengan persamaan (4).

$$y_k = f(y_{net_k}) = \frac{1}{1 + e^{-y_{net_k}}} \quad (4)$$

**Fase 2** : **Backpropagation**

**Tahap 6** : Hitung faktor  $\delta$  unit keluaran *error*  $y_k, (k = 1, 2, \dots, m)$ .

$$\delta_k = (t_k - y_k) f'(y_{net_k}) = (t_k - y_k) y_k (1 - y_k) \quad (5)$$

Di mana:  $t_k$  = target keluaran,  $\delta$  = unit keluaran yang digunakan untuk memperbarui bobot. Hitung perubahan bobot baru dan bias  $w_{kj}$ , dengan nilai laju pembelajaran  $\alpha$ .

$$\delta_{w_{ji}} = \alpha \delta_k \cdot z_j, (k = 1, 2, \dots, m; j = 0, 1, \dots, p) \quad (6)$$

**Tahap 7** : Hitung faktor  $\delta$  dalam unit lapisan tersembunyi berdasarkan nilai *error*  $Z_j, (j = 1, 2, \dots, p)$ .

$$\delta_{net_j} = \sum_{k=1}^m \delta_k w_{kj} \quad (7)$$

Hitung faktor  $\delta$  dalam unit lapisan tersembunyi.

$$\delta_j = \delta_{net_j} f'(z_{net_j}) = \delta_{net_j} \cdot z_j (1 - z_j) \quad (8)$$

Hitung nilai perubahan bobot dan bias  $v_{ji}$ .

$$\delta y_{ji} = \alpha \delta_k \cdot z_j, (k = 1, 2, \dots, p; j = 0, 1, \dots, n) \quad (9)$$

**Fase 3** : **Perubahan bobot**

**Tahap 8** : Hitung semua perubahan bobot dan bias yang mengarah ke unit keluaran.

$$w_{kj}(new) = w_{kj}(old) + \delta w_{ji}; \quad (k = 1, 2, \dots, p; j = 0, 1, \dots, n) \quad (10)$$



Hitung perubahan bobot dan bias di dalam lapisan tersembunyi dengan persamaan (11).

$$v_{kj}(new) = v_{kj}(old) + \delta v_{ji}; \quad (j = 1,2, \dots, p; w = 0,1, \dots, n) \quad (11)$$

## 2.2. Metode Evaluasi Model

Pengukuran suatu model peramalan dapat menggunakan beberapa metode statistik di antaranya *Mean Absolute Error* (MAE), *Mean Square Error* (MSE), *Root Mean Squared Error* (RMSE) dan *Normalize Mean Square Error* (NMSE). Tujuan dari pengukuran metode peramalan adalah mendapatkan nilai terbaik antaran nilai asli dan nilai hasil peramalan (Haviluddin et al., 2015; Huang et al., 2015). Metode untuk mengukur akurasi peramalan yang digunakan dalam penelitian ini adalah MSE (Haviluddin et al., 2016; Simanungkalit et al., 2020). Rumus MSE terlihat pada persamaan (12).

$$MSE = \frac{1}{M} \sum_{t=1}^M (x_t - \hat{x}_t)^2 \quad (12)$$

Di mana,  $x_t$  adalah nilai observasi data;  $x_t - \hat{x}_t$  adalah nilai hasil prediksi;  $M$  adalah nilai dari deret waktu.

## 2.3. Data Peramalan

Data yang digunakan dalam penelitian ini adalah data pelayanan *service* mobil (*after-sale*) yang diperoleh dari PT. Samekarindo Indah Samarinda, Kalimantan Timur sebanyak 60 (enam puluh) bulan, dimulai dari bulan Januari 2014 - Desember 2018. Data penelitian dapat dilihat pada Tabel 1.

**Tabel 1. Data After-Sale.**

Bulan/Tahun	2014	2015	2016	2017	2018
Januari	840 unit	860 unit	816 unit	783 unit	720 unit
Februari	792 unit	776 unit	808 unit	626 unit	631 unit
Maret	858 unit	838 unit	813 unit	699 unit	663 unit
April	852 unit	803 unit	789 unit	639 unit	640 unit
Mei	868 unit	781 unit	854 unit	691 unit	649 unit
Juni	853 unit	819 unit	875 unit	648 unit	582 unit
Juli	822 unit	903 unit	760 unit	696 unit	691 unit
Agustus	872 unit	824 unit	777 unit	650 unit	648 unit
September	847 unit	787 unit	743 unit	591 unit	644 unit
Oktober	831 unit	843 unit	761 unit	663 unit	674 unit
November	852 unit	794 unit	742 unit	625 unit	624 unit
Desember	855 unit	803 unit	796 unit	673 unit	708 unit

*Software* yang digunakan untuk melakukan analisis data dalam penelitian ini adalah MATLAB. Berdasarkan prinsip sistem cerdas maka data *after-sale* dilakukan normalisasi terlebih dahulu. Data dalam penelitian ini dinormalisasi dalam ordo [0,1] menggunakan rumus normalisasi seperti terlihat pada persamaan (13).

$$\bar{x} = \frac{0.8(x-x_{min})}{(x_{max}-x_{min})} + 0.1 \quad (13)$$

Di mana,  $\bar{x}$  adalah nilai data deret waktu normalisasi dari  $x$ ;  $x$  adalah nilai data deret waktu asli;  $x_{max}$  adalah nilai maksimum data deret waktu, dan  $x_{min}$  adalah nilai minimum data deret waktu.

Lebih lanjut, untuk mengembalikan ke data asli maka dilakukan proses denormalisasi. Data *after-sale* yang telah dinormalisasi dapat dilihat pada Tabel 2.



Tabel 2. Data *After-Sale* Setelah Dinormalisasi.

Bulan/Tahun	Training			Testing	
	2014	2015	2016	2017	2018
Januari	0,7429	0,7928	0,6831	0,6009	0,4439
Februari	0,6233	0,5834	0,6632	0,2096	0,2221
Maret	0,7878	0,7380	0,6757	0,3915	0,3018
April	0,7729	0,6507	0,6158	0,2420	0,2445
Mei	0,8127	0,5959	0,7778	0,3716	0,2669
Juni	0,7753	0,6906	0,8302	0,2644	0,1
Juli	0,6981	0,9	0,5436	0,3841	0,3716
Agustus	0,8227	0,7031	0,5859	0,2694	0,2644
September	0,7604	0,6109	0,5012	0,1224	0,2545
Oktober	0,7205	0,7504	0,5461	0,3018	0,3292
November	0,7728	0,6283	0,4987	0,2071	0,2046
Desember	0,7803	0,6507	0,6333	0,3267	0,4140

### 3. HASIL DAN PEMBAHASAN

Pada bagian ini menjelaskan tentang hasil pengujian metode BPNN yang diujikan pada data *after-sale*. Berdasarkan aturan pelatihan dalam JST, data *after-sale* sebanyak 60 data sampel dibagi menjadi dua bagian. Data pelatihan (*training*) sebanyak 48 sampel data, dan data pengujian (*testing*) sebanyak 12 data. Lebih lanjut, jumlah masukan (*input layer*) yang digunakan sebanyak lima neurons  $P = [p(t-5), p(t-4), p(t-3), p(t-2), p(t-1)]$  dan jumlah keluaran (*output layer*) sebanyak satu neuron [*output neurons T*].

Arsitektur metode BPNN yang digunakan dalam penelitian ini terdiri dari dua kategori, yaitu satu lapisan tersembunyi dan dua lapisan tersembunyi. Fungsi aktivasi yang digunakan; untuk satu lapisan tersembunyi; dari lapisan masukan ke lapisan tersembunyi adalah *tansig*, dan dari lapisan tersembunyi ke lapisan keluaran adalah *purelin*. Lebih lanjut, untuk dua lapisan tersembunyi; dari lapisan masukan ke lapisan tersembunyi adalah *tansig* dan *logsig*, dan dari lapisan tersembunyi ke lapisan keluaran adalah *purelin*. Algoritma pelatihan yang digunakan untuk semua arsitektur metode BPNN adalah *gradient descent (traingdm)*. Metode pengujian akurasi peramalan adalah MSE yang digunakan untuk membandingkan nilai asli dan nilai hasil peramalan. Nilai MSE yang mendekati 1 merupakan nilai hasil pelatihan metode BPNN dengan arsitektur yang terbaik. Model arsitektur BPNN yang diuji terlihat pada Tabel 3.

Tabel 3. Model Arsitektur BPNN.

Variabel	Nilai
Arsitektur Hidden Layer	3, 3, 3, 2, 2, 2 5-5-5-5-1; 5-5-6-6-1; 5-5-10-10-1; 5-10-5-1; 5-10-7-1; 5-10-9-1 trainlm; traingd; traingdx
Fungsi Pembelajaran	logsig; purelin
Fungsi Aktivasi	0,2 0,3 0,9

Tahap selanjutnya adalah melakukan pengujian model arsitektur yang telah ditentukan. Pengujian ini dilakukan dengan menggunakan metode *try-and-error*. Hal ini berarti nilai akhir setiap kali pengujian disimpan dan menjadi perbandingan model sampai mendapatkan model arsitektur yang baik untuk prediksi. Hasil pengujian model arsitektur BPNN dapat dilihat pada Tabel 4.

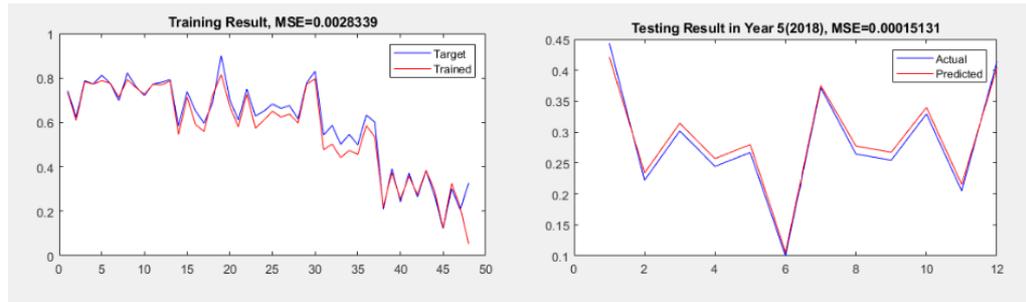


Tabel 4. Hasil Pengujian Model Arsitektur BPNN.

No.	Parameter		MSE		
	Model Arsitektur	Fungsi Pembelajaran	LR	Pelatihan	Pengujian
1		TRAINLM	0,2	0,0021487	0,00053394
2		TRAINGD	0,2	0,0095134	0,0042284
3		TRAINGD	0,2	0,011162	0,0076683
4		TRAINLM	0,3	0,0037026	0,01697
5	5-5-5-1	TRAINGD	0,3	0,0085026	0,0001958
6		TRAINGD	0,3	0,011132	0,00064377
7		TRAINLM	0,9	0,0028339	0,00015131
8		TRAINGD	0,9	0,0095473	0,0085697
9		TRAINGD	0,9	0,0061053	0,010013
10		TRAINLM	0,2	0,0038534	0,0026907
11		TRAINGD	0,2	0,014971	0,0046498
12		TRAINGD	0,2	0,008699	0,0086819
13		TRAINLM	0,3	0,0051347	0,0082093
14	5-5-6-6-1	TRAINGD	0,3	0,010365	0,003274
15		TRAINGD	0,3	0,011298	0,0004512
16		TRAINLM	0,9	0,0070278	0,0011997
17		TRAINGD	0,9	0,0086084	0,003397
18		TRAINGD	0,9	0,0097336	0,0058092
19		TRAINLM	0,2	0,0088314	0,0010997
20		TRAINGD	0,2	0,0094109	0,0032566
21		TRAINGD	0,2	0,0078561	0,0027335
22		TRAINLM	0,3	0,0080935	0,0033326
23	5-5-10-10-1	TRAINGD	0,3	0,009594	0,0013082
24		TRAINGD	0,3	0,010393	0,0038013
25		TRAINLM	0,9	0,0056065	0,0085379
26		TRAINGD	0,9	0,018702	0,0024079
27		TRAINGD	0,9	0,0093755	0,0088095
28		TRAINLM	0,2	0,0016632	0,00025718
29		TRAINGD	0,2	0,009171	0,00094251
30		TRAINGD	0,2	0,017845	0,00067779
31		TRAINLM	0,3	0,00083401	0,0016793
32	5-10-5-1	TRAINGD	0,3	0,00794141	0,0027773
33		TRAINGD	0,3	0,0092636	0,0070622
34		TRAINLM	0,9	0,0032714	0,0048179
35		TRAINGD	0,9	0,0099905	0,0041116
36		TRAINGD	0,9	0,031785	0,00064535
37		TRAINLM	0,2	0,0094055	0,0053769
38		TRAINGD	0,2	0,0093389	0,008574
39		TRAINGD	0,2	0,0070697	0,009047
40		TRAINLM	0,3	0,0081332	0,0018103
41	5-10-7-1	TRAINGD	0,3	0,0085616	0,0006072
42		TRAINGD	0,3	0,008826	0,0024631
43		TRAINLM	0,9	0,0014591	0,0036185
44		TRAINGD	0,9	0,0076877	0,0007095
45		TRAINGD	0,9	0,010232	0,0044151
46		TRAINLM	0,2	0,0040588	0,0010082
47		TRAINGD	0,2	0,0080645	0,0097648
48		TRAINGD	0,2	0,021152	0,0004501
49		TRAINLM	0,3	0,0068399	0,002719
50	5-10-9-1	TRAINGD	0,3	0,0068869	0,007943
51		TRAINGD	0,3	0,0088565	0,0012978
52		TRAINLM	0,9	0,006209	0,0024579
53		TRAINGD	0,9	-	-
54		TRAINGD	0,9	0,013497	0,0042891

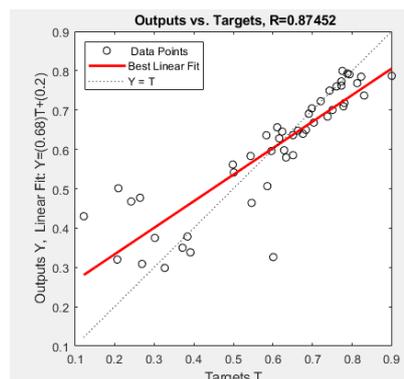
Plot grafik hasil pelatihan dan pengujian dapat dilihat pada Gambar 1. Dari Gambar 1 terlihat bahwa nilai MSE hasil pelatihan yang didapatkan sebesar 0,0028339. Hal ini menunjukkan bahwa hasil pelatihan hampir mendekati dengan target yang diinginkan.





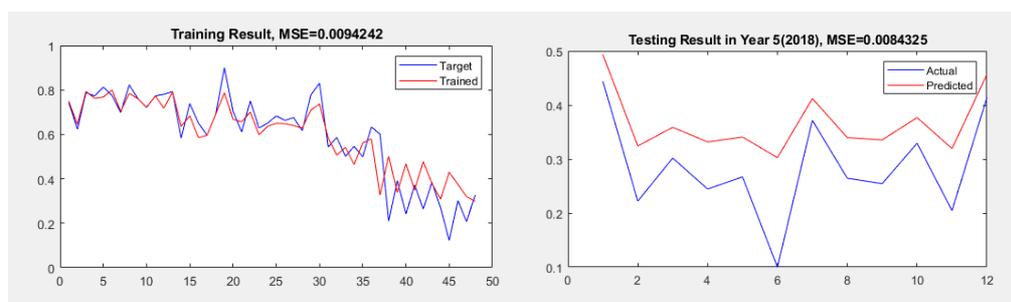
**Gambar 1. Plot Grafik Hasil Pelatihan dan Pengujian Model Arsitektur.**

Lebih lanjut, plot grafik hasil pengujian menghasilkan nilai MSE sebesar 0,00015131. Hal ini menunjukkan bahwa hasil pengujian hampir mendekati dengan target yang diinginkan. Tahap selanjutnya adalah memilih model terbaik dari hasil uji coba untuk meramalkan pelayanan *service* mobil (*after-sale*) pada tahun berikutnya (tahun 2019). Berdasarkan hasil uji coba menggunakan *Learning Rate* (LR) 0,3 dengan 5 inputan, 3 *hidden layer* dengan masing-masing *neuron* yaitu 5-5-5 dan 1 *output*, *epoch* maksimal sebesar 1000, *epoch* berhenti pada iterasi ke lima puluh dengan waktu eksekusi selama kurang dari satu detik didapatkan analisis pasca pelatihan (*post-training analysis*) seperti pada Gambar 2. yang menghasilkan koefisien kecocokan sebesar 0,68 dari nilai aktual dan nilai prediksi. Sedangkan, intersep kesesuaian dari nilai aktual dan nilai prediksi sebesar 0,2. Koefisien korelasi antara nilai aktual dan nilai prediksi didapatkan sebesar 0,87452.



**Gambar 2. Post Training Analysis Jaringan BPNN (*learning rate* 0,3).**

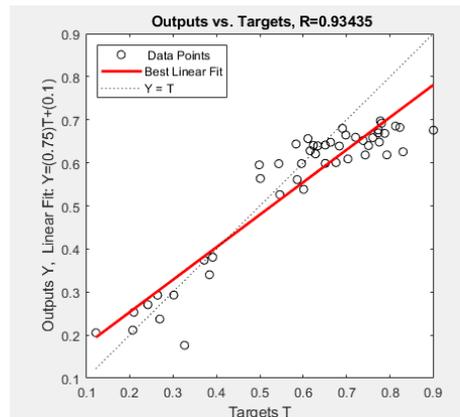
Dari Gambar 2 terdapat garis putus-putus yang merupakan target dan garis merah yang merupakan hasil prediksi dengan sedikit kesamaan dikarenakan koefisien relasinya kurang mendekati angka 1 yang berarti hasil prediksi kurang mendekati target. Plot grafik hasil pelatihan dan pengujian jaringan menggunakan arsitektur terbaik pada pengujian jaringan BPNN dengan LR 0,3 ditunjukkan oleh Gambar 3. dengan nilai MSE sebesar 0,0094242 untuk hasil pelatihan dan 0,0084325 untuk hasil pengujian.



**Gambar 3. Plot Grafik Hasil Pelatihan dan Pengujian Jaringan BPNN (*learning rate* 0,3).**

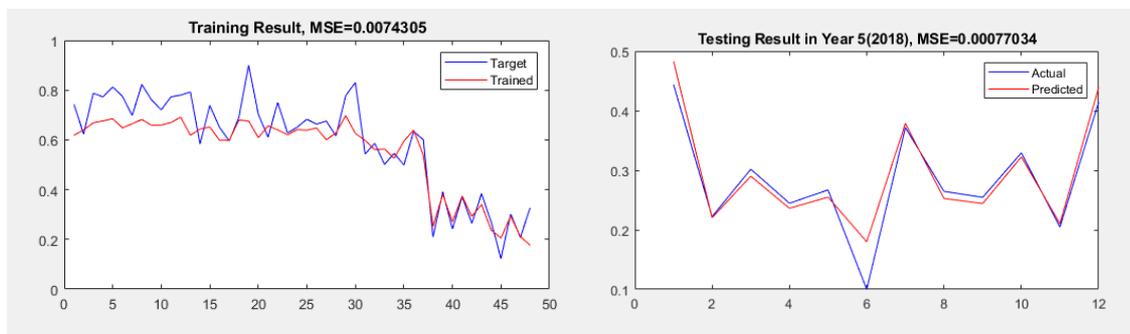


Uji coba berikutnya menggunakan LR 0,9 dengan 5 inputan, 3 *hidden layer* dengan masing-masing *neuron* yaitu 5-5-5 dan 1 *output*, serta *epoch* maksimal sebesar 1000, membuat *epoch* berhenti pada iterasi ke satu dengan waktu eksekusi selama kurang dari satu detik. Analisis pasca pelatihan (*post-training analysis*) dengan LR 0,9 menghasilkan koefisien kecocokan sebesar 0,75 dari nilai aktual dan nilai prediksi seperti yang terlihat pada Gambar 4. Intersep kesesuaian dari nilai aktual dan nilai prediksi sebesar 0,1. Koefisien korelasi antara nilai aktual dan nilai prediksi sebesar 0,93435. Dilihat dari Gambar 4 target yang digambarkan sebagai garis putus-putus dan hasil prediksi yang digambarkan dengan garis berwarna merah, memiliki sedikit perbedaan dan dikarenakan koefisien korelasinya mendekati 1 sehingga hasil prediksi cukup baik.



**Gambar 4. Post Training Analysis Jaringan BPNN (learning rate 0,9).**

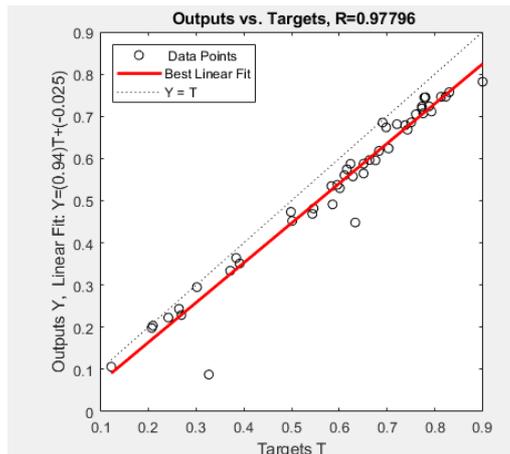
Plot grafik hasil pelatihan jaringan menggunakan arsitektur terbaik pada pengujian jaringan BPNN dengan LR 0,9 ditunjukkan oleh Gambar 5. Plot grafik hasil pelatihan dan pengujian menggunakan arsitektur terbaik didapatkan nilai MSE pelatihan sebesar 0,0074305 dan nilai MSE pengujian sebesar 0,00077034.



**Gambar 5. Plot Grafik Hasil Pelatihan dan Pengujian Jaringan BPNN (learning rate 0,9).**

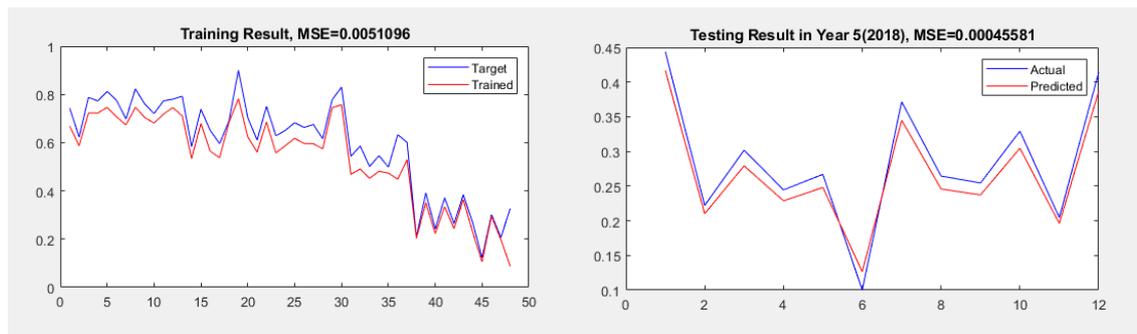
Lebih lanjut, hasil uji coba menggunakan LR 0,2 dengan 5 inputan, 2 *hidden layer* dengan masing-masing *neuron* yaitu 10-5 dan 1 *output*, *epoch* maksimal sebesar 1000 membuat *epoch* berhenti pada iterasi ke dua dengan waktu eksekusi selama kurang dari satu detik. Analisis pasca pelatihan (*post-training analysis*) dengan LR 0,2 menghasilkan koefisien kecocokan sebesar 0,94 dari nilai aktual dan nilai prediksi seperti yang terlihat pada Gambar 6. Intersep kesesuaian dari nilai aktual dan nilai prediksi didapatkan sebesar 0,025. Koefisien korelasi antara nilai aktual dan nilai prediksi sebesar 0,97796.





**Gambar 6. Post Training Analysis Jaringan BPNN (*learning rate 0,2*).**

Dilihat dari Gambar 6, target yang digambarkan sebagai garis putus-putus dan hasil prediksi yang digambarkan dengan garis berwarna merah, memiliki sedikit perbedaan dan dikarenakan koefisien korelasinya mendekati 1 sehingga hasil prediksi dengan baik. Plot grafik hasil pelatihan jaringan menggunakan arsitektur terbaik pada pengujian jaringan BPNN dengan LR 0,2 ditunjukkan oleh Gambar 7. Pada Gambar 7, plot grafik hasil pelatihan menggunakan arsitektur terbaik mendapatkan nilai MSE sebesar 0,0051096 dan 0,00045581 untuk hasil pengujian.



**Gambar 7. Plot Grafik Hasil Pelatihan dan Pengujian Jaringan BPNN (*learning rate 0,2*).**

Dari hasil pelatihan model pada data *after-sale* yang ditunjukkan oleh Tabel 5, memperlihatkan arsitektur terbaik pada pengujian variabel dengan menggunakan LR 0,3 didapatkan nilai MSE sebesar 0,0084325; LR 0,9 didapatkan MSE 0,00077034 dan untuk LR 0,2 didapatkan nilai MSE 0,00045581. Dalam percobaan ini, arsitektur terbaik untuk memprediksi pelayanan *service* mobil (*after-sale*) tahun 2019 yaitu 5-10-5-1; *trainlm*; LR 0,2 dengan nilai MSE sebesar 0,00045581.

**Tabel 5. Nilai MSE Peramalan.**

Model Arsitektur	Fungsi Pembelajaran	Learning Rate	MSE (Pengujian)
5-5-5-1	TRAINGD	0,3	0,0084325
5-5-5-1	TRAINILM	0,9	0,00077034
5-10-5-1	TRAINLM	0,2	0,00045581

Selanjutnya, Tabel 6 menampilkan hasil prediksi pelayanan *service* mobil (*after-sale*) pada tahun 2018. Lebih lanjut dilakukan perbandingan antara data aktual dan data hasil prediksi. Dari hasil perbandingan data aktual diperoleh total *error* sebesar 0,005469693 dan rata-rata *error* yang diperoleh adalah 0,000455808.



Tabel 6. Hasil Prediksi Berdasarkan Data Aktual.

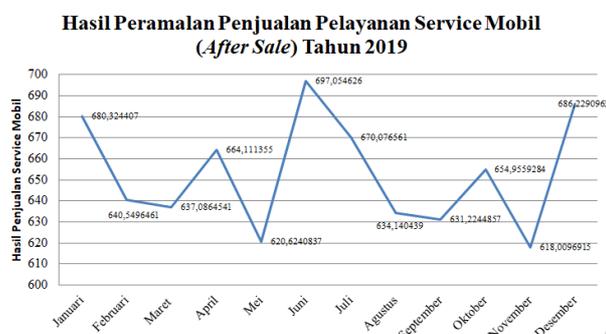
Bulan	Prediksi 2018 ( $Y'$ )	Prediksi 2018 ( $Y$ )	Aktual 2018 ( $X'$ )	Aktual 2018 ( $X$ )	Selisih ( $X' - Y'$ )	Error ( $(X' - Y')^2$ )
Januari	0,416768303	709	0,443925234	720	0,02715693	0,000737499
Februari	0,210174189	626	0,22211838	631	0,011944191	0,000142664
Maret	0,27929194	653	0,301869159	663	0,022577219	0,000509731
April	0,228784639	633	0,244548287	640	0,015763648	0,000248493
Mei	0,248094925	641	0,266978193	649	0,018883268	0,000356578
Juni	0,126513711	592	0,1	582	0,026513711	0,000702977
Juli	0,345045251	680	0,37165109	691	0,026605839	0,000707871
Agustus	0,245918121	640	0,264485981	648	0,01856786	0,000344765
September	0,237287113	637	0,254517134	644	0,017230021	0,000296874
Oktober	0,304638891	664	0,329283489	674	0,024644598	0,000607356
November	0,196259399	620	0,204672897	624	0,008413499	0,000070786
Desember	0,386740501	697	0,414018692	708	0,027278191	0,0007441
<b>Total</b>	3.225516982	-	3.418068536	-	0,192551554	0,005469693
<b>Rata-Rata</b>	0,268793082	-	0,284839045	-	0,01604596	0,000455808

Hasil peramalan pelayanan *service* mobil (*after-sale*) pada tahun 2019 dapat dilihat pada Tabel 7 dan plot hasil peramalan terlihat pada Gambar 8.

Tabel 7. Hasil Peramalan *After-Sale* Tahun 2019.

Bulan	Jumlah Mobil (Unit)
Januari	680
Februari	640
Maret	637
April	664
Mei	620
Juni	697
Juli	670
Agustus	634
September	631
Oktober	654
November	618
Desember	686

Gambar 8 menunjukkan bahwa pelayanan *service* mobil (*after-sale*) pada bulan Februari dan Maret akan mengalami penurunan dan pada bulan April akan mengalami peningkatan, lalu pada bulan Mei akan terjadi penurunan terendah. Pada bulan Juni akan mengalami peningkatan sampai dengan bulan Juli, di bulan Agustus dan September akan mengalami penurunan, lalu dibulan Oktober akan mengalami kenaikan sedikit, dan kembali menurun di bulan November, lalu pada bulan Desember akan mengalami kenaikan kembali.

Gambar 8. Grafik Hasil Peramalan Data *After-Sale* tahun 2019

Kenaikan pelayanan *service* mobil (*after-sale*) terjadi apabila memasuki bulan liburan, karena saat memasuki liburan banyak konsumen yang melakukan *service*, sehingga pelayanan *service* mobil (*after-sale*) menjadi meningkat. Faktor penurunan terjadi karena faktor ekonomi, jika pertumbuhan ekonomi menurun, maka mengakibatkan penurunan daya beli karena konsumen lebih mementingkan kebutuhan primernya.

#### 4. KESIMPULAN

Prediksi pelayanan *service* mobil (*after-sale*) dengan menggunakan metode *Backpropagation Neural Network* (BPNN) telah dilakukan. Hasil peramalan menunjukkan bahwa pelayanan *service* mobil (*after-sale*) terbesar terdapat di bulan Juni 2019 sebesar 697 unit mobil dan jumlah hasil peramalan terendah pada bulan Mei 2019 yaitu sebesar 620 unit mobil. Berdasarkan hasil percobaan, metode BPNN dengan parameter arsitektur 5-10-1; fungsi pembelajaran adalah *trainlm*; fungsi aktivasi adalah *logsig* dan *purelin*; laju pembelajaran adalah 0,2 mampu menghasilkan tingkat kesalahan prediksi yang baik dengan nilai MSE sebesar 0,000455808. Hal ini menunjukkan bahwa metode BPNN dapat menjadi alternatif metode dalam memprediksi pelayanan *service* mobil (*after-sale*) dengan data yang berjenis *time series*.

#### DAFTAR PUSTAKA

- Aini, H., Haviluddin, H., Budiman, E., Wati, M., & Puspitasari, N. (2019). Prediksi Produksi Minyak Kelapa Sawit Menggunakan Metode Backpropagation Neural Network. *Sains, Aplikasi, Komputasi Dan Teknologi Informasi*, 1(1), 24. <https://doi.org/10.30872/jsakti.v1i1.2261>
- Alfajriani, A., Wati, M., & Puspitasari, N. (2020). Penerapan Metode Fuzzy Time Series Chen dan Hsu dalam Memprediksi Kunjungan Wisatawan di Museum Mulawarman. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, 4(2), 144–153. <https://doi.org/http://dx.doi.org/10.30872/jurti.v4i2.5802>
- Ardianto, C., Haryanto, H., & Mulyanto, E. (2018). Prediksi Tingkat Kerawanan Kebakaran di Daerah Kudus Menggunakan Fuzzy Tsukamoto. *Creative Information Technology Journal*, 4(3), 186. <https://doi.org/10.24076/citec.2017v4i3.109>
- Bisht, K., & Kumar, S. (2016). Fuzzy time series forecasting method based on hesitant fuzzy sets. *Expert Systems with Applications*, 64, 557–568. <https://doi.org/10.1016/j.eswa.2016.07.044>
- Gadaleta, D., Manganelli, S., Manganaro, A., Porta, N., & Benfenati, E. (2016). A knowledge-based expert rule system for predicting mutagenicity (Ames test) of aromatic amines and azo compounds. *Toxicology*, 370, 20–30. <https://doi.org/10.1016/j.tox.2016.09.008>
- GAIKINDO. (2018). *Domestic Auto Production By Category 2018*. Gabungan Industri Kendaraan Bermotor Indonesia (GAIKINDO). <https://www.gaikindo.or.id/indonesian-automobile-industry-data/>
- Haviluddin, ., Alfred, R., Obit, J. H., Ahmad Hijazi, M. H., & Ag Ibrahim, A. A. (2015). A Performance Comparison of Statistical and Machine Learning Techniques in Learning Time Series Data. *Advanced Science Letters*, 21(10), 3037–3041. <https://doi.org/10.1166/asl.2015.6490>
- Haviluddin, & Dengen, N. (2016). Comparison of SARIMA, NARX and BPNN models in forecasting time series data of network traffic. *2016 2nd International Conference on Science in Information Technology (ICSITech)*, 264–269. <https://doi.org/10.1109/ICSITech.2016.7852645>
- Haviluddin, H., Arifin, Z., Kridalaksana, A. H., & Cahyadi, D. (2016). Prediksi Kedatangan Turis Asing ke Indonesia Menggunakan Backpropagation Neural Networks. *Jurnal Teknologi Dan Sistem Komputer*, 4(4), 485. <https://doi.org/10.14710/jtsiskom.4.4.2016.485-490>
- Huang, H.-X., Li, J.-C., & Xiao, C.-L. (2015). A proposed iteration optimization approach integrating backpropagation neural network with genetic algorithm. *Expert Systems with Applications*, 42(1), 146–155. <https://doi.org/10.1016/j.eswa.2014.07.039>
- Ma, E., Liu, Y., Li, J., & Chen, S. (2016). Anticipating Chinese tourists arrivals in Australia: A time series analysis. *Tourism Management Perspectives*, 17, 50–58. <https://doi.org/10.1016/j.tmp.2015.12.004>
- Mahanggara, A., & Laksito, A. D. (2019). PREDIKSI PENGUNDURAN DIRI MAHASISWA UNIVERSITAS AMIKOM YOGYAKARTA MENGGUNAKAN METODE NAIVE BAYES. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 10(1), 273–280.



<https://doi.org/10.24176/simet.v10i1.2967>

- Majhi, B., Rout, M., & Baghel, V. (2014). On the development and performance evaluation of a multiobjective GA-based RBF adaptive model for the prediction of stock indices. *Journal of King Saud University - Computer and Information Sciences*, 26(3), 319–331. <https://doi.org/10.1016/j.jksuci.2013.12.005>
- Minarni, M., & Aldyanto, F. (2016). Prediksi Jumlah Produksi Roti Menggunakan Metode Logika Fuzzy (Studi Kasus: Roti Malabar Bakery). *Jurnal TEKNOIF*, 4(2), 59–65.
- Mislan, Gaffar, A. F. O., Haviluddin, & Puspitasari, N. (2018). Water Level Prediction of Lake Cascade Mahakam Using Adaptive Neural Network Backpropagation (ANNBP). *IOP Conference Series: Earth and Environmental Science*, 144(1), 012009. <https://doi.org/10.1088/1755-1315/144/1/012009>
- Purnawansyah, & Haviluddin. (2014). Comparing performance of Backpropagation and RBF neural network models for predicting daily network traffic. *2014 Makassar International Conference on Electrical Engineering and Informatics (MICEEI)*, 166–169. <https://doi.org/10.1109/MICEEI.2014.7067332>
- Puspitasari, N., Tejawati, A., & Prakoso, F. (2019). Estimasi Stok Penerimaan Bahan Bakar Minyak Menggunakan Metode Fuzzy Tsukamoto. *JRST (Jurnal Riset Sains Dan Teknologi)*, 3(1), 9. <https://doi.org/10.30595/jrst.v3i1.3112>
- Sakinah, N. P., Cholissodin, I., & Widodo, A. W. (2018). Prediksi Jumlah Permintaan Koran Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK)*, 2(7), 2612–2618.
- Simanungkalit, J. R., Haviluddin, H., Pakpahan, H. S., Puspitasari, N., & Wati, M. (2020). Algoritma Backpropagation Neural Network dalam Memprediksi Harga Komoditi Tanaman Karet. *ILKOM Jurnal Ilmiah*, 12(1), 32–38. <https://doi.org/10.33096/ilkom.v12i1.521.32-38>
- Solihat, A. (2019). Layanan Purna Jual pada Produk Otomotif. *Business Innovation and Entrepreneurship Journal*, 1(1), 6–10. <https://doi.org/10.35899/biej.v1i1.5>
- Syafiq, M., Hartama, D., Kirana, I. O., Gunawan, I., & Wanto, A. (2020). Prediksi Jumlah Penjualan Produk di PT Ramayana Pematangsiantar Menggunakan Metode JST Backpropagation. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 175. <https://doi.org/10.30865/jurikom.v7i1.1963>
- Untari, S. N., Djaja, S., & Widodo, J. (2018). STRATEGI PEMASARAN MOBIL MEREK DAIHATSU PADA DEALER DAIHATSU JEMBER. *JURNAL PENDIDIKAN EKONOMI: Jurnal Ilmiah Ilmu Pendidikan, Ilmu Ekonomi Dan Ilmu Sosial*, 11(2), 82. <https://doi.org/10.19184/jpe.v11i2.6451>
- Widians, J. A., Puspitasari, N., & Sari, A. F. A. (2019). The Prediction Of Tourist Visiting With Average Based Fuzzy Time Series Method. *International Journal of Engineering and Advanced Technology*, 8(5C), 1467–1469. <https://doi.org/10.35940/ijeat.E1215.0585C19>



## Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma *Naïve Bayes*

Ami Natuzzuhriyah <sup>(1)\*</sup>, Nisa Nafisah <sup>(2)</sup>, Rini Mayasari <sup>(3)</sup>

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa, Karawang  
e-mail : ami.natuzzuhriyah17051@student.unsika.ac.id, nisa.nafisah17162@unsika.ac.id,  
rini.mayasari@staff.unsika.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 25 Maret 2021, direvisi 9 Juni 2021, diterima 9 Juni 2021, dan dipublikasikan 22 September 2021.

### Abstract

*Since the spread of Covid-19 in Indonesia, in early March 2020, the activities of Educational Institutions have not been disrupted. As conventional learning. Learning at Singaperbangsa University began with regulation from the Ministry of Education and Culture of the Republic of Indonesia, from learning that boldly affects concentration, influences concentration, such as signals, learning atmosphere, and teaching methods, so that factors affect the level of student satisfaction in learning. This study aims to determine the level of student satisfaction with learning who dares to use the Bayes naive algorithm using RapidMiner tools with results obtained with an accuracy rate of 76.92%, class precision of 100.00%, class recall 57.14%, and an AUC value of 0.881 or close to, so the resulting model is good. In other words, the results obtained using the Naïve Bayes algorithm can be used as material for making decisions about the level of online learning satisfaction.*

**Keywords: Online Learning, CRISP-DM, Naïve Bayes, Classification, Data Mining**

### Abstrak

Semenjak penyebaran Covid-19 di Indonesia, pada awal Maret 2020 menyebabkan aktivitas lembaga pendidikan terganggu. Sebagai gantinya kegiatan pembelajaran konvensional menjadi daring. Pembelajaran daring di Universitas Singaperbangsa dimulai sejak adanya peraturan dari Kemdikbud RI, dari pembelajaran secara daring tersebut mempengaruhi konsentrasi, kendala seperti signal, suasana pembelajaran dan cara mengajar, sehingga faktor mempengaruhi tingkat kepuasan mahasiswa dalam pembelajaran. Penelitian ini bertujuan untuk mengetahui tingkat kepuasan mahasiswa terhadap pembelajaran daring menggunakan algoritma *Naïve Bayes* dengan menggunakan *tools* RapidMiner dengan hasil yang didapatkan yaitu tingkat akurasi sebesar 76,92%, *class precision* 100.00%, *class recall* 57.14% serta nilai AUC 0.881 atau mendekati angka 1, jadi model yang dihasilkan baik. Dengan kata lain, dari hasil yang didapatkan menggunakan algoritma *Naïve Bayes* dapat digunakan sebagai bahan untuk pengambilan keputusan tingkat kepuasan pembelajaran secara daring.

**Kata Kunci: Pembelajaran Daring, CRISP-DM, Naïve Bayes, Klasifikasi, Data Mining**

## 1. PENDAHULUAN

Semenjak adanya *coronavirus (severe acute respiratory syndrome coronavirus 2)* yaitu penyakit karena infeksi *virus* yang biasanya disebut dengan *Covid-19* merupakan virus yang menyerang ke sistem pernapasan manusia dan dapat menular ke siapa saja seperti anak-anak, golongan usia lanjut atau lansia, ibu hamil, ibu menyusui bahkan bayi dengan masing-masing gejala yang didapatkan berbeda seperti gejala ringan, sedang dan pemulihan atau karantina tanpa perlu ke rumah sakit. Definisi lainnya *coronavirus* merupakan penyakit menular yang disebabkan oleh sindrom pernapasan akut *coronavirus 2 (Sars-CoV-2)*. Penyakit ini pertama kali ditemukan pada Desember 2019 di Wuhan, China dan sejak itu menyebar secara global, mengakibatkan pandemi *coronavirus 2019-2020* (Supriatna, 2020).

Pandemi ini memberikan dampak untuk seluruh kehidupan manusia pada aktivitas yang sering dilaksanakan pada masa sebelum wabah ini seperti pendidikan yang awalnya berinteraksi dengan guru dilakukan secara tatap muka sedangkan karena adanya pandemi ini pendidikan



sangat terganggu sehingga pemerintah harus membuat dan mengambil keputusan untuk menutup sekolah untuk sementara dengan pembelajaran dilakukan secara daring. Pembelajaran daring ini merupakan sistem belajar yang terbuka dan tersebar dengan menggunakan perangkat pedagogi (alat bantu pendidikan), yang dimungkinkan melalui internet dan teknologi berbasis jaringan untuk memfasilitasi pembentukan proses belajar dan pengetahuan melalui aksi dan interaksi yang berarti (Pakpahan & Fitriani, 2020). Penutupan sementara lembaga pendidikan untuk mencegah penyebaran virus ini berdampak pada pelajar, di mana proses dalam pembelajaran berlangsung mempengaruhi psikologis anak didik dan menurunnya kualitas keterampilan murid dengan pelaksanaan pembelajaran berlangsung secara *online* atau daring yang skalanya belum pernah terukur dan teruji sebelumnya (Aji, 2020). Sehingga, dampak yang dialami dari tingkat kepuasan maupun pemahaman pada materi yang telah disampaikan yaitu berdampak pada nilai, pengetahuan, kerja sama dan keterampilan yang didapatkan oleh pelajar di lembaga pendidikan. Seperti, perguruan tinggi, SMA/SMK, SMP bahkan sampai dengan sekolah dasar.

Untuk memutus rantai penyebaran *Covid-19* perguruan tinggi negeri di Karawang yaitu Universitas Singaperbangsa Karawang melaksanakan pembelajaran di rumah dengan jarak jauh atau daring dengan berinteraksi antara dosen pengampu dan mahasiswa dilakukan melalui *Google Classroom*, *Whatsapp*, *Zoom*, *Google Meet* dan media lainnya yang mendukung pelajaran. Dalam proses belajar tersebut tingkat kepuasan dan pemahaman mahasiswa terhadap mata kuliah sangat penting untuk menghasilkan mahasiswa yang berkualitas dari materi berkualitas dari materi atau pembelajaran yang disampaikan. Menurut Adijaya (2018) pembelajaran daring merupakan paradigma baru dalam proses pembelajaran yang dapat dilakukan dengan cara yang sangat mudah tanpa harus bertatap muka di suatu ruang kelas dan hanya mengandalkan sebuah aplikasi berbasis koneksi internet maka proses pembelajaran dapat berlangsung. Pembelajaran daring adalah sebuah jenis proses pembelajaran yang mengandalkan koneksi internet untuk mengadakan proses pembelajaran dengan jarak jauh. Pada penelitian sebelumnya telah dilakukan oleh (Siddik et al., 2020) menggunakan algoritma *Naïve Bayes* untuk klasifikasi kepuasan mahasiswa terhadap pelayanan perguruan tinggi dengan hasil menunjukkan dari 213 mahasiswa hasil dari pengujian klasifikasi menggunakan *naïve bayes* yang dilakukan *accuracy* 96,24%, *precision* 93,14 %, dan *recall* 98,96%. Penelitian lain yang dilakukan oleh (Silvia Siltonga et al., 2019) yaitu prediksi tingkat pemahaman mahasiswa terhadap mata kuliah berdasarkan posisi duduk dengan menggunakan algoritma *Naïve Bayes* hasil pengujian menunjukkan akurasi yang didapatkan sebesar 88.24% yaitu dengan 8 responden menyatakan ketidakpahaman sedangkan 60 responden lainnya menyatakan paham terhadap tingkat pemahaman mahasiswa terhadap mata kuliah. Berdasarkan uraian tersebut, penelitian ini menggunakan metode klasifikasi yaitu algoritma *Naïve Bayes* untuk menghasilkan tingkat kepuasan dari mahasiswa terhadap mata kuliah serta tujuan yang diharapkan dari hasil ini dapat memberikan masukan serta perbaikan dalam pembelajaran yang diberikan kepada mahasiswa dalam menerima materi.

## 2. METODE PENELITIAN

### 2.1. Pengumpulan Data

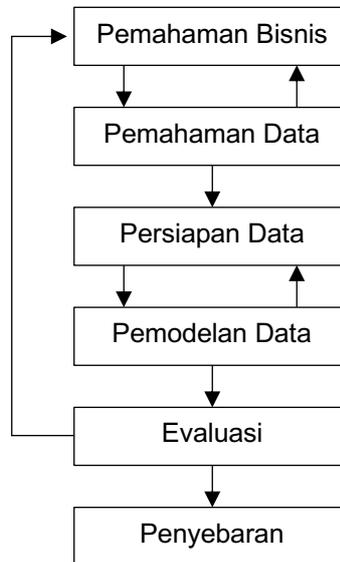
Teknik pengumpulan data yang digunakan pada penelitian ini yaitu kuesioner/angket daring dengan menggunakan Google Form untuk membuat pertanyaan kepada responden kemudian mengisi kuesioner yang telah dibagikan, kemudian menggunakan metode studi pustaka yaitu dengan mengumpulkan referensi dari beberapa literatur yang dibutuhkan atau dapat mendukung penelitian untuk landasan teori dalam pembuatan laporan.

### 2.2. Analisis Data

*Data mining* merupakan penambangan data atau upaya untuk menggali informasi yang berharga dan berguna pada *database* yang sangat besar. Dalam *data mining* terdapat beberapa metode yang digunakan untuk proses tahapan *data mining*, salah satu metode yang sering dipakai yaitu CRISP-DM, yaitu proses standar *data mining* dari portal industri yang sudah dikembangkan pada tahun 1996. Proses tersebut dilakukan untuk mendapatkan strategi memecahkan masalah dari



sebuah industri dan juga sebagai wadah pengetahuan bagi para peneliti (Amri, 2020). Menurut Purnama et al. (2012) metode analisis data *Cross-Industry Standard Process for Data Mining* (CRISP-DM) mempunyai 6 tahapan seperti pada Gambar 1.



**Gambar 1. Tahapan CRISP-DM**

Enam tahapan metode CRISP-DM pada Gambar 1 yaitu:

- 1) Pemahaman bisnis (*bussiness understanding*) yaitu untuk menentukan sasaran maupun tujuan bisnis atau memahami situasi untuk menentukan tujuan *data mining* dan membuat perencanaan serta jadwal penelitian yang dilakukan.
- 2) Pemahaman data (*data understanding*) yaitu pada tahap ini mempelajari data awal yang telah dikumpulkan untuk mengenal data yang akan digunakan atau mengidentifikasi masalah yang berkaitan dengan kualitas dari suatu data untuk membuat hipotesis awal untuk *data understanding*.
- 3) Persiapan data (*data preparation*) yaitu tahap ini dilakukan untuk pemilihan *table* dan *field* untuk bahan *data mining*.
- 4) Pemodelan data (*modeling*) yaitu untuk menentukan *data mining tools*, teknik data mining yang digunakan pada penelitian ini yaitu *Naïve Bayes*.
- 5) Evaluasi (*evaluation*) yaitu untuk interpretasi hasil *data mining* yang telah didapatkan untuk pemodelan dan fase sebelumnya untuk menyesuaikan model agar sesuai dengan tujuan yang ingin dicapai dari awal.
- 6) Penyebaran (*deployment*) yaitu tahap akhir dengan penyusunan laporan atau presentasi dari evaluasi *data mining* yang telah didapatkan pada fase-fase yang telah digunakan sebelumnya.

### 2.3. Penggunaan Algoritma

Metode klasifikasi yaitu terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah (Rezki et al., 2020). Salah satu algoritma yang digunakan untuk metode klasifikasi ini yaitu Algoritma *Naïve Bayes* yaitu metode probabilitas dan statistik, metode ini diperkenalkan pertama kali oleh Thomas Bayes seorang ilmuwan dari Inggris untuk memprediksi peluang yang terjadi di masa depan dengan pengalaman sebelumnya atau dinamakan dengan teorema bayes (Dhany & Izhari, 2019). Tahapan dari proses algoritma *Naïve Bayes* ini menurut Wulandari et al. (2020) yaitu:

- 1) Menghitung jumlah kelas/label
- 2) Menghitung jumlah kasus per kelas
- 3) Kalikan semua variabel kelas
- 4) Bandingkan hasil per kelas



## 2.4. RapidMiner

RapidMiner adalah RapidMiner merupakan perangkat lunak yang dapat diakses oleh siapa saja dan bersifat terbuka (*open source*). RapidMiner ini dijadikan sebuah solusi untuk menganalisa terhadap *data processing*. Pada RapidMiner ini digunakan berbagai teknik seperti teknik deskriptif dan prediksi. RapidMiner ini menggunakan bahasa pemrograman Java untuk pengoperasiannya (Sari et al., 2020).

## 3. HASIL DAN PEMBAHASAN

### 3.1. Business Understanding

Pada awal tahun 2020 aktivitas lembaga pendidikan diberhentikan sementara untuk mencegah penyebaran virus yang semakin meningkat setiap harinya sehingga pembelajaran dilakukan secara daring. Dari pemahaman tersebut maka tujuan dari penelitian untuk mengetahui dan mengelompokkan tingkat kepuasan pembelajaran secara daring dimasa pandemi ini.

### 3.2. Data Understanding

Tahapan-tahapan dalam melakukan *data understanding* sebagai berikut:

- 1) Mendeskripsikan data yaitu *dataset* mahasiswa baru yang didapatkan dengan menyebarkan kuesioner dengan menggunakan Google Form. Total responden yang mengisi kuesioner sebanyak 51 mahasiswa dari beberapa pertanyaan yang dibuat dan terdiri dari data nominal.
- 2) Pemilihan atribut yang digunakan yaitu jenis kelamin, penggunaan media belajar daring, mudah berkonsentrasi, kendala (*signal*), suasana pembelajaran, cara mengajar, komunikasi dan penyampaian materi.
- 3) Pertanyaan mengenai kuesioner dengan menggunakan Google Form yaitu:
  - a) Penyampaian Materi, dengan pilihan Puas dan Tidak Puas
  - b) Jenis Kelamin, dengan pilihan L (Laki-laki) dan P (Perempuan)
  - c) Penggunaan Media Daring, dengan pilihan S (Setuju), KS (Kurang Setuju) dan TS (Tidak Setuju).
  - d) Mudah Berkonsentrasi, dengan pilihan S (Setuju), KS (Kurang Setuju) dan TS (Tidak Setuju).
  - e) Kendala, dengan pilihan S (Setuju), KS (Kurang Setuju) dan TS (Tidak Setuju).
  - f) Suasana Pembelajaran, dengan pilihan Membosankan dan mendukung.
  - g) Cara Mengajar, dengan pilihan Serius, Menyenangkan dan Menyenangkan & Serius.
  - h) Komunikasi, dengan pilihan S (Setuju), KS (Kurang Setuju) dan TS (Tidak Setuju).

### 3.3. Data Preparation

Agar mendapatkan *data modeling* yang akurat, maka pada tahap ini data sebanyak 51 dengan jumlah atribut sebanyak 8 akan dilakukan pemeriksaan data. Hasil yang didapatkan terdapat data yang tidak konsisten pada atribut cara mengajar yaitu *menyenagkan* menjadi *menyenangkan*.

### 3.4. Modeling

Dari data yang telah didapatkan pada tahap *preparation*, maka pemodelan yang digunakan yaitu metode klasifikasi dengan menggunakan algoritma *Naïve Bayes*. Pada tahapan ini terdapat perhitungan manual *Naïve Bayes* dan menggunakan *tools* sebagai pemodelannya.

#### 3.4.1. Perhitungan Manual

Perhitungan manual yang dilakukan yaitu mencari probabilitas puas dan tidak puas dari atribut yang digunakan dengan menggunakan rumus perhitungan *Naïve Bayes* (Annur, 2018) seperti berikut:



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

Di mana  $X$  adalah data dengan *class* yang belum diketahui,  $H$  adalah hipotesis data  $X$  merupakan suatu *class* spesifik,  $P(H|X)$  adalah probabilitas hipotesis  $H$  berdasarkan kondisi  $x$  (posteriori prob.),  $P(H)$  adalah probabilitas hipotesis  $H$  (prior prob.),  $P(X|H)$  adalah probabilitas  $X$  berdasarkan kondisi tersebut, dan  $P(X)$  adalah probabilitas dari  $X$ .

1) Penyampaian Materi

**Tabel 1. Perhitungan Atribut Penyampaian Materi.**

	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
Puas	26	25	26/51	25/51
Tidak Puas	25	26	25/51	26/51
	51	51	100%	100%

2) Jenis Kelamin

**Tabel 2. Perhitungan Atribut Jenis Kelamin.**

	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
L	18	19	18/26	19/25
P	8	6	6/26	6/25
	26	25	100%	100%

3) Penggunaan Media Daring

**Tabel 3. Perhitungan Atribut Penggunaan Media Daring.**

	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
S	22	20	22/26	20/25
KS	3	4	3/26	4/25
TS	1	1	1/26	1/25
	26	25	100%	100%

4) Mudah Berkonsentrasi

**Tabel 4. Perhitungan Atribut Mudah Berkonsentrasi.**

	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
S	5	3	5/26	3/25
KS	16	12	16/26	12/25
TS	5	10	5/26	20/25
	26	25	100%	100%

5) Kendala

**Tabel 5. Perhitungan Atribut Kendala.**

	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
S	25	25	25/26	25/25
KS	1	0	1/26	0/0
TS	0	0	0/26	0/25
	26	25	100%	100%



## 6) Suasana Pembelajaran

Tabel 6. Perhitungan Atribut Suasana Pembelajaran.

	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
<b>Membosankan</b>	13	13	13/26	13/26
<b>Mendukng</b>	13	12	13/26	12/26
	26	25	100%	100%

## 7) Cara Mengajar

Tabel 7. Perhitungan Atribut Cara Mengajar.

	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
<b>Serius</b>	10	9	10/26	9/25
<b>Menyenangkan</b>	3	6	3/26	6/25
<b>Menyenangkan dan Serius</b>	13	10	13/26	10/25
	26	25	100%	100%

## 8) Komunikasi

Tabel 8. Perhitungan Atribut Komunikasi.

	Puas	Tidak Puas	P (Puas)	P (Tidak Puas)
<b>S</b>	15	14	15/26	14/25
<b>KS</b>	9	7	9/26	7/25
<b>S</b>	2	4	2/26	4/25
	26	25	100%	100%

Jadi, hasil dari hitungan probabilitas *posterior* dari masing-masing tabel atau atribut yang digunakan pada persamaan (1).

⇒ For Class Puas

$$\begin{aligned}
 &= \left( \frac{18}{26} \times \frac{22}{26} \times \frac{5}{26} \times \frac{25}{26} \times \frac{13}{26} \times \frac{10}{26} \times \frac{15}{26} \right) \left( \frac{26}{51} \right) \\
 &= \left( \frac{928125}{77228944} \right) (0,5098) \\
 &= 0,0120 \times 0,5098 \\
 &= 0,0061
 \end{aligned}$$

⇒ For Class Tidak Puas

$$\begin{aligned}
 &= \left( \frac{19}{25} \times \frac{20}{25} \times \frac{3}{25} \times \frac{25}{25} \times \frac{13}{26} \times \frac{6}{25} \times \frac{14}{25} \right) \left( \frac{25}{51} \right) \\
 &= \left( \frac{9576}{1953125} \right) (0,4901) \\
 &= 0,0049 \times 0,4901 \\
 &= 0,0024
 \end{aligned}$$

Jadi, dari hasil yang didapatkan probabilitas Puas > Tidak Puas



### 3.4.2. Pengujian Menggunakan RapidMiner

Pengujian menggunakan RapidMiner dengan menggunakan algoritma *Naïve Bayes* menggunakan *data training* dan *data testing* untuk dilakukan proses klasifikasi.

Row No.	Hasil Pembe...	Jenis Kelamin	Penggunaan...	Mudah Berk...	Kendala (sig...	Suasana Pe...	Cara Mengaj...	Komunikasi
1	Puas	P	S	KS	S	Membosankan	Serius	T
2	Tidak Puas	P	S	KS	S	Mendukung	Menyenangk...	KT
3	Tidak Puas	P	KS	KS	S	Mendukung	Serius dan m...	T
4	Tidak Puas	P	KS	S	S	Membosankan	Serius dan m...	T
5	Puas	P	S	KS	KS	Membosankan	Serius dan m...	T
6	Tidak Puas	P	S	KS	S	Membosankan	Serius dan m...	TT
7	Tidak Puas	P	S	TS	S	Membosankan	Serius	T
8	Tidak Puas	P	S	KS	S	Membosankan	Serius	T
9	Puas	L	S	KS	S	Mendukung	Serius dan m...	T
10	Puas	P	S	KS	S	Mendukung	Serius dan m...	T
11	Puas	P	S	KS	S	Membosankan	Serius dan m...	T
12	Tidak Puas	P	KS	TS	S	Membosankan	Menyenangkan	T
13	Puas	P	S	KS	S	Membosankan	Serius	T
14	Puas	P	S	TS	S	Mendukung	Serius dan m...	TT
15	Puas	L	S	TS	S	Membosankan	Serius	TT
16	Puas	P	S	KS	S	Membosankan	Serius	KT

ExampleSet (51 examples, 1 special attribute, 7 regular attributes)

Gambar 2. Data Training,

Row No.	Hasil Pembe...	Jenis Kelamin	Penggunaan...	Mudah Berk...	Kendala (sig...	Suasana Pe...	Cara Mengaj...	Komunikasi
1	Puas	P	S	KS	S	Membosankan	Serius	T
2	Tidak Puas	P	S	KS	S	Mendukung	Menyenangk...	KT
3	Tidak Puas	P	KS	KS	S	Mendukung	Serius dan m...	T
4	Tidak Puas	P	KS	S	S	Membosankan	Serius dan m...	T
5	Puas	P	S	KS	KS	Membosankan	Serius dan m...	T
6	Tidak Puas	P	S	KS	S	Membosankan	Serius dan m...	TT
7	Tidak Puas	P	S	TS	S	Membosankan	Serius	T
8	Tidak Puas	P	S	KS	S	Membosankan	Serius	T
9	Puas	L	S	KS	S	Mendukung	Serius dan m...	T
10	Puas	P	S	KS	S	Mendukung	Serius dan m...	T
11	Puas	P	S	KS	S	Membosankan	Serius dan m...	T
12	Tidak Puas	P	KS	TS	S	Membosankan	Menyenangk...	T
13	Puas	P	S	KS	S	Membosankan	Serius	T

Gambar 3. Data Testing.

Setelah dilakukan pemodelan, maka hasil pengujian menggunakan RapidMiner dengan mengklasifikasi 2 *class* yaitu Puas dan Tidak Puas, selain *accuracy* ada dua hal yang dihasilkan dalam klasifikasi yaitu *class precision* merupakan perbandingan dokumen dengan hasil yang relevan pada hasil *query* dan *class recall* perbandingan jumlah dokumen relevan terambil terhadap total dokumen relevan. Akurasi yang didapatkan sebesar 76,92%, *class precision* menghasilkan 66.67%, dan *class recall* 100%. Rumus yang digunakan untuk mencari akurasi, *precision*, dan *recall* (Pratama et al., 2020) dijelaskan sebagai berikut.

Tabel 9. Nilai Prediksi.

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP (True Positive) Correct Result	FP (False Positive) Unexpected Result
	FALSE	FN (False Negative) Missing Result	TN (True Negative)



$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

Hasil yang didapatkan dari persamaan di atas yaitu:

1) *Precision*

Rumus dari *precision* pada persamaan (2) yaitu:

$$Precision = \frac{6}{6+3} = \frac{6}{9} = 66,67\%$$

2) *Recall*

Rumus dari *precision* pada persamaan (3) yaitu:

$$Recall = \frac{6}{6+0} = \frac{6}{6} = 100\%$$

3) Akurasi

Rumus dari *precision* pada persamaan (4) yaitu:

$$Akurasi = \frac{6+4}{6+4+3+0} = \frac{10}{13} = 76,92\%$$

Akan tetapi dari hasil yang telah didapatkan perlu ditinjau ulang kembali dari jumlah *dataset*.

accuracy: 76.92%

	true Puas	true Tidak Puas	class precision
pred. Puas	6	3	66.67%
pred. Tidak Puas	0	4	100.00%
class recall	100.00%	57.14%	

Gambar 4. Hasil Akurasi.

Dengan hasil *confidence* Puas dan Tidak Puas yaitu dapat dilihat pada Gambar 5.

Row No.	Hasil Pembe...	prediction(H...	confidence(...	confidence(...	Jenis Kelamin	Penggunaan...	Mudah Berk...	Kendala (sig...	Suasana Pe...	Cara Mengaj...	Komunikasi
1	Puas	Puas	0.547	0.453	P	S	KS	S	Membosankan	Serius	T
2	Tidak Puas	Tidak Puas	0.026	0.974	P	S	KS	S	Mendukung	Menyenangk...	KT
3	Tidak Puas	Puas	0.511	0.489	P	KS	KS	S	Mendukung	Serius dan m...	T
4	Tidak Puas	Puas	0.661	0.339	P	KS	S	S	Membosankan	Serius dan m...	T
5	Puas	Puas	0.987	0.013	P	S	KS	KS	Membosankan	Serius dan m...	T
6	Tidak Puas	Tidak Puas	0.398	0.602	P	S	KS	S	Membosankan	Serius dan m...	TT
7	Tidak Puas	Tidak Puas	0.329	0.671	P	S	TS	S	Membosankan	Serius	T
8	Tidak Puas	Puas	0.547	0.453	P	S	KS	S	Membosankan	Serius	T
9	Puas	Puas	0.683	0.317	L	S	KS	S	Mendukung	Serius dan m...	T
10	Puas	Puas	0.605	0.395	P	S	KS	S	Mendukung	Serius dan m...	T
11	Puas	Puas	0.585	0.415	P	S	KS	S	Membosankan	Serius dan m...	T
12	Tidak Puas	Tidak Puas	0.185	0.815	P	KS	TS	S	Membosankan	Menyenagkan	T
13	Puas	Puas	0.547	0.453	P	S	KS	S	Membosankan	Serius	T

Gambar 5. Hasil Confidence.

Model distribusi dengan label Penyampaian Materi yaitu *class* Puas : 7 distribusi dan *class* Tidak Puas : 7 distribusi atau hasil dari klasifikasi dengan 2 kelas yaitu Puas dan Tidak Puas didapatkan untuk *class* Puas (0,510) dan *class* Tidak Puas (0,490).



## SimpleDistribution

Distribution model for label attribute Hasil Pembelajaran

Class Puas (0.510)  
7 distributions

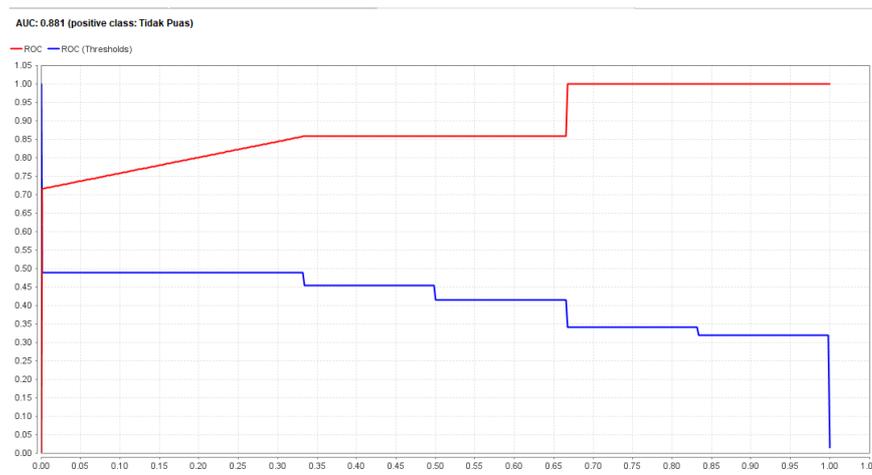
Class Tidak Puas (0.490)  
7 distributions

**Gambar 6. Hasil SimpleDistribution.**

### 3.5. Data Evaluasi

Pada tahap evaluasi data dengan merepresentasikan grafis yang menentukan klasifikasi mana yang lebih baik, maka digunakan metode yang menghitung luas daerah di bawah kurva ROC yang disebut AUC (*Area Under the ROC Curve*) atau probabilitas. Sedangkan *Receiver Operating Characteristic (ROC) Curve* untuk evaluasi dari metode klasifikasi yaitu algoritma *Naïve Bayes*. Berikut adalah nilai atau evaluasi dari kurva AUC (Hastuti, 2012):

- 1) Performance Klasifikasi 0,90 –1,00 Paling baik
- 2) Performance Klasifikasi 0,80 –0,90 Baik
- 3) Performance Klasifikasi 0,70 –0,80 Adil atau sama
- 4) Performance Klasifikasi 0,60 –0,70 Rendah
- 5) Performance Klasifikasi 0,50 –0,60 Gagal



**Gambar 7. Kurva ROC.**

Jadi dari 2 kurva dengan warna biru dan merah didapatkan kinerja dari warna biru lebih bagus atau dari gambar dapat disimpulkan luas area kurva berwarna biru lebih besar dibanding luas area kurva berwarna merah dengan nilai AUC 0.881, mendekati angka 1, maka model yang dihasilkan baik.

## 4. KESIMPULAN

Berdasarkan perhitungan *data mining* dengan metode klasifikasi menggunakan algoritma *Naïve Bayes* kesimpulan yang didapatkan dari *dataset* yang diperoleh dengan 2 *class* Puas dan Tidak Puas bahwa hasil pengujian algoritma *Naïve Bayes* ini diperoleh tingkat akurasi sebesar 76,92% dan *class precision* menghasilkan 100.00%, *class recall* 57.14%, dan nilai AUC 0.881, mendekati angka 1, jadi model yang dihasilkan baik. Jadi, dari hasil yang dapatkan tersebut teknik klasifikasi yang diuji coba dapat digunakan sebagai bahan untuk pengambilan keputusan tingkat kepuasan pembelajaran secara daring.



## DAFTAR PUSTAKA

- Adijaya, N. (2018). Persepsi Mahasiswa dalam Pembelajaran Online\* \*Pengembangan teori dari penelitian berjudul "persepsi mahasiswa terhadap materi ajar pada pembelajaran online" yang telah dipublikasi di Jurnal Eduscience Vol. 3/1. *Wanastra: Jurnal Bahasa Dan Sastra*, 10(2), 105–110. <https://doi.org/10.31294/w.v10i2.3931>
- Aji, R. H. S. (2020). Dampak Covid-19 pada Pendidikan di Indonesia: Sekolah, Keterampilan, dan Proses Pembelajaran. *SALAM: Jurnal Sosial Dan Budaya Syar-I*, 7(5), 395–402. <https://doi.org/10.15408/SJSBS.V7I5.15314>
- Amri, S. (2020). Perbandingan Kerangka Model Klasifikasi untuk Pemilihan Metode Kontrasepsi dengan Pendekatan CRIPS-DM. *Information Science and Library*, 1(1), 14–23. <https://doi.org/10.26623/JISL.V1I1.2488>
- 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>
- Dhany, H. W., & Izhari, F. (2019). ANALISIS ALGORITHMS SUPPORT VECTOR MACHINE DENGAN NAIVE BAYES KERNEL PADA KLASIFIKASI DATA. *Jurnal Teknik Dan Informatika*, 6(2), 30–35.
- Hastuti, K. (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif. *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012*, 14(1), 241–249.
- Pakpahan, R., & Fitriani, Y. (2020). ANALISA PEMANFAATAN TEKNOLOGI INFORMASI DALAM PEMBELAJARAN JARAK JAUH DI TENGAH PANDEMI VIRUS CORONA COVID-19. *Journal of Information System, Applied, Management, Accounting and Research*, 4(2), 30–36.
- Pratama, A., Midyanti, D. M., & Bahri, S. (2020). PENERAPAN NAÏVE BAYES CLASSIFIER DENGAN ALGORITMA STEMMING NAZIEF DAN ADRIANI UNTUK APLIKASI DETEKSI UJARAN KEBENCIAN BERBASIS WEB. *Coding Jurnal Komputer Dan Aplikasi*, 8(1), 227–236. <https://doi.org/10.26418/CODING.V8I1.39457>
- Purnama, I., Saputra, R., & Wibowo, A. (2012). IMPLEMENTASI DATA MINING MENGGUNAKAN CRISP-DM PADA SISTEM INFORMASI EKSEKUTIF DINAS KELAUTAN DAN PERIKANAN PROVINSI JAWA TENGAH. *Prosiding Seminar Nasional Ilmu Komputer UNDIP 2012*.
- Rezki, M., Kholifah, D. N., Faisal, M., Priyono, P., & Suryadithia, R. (2020). Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting Menggunakan Algoritma Naïve Bayes. *Jurnal Infortech*, 2(2), 264–270. <https://doi.org/10.31294/infortech.v2i2.9286>
- Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 192. <https://doi.org/10.24114/cess.v5i2.18519>
- Siddik, M., Hendri, H., Putri, R. N., Desnelita, Y., & Gustientiedina, G. (2020). Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes. *INTECOMS: Journal of Information Technology and Computer Science*, 3(2), 162–166. <https://doi.org/10.31539/INTECOMS.V3I2.1654>
- Silvia Siltonga, D., Saifullah, & Dewi, R. (2019). Analisis Metode Naive Bayes dalam Memprediksi Tingkat Pemahaman Mahasiswa Terhadap Mata Kuliah Berdasarkan Posisi Duduk. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 427–436.
- Supriatna, E. (2020). Wabah Corona Virus Disease Covid 19 Dalam Pandangan Islam. *SALAM: Jurnal Sosial Dan Budaya Syar-I*, 7(6), 555–564. <https://doi.org/10.15408/SJSBS.V7I6.15247>
- Wulandari, F., Jusia, P. A., & Jasmir, J. (2020). Klasifikasi Data Mining Untuk Mendiagnosa Penyakit ISPA Menggunakan Metode Naïve Bayes Pada Puskesmas Jambi Selatan. *Jurnal Manajemen Teknologi Dan Sistem Informasi (JMS)*, 2(3), 214–227.



## Implementasi Algoritma RC4 pada Sistem Pengamanan Dokumen Digital Soal Ujian

Fauziah Suwarsita Febriyani <sup>(1)\*</sup>, Arief Arfriandi <sup>(2)</sup>

Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang, Semarang  
e-mail : fauziyahsuwarsita@gmail.com, arfriandi@mail.unnes.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 24 Mei 2021, direvisi 18 September 2021, diterima 19 September 2021, dan dipublikasikan 22 September 2021.

### Abstract

*The development of science and technology has led to changes in the use of documents in life to become digital data. However, this can cause problems, namely regarding data security and confidentiality. To increase security and confidentiality can be done with cryptographic algorithm RC4. The research method uses the Waterfall method. The result of this research is a website that can secure document files with \* doc extension using the RC4 algorithm. The test was carried out using the blackbox test and the CrackStation test for encryption testing. The results of the test show that the website can run well and successfully implements the RC4 algorithm.*

**Keywords:** Digital Data, Cryptography, RC4 Algorithm, Waterfall, Website

### Abstrak

Dokumen dalam kehidupan sehari-hari telah berganti dari yang manual menjadi dokumen digital. Namun hal tersebut dapat menimbulkan masalah yaitu mengenai keamanan dan kerahasiaan data. Untuk meningkatkan keamanan dan kerahasiaan dapat dilakukan dengan kriptografi algoritma RC4. Metode penelitian menggunakan model *waterfall*. Hasil dari penelitian yaitu sebuah *website* yang dapat mengamankan file dokumen berekstensi *\*doc* menggunakan algoritma RC4. Pengujian dilakukan dengan menggunakan uji *blackbox* dan uji CrackStation untuk pengujian enkripsi. Hasil dari pengujian menunjukkan sistem dapat berjalan dengan baik dan berhasil mengimplementasikan algoritma RC4.

**Kata Kunci:** Data Digital, Kriptografi, Algoritma RC4, Waterfall, Website

## 1. PENDAHULUAN

Penggunaan dokumen digital saat ini telah banyak digunakan, salah satunya yaitu pada pembuatan soal ujian sekarang telah berkembang menjadi data digital. Sehingga menjadi lebih mudah diakses baik oleh guru maupun siswa.

Kemudahan pengaksesan informasi, tentunya berdampak pada munculnya resiko dan ancaman keamanan dan integritas data (Sumarno, 2018). Menurut DataLossDB (2016), pada tahun 2014 kebocoran data sekitar 50% terjadi pada sektor bisnis, sekitar 20% terjadi pada sektor pemerintahan dan sekitar 30% terjadi pada sektor kesehatan dan pendidikan. Kebocoran data juga terjadi pada pengguna pribadi, tetapi sulit untuk mengetahui jumlah persis dan tingkat keparahan kebocoran pada data pribadi (Alneyadi et al., 2016).

Untuk menjaga keamanan data dan menghindari kebocoran data dapat dilakukan berbagai cara salah satunya yaitu dengan menggunakan kriptografi. Kriptografi merupakan ilmu yang menggunakan teknik-teknik matematika yang berkaitan dengan aspek keamanan informasi seperti kerahasiaan, integritas data, autentikasi entitas, dan autentikasi asal data (Menezes et al., 2001). Dalam ilmu kriptografi terdapat berbagai macam algoritma salah satunya yaitu RC4.

Berdasarkan cara kerjanya, algoritma kriptografi RC4 merupakan jenis *stream cipher*. Kelebihannya adalah dengan menggunakan cara ini enkripsi algoritma RC4 dapat dilakukan pada data dengan panjang yang beragam. Algoritma RC4 dinilai sangat cepat dalam prosesnya, kurang lebih 10 kali lebih cepat dari DES (Hakim et al., 2014). Algoritma dari metode RC4 *stream cipher* ini terdiri dari dua bagian, yaitu: *key setup* dan *stream generation*. Pada *key setup* terdapat

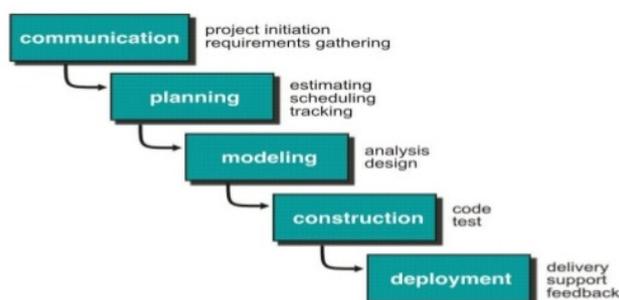


tiga tahapan proses di dalamnya, yaitu Inisialisasi S-Box, menyimpan *key* dalam *key byte array*, permutasi pada S-Box. Pada *stream generation* akan menghasilkan nilai *pseudorandom* yang akan dikenakan operasi XOR untuk menghasilkan *ciphertext* ataupun sebaliknya yaitu menghasilkan *plaintext* (Saragi et al., 2020).

Penelitian terkait implementasi algoritma RC4 pernah dilakukan oleh peneliti terdahulu. Seperti penelitian yang dilakukan oleh Ridho & Jemakmun (2020), pada penelitian tersebut sistem bertujuan untuk mengamankan *database* pegawai. Waluyo & Kanahebi (2021) dan Purba et al. (2020) telah melakukan penelitian dengan mengimplementasikan algoritma untuk mengamankan file teks. Zebua & Ndruru (2017) juga melakukan penelitian untuk menguraikan pengamanan citra digital berdasarkan modifikasi algoritma RC4. Penelitian lainnya yaitu penelitian yang dilakukan oleh Subhan et al. (2017) dengan mengimplementasikan algoritma RC4 berbasis android.

## 2. METODE PENELITIAN

Desain atau metode penelitian yang digunakan pada penelitian ini adalah model *waterfall*. Menurut Pressman (2003) model *waterfall* adalah model klasik yang bersifat sistematis, berurutan dalam membangun *software*. Terdapat lima tahapan pada metode penelitian *waterfall*, antara lain komunikasi, perencanaan, pemodelan, konstruksi, dan *deployment*.



Gambar 1. Tahapan Model *Waterfall* (Pressman, 2003).

### 2.1. Komunikasi

Langkah awal dalam membangun sistem yaitu menganalisis dan mengkomunikasikan kebutuhan-kebutuhan sistem berdasarkan hasil pengumpulan data. Dari tahap komunikasi diperoleh kesimpulan dibutuhkan adanya sistem untuk mengamankan dokumen ujian SMK Negeri 1 Slawi.

### 2.2. Perencanaan

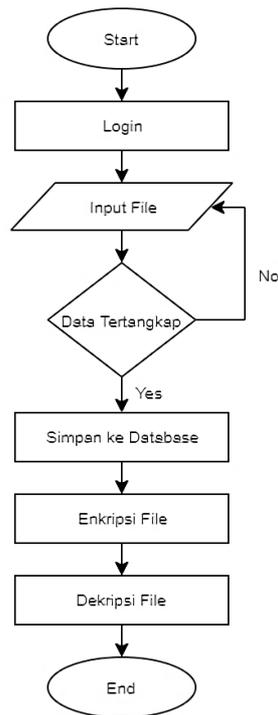
Tahap perencanaan menjabarkan tentenag tugas-tugas teknik yang harus dilakukan mulai dari pengumpulan data sampai pengujian sistem, resiko yang mungkin terjadi pada saat menjalankan tugas-tugas tersebut, dan hasil yang ingin diperoleh yaitu terciptanya sebuah sistem pengamanan dokumen digital soal ujian dengan menggunakan algoritma RC4.

### 2.3. Pemodelan

Tahap pemodelan berfokus pada perancangan skema sistem pengamanan dokumen digital soal ujian kemudian merancang *database* menggunakan *MySQL* sebagai tempat penyimpanan data sistem. Selanjutnya merancang antarmuka sistem pengamanan dokumen digital soal ujian. *Flowcart* Sistem dapat dilihat pada Gambar 2.



Flowchart Sistem Pengamanan Data Digital



Gambar 2. Flowchart Sistem

Sistem dimulai dengan *login* yaitu *user* memasukkan *username* dan *password*. Kemudian, *user* dapat menginputkan file ke dalam sistem, yang selanjutnya sistem akan menangkap file tersebut dan disimpan ke dalam *database*. Selanjutnya file tersebut dapat dienkripsi dan kemudian didekripsikan kembali jika diperlukan oleh *user*.

#### 2.4. Konstruksi

Konstruksi merupakan tahap pembuatan kode, sehingga tercipta sistem yang telah dirancang sebelumnya. Bahasa pemrograman yang digunakan untuk membangun sistem adalah PHP dan menggunakan *database MySQL* sebagai penyimpanan datanya. Algoritma RC4 digunakan untuk mengamankan data yang dimasukkan ke dalam *database*. Dan kemudian dapat dikembalikan lagi seperti semula saat diperlukan.

Setelah proses pengkodean selesai, maka akan dilakukan pengujian terhadap sistem yang telah dibuat. Pengujian sistem dilakukan menggunakan *blackbox testing*. Pengujian *blackbox* dilakukan untuk mengetahui apakah fungsi-fungsi pada program dapat berjalan dengan baik, mulai dari menerima input, memproses, dan memberikan *output*. Pengujian selanjutnya menggunakan *software* penyerang yang berfokus terhadap hasil implementasi enkripsi algoritma RC4 pada isi file dokumen. *Software* yang digunakan adalah CrackStation.

#### 2.5. Deployment

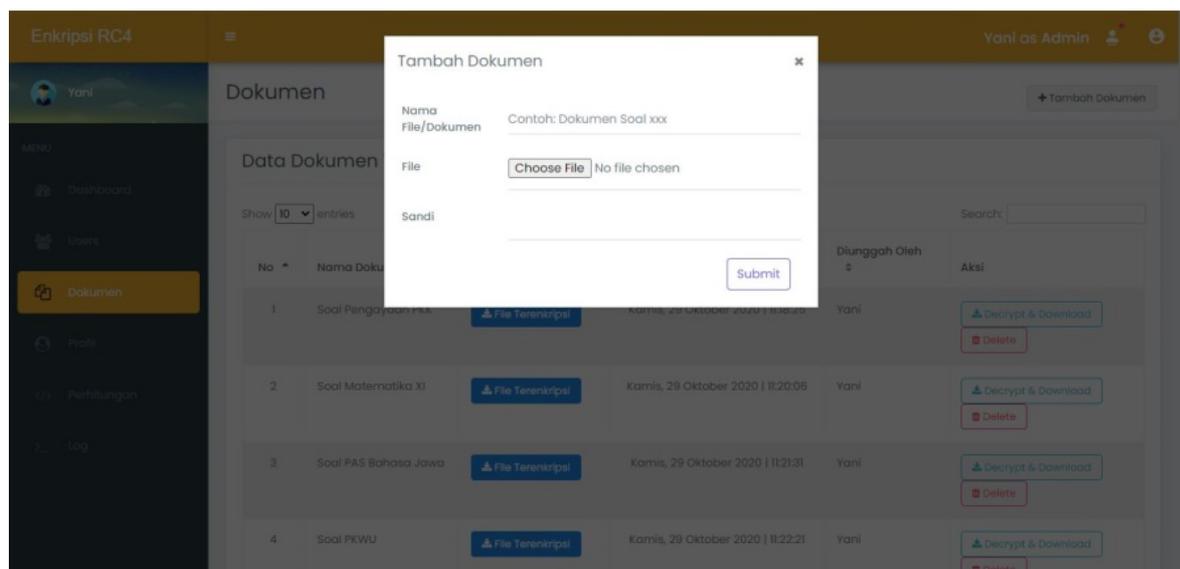
Tahap *deployment* merupakan tahapan dimana sistem telah siap digunakan oleh pengguna. Kemudian untuk menjaga sistem tetap berjalan dengan baik maka perlu dilakukan pemeliharaan secara berkala sesuai dengan kebutuhan.



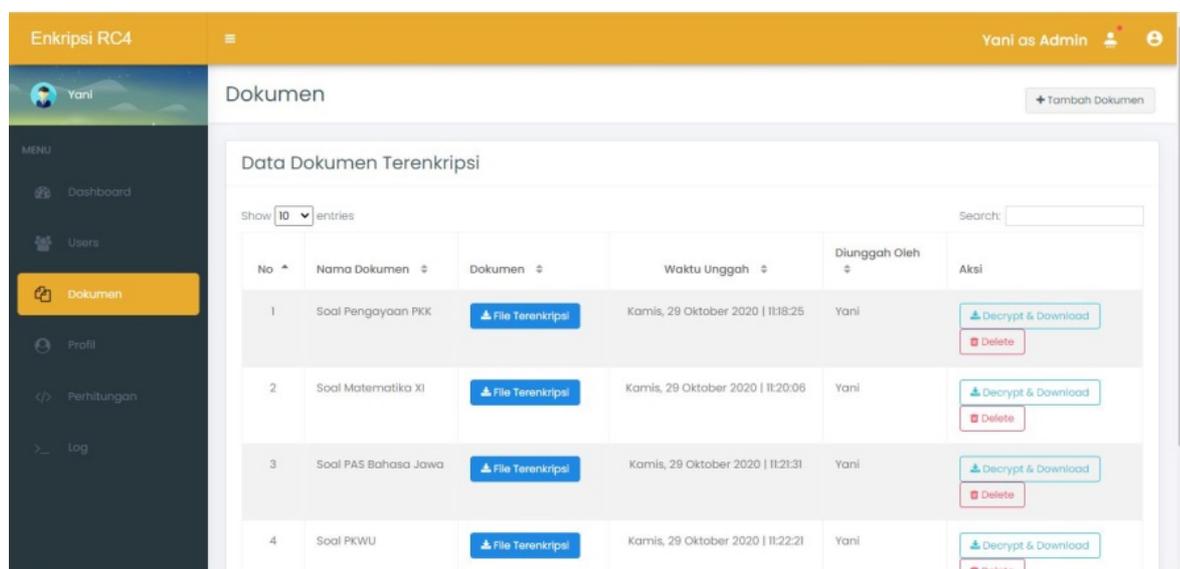
### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil Implementasi

Sistem dapat mengenkripsi dan mendekripsikan kembali file dokumen dengan format file *doc*. menggunakan algoritma RC4 sehingga dapat mengamankan dokumen soal ujian. Untuk mengamankan dokumen, *user* harus menginputkan dokumen beserta kuncinya seperti yang dapat dilihat pada Gambar 3. File yang berhasil diamankan akan masuk ke dalam *database* dan ditampilkan pada menu dokumen seperti pada Gambar 4. Kemudian jika ingin mendekripsikannya klik tombol *decrypt & download* kemudian *user* harus menginputkan kunci yang sama seperti saat mengamankan dokumen.



Gambar 3. Form Tambah Dokumen



Gambar 4. Halaman Daftar Dokumen Terenkripsi

#### 3.2. Hasil Uji *Blackbox*

Hasil pengujian *blackbox* menunjukkan bahwa sistem dapat berjalan dengan baik sesuai dengan input yang diberikan. Hasil pengujian *blackbox* dapat dilihat pada Tabel 1.



Artikel ini didistribusikan mengikuti lisensi Atribusi-NonKomersial CC BY-NC sebagaimana tercantum pada <https://creativecommons.org/licenses/by-nc/4.0/>.

Tabel 1. Hasil Uji *Blackbox*.

No.	Test Case	Skenario Pengujian	Hasil yang Diharapkan	Hasil Pengujian
1.	Memasukkan <i>username</i> dan <i>password</i> .	Mengisi <i>form login</i> pada halaman <i>login</i> kemudian pilih tombol <i>login</i> .	Saat data yang sudah terisi sesuai dengan data pada <i>database</i> maka akan masuk menuju halaman <i>dashboard</i> . Apabila tidak sesuai atau kosong maka akan muncul peringatan.	Sesuai
2.	Melihat data <i>users</i> .	Pilih menu <i>users</i> .	Sistem menampilkan daftar semua <i>user</i> yang ada dalam <i>database</i> .	Sesuai
3.	Menambah <i>user</i> .	Klik tombol tambah <i>user</i> .	Ketika klik tombol tambah <i>user</i> maka muncul <i>form</i> "Add New User" data yang harus diisi. Ketika <i>form</i> sudah diisi, data akan tersimpan dan muncul pesan berhasil.	Sesuai
4.	Mengubah dan menghapus data <i>user</i> .	Memfungsikan tombol "Edit" dan "Delete" pada daftar <i>user</i> .	Ketika klik "Edit" sistem menampilkan <i>form edit</i> data <i>user</i> yang dipilih. Ketika klik "Delete" sistem menghapus data <i>user</i> yang dipilih.	Sesuai
5.	Mencari data <i>user</i> .	Mengetik nama <i>user</i> yang dicari pada kolom pencarian.	Sistem menampilkan hasil pencarian sesuai dengan nama <i>user</i> yang dimasukkan pada kolom pencarian.	Sesuai
6.	Melihat data data dokumen.	Pilih menu dokumen.	Sistem menampilkan daftar semua dokumen.	Sesuai
7.	Menambah dokumen (enkripsi dokumen).	- Klik tombol tambah dokumen. - Mengisi semua <i>field</i> termasuk sandi untuk enkripsi.	Ketika klik tombol tambah dokumen maka muncul <i>form</i> "Tambah Dokumen". Setelah semua <i>field</i> terisi data dokumen akan tersimpan dan terenkripsi dan muncul pesan sistem berhasil.	Sesuai
8.	Men-download dan mendekripsi dokumen.	Klik tombol "Decrypt & Download" pada daftar dokumen.	Ketika klik tombol "Decrypt & Download" dokumen akan ter-download dalam keadaan sudah terdekripsi.	Sesuai
9.	Melihat profil.	Pilih menu profil.	Sistem menampilkan detail data <i>user</i> dan riwayat dari <i>user</i> .	Sesuai
10.	Mengubah data profil <i>user</i> .	Pilih submenu "Edit Profil" pada menu profil.	Ketika klik "Edit Profil" maka sistem menampilkan <i>field</i> edit data profil <i>user</i> .	Sesuai
11.	Melakukan perhitungan	Pilih menu perhitungan kemudian mengisi <i>form</i> perhitungan.	Setelah <i>form</i> perhitungan terisi maka muncul alur perhitungan enkripsi algoritma RC4 sesuai dengan teks dan sandi yang dimasukkan.	Sesuai
12.	Melihat <i>log</i> aktivitas sistem	Pilih menu <i>log</i>	Sistem menampilkan <i>log</i> aktivitas yang terjadi pada sistem.	Sesuai



### 3.3. Hasil Uji CrackStation

Hasil pengujian terhadap 10 sampel diketahui bahwa algoritma RC4 tidak berhasil dipecahkan. Hasil uji CrackStation dapat dilihat pada Gambar 5.

Hash	Type	Result
ÁpİöööÁöÜb%obö·öimö(ö?ñ¿í	Unknown	Unrecognized hash format.
öIäÉöööööÜööÖöY!o%2«öiP	Unknown	Unrecognized hash format.
öÉäiöööxýB(ö[2!Vööö{W·ú3´ööö~	Unknown	Unrecognized hash format.
öIäÉöööxýööYUa5rpö`ö	Unknown	Unrecognized hash format.
öÉäiöööxýB(ö[2!Vööö{W·ú3´ööö°	Unknown	Unrecognized hash format.
öIäÉööööööIäöY/mdö	Unknown	Unrecognized hash format.
ÁpÁö5öÜöööÁö\`Ld«öiw»ö°	Unknown	Unrecognized hash format.
öéÄé- WäÜ'In`aö!D*ö²Zöö=ö;ö	Unknown	Unrecognized hash format.
ÁpÁö5öÜöööÁö\`Cöö`öiööö%öööö	Unknown	Unrecognized hash format.
öÉäiö öP±ööööZ`rd»xöB¥ö²ö¿	Unknown	Unrecognized hash format.

Color Codes: **Green**: Exact match, **Yellow**: Partial match, **Red**: Not found.

Gambar 5. Hasil Uji CrackStation

### 3.4. Pembahasan

Penelitian ini menggunakan metode penelitian model *waterfall* dimana setiap tahap dilakukan dengan urut tanpa saling tumpang tindih sehingga meminimalisasi terjadinya kesalahan. Setiap tahapan pada model *waterfall* juga memiliki spesifikasinya sendiri, sehingga sebuah sistem dapat dikembangkan sesuai dengan apa yang dikehendaki. Software yang dikembangkan dengan metode ini biasanya menghasilkan kualitas yang baik (Widiyanto, 2018).

Penelitian ini mengimplementasikan algoritma RC4 pada pengamanan sistem soal ujian kerana berdasarkan cara kerjanya algoritma RC4 merupakan algoritma jenis *stream cipher* yang memproses input data pada satu waktu. Input data pada umumnya berbentuk *byte* atau bahkan bit (*byte* dalam hal RC4) sehingga dengan menggunakan algoritma RC4 proses enkripsi dan dekripsi dapat diimplementasikan pada file dokumen soal ujian dengan panjang yang bervariasi. Algoritma RC4 juga tidak perlu menunggu sejumlah input data atau pesan tertentu sebelum diproses, atau menambahkan *byte* tambahan untuk melakukan enkripsi (Nugroho et al., 2016).

Algoritma RC4 dirancang supaya dapat diimplementasikan di *software* dengan sangat efisien. Sehingga membuat algoritma RC4 sangat populer untuk aplikasi internet, antara lain RC4 digunakan dalam standar TLS (*Transport Layer Security*), dan WEP (*Wireless Equivalent Privacy*) (Agung & Budiman, 2015).

Pada pengujian CrackStation menggunakan 10 sampel sudah dapat membuktikan bahwa algoritma RC4 dapat diimplementasikan untuk enkripsi, ditunjukkan dengan *ciphertext* hasil dari proses enkripsi tidak dapat dipecahkan.

## 4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dihasilkan sebuah sistem pengamanan dokumen digital dengan mengimplementasikan algoritma RC4. Sebagaimana pengujian yang telah dilakukan pada pengujian *blackbox* menunjukkan bahwa sistem dapat berjalan sesuai dengan inputan yang diberikan. Kemudian pada pengujian *software* CrackStation untuk menguji 10 sampel *ciphertext* menunjukkan bahwa *ciphertext* 100% tidak dapat dipecahkan.



## DAFTAR PUSTAKA

- Agung, H., & Budiman, B. (2015). IMPLEMENTASI AFFINE CHIPER DAN RC4 PADA ENKRIPSI FILE TUNGGAL. *Prosiding SNATIF*, 0(0), 243–250.
- Alneyadi, S., Sithirasanen, E., & Muthukkumarasamy, V. (2016). A survey on data leakage prevention systems. *Journal of Network and Computer Applications*, 62(C), 137–152. <https://doi.org/10.1016/j.jnca.2016.01.008>
- DataLossDB. (2016, February 11). *2015 Reported Data Breaches Surpasses All Previous Years*. <https://datalossdbdotorg.wordpress.com/>
- Hakim, E. L., Khairil, K., & Utami, F. H. (2014). APLIKASI ENKRIPSI DAN DESKRIPSI DATA MENGGUNAKAN ALGORITMA RC4 DENGAN MENGGUNAKAN BAHASA PEMROGRAMAN PHP. *JURNAL MEDIA INFOTAMA*, 10(1), 1–7. <https://doi.org/10.37676/JMI.V10I1.226>
- Menezes, A. J., van Oorschot, P. C., & Vanstone, S. a. (2001). Handbook of Applied Cryptography. In *Handbook of Applied Cryptography* (Issue 9). CRC Press.
- Nugroho, N. B., Azmi, Z., & Arif, S. N. (2016). APLIKASI KEAMANAN EMAIL MENGGUNAKAN ALGORITMA RC4. *Jurnal SAINTIKOM*, 15(3), 81–88.
- Pressman, R. S. (2003). *Rekayasa Perangkat Lunak Pendekatan Praktisi (Buku II)*. Andi.
- Purba, B., Gulo, F. A., Utami, N. I., & Sihotang, Y. A. (2020). Pengamanan File Teks Menggunakan Algoritma RC4. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 420–425.
- Ridho, M., & Jemakmun. (2020). *IMPLEMENTASI KEAMANAN BASI DATA PADA DATA PEGAWAI DI PT TIRTA MUSI MENGGUNAKAN METODE RC4*. Universitas Bina Darma.
- Saragi, D. R., Gultom, J. M., Tampubolon, J. A., & Gunawan, I. (2020). Pengamanan Data File Teks (Word) Menggunakan Algoritma RC4. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 1(2), 114–119. <https://doi.org/10.30865/json.v1i2.1745>
- Subhan, S., Amini, S., & Ariyani, P. F. (2017). IMPLEMENTASI PENGAMANAN DATA ENKRIPSI SMS DENGAN ALGORITMA RC4 BERBASIS ANDROID. *Prosiding SENIATI*, A29.6.
- Sumarno, S. (2018). Analisis Kinerja Kombinasi Algoritma Message-Digest Algoritim 5 (MD5), Rivest Shamir Adleman (RSA ) dan Rivest Cipher 4 (RC4) Pada Keamanan E-Dokumen. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(1), 41–48.
- Waluyo, S., & Kanahebi, D. V. (2021). Sistem Pengamanan File Menggunakan Algoritma Rc4 Berbasis Webbase Studi Kasus : Pt. Tjipta Jaya Bersama. *Semnas Ristek (Seminar Nasional Riset Dan Inovasi Teknologi)*, 5(1), 803–808.
- Widiyanto, W. W. (2018). ANALISA METODOLOGI PENGEMBANGAN SISTEM DENGAN PERBANDINGAN MODEL PERANGKAT LUNAK SISTEM INFORMASI KEPEGAWAIAN MENGGUNAKAN WATERFALL DEVELOPMENT MODEL, MODEL PROTOTYPE, DAN MODEL RAPID APPLICATION DEVELOPMENT (RAD). *Jurnal Informa : Jurnal Penelitian Dan Pengabdian Masyarakat*, 4(1), 34–40. <https://doi.org/10.46808/INFORMA.V4I1.34>
- Zebua, T., & Ndruru, E. (2017). Pengamanan Citra Digital Berdasarkan Modifikasi Algoritma RC4. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(4), 275–282. <https://doi.org/10.25126/jtiik.201744474>



## Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan *Start-up*

Adhitya Prayoga Permana <sup>(1)</sup>, Kurniyatul Ainiyah <sup>(2)\*</sup>, Khadijah Fahmi Hayati Holle <sup>(3)</sup>

Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim, Malang  
e-mail : {18650086,18650088}@student.uin-malang.ac.id, khadijah.holle@uin-malang.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 17 Juni 2021, direvisi 17 Agustus 2021, diterima 24 Agustus 2021, dan dipublikasikan 22 September 2021.

### Abstract

*Start-ups have a very important role in economic growth, the existence of a start-up can open up many new jobs. However, not all start-ups that are developing can become successful start-ups. This is because start-ups have a high failure rate, data shows that 75% of start-ups fail in their development. Therefore, it is important to classify the successful and failed start-ups, so that later it can be used to see the factors that most influence start-up success, and can also predict the success of a start-up. Among the many classifications in data mining, the Decision Tree, kNN, and Naive Bayes algorithms are the algorithms that the authors chose to classify the 923 start-up data records that were previously obtained. The test results using cross-validation and T-test show that the Decision Tree Algorithm is the most appropriate algorithm for classifying in this case study. This is evidenced by the accuracy value obtained from the Decision Tree algorithm, which is greater than other algorithms, which is 79.29%, while the kNN algorithm has an accuracy value of 66.69%, and Naive Bayes is 64.21%.*

**Keywords:** Classification, Decision Tree, kNN, Naive Bayes, Start-up

### Abstrak

*Start-up memiliki peran yang sangat penting dalam pertumbuhan ekonomi, adanya start-up dapat membuka banyak lapangan kerja baru. Namun, tidak semua start-up yang sedang berkembang dapat menjadi sebuah start-up yang sukses. Hal ini dikarenakan start-up memiliki tingkat kegagalan yang tinggi, data menunjukkan sebanyak 75% start-up mengalami kegagalan dalam perkembangannya. Oleh karena itu, penting untuk melakukan pengklasifikasian start-up sukses dan gagal, sehingga nantinya dapat digunakan untuk melihat faktor-faktor yang paling mempengaruhi keberhasilan start-up, dan juga dapat memprediksi keberhasilan suatu start-up. Di antara banyaknya metode klasifikasi dalam data mining, algoritma Decision Tree, kNN, dan Naive Bayes merupakan algoritma yang penulis pilih untuk melakukan klasifikasi terhadap 923 record data start-up yang telah didapatkan sebelumnya. Hasil pengujian menggunakan cross validation dan T-test menunjukkan algoritma Decision Tree merupakan algoritma paling tepat untuk melakukan klasifikasi dalam studi kasus ini. Hal ini dibuktikan dengan nilai akurasi yang diperoleh oleh algoritma Decision Tree lebih besar diantara algoritma lainnya, yaitu sebesar 79,29%, sedangkan algoritma kNN memiliki nilai akurasi 66,69%, dan Naive Bayes sebesar 64,21%.*

**Kata Kunci:** Decision Tree, Klasifikasi, kNN, Naive Bayes, Start-up

## 1. PENDAHULUAN

Perkembangan teknologi yang terjadi tentunya memberikan dampak dalam berbagai bidang, salah satunya ekonomi. Banyak perusahaan-perusahaan baru dalam bidang teknologi yang bermunculan dan menarik perhatian masyarakat, salah satunya *start-up*. *Start-up* merupakan organisasi yang dibangun untuk mendapatkan keuntungan yang maksimal dengan rancangan model bisnis yang tepat (Blank, 2013 dalam Afdi & Purwanggono, 2017). *Start-up* hadir dengan ide-ide dan inovasi baru yang dapat menciptakan banyak lapangan pekerjaan bagi masyarakat. Oleh karena itu, dapat dikatakan, *start-up* memiliki peran yang sangat penting dalam pertumbuhan ekonomi karena dapat menggerakkan roda perekonomian.



Di balik peran pentingnya, *start-up* memiliki tingkat kegagalan yang cukup tinggi. Menurut penelitian yang dilakukan Sikhar Ghosh, dosen senior di Harvard Business School, terhadap 2000 perusahaan dalam periode 2004 – 2010, penelitian dilakukan terhadap *start-up - start-up* yang telah menerima bantuan pendanaan, sebanyak 75% *start-up* mengalami kegagalan dalam perkembangannya (Ghosh, 2013). Kegagalan sebuah *start-up* tentunya akan menimbulkan kerugian bagi semua pihak yang berkontribusi, salah satunya yaitu investor. Oleh karena itu, para investor harus cermat dalam memilih dan melihat faktor-faktor yang mempengaruhi keberhasilan maupun kegagalan suatu *start-up*.

Tujuan utama investor adalah menemukan *start-up* yang memiliki potensi untuk berkembang pesat dan dapat memberikan keuntungan. Dengan ini, maka penulis memutuskan untuk melakukan proses klasifikasi terhadap faktor-faktor yang mempengaruhi keberhasilan maupun kegagalan sebuah *start-up*. Sehingga nantinya para investor dapat memprediksi *start-up* mana yang memiliki potensi besar untuk berhasil, dan juga melihat faktor apa saja yang sangat mempengaruhi keberhasilan *start-up*. Algoritma dalam *data mining* yang umum digunakan untuk melakukan klasifikasi yaitu Naïve Bayes, K-Nearest Neighbor, Decision Tree, Random Forest, dan Support Vector Machines (Wibisono & Fahrurrozi, 2019). Pada penelitian ini, penulis menggunakan beberapa metode klasifikasi yang kemudian akan dilakukan perbandingan terhadap nilai performa (akurasi, presisi, dan *recall*) yang dihasilkan masing-masing algoritma.

Sebelumnya, Rahman et al. (2018) telah melakukan penelitian yaitu melakukan komparasi antara algoritma kNN dengan Naive Bayes untuk mengklasifikasi kualitas air bersih. Hasilnya k-NN memiliki nilai rata-rata akurasi lebih tinggi, yaitu sebesar 82,42%, sedangkan Naive Bayes hanya mendapatkan nilai rata-rata akurasi sebesar 70,32%. Penelitian sejenis lainnya dilakukan oleh Marutho (2019) yaitu melakukan perbandingan antara metode Naive Bayes, k-NN, dan Decision Tree dalam studi kasus laporan level ketinggian air di Jakarta. Hasilnya menunjukkan dengan menggunakan evaluasi model *cross validation* dengan *k-fold* 10 menghasilkan nilai akurasi sebesar 96,56% untuk algoritma Decision Tree, kemudian Naïve Bayes sebesar 94,32%, dan terakhir K-NN sebesar 95,98%. Sehingga kesimpulannya adalah Decision Tree merupakan metode terbaik untuk diimplementasikan ke dalam dataset ketinggian air di Jakarta. Setiyorini & Asmono (2018) juga melakukan komparasi algoritma antara Decision Tree, Naïve Bayes, dan kNN untuk mengklasifikasi kinerja siswa. Hasil penelitian yang dilakukan pada *dataset student performance* ini menunjukkan algoritma kNN memiliki nilai akurasi tertinggi yaitu sebesar 79,31%, selanjutnya Decision Tree dengan nilai akurasi 78,85%, sedangkan algoritma Naïve Bayes nilai akurasinya sebesar 77,69%.

Penelitian lainnya dilakukan oleh Dellermann et al. (2018) yang memprediksi kesuksesan *start-up* di tahap awal melalui metode kecerdasan hibrida. Dalam penelitian ini penulis menggunakan *Hybrid Intelligence Method* di mana memperoleh kesimpulan bahwa manusia dapat melengkapi kelemahan ketika mesin gagal. Terutama mengenai informasi tersembunyi dan resiko yang tidak diketahui. Penelitian sejenis juga dilakukan oleh Glupker et al. (2019) melakukan prediksi kesuksesan investor menggunakan teori grafik dan pembelajaran mesin. Data yang digunakan bersumber dari basis *Crunch*, termasuk di dalamnya adalah karakteristik perusahaan rintisan, investor, dan individu di seluruh dunia di seluruh abad ke-20 dan hingga tahun 2013. Hasilnya akurasi keseluruhan yang diperoleh berkisar antara 55% dan 75%. Selanjutnya yaitu penelitian untuk identifikasi investor yang sukses di ekosistem *start-up* yang dilakukan oleh Gupta et al. (2015). Dalam penelitian ini, peneliti membuat *platform Investor Rank* dengan menggunakan pendekatan metode heuristik. Di mana hasilnya investor-investor potensial dapat diidentifikasi lebih dulu dan diverifikasi pengaruh jumlah pendanaan dan waktu pendanaan selama *start-up* berdiri.

Setiap metode yang klasifikasi memiliki kelebihan dan kekurangan masing-masing. Kelebihan yang dimiliki Decision Tree adalah sifatnya yang fleksibel sehingga mampu meningkatkan kualitas keputusan yang dihasilkan, sedangkan kekurangan dari algoritma ini adalah akan terjadi *overlap* jika menggunakan data yang memiliki kelas dan kriteria dengan jumlah yang sangat banyak. Sedangkan kelebihan dari metode Naïve Bayes adalah proses perhitungannya sederhana sehingga prosesnya lebih cepat dan efisien, namun dengan fakta bahwa masing-



masing variabel bersifat independen, hal ini dapat mengurangi besar akurasi yang dihasilkan. Terakhir adalah kNN, kelebihan metode ini yaitu dapat diterapkan pada data yang besar secara efektif dengan hasil yang akurat, namun kekurangannya yaitu membutuhkan biaya komputasi yang cukup tinggi karena harus melakukan perhitungan jarak pada setiap *query instance* secara bersama-sama.

Melihat penelitian-penelitian yang telah dilakukan sebelumnya, dan mempertimbangkan kelebihan dan kekurangan masing-masing metode, maka penulis memutuskan untuk melakukan perbandingan antara metode klasifikasi Decision Tree, Naive Bayes, dan kNN (K-Nearest Neighbor). Hal inilah yang juga menjadi keunikan dari penelitian ini, ketiga metode tersebut akan diimplementasikan kedalam dataset *start-up* sebanyak 923 *record* data dengan 19 atribut yang telah dipilih. Perbandingan ini dilakukan untuk menemukan algoritma terbaik yang dapat digunakan untuk melakukan klasifikasi *start-up*. Dataset tersebut akan dianalisis dengan menggunakan tahapan-tahapan yang ada di dalam proses CRISP-DM (*Cross Industry Standard Process for Data Mining*).

## 2. METODE PENELITIAN

### 2.1. Dataset

*Dataset* yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder merupakan data yang didapatkan secara tidak langsung, data-data tersebut dapat bersumber dari buku, jurnal, dokumentasi, literatur, dan sumber informasi lainnya yang berhubungan dengan topik penelitian (Sabna & Muhandi, 2016). Dataset *start-up* yang digunakan dalam penelitian ini diperoleh dari website <https://www.kaggle.com/manishkc06/startup-success-prediction>.

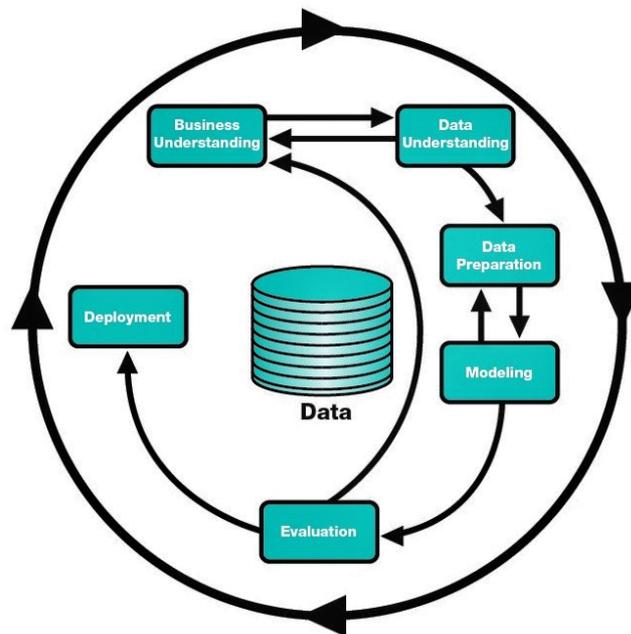
### 2.2. Variabel (Atribut) Penelitian

Variabel yang digunakan terdiri dari 19 atribut yang dipilih dari 49 atribut yang ada pada *dataset*. 19 atribut ini dipilih berdasar atribut yang paling signifikan karena ada beberapa atribut seperti *category\_code* yang dapat mewakili beberapa atribut lain seperti *is\_software*, *is\_web*, dll. 19 atribut yang dipilih yaitu 1d sebagai nilai primer, 17 atribut biasa dan 1 atribut sebagai label. 17 atribut biasa ini terdiri dari *Age\_first\_funding\_year* yaitu jumlah pendanaan pertama *start-up* tersebut, *Age\_last\_funding\_year* adalah jumlah pendanaan terakhir *start-up*, *Age\_first\_milestone\_year* merupakan rekor pendanaan pada waktu tersebut, *Age\_last\_milestone\_year* adalah rekor pendanaan pada waktu terakhir kali, *Relationships* menggambarkan jumlah relasi kerjasama yang dilakukan *start-up* tersebut, *Funding\_total\_usd* merupakan total pendanaan *start-up*, *Milestones* menandakan kategori *event* penting ke sekian pada saat *start-up* berdiri, *Category\_code* merupakan jenis kategori *start-up* tersebut. Sedangkan *Funding\_rounds*, *Has VC*, *Has angel*, *Has roundA*, *Has roundB*, *Has roundC*, *Has roundD*, merupakan tahapan perputaran pendanaan uang yang harus dilalui sebuah *start-up* untuk mencapai tahapan tertentu. *Avg\_participants* merupakan rata-rata peserta ataupun pendukung pelaku *start-up*, *Is top 500* menandakan apakah *start-up* tersebut termasuk 500 besar *start-up*, dan terakhir atribut yang berperan sebagai label adalah Status (*closed* dan *acquired*).

### 2.3. Proses Analisis Data

Metode yang digunakan dalam analisis data mengacu pada tahapan proses CRISP-DM dengan menggunakan *tools* RapidMiner. Pada Gambar 1 berikut ini merupakan tahapan dalam proses CRISP-DM.





Gambar 1. Alur Proses CRISP-DM.

### 2.3.1. Pemahaman Bisnis (*Business Understanding*)

Tahapan ini merupakan pemahaman terhadap permasalahan yang akan diteliti. Dalam penelitian ini permasalahan yang dibahas yaitu besarnya persentase kegagalan yang dimiliki sebuah *start-up* yang sedang berkembang, membuat investor atau tim pengembang *start-up* harus dapat mengetahui faktor apa saja yang paling berpengaruh dalam penentuan keberhasilan sebuah *start-up*. Selain itu, dengan dataset yang ada, investor juga dapat memprediksi *start-up* yang sedang berkembang dapat menjadi *start-up* sukses atau gagal.

### 2.3.2. Pemahaman Data (*Data Understanding*)

Data yang diperoleh dari *website kaggle.com* memiliki *record* sebanyak 923 data *start-up*. Atribut yang digunakan sebanyak 19, yaitu *Status*, *Id*, *Age first funding year*, *Age last funding year*, *Age first milestone year*, *Age last milestone year*, *Relationships*, *Funding rounds*, *Funding total usd*, *Milestones*, *Category code*, *Has VC*, *Has angel*, *Has roundA*, *Has roundB*, *Has roundC*, *Has roundD*, *Avg participants*, *Is top 500*. Data ini nantinya akan digunakan untuk melakukan prediksi keberhasilan *start-up*, sekaligus melihat faktor-faktor yang paling mempengaruhi keberhasilan tersebut.

### 2.3.3. Persiapan Data (*Data Preparation*)

Dari 923 data *start-up* memiliki 322 *missing value*, sehingga diperlukannya proses *data preparation* agar dapat memperbaiki kualitas data. *Missing value* dipecahkan dengan menambahkan data dengan nilai rata-rata pada atribut-atribut yang datanya hilang ataupun kosong (Praningki & Budi, 2018). Kemudian pada tahap ini juga menggunakan operator *filter example*, filter yang digunakan adalah *label is not missing* yang diterapkan pada atribut label.

### 2.3.4. Pemodelan (*Modeling*)

Tahapan pemodelan ini dilakukan untuk membangun data *training* dengan menggunakan pelatihan algoritma sehingga menghasilkan sejumlah aturan (Sabna & Muhardi, 2016). Pada penelitian ini, menggunakan algoritma Decision Tree, selain itu, algoritma lain yang digunakan untuk melakukan perbandingan, yaitu algoritma kNN dan Naive Bayes.



## 1) Decision Tree

Algoritma Decision Tree bersifat sangat kuat, populer, berbasis logika, dan mudah dipahami (Lakshmi et al., 2016). Hal yang menarik dari Decision Tree adalah penggunaan struktur pohon (*tree*) yang berfungsi untuk merepresentasikan aturan yang terbentuk dari hasil klasifikasi. Dalam *tree*, atribut direpresentasikan oleh sebuah *node*, dan kelas direpresentasikan oleh daun (*leaf*). Setiap pohon memiliki akar (*root*) yaitu *node* yang berada di paling atas (Anam & Santoso, 2018).

## 2) K-Nearest Neighbor

K-Nearest Neighbor merupakan salah satu metode klasifikasi dalam data mining yang termasuk ke dalam *supervised learning*. Pengklasifikasian yang dilakukan berdasarkan atribut dan data *training*, sehingga proses pengklasifikasian data baru dilakukan berdasarkan perbandingan kemiripan mayoritas pada data *training*. Dalam kNN nilai jarak ditentukan dengan pengujian data *testing* terhadap data *training* kemudian menggunakan nilai terkecil dari nilai ketetanggaan terdekat (Krisandi et al., 2015). Penghitungan jarak umumnya menggunakan jarak *Euclidean Distance* sebagai berikut.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

Di mana  $d(x_i, x_j)$  merupakan jarak Euclidean,  $x_i$  dan  $x_j$  adalah dua titik yang akan dihitung jaraknya, dengan  $x_i$  merupakan *record* data ke  $i$  dan  $x_j$  adalah *record* data ke  $j$ , serta  $a_r$  adalah data ke- $r$  dengan  $i, j$  bernilai  $1, 2, 3, \dots, n$ .

## 3) Naïve Bayes

Naïve Bayes merupakan algoritma yang berdasarkan teorema Bayes, di mana antar atributnya tidak memiliki hubungan atau ketergantungan, sehingga setiap atribut bersifat saling bebas. Klasifikasi Naïve Bayes merupakan metode klasifikasi yang menghitung probabilitas suatu kejadian berdasarkan kondisi tertentu, berikut persamaan dalam Naïve Bayes (Lukito & Chrismanto, 2015).

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)} \quad (2)$$

Dimana B adalah data yang akan dicari *class*-nya, sedangkan A merupakan hipotesis *class*.  $P(A|B)$  merupakan peluang terjadinya A terhadap kondisi B,  $P(A)$  merupakan peluang terjadinya A, dan  $P(B)$  adalah peluang terjadinya B.

### 2.3.5. Evaluasi (*Evaluation*)

Pada tahapan evaluasi, dilakukan pengujian menggunakan *cross validation* dari algoritma yang digunakan untuk klasifikasi. Hasil pengujian tersebut akan dibandingkan nilai akurasi, presisi, dan recall dari algoritma Decision Tree, kNN, dan Naive Bayes. Setelah itu, pengujian juga dilanjutkan dengan metode T-test untuk melihat algoritma yang terbaik.

## 1) Cross Validation

*Cross validation* adalah sebuah pengujian standar yang berfungsi untuk memprediksi *error rate*. Jumlah data training dan data testing yang mewakili setiap kelas harus memiliki jumlah yang sama, pembagian data dilakukan secara acak dengan perbandingan yang sama pada setiap kelasnya. Tingkat kesalahan pada setiap tingkat iterasi akan dihitung rata-ratanya sehingga dapat menghasilkan *error rate* keseluruhan (Hastuti, 2012).

## 2) T-Test

T-test merupakan metode pengujian hipotesis yang memperlakukan satu individu (objek penelitian) dengan dua perlakuan berbeda. Sampel yang berasal dari objek yang sama akan dibagi menjadi dua data, data pertama menggunakan perlakuan pertama, dan data kedua mendapat perlakuan kedua. Nilai *performance* didapatkan dengan membandingkan



kondisi objek pertama dan kondisi objek kedua (Hastuti, 2012). Rumus perhitungan pada T-test adalah sebagai berikut (Kadafi, 2018).

$$T_{hitung} = \frac{X - \mu_0}{\frac{\sigma}{\sqrt{n}}} \quad (3)$$

Di mana  $T_{hitung}$  adalah nilai T yang dicari dan menunjukkan standar deviasi pada distribusi normal (tabel t), kemudian X merupakan nilai rata-rata dari data yang diolah,  $\mu_0$  adalah rata-rata nilai yang menjadi sampel, dan s adalah standar deviasi dari populasi yang telah diketahui, dengan n adalah jumlah populasi yang digunakan dalam penelitian. Jika nilai uji T adalah lebih kecil dari  $\alpha$  maka  $H_0$  ditolak (Huda, 2013).

### 2.3.6. Penyebaran (Deployment)

Selanjutnya, hasil pemodelan atau hasil klasifikasi pada tahap sebelumnya dapat digunakan sebagai acuan untuk melakukan prediksi terhadap *start-up* yang sedang berkembang dapat menjadi *start-up* sukses atau gagal.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Data Understanding

Pembahasan mengenai hasil dari penelitian ini dimulai dari tahapan pertama dalam proses CRISP-DM yaitu *data understanding*. Gambar 2 berikut merupakan tampilan dari *dataset* yang digunakan dalam penelitian ini.

Row No.	status	id	age_first_fu...	age_last_fu...	age_first_m...	age_last_mi...	relationships	funding_rou...	funding_tota...	milestones
1	acquired	c:6669	2.249	3.003	4.668	6.704	3	3	375000	3
2	acquired	c:16283	5.126	9.997	7.005	7.005	9	4	40100000	1
3	acquired	c:65620	1.033	1.033	1.458	2.205	5	1	2600000	2
4	acquired	c:42668	3.131	5.315	6.003	6.003	5	3	40000000	1
5	closed	c:65806	0	1.669	0.038	0.038	2	2	1300000	1
6	closed	c:22898	4.545	4.545	5.003	5.003	3	1	7500000	1
7	acquired	c:16191	1.720	5.211	3	6.608	6	3	26000000	2
8	acquired	c:5192	1.647	6.762	5.606	7.362	25	3	34100000	3
9	acquired	c:1043	3.586	11.112	8.005	9.995	13	3	9650000	4
10	acquired	c:498	1.671	4.685	2.918	6.115	14	3	5750000	4
11	acquired	c:3949	4.627	9.449	10.134	10.649	22	3	27500000	3
12	closed	c:4829	1.085	5.337	-0.616	4.608	8	5	10400000	2
13	closed	c:30290	4.904	4.904	?	?	0	1	350000	0
14	acquired	c:1491	0.019	2.436	0.794	4.378	15	3	9950000	3
15	acquired	c:15645	4.666	8.997	8.838	8.838	12	5	10700000	1
16	closed	c:54177	6.608	6.608	?	?	0	1	200000	0
17	closed	c:16770	2.586	6.764	5.501	5.501	8	5	49000000	1
18	acquired	c:107	4.592	7.173	-0.499	12.680	7	4	25000000	3
19	acquired	c:50727	0.743	1.581	1.285	3.003	10	3	4575000	3

Gambar 2. *Dataset Start-up.*

### 3.2. Data Preparation

Tahapan selanjutnya yaitu *data preparation*, pada tahap ini dilakukan pembersihan data menggunakan operator *Replace Missing Value* yang ada di dalam RapidMiner, sehingga akan menghasilkan data yang bersih dari *missing value*. Dengan operator ini *missing value* dipecahkan dengan menambahkan data dengan nilai rata-rata. Tetapi operator *Replace Missing Value* tidak dapat merubah data dari label yang kosong, sehingga diperlukan lagi satu operator yaitu *Filter Example*, dengan filter label *is not missing*. Hasil dari tahapan data preparation ini dapat dilihat pada Gambar 3 berikut tepatnya pada kolom *Missing*, nilai yang ada di dalam kolom tersebut



adalah 0 untuk setiap data atributnya yang menandakan sudah tidak ada data yang terdapat *missing value* di dalamnya.

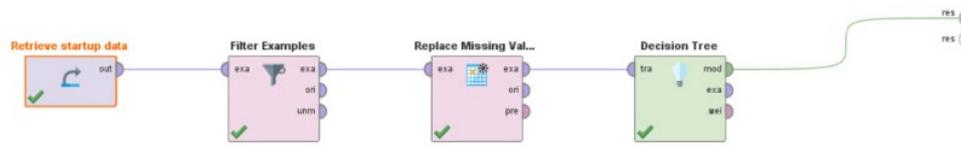
Name	Type	Missing	Statistics	Filter (19 / 19 attributes)	Search for Attributes
status	Polynomial	0	Least closed (326)	Most acquired (596)	Values acquired (596), closed (326)
id	Polynomial	0	Least c:163104 (0)	Most c:28482 (2)	Values c:28482 (2), c:10054 (1), ...[920 more]
age_first_funding_year	Real	0	Min -9.047	Max 21.896	Average 2.238
age_last_funding_year	Real	0	Min -9.047	Max 21.896	Average 3.936
age_first_milestone_year	Real	0	Min -14.170	Max 24.685	Average 3.059
age_last_milestone_year	Real	0	Min -7.005	Max 24.685	Average 4.759
relationships	Integer	0	Min 0	Max 63	Average 7.710
funding_rounds	Integer	0	Min 1	Max 10	Average 2.312
funding_total_usd	Integer	0	Min 11000	Max 5700000000	Average 25445421.271
milestones	Integer	0	Min 0	Max 8	Average 1.842
category_code	Polynomial	0	Least sports (1)	Most software (153)	Values software (153), web (144), ...[33 more]
has_VC	Integer	0	Min 0	Max 1	Average 0.326
has_angel	Integer	0	Min 0	Max 1	Average 0.255
has_roundA	Integer	0	Min 0	Max 1	Average 0.509
has_roundB	Integer	0	Min 0	Max 1	Average 0.393
has_roundC	Integer	0	Min 0	Max 1	Average 0.233
has_roundD	Integer	0	Min 0	Max 1	Average 0.100
avg_participants	Real	0	Min 1	Max 16	Average 2.836
is_top500	Integer	0	Min 0	Max 1	Average 0.609

Gambar 3. Hasil *Data Preparation*.

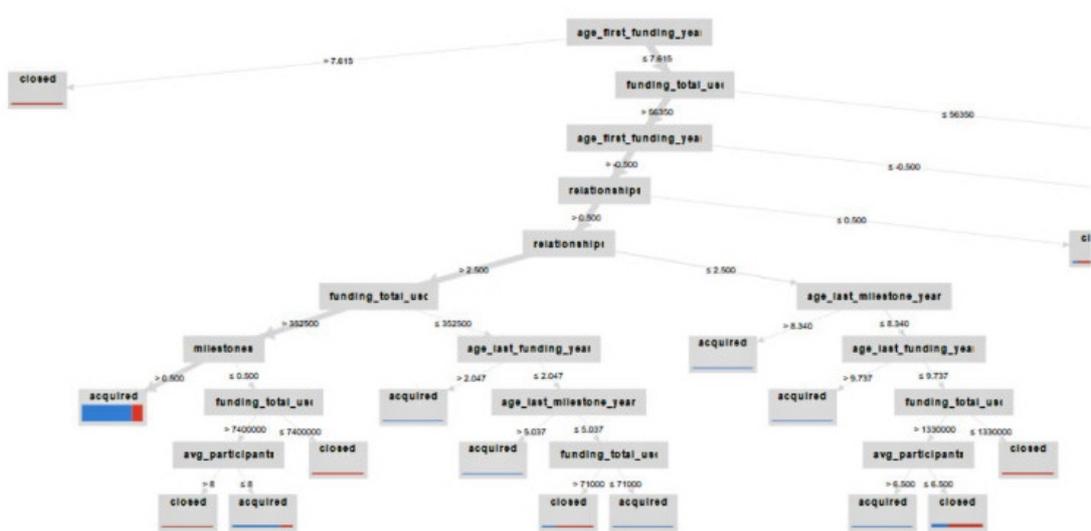
### 3.3. Modeling

Kemudian memasuki tahapan selanjutnya yaitu *modeling*, pada tahap ini data training diklasifikasikan menggunakan algoritma Decision Tree. Hasil pembentukan model ini dapat diketahui faktor apa saja yang sangat mempengaruhi kesuksesan sebuah *start-up*. Pemodelan yang dilakukan pada penelitian ini dapat dilihat pada Gambar 4 berikut, kemudian hasil klasifikasinya yang berupa *tree* dapat dilihat pada Gambar 5.





Gambar 4. Modeling.

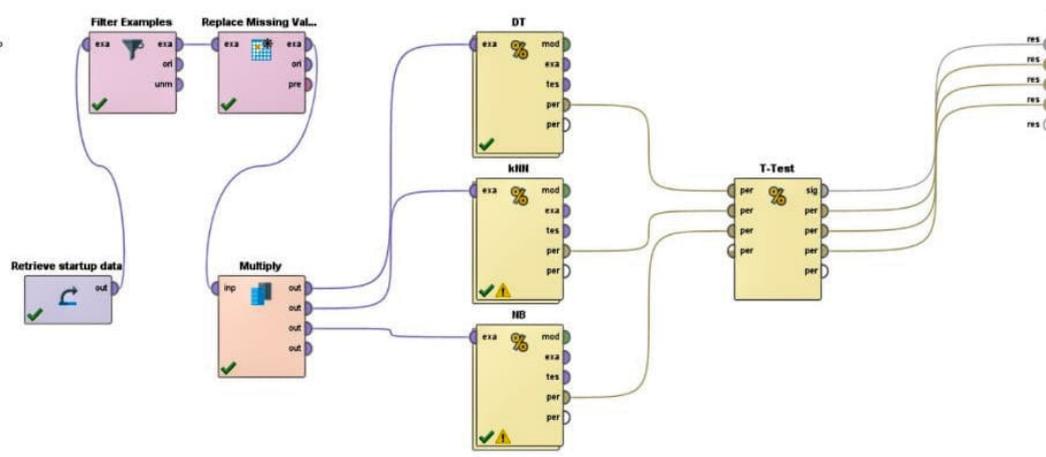


Gambar 5. Tree Hasil Modeling.

Dari hasil *tree* diatas, dapat ditarik kesimpulan bahwa faktor-faktor penting yang mempengaruhi kesuksesan sebuah *start-up* adalah *age\_first\_funding\_year*, *total\_funding*, serta *relationship*. Variabel *age\_first\_funding\_year*, *total\_funding*, serta *relationship* yang semakin besar maka semakin besar pula kesempatan sebuah *start-up* tersebut akan semakin sukses.

### 3.4. Evaluation

Tahap berikutnya yaitu evaluasi, pada tahap ini melakukan perbandingan hasil pengujian algoritma Decision Tree, kNN, dan Naive Bayes menggunakan *cross validation* dengan 10-fold *cross validation* dan dilanjutkan dengan pengujian T-test dengan  $\alpha = 0,05$ . Desain dari model pengukuran *performance* dapat dilihat pada Gambar 6 dibawah ini. Sedangkan hasil perbandingan nilai *performance* menggunakan *cross validation* dan T-test ada pada Tabel 1 dan Tabel 2 berikut.



Gambar 6. Desain Model Pengujian.



Tabel 1. Hasil Pengujian dengan *Cross Validation*.

No.	Algoritma	Akurasi	Presisi	Recall
1	Decision Tree	79,29%	78,99%	56,27%
2	kNN	66,69%	55,13%	40,14%
3	Naive Bayes	64,21%	51,32%	79,16%

Dari hasil perbandingan performa akurasi, terlihat algoritma Decision Tree memiliki nilai persentase paling tinggi yaitu 79,29%, sedangkan algoritma kNN dengan 66,69%, dan Naive Bayes dengan 64,21%. Pada performa presisi, algoritma Decision Tree masih menunjukkan nilai terbaik dengan 78,99%, diikuti algoritma kNN dengan 55,13%, dan Naive Bayes 51,32%. Dari hasil performa *recall*, ternyata algoritma Naive Bayes menunjukkan hasil paling baik dengan 79,16%, sedangkan Decision Tree 56,27% dan kNN dengan 40,14%. Dari keseluruhan hasil belum ada yang melebihi 80% sehingga belum bisa dikatakan baik, akan tetapi algoritma Decision Tree memiliki 2 nilai tertinggi pada performa akurasi dan presisi. Sehingga dari ketiga algoritma tersebut pada pengujian dengan *cross validation* ini, algoritma Decision Tree adalah pilihan terbaik.

Tabel 2. Hasil Pengujian dengan T-test.

	DT	kNN	NB
DT		0,000	0,000
kNN	0,000		0,484
NB	0,000	0,484	

Berdasarkan tabel hasil T-test di atas, menunjukkan hasil bahwa algoritma Decision Tree merupakan algoritma yang paling dominan terhadap algoritma yang lain. Suatu algoritma bisa dikatakan dominan apabila nilai hasil perbandingan performa T-test dengan algoritma lain lebih kecil daripada nilai *alpha*, sedang nilai *alpha* disini adalah 0,050. Sehingga berdasarkan perbandingan nilai performa dan hasil T-test, dapat disimpulkan bahwa algoritma Decision Tree adalah algoritma terbaik yang dapat digunakan pada studi kasus pengklasifikasian *start-up*.

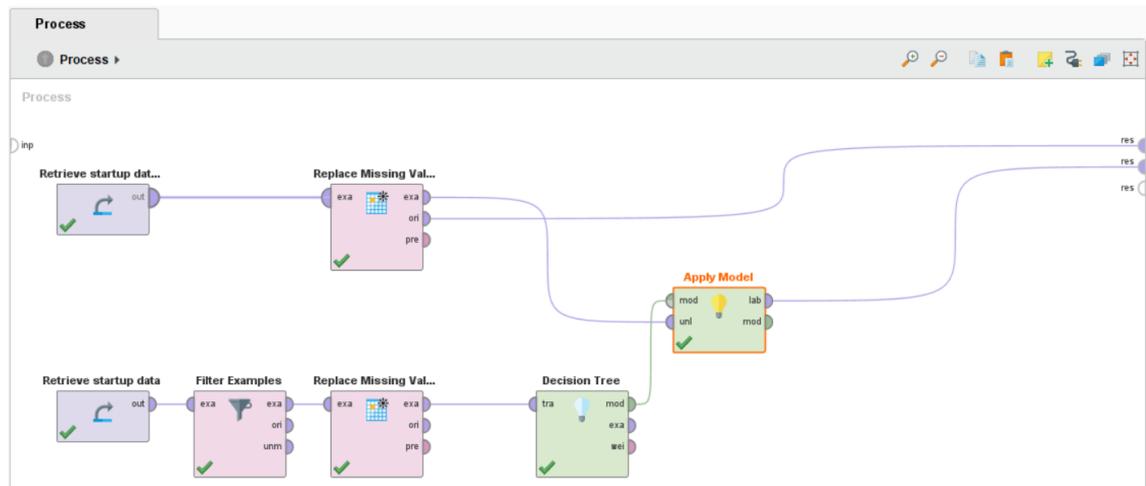
### 3.5. Deployment

Tahapan terakhir yaitu *deployment*, pada tahap ini model yang telah dibentuk pada tahapan *modeling* digunakan untuk melakukan prediksi keberhasilan *start-up*. Untuk data *testing* yang akan digunakan dapat dilihat pada Gambar 7. Desain model untuk melakukan prediksi dapat dilihat pada Gambar 8 berikut, kemudian perhatikan Gambar 9 di bawah ini untuk hasil prediksi yang telah dilakukan.

id_tota...	milestones	category_co...	has_VC	has_angel	has_roundA	has_roundB	has_roundC	has_roundD	avg_particip...	is_top500	status ↓
i000	2	software	1	0	1	1	0	0	1.667	1	?
i000	2	web	1	1	1	0	0	0	1.750	1	?
i0	0	public_relatio...	1	0	0	0	0	0	1	0	?
i00	1	web	0	0	1	0	0	0	2	0	?
i0	2	mobile	0	1	0	0	0	0	1	0	?
i000	1	public_relatio...	1	0	0	0	0	1	5	1	?
i00	2	cleantech	0	0	0	1	0	0	1	1	?
i0	2	social	0	1	0	0	0	0	3	1	?
i00	1	web	0	0	1	0	0	0	1	0	?
i000	1	software	1	0	0	0	1	0	2.500	1	?
i000	0	biotech	0	0	0	0	0	1	6	1	?
i0	0	manufacturing	0	0	1	0	0	0	3	0	?

Gambar 7. Data Testing.





Gambar 8. Desain Model Prediksi.

Row No.	prediction(s...	confidence(...	confidence(...	id	age_first_fu...	age_last_fu...	age_first_m...	age_last_mi...	relationships	funding_rou...	funding_...
1	acquired	0.825	0.175	c:16191	1.720	5.211	3	6.608	6	3	2600000
2	acquired	0.825	0.175	c:4829	1.085	5.337	-0.616	4.608	8	5	1040000
3	closed	0.097	0.903	c:54177	6.608	6.608	2.081	4.229	0	1	200000
4	closed	0.326	0.674	c:15888	5.490	5.490	0	0	1	1	3170000
5	closed	0.273	0.727	c:149809	0	0.584	0.584	0.595	9	2	125000
6	acquired	0.825	0.175	c:22027	5.186	6.814	7.334	7.334	9	2	2900000
7	acquired	0.825	0.175	c:25325	2.644	2.644	0	7.392	4	1	5000000
8	acquired	0.825	0.175	c:70586	4.285	4.285	2.416	5.515	7	1	400000
9	acquired	0.825	0.175	c:34028	1.956	1.956	1.003	1.003	5	1	5000000
10	acquired	0.825	0.175	c:45611	5.685	6.315	5.005	5.005	4	2	1505000
11	closed	0.326	0.674	c:31754	4.238	4.238	2.081	4.229	1	1	5000000
12	closed	0.097	0.903	c:158294	4.370	4.370	2.081	4.229	0	1	662000

Gambar 9. Hasil Prediksi.

#### 4. KESIMPULAN

Berdasarkan hasil perbandingan antara algoritma Decision Tree, kNN, dan Naive Bayes, untuk melakukan klasifikasi terhadap 923 data *start-up*, menunjukkan algoritma Decision Tree merupakan algoritma yang paling cocok untuk digunakan di antara algoritma kNN dan Naive Bayes. Hasil akurasi Decision Tree adalah sebesar 79,29%, sedangkan algoritma kNN dengan 66,69%, dan Naive Bayes dengan 64,21%. Selanjutnya untuk nilai presisinya, Decision Tree masih lebih unggul dengan nilai 78,99%, diikuti algoritma kNN dengan 55,13%, dan Naive Bayes 51,32%. Dari hasil performa *recall*, ternyata algoritma Naive Bayes menunjukkan hasil paling baik dengan 79,16%, sedangkan Decision Tree 56,27% dan kNN dengan 40,14%. Hasil pengujian T-test juga menunjukkan algoritma Decision Tree adalah algoritma paling dominan di antara algoritma yang lain. Selain itu, faktor-faktor yang sangat mempengaruhi kesuksesan sebuah *start-up* adalah *age\_first\_funding\_year*, *total\_funding*, serta *relationship*. Variabel *age\_first\_funding\_year*, *total\_funding*, serta *relationship* yang semakin besar maka semakin besar pula kesempatan sebuah *start-up* tersebut akan sukses.

Dikarenakan pada penelitian ini penulis melakukan seleksi fitur/atribut secara manual, maka penulis menyarankan bagi penelitian selanjutnya untuk menggunakan metode *feature selection* untuk mengoptimalkan jumlah *feature/attributes* sehingga dapat meningkatkan nilai akurasi yang diperoleh. Selain itu penulis juga menyarankan penelitian selanjutnya untuk melakukan



percobaan dengan metode klasifikasi lain seperti Support Vector Machine, Neural Network, dan sebagainya.

## DAFTAR PUSTAKA

- Afdi, Z., & Purwanggono, B. (2017). Perancangan Strategi berbasis Metodologi Lean Startup untuk Mendorong Pertumbuhan Perusahaan Rintisan berbasis Teknologi di Indonesia. *Industrial Engineering Online*, 6(4), 1–13.
- Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Jurnal Ilmiah Ilmu-Ilmu Teknik*, 8(1), 13–19.
- Blank, S. (2013, May). *Why the Lean Start-Up Changes Everything*. Harvard Business Review. <https://hbr.org/2013/05/why-the-lean-start-up-changes-everything>
- Dellermann, D., Ebel, P., Lipusch, N., Popp, K. M., & Leimeister, J. M. (2017). Finding the Unicorn: Predicting Early Stage Startup Success Through a Hybrid Intelligence Method. *International Conference on Information Systems (ICIS)*, 1–12. <https://doi.org/https://dx.doi.org/10.2139/ssrn.3159123>
- Glupker, J., Nair, V., Richman, B., Riener, K., & Sharma, A. (2019). Predicting investor success using graph theory and machine learning. *Journal of Investment Management*, 17(1), 92–103.
- Gupta, S., Pienta, R., Tamersoy, A., Chau, D. H., & Basole, R. C. (2015). Identifying Successful Investors in the Startup Ecosystem. *Proceedings of the 24th International Conference on World Wide Web*, 39–40. <https://doi.org/10.1145/2740908.2742743>
- Hastuti, K. (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif. *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012*, 14(1), 241–249.
- Huda, F. A. (2013). *t-Test*.
- Kadafi, A. R. (2018). Perbandingan Algoritma Klasifikasi Untuk Penjurusan Siswa SMA. *Jurnal ELTIKOM*, 2(2), 67–77. <https://doi.org/10.31961/eltikom.v2i2.86>
- Krisandi, N., Helmi, & Prihando, B. (2015). Acute toxicity of zinc oxide nanoparticles and bulk ZnCl<sub>2</sub> to rats. In *Information Technology* (Vol. 2, Issue 1, pp. 123–126). CRC Press. <https://doi.org/10.1201/b18776-23>
- Lakshmi, B. N., Indumathi, T. S., & Ravi, N. (2016). A Study on C.5 Decision Tree Classification Algorithm for Risk Predictions During Pregnancy. *Procedia Technology*, 24, 1542–1549. <https://doi.org/10.1016/j.protcy.2016.05.128>
- Lukito, Y., & Chrismanto, A. R. (2015). Perbandingan Metode-Metode Klasifikasi untuk Indoor Positioning System. *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(2), 123–131. <https://doi.org/10.28932/jutisi.v1i2.373>
- Marutho, D. (2019). Perbandingan Metode Naive Bayes , KNN , Decision Tree Pada Laporan Water Level Jakarta. *Manajemen Informatika AMIK JTC Semarang*, 15(2), 90–97. <https://doi.org/https://doi.org/10.53845/infokam.v15i2.175>
- Praningki, T., & Budi, I. (2018). Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN. *Creative Information Technology Journal*, 4(2), 83. <https://doi.org/10.24076/citec.2017v4i2.100>
- Rahman, M. A., Hidayat, N., & Afif Supianto, A. (2018). Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naive Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(12), 6346–6353.
- Sabna, E., & Muhardi, M. (2016). Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, dan Hasil Belajar. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 2(2), 41. <https://doi.org/10.24014/coreit.v2i2.2392>
- Setiyorini, T., & Asmono, R. T. (2018). Komparasi Metode Decision Tree, Naive Bayes Dan K-Nearest Neighbor Pada Klasifikasi Kinerja Siswa. *Jurnal Techno Nusa Mandiri*, 15(2), 85. <https://doi.org/10.33480/techno.v15i2.889>
- Wibisono, A. B., & Fahrurozi, A. (2019). PERBANDINGAN ALGORITMA KLASIFIKASI DALAM PENGKLASIFIKASIAN DATA PENYAKIT JANTUNG KORONER. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 24(3), 161–170. <https://doi.org/10.35760/tr.2019.v24i3.2393>



## Perbandingan Faktor-Faktor Yang Mempengaruhi Penggunaan *Electronic-Know Your Customer (e-KYC)*

Mr. Fitree Tahe <sup>(1)</sup>, Maria Ulfah Siregar <sup>(2)\*</sup>, Mohammad Farhan Qudratullah <sup>(3)</sup>

<sup>1,2</sup>Magister Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga, Yogyakarta

<sup>2</sup>Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga, Yogyakarta

<sup>3</sup>Matematika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga, Yogyakarta

e-mail : fitreetah04@gmail.com, {maria.siregar,mohammad.qudratullah}@uin-suka.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 2 September 2021, direvisi 19 September 2021, diterima 20 September 2021, dan dipublikasikan 22 September 2021.

### Abstract

*There are many technological developments in banks, one of which is online transactions. To get these transactions, an account should be opened using the electronic know you customer (e-KYC) verification system at banks. This research wants to know the differences in the factors that influence behavioral intentions to use e-KYC at SCB (The Siam Commercial Bank) Thailand and Bank Mandiri Indonesia. This is a quantitative research using a survey. The researcher has prepared a questionnaire of 160 respondents: 80 for Bank Mandiri and 80 for SCB. The results indicate that the willingness of Bank Mandiri Indonesia's respondents to use electronic identity verification services is influenced by smaller factors compared to SCB Thailand's respondents. In total, all of the factors defined in this research is influenced the willingness.*

**Keywords:** *Online Transaction, Verification System, Bank Account, Behavioral Intentions, Questionnaire*

### Abstrak

Terdapat banyak perkembangan teknologi di bank salah satunya adalah transaksi online. Untuk melaksanakan transaksi tersebut maka rekening di suatu bank perlu dibuka dengan menggunakan sistem verifikasi *electronic know you customer (e-KYC)*. Penelitian ini ingin mengetahui perbedaan faktor-faktor yang mempengaruhi niat perilaku untuk menggunakan e-KYC di Bank SCB (The Siam Commercial Bank) dan Bank Mandiri Indonesia. Metode penelitian ini adalah kuantitatif dengan menggunakan metode survei. Peneliti telah menyiapkan kuesioner untuk 160 responden: 80 untuk Bank Mandiri dan 80 untuk SCB. Hasil penelitian ini adalah bahwa kesediaan responden Bank Mandiri Indonesia untuk menggunakan layanan verifikasi identitas elektronik dipengaruhi oleh lebih sedikit faktor dibandingkan dengan responden SCB Thailand. Secara total, keseluruhan faktor yang digunakan di dalam penelitian ini mempengaruhi kesediaan tersebut.

**Kata Kunci:** *Transaksi Online, Sistem Verifikasi, Akun Bank, Behavioral Intentions, Kuesioner*

## 1. PENDAHULUAN

Pembukaan rekening untuk menerima setoran atau menerima uang dari warga negara adalah titik awal di suatu lembaga keuangan untuk memberikan layanan keuangan kepada nasabah. Lembaga keuangan haruslah memiliki proses untuk mengetahui pelanggan (*Know Your Customer: KYC*) yang dapat diidentifikasi dan verifikasi identitas pelanggan secara akurat dan efektif yang akan mendukung sistem lembaga keuangan memiliki kredibilitas dan stabilitas yang berkelanjutan.

Terkait layanan e-KYC, diperlukan adanya kesiapan nasabah dalam menerima implementasi teknologi di dalam perbankan. *Technology Readiness* atau tingkat kesiapan mengacu kepada kecenderungan seseorang dalam menggunakan dan memanfaatkan teknologi baru dalam mencapai tujuan mereka baik dalam kehidupan sehari-hari, maupun dalam dunia pekerjaan (Parasuraman & Colby, 2015). Infiltrasi teknologi ke dalam dunia perbankan memberikan tanggung jawab yang mengikuti yaitu melindungi keamanan pelanggan dalam transaksi *online*.



Hal ini mempengaruhi keamanan dan kepercayaan pelanggan. Sehingga perlunya ada perlindungan yang secara teknis dibagi menjadi dua: perlindungan keamanan dan perlindungan privasi (Shin, 2010). Penggunaan teknologi ini menghadirkan eksternalitas jaringan dari *tools* yang digunakan pengguna yaitu peningkatan nilai yang dirasakan atas individu dalam mengkonsumsi suatu produk atau layanan (Katz & Shapiro, 1994). Proses transaksi elektronik adalah berbeda dengan transaksi tradisional karena infrastruktur transaksi pada dasarnya berbeda. Isu yang muncul adalah keamanan, dan kekhawatiran tentang penggunaannya (Linck et al., 2006). Kesadaran akan keamanan adalah mengacu pada tingkat kepercayaan yang dimiliki pengguna layanan terhadap penyedia layanan: Jika tingkat keamanan yang dirasakan dalam suatu transaksi keuangan terlalu rendah, konsumen tidak akan berpartisipasi dalam transaksi tersebut (C. Kim et al., 2010). Kepercayaan (*trust*) mengacu pada keyakinan seseorang untuk menyesuaikan diri dengan tindakan dari pihak layanan. Meskipun pengguna layanan menggunakan sistem yang tidak sempurna, namun pengguna layanan yakin bahwa bank tidak akan menyalahgunakan data pribadi (Abrazhevich, 2004; Mayer et al., 1995). Minat untuk menggunakan merupakan kecenderungan perilaku dari pengguna untuk tetap menggunakan suatu teknologi yang diberikan. Dalam melakukan transaksi online ditemukan bahwa kepercayaan pada sistem mempengaruhi niat menggunakannya, ketika pengguna layanan memiliki kepercayaan pada sistem yang tinggi (D. J. Kim et al., 2008; Lin et al., 2005). Layanan teknologi yang diberikan suatu bank haruslah berkualitas dan difokuskan pada nasabah, layanan, kualitas serta level (Wijatmoko & Siregar, 2020). Sehingga bisa dikatakan bahwa kehidupan manusia sangat didukung oleh kemajuan teknologi (Ardiansyah & Fatwanto, 2020), tidak terlepas adalah dalam transaksi perbankan .

SCB (The Siam Commercial Bank) merupakan bank komersial pertama dan salah satu dari lima bank terbesar di Thailand. Didirikan oleh Bantuan Kerajaan Raja Chulalongkorn pada tahun 1906, sekarang memiliki 1.070 cabang. SCB juga memiliki *e-KYC*.

Mandiri adalah bank yang berkantor pusat di Jakarta, dan merupakan bank terbesar di Indonesia dalam hal aset, pinjaman, dan deposit. Bank ini berdiri pada tanggal 2 Oktober 1998 sebagai bagian dari program restrukturisasi perbankan yang dilaksanakan oleh Pemerintah Indonesia.

Kedua perusahaan tersebut terus melakukan pengembangan untuk meningkatkan kepuasan nasabahnya. Adanya kepuasan nasabah adalah upaya untuk mempertahankan nasabah agar tetap menggunakan jasa layanan. Umumnya pengguna akan tetap menggunakan suatu sistem jika sistem tersebut sesuai dengan keinginan pengguna (Ardiani, 2020). Berdasarkan latar belakang tersebut penulis tertarik dan hendak mengetahui perbedaan faktor-faktor yang mempengaruhi niat perilaku untuk menggunakan *e-KYC* untuk membuka rekening di Bank SCB dan Bank Mandiri.

## 2. METODE PENELITIAN

Penelitian ini adalah penelitian kuantitatif dengan menggunakan metode survei dari kelompok sampel melalui kuesioner. Langkah-langkah penelitian adalah sebagai berikut:

### 2.1. Pembagian Responden

Jumlah populasi yang besar menyebabkan peneliti menggunakan *purposive sampling* dalam menentukan responden. Sehingga responden adalah mereka yang berniat menggunakan layanan sistem verifikasi identitas elektronik untuk membuka rekening dan yang tertarik dengan sistem *e-KYC*. Dengan digunakannya *purposive sampling* diharapkan data yang diperoleh adalah data yang baik, yang mendukung tercapainya tujuan dari penelitian. Responden dibagi menjadi 2 yaitu dari Indonesia dan dari Thailand.

Peneliti menentukan ukuran sampel dengan menggunakan rasio antara unit sampel dengan jumlah pertanyaan (Thangrattanasuwan, 2017). Pada penelitian ini, unit sampel ada 5, yaitu kriteria pembobotan dari Sangat Setuju sampai Sangat Tidak Setuju; jumlah pertanyaan ada 32 yang diadaptasi dari berbagai sumber relevan (ditunjukkan di Tabel 1). Sehingga, peneliti peroleh sebagai berikut:



5 (unit sampel) x 32 (jumlah pertanyaan).

$$n = 160 \text{ sampel}$$

Sehingga, sampel yang ideal adalah 160 buah. Berdasarkan jumlah sampel di atas, peneliti membagi sampel menjadi 80 responden untuk analisis faktor yang mempengaruhi penggunaan e-KYC di Bank Mandiri Indonesia dan 80 responden untuk analisis yang sama di Bank Siam Commercial Thailand. Sehingga teknik sampling yang digunakan adalah kombinasi *purposive sampling* dan *quota sampling* (Setiawan, 2005).

Penentuan jumlah sampel di atas adalah sesuai dengan yang disampaikan oleh Borg dan Gall (Alwi, 2015) yaitu di dalam penelitian eksperimen dan komparatif, diperlukan sampel sejumlah 15-30 untuk setiap kelompok. Penelitian yang diajukan adalah komparatif, populasi sangat besar, sehingga peneliti mendapatkan jumlah sampel 160 yang dibagi ke dalam dua kelompok.

## 2.2. Instrumen Penelitian

Agar kuesioner sesuai dengan kerangka konseptual penelitian ini dan dapat menjawab tujuan penelitian secara lengkap, maka kuesioner ini disiapkan dengan mengacu pada tinjauan pustaka, teori dan konsep terkait. Tabel 1 menunjukkan sumber yang dirujuk dalam penyusunan instrumen:

**Tabel 1. Sumber Penyusunan Instrumen.**

Variabel/ Faktor	Sumber
Kesiapan Teknologi ( <i>Technology Readiness</i> )	(Parasuraman & Colby, 2015)
Efek Jaringan ( <i>Network Externalities</i> )	(Katz & Shapiro, 1994)
Perlindungan Teknis ( <i>Technical Protection</i> )	(Shin, 2010)
Tata Cara Transaksi ( <i>Transaction Procedures</i> )	(Linck et al., 2006)
Keamanan yang Dirasakan ( <i>Perceived Security</i> )	(C. Kim et al., 2010)
Kepercayaan yang Dirasakan ( <i>Perceived Trust</i> )	(Abrazhevich, 2004; Mayer et al., 1995)
Niat untuk Menggunakan ( <i>Intention to Use</i> )	(Lin et al., 2005)

Peneliti melakukan survei selama 1 bulan yang dibagi menjadi 3 bagian sebagai berikut:

**Bagian 1:** Penyaringan pertanyaan untuk target audien yang memiliki niat untuk menggunakan layanan sistem identifikasi elektronik dan tertarik untuk menggunakannya.

**Bagian 2:** Pertanyaan tentang informasi umum responden tentang penggunaan sistem identifikasi elektronik dengan karakteristik kuesioner pilihan ganda (*multiple choices*), seperti jenis kelamin, usia tingkat pendidikan, pekerjaan, pendapatan rata-rata, bank yang dipilih alasan memilih layanan sistem identifikasi elektronik.

**Bagian 3:** Informasi kesiapan teknologi sebagai efek eksternal jaringan, keselamatan teknis proses transaksi, kesadaran akan keamanan dan keyakinan yang mempengaruhi niat untuk menggunakan layanan sistem identifikasi elektronik.

Kuesioner tersebut memiliki *rating scaled* dengan kriteria penentuan bobot penilaian menjadi 5 tingkatan sebagai berikut:

Skor 1: sangat tidak setuju.

Skor 2: tidak setuju.

Skor 3: cukup setuju.

Skor 4: setuju.

Skor 5: sangat setuju.

Untuk penelitian ini, kriteria evaluasi ditetapkan. Peneliti menggunakan prinsip pemisahan interval interpretasi menurut prinsip interval kelas.

$$\frac{\text{Skor Tertinggi} - \text{Skor Terendah}}{\text{Banyak kelas}} = \frac{5 - 1}{5} = 0.8$$



Oleh karena itu, kriteria penilaian rata-rata ditetapkan pada berbagai tingkat opini sebagai berikut:

- Skor rata-rata 4.21-5.00 berarti pengguna jasa setuju dengan berbagai faktor pada level tertinggi.
- Skor rata-rata 3.41-4.20 berarti pengguna jasa setuju dengan berbagai faktor pada level tinggi.
- Skor rata-rata 2.61-3.40 berarti pengguna jasa setuju dengan berbagai faktor pada level sedang.
- Skor rata-rata 1.81-2.60 berarti pengguna jasa setuju dengan berbagai faktor pada level rendah.
- Skor rata-rata 1.00-1.80 berarti bahwa pengguna jasa setuju dengan berbagai faktor pada level terendah.

### 2.3. Pengumpulan Data

Pengumpulan data dibagi menjadi 2 bagian sebagai berikut:

- 1) Data primer (*primary data*) adalah kumpulan data yang diperoleh dari kuesioner yang bermaksud menggunakan layanan untuk analisis data lebih lanjut.
- 2) Data sekunder (*secondary data*) adalah kumpulan data yang diperoleh dari studi penelitian, tinjauan literatur, konsep, teori dan penelitian terkait, termasuk pencarian dari Internet dan dokumen terkait, dan lain-lain.

### 2.4. Analisis Data

Setelah mengumpulkan kuesioner, peneliti menganalisis data penelitian. Analisis data yang dilakukan adalah analisis statistik deskriptif. Analisis statistik deskriptif (*Descriptive Statistics*) menunjukkan karakteristik data umum responden dan menyajikan data dalam berbagai bentuk seperti tabel distribusi frekuensi, persentase, mean, dan standar deviasi dengan serta penjelasannya. Peneliti tidak bisa menggunakan analisis dengan statistika inferensial disebabkan teknik sampling yang peneliti gunakan adalah gabungan *Purposive Sampling* dan *Quota Sampling* yang termasuk ke dalam *non-probabilistic sampling* (Setiawan, 2005).

## 3. HASIL DAN PEMBAHASAN

### 3.1. Faktor yang Mempengaruhi Niat Menggunakan e-KYC Bank Mandiri Indonesia

#### 3.1.1. Uji Realibilitas

Tabel 2 menunjukkan keandalan alat (*reliability*) dari instrumen responden Bank Mandiri Indonesia. Keandalan alat diverifikasi dengan menganalisis koefisien Cronbach's Alpha. Koefisien keandalan mengukur kualitas kepercayaan pada kuesioner dengan kriteria yang sesuai. Untuk menentukan reliabilitas instrumen, koefisien Cronbach's Alpha seharusnya memiliki nilai lebih besar dari 0.7 (Aoki & Downes, 2003).

**Tabel 2. Pengujian Keandalan Alat Ukur (*Reliability*).**

Variabel	Pertanyaan	Cronbach's Alpha	N of Items
<b>Kesiapan Teknologi</b>			
1. Optimisme	OT1-OT4	0.865	4
2. Inovasi	INV1-INV3	0.668	3
3. Perasaan tidak nyaman saat digunakan	DI1-DI3	0.746	3
4. Merasa tidak aman	INS1-INS3	0.918	3
<b>Efek Eksternal dari Jaringan</b>	NE1-NE3	0.754	3
<b>Perlindungan Teknis</b>	TP1-TP3	0.877	3
<b>Prosedur Transaksi</b>	TR1-TR3	0.722	3
<b>Kesadaran Keselamatan</b>	PS1-PS3	0.740	3
<b>Persepsi Percaya Diri</b>	PT1-PT3	0.899	3
<b>Niat untuk Menggunakan e-KYC</b>	IU1-IU3	0.920	4



Dari pengecekan reliabilitas koefisien Cronbach's Alpha pada semua variabel, diketahui bahwa masing-masing variabel memiliki koefisien Cronbach's Alpha yang hanya faktor Inovasi memiliki nilai lebih kecil dari 0.7. Ini menyatakan bahwa variabel yang digunakan dalam penelitian ini adalah sangat reliabel karena memiliki korelasi di atas 0.5.

### 3.1.2. Analisis Faktor

Untuk menganalisis faktor-faktor atau variabel, beberapa metode analisis dilakukan, yaitu:

#### 1) Kaiser-Meyer-Olkin (KMO) dan Bartlett's Test

Berdasarkan Tabel 3 berikut, dapat diartikan bahwa analisis faktor tepat menjadi 10 kelompok.

**Tabel 3. Hasil KMO and Bartlett's Test.**

<b>KMO and Bartlett's Test</b>	
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.744
Bartlett's Test of Sphericity Approx. Chi-Square	2076.963
df	496
Sig.	.000

Diperoleh nilai KMOMSA (Kaiser-Meyer-Olkin *Measure of Sampling Adequacy*) sebesar 0.744 dan nilai signifikansi Bartlett's Test of Sphericity adalah 0.000 sehingga disimpulkan bahwa kelompok-kelompok faktor adalah tepat untuk merepresentasikan layanan sistem e-KYC.

#### 2) Measure of Sampling Adequacy (MSA)

Tabel 4 dan 5 menunjukkan hasil dari pengukuran kecukupan *sampling*.

**Tabel 4. Measure of Sampling Adequacy (MSA).**

No	Variabel	Komponen
1	OT1 Menurut Anda teknologi baru membantu untuk memiliki kualitas hidup yang lebih baik.	.906 <sup>a</sup>
2	OT2 Menurut Anda menggunakan teknologi memberi Anda lebih banyak kebebasan untuk menjalani hidup Anda.	.680 <sup>a</sup>
3	OT3 Menurut Anda teknologi dapat memberi Anda lebih banyak kendali atas kehidupan sehari-hari Anda. Misalnya, lebih banyak waktu tanpa pergi ke bank?	.710 <sup>a</sup>
4	OT4 Menurut Anda teknologi telah membuatnya lebih efektif dalam kehidupan pribadi Anda.	.751 <sup>a</sup>
5	INV1 Anda dapat merujuk teman Anda tentang teknologi baru.	.895 <sup>a</sup>
6	INV2 Anda sering dapat menggunakan layanan teknologi baru tanpa bantuan orang lain.	.768 <sup>a</sup>
7	INV3 Anda selalu mengikuti perkembangan teknologi baru di bidang yang Anda minati.	.629 <sup>a</sup>
8	DI1 Anda berpikir bahwa sistem teknologi baru tidak cocok untuk Anda.	.598 <sup>a</sup>
9	DI2 Anda merasa tidak nyaman. Saat mencari saran tentang cara menggunakan teknologi baru.	.663 <sup>a</sup>
10	DI3 Anda merasa tidak nyaman. Jika teknologi baru sulit dipahami.	.711 <sup>a</sup>
11	INS1 Bila melakukan transaksi, Anda akan merasa tidak aman tentang transaksi elektronik.	.698 <sup>a</sup>
12	INS2 Bila melakukan transaksi Anda akan merasa tidak percaya dengan akurasi dan kemampuan menggunakan teknologi.	.619 <sup>a</sup>
13	INS3 Bila melakukan transaksi Anda akan merasa tidak aman tentang keamanan teknologi baru.	.530 <sup>a</sup>



14	NE1	Menurut Anda sebagian besar teman Anda akan menggunakan sistem verifikasi identitas elektronik (e-KYC).	.696a
----	-----	---	-------

**Tabel 5. Measure of Sampling Adequacy (MSA) (lanjutan).**

No	Variabel	Komponen
15	NE2 Menurut Anda keluarga/teman Anda akan menggunakan sistem verifikasi identitas elektronik (e-KYC).	.585a
16	NE3 Anda pikir banyak teman akan menggunakan sistem verifikasi identitas elektronik (e-KYC) di masa depan.	16
17	TP1 Anda merasa bahwa informasi pribadi Anda tidak akan diberikan kepada individu.	.639a
18	TP2 Anda merasa data transaksi internet Anda terlindungi dan membuat Anda merasa aman.	.755a
19	TP3 Anda merasa bahwa sistem memiliki mekanisme teknis untuk melindungi informasi pribadi Anda dalam transaksi online.	.688a
20	TR1 Anda merasa bahwa Anda dapat mengubah informasi pribadi Anda sebelum mengkonfirmasi transaksi untuk membuka rekening deposit melalui sistem e-KYC.	.844a
21	TR2 Anda berpikir bahwa setelah menyelesaikan proses verifikasi identitas elektronik (e-KYC), sistem akan meminta Anda untuk memverifikasi informasi pribadi Anda sebelum konfirmasi. Seperti, nama depan dan belakang Anda.	.674a
22	TR3 Anda mengira bahwa ketika proses verifikasi identitas elektronik (e-KYC) selesai, sistem akan menampilkan nomor rekening Anda.	.599a
23	PS1 Menurut Anda sistem verifikasi identitas elektronik (e-KYC) aman.	.878a
26	PT1 Anda mempercayai bank yang menyediakan sistem identifikasi elektronik (e-KYC).	.840 <sup>a</sup>
27	PT2 Anda mempercayai layanan sistem identifikasi elektronik (e-KYC).	.848 <sup>a</sup>
28	PT3 Anda mempercayai informasi yang diberikan selama proses verifikasi identitas elektronik (e-KYC).	.831 <sup>a</sup>
29	IU1 Jika diberi kesempatan, Anda akan menggunakan sistem identifikasi elektronik (e-KYC).	.742 <sup>a</sup>
30	IU2 Anda kemungkinan akan menggunakan sistem identifikasi elektronik (e-KYC) di masa depan.	.812 <sup>a</sup>
31	IU3 Anda dipersiapkan untuk menggunakan layanan sistem identifikasi elektronik (e-KYC).	.791 <sup>a</sup>
32	IU4 Anda berniat mencoba layanan sistem identifikasi elektronik (e-KYC).	.874 <sup>a</sup>

Berdasarkan Tabel 4 dan 5, dapat dilihat bahwa 32 variabel ini mempunyai nilai MSA lebih besar dari 0.5. Sehingga dapat disimpulkan bahwa variabel sudah memadai untuk dianalisis lebih lanjut.

### 3) Total Variance Explained

Tabel 6 menjelaskan variansi pada faktor-faktor. Kriteria pertama yang digunakan adalah nilai Eigen. Faktor yang mempunyai nilai Eigen lebih dari 1 akan dipertahankan, sebaliknya faktor tidak akan diikutsertakan dalam model. Berdasarkan Tabel 6 di bawah diperoleh nilai Eigen yang lebih besar dari 1 pada faktor 1, 2, 3, 4, 5, dan 6. Dengan kriteria ini diperoleh jumlah faktor yang digunakan adalah enam. Variabel 9-32 tidak ditampilkan pada Tabel 6 karena keterbatasan halaman.

Kriteria kedua adalah penentuan berdasarkan nilai persentase variansi total yang dapat dijelaskan oleh banyaknya faktor yang akan dibentuk. Pada Tabel 6 dapat dilakukan interpretasi yang berkaitan dengan variansi total kumulatif sampel. Jika variabel-variabel itu diringkas menjadi beberapa faktor, maka nilai total variansi yang dapat dijelaskan adalah sebagai berikut.

- Jika ke-32 variabel diekstraksi menjadi 1 faktor, diperoleh variansi total yang dapat dijelaskan adalah  $10.686/32 \times 100\% = 33.3938\%$ . Seperti ditunjukkan pada Tabel 2, 1



faktor yang dibentuk adalah faktor Optimisme, yaitu pertanyaan dengan variabel OT1 sampai OT 4.

- Jika ke-32 variabel diekstraksi menjadi 2 faktor, diperoleh  $4.562/32 \times 100\% = 14.25625\%$ . Pada Tabel 2, 2 faktor yang dibentuk adalah Optimisme (OT1-OT4), dan Inovasi (variabel INV1-INV3).
- dan seterusnya.

**Tabel 6. Total Variance Explained.**

Component	Total Variance Explained					
	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.686	33.395	33.395	10.686	33.395	33.395
2	4.562	14.256	47.651	4.562	14.256	47.651
3	2.950	9.218	56.869	2.950	9.218	56.869
4	2.059	6.434	63.303	2.059	6.434	63.303
5	1.274	3.982	67.285	1.274	3.982	67.285
6	1.199	3.748	71.033	1.199	3.748	71.033
7	0.980	3.063	74.095			
8	0.937	2.927	77.022			

- Jika ke-32 variabel diekstraksi menjadi 6 faktor, diperoleh variansi total yang dapat dijelaskan adalah  $1.199/32 \times 100\% = 3.746875\%$ . Seperti ditunjukkan pada Tabel 2, 6 faktor yang dibentuk adalah Optimisme (OT1-OT4), Inovasi (INV1-INV3), Perasaan tidak nyaman saat digunakan (DI1-DI3), Merasa tidak aman (INS1-INS3), Efek eksternal dari jaringan (NE1-NE3). Kumulatif untuk 6 faktor adalah 71.03125%.

#### 4) Rotated Component Matrix

*Rotated component matrix* menampung nilai faktor *loading* dari tiap-tiap variabel. Faktor *loading* merupakan besarnya korelasi antara faktor yang terbentuk dengan variabel tersebut.

Berdasarkan Tabel 7, korelasi antara variabel OT1 dengan faktor 4 (.703) adalah yang tertinggi sehingga dapat dikatakan bahwa variabel OT masuk ke dalam faktor 4. Dengan cara yang sama, diperoleh faktor untuk setiap variabel. Sehingga 32 variabel bisa direduksi ke dalam enam faktor, yaitu:

- Faktor 1: INV2-INV3, dan IU1-IU4, diberikan nama faktor Inovasi dan Niat untuk menggunakan *e-KYC*.
- Faktor 2: TP1-TP3, TR1-TR2, dan PS2-PS3, dengan nama Perlindungan Teknis, Prosedur Transaksi, dan Kesadaran Keselamatan.
- Faktor 3: DI1-DI3, dan INS1-INS3, dengan nama Perasaan tidak nyaman saat digunakan, dan Merasa tidak aman.
- Faktor 4: OT1-OT4, INV1, dan TR3, diberikan nama Optimisme, Inovasi, dan Prosedur Transaksi.
- Faktor 5: PS1, dan PT1-PT3, diberikan nama Kesadaran Keselamatan, dan Persepsi Percaya Diri.
- Faktor 6: NE1-NE3, diberikan nama Efek Eksternal dari Jaringan.

#### 5) Component Transformation Matrix

Tabel 8 menunjukkan matriks transformasi dari komponen atau faktor. Dari Tabel 7 diketahui bahwa banyak komponen/faktor adalah enam.

Matriks transformasi komponen menunjukkan nilai korelasi untuk setiap komponen yang lebih besar dari .5, yaitu: .497, .896, .519, .544, .315, dan .453, untuk komponen 1 sampai 6 berurutan. Disimpulkan bahwa komponen 1, 5, 6 tidak valid dan komponen 2 (Perlindungan Teknis, Prosedur Transaksi, dan Kesadaran Keselamatan), 3 (Perasaan tidak nyaman saat digunakan,



dan Merasa tidak aman), dan 4 (Optimisme, Inovasi, dan Prosedur Transaksi) valid, maka ketiga faktor yang terbentuk ini disimpulkan layak untuk merangkum ke-32 variabel. Ini menyatakan hubungan antara pertanyaan-pertanyaan dari ketiga factor (Siregar et al., 2020).

**Tabel 7. Rotated Component Matrix.**

Variabel	Rotated Component Matrix <sup>a</sup>					
	Component					
	1	2	3	4	5	6
OT1	.384	-.076	-.129	.703	.206	.120
OT2	.493	.130	-.056	.532	.234	-.085
OT3	.283	.075	-.288	.638	.259	.074
OT4	.186	-.082	-.097	.803	.062	.226
INV1	.184	.131	.034	.517	.123	.618
INV2	.414	.076	.042	.405	.316	.259
INV3	.559	.319	.082	.354	-.193	-.173
DI1	-.022	.205	.720	-.292	.042	-.059
DI2	.089	.271	.760	-.244	.204	.081
DI3	.091	-.068	.690	.308	.057	-.027
INS1	.125	.066	.868	.089	.010	-.022
INS2	.043	.025	.889	-.121	-.066	.158
INS3	.015	-.132	.875	-.063	-.029	.080
NE1	.108	.120	.197	.180	.360	.644
NE2	.161	.474	.003	.018	.023	.740
NE3	.307	.249	-.029	.401	.263	.470
TP1	.178	.848	.010	-.007	.049	.127
TP2	.054	.852	.002	-.055	.175	.198
TP3	.258	.710	-.074	.105	.320	-.037
TR1	.127	.382	.371	.315	.271	.119
TR2	-.102	.421	.125	.527	.283	.271
TR3	-.064	.451	.330	.472	.228	-.031
PS1	.268	.494	.011	.281	.603	.078
PS2	.376	.610	.158	.082	.291	.131
PS3	-.059	.646	.201	.020	.131	.380
PT1	.129	.274	.005	.254	.706	.444
PT2	.246	.236	.120	.225	.765	.133
PT3	.326	.342	.026	.200	.756	.124
IU1	.879	.083	.088	.065	.247	.068
IU2	.806	.043	.097	.110	.220	.106
IU3	.809	.288	.030	.160	-.013	.193
IU4	.817	.045	.129	.217	.219	.164

**Tabel 8. Component Transformation Matrix.**

Component Transformation Matrix						
Component	1	2	3	4	5	6
1	.497	.485	.170	.427	.450	.322
2	-.161	.198	.896	-.362	-.041	.014
3	.519	-.681	.361	.297	-.136	-.171
4	-.645	-.294	.171	.544	.184	.371
5	-.199	.176	.079	.310	.315	-.853
6	-.027	.379	.059	.453	-.803	-.041



### 3.2. Faktor yang Mempengaruhi Niat Menggunakan e-KYC SCB Thailand

#### 3.2.1. Uji Reliabilitas

Tabel 9 menunjukkan keandalan alat dari instrumen responden SCB Thailand. Nilai koefisien Cronbach's Alpha dari semua variabel adalah lebih besar dari 0.7. Sehingga kuesioner ini adalah reliabel.

**Tabel 9. Pengujian Keandalan Alat Ukur (*Reliability*).**

Variabel	Pertanyaan	Cronbach's Alpha	N of Items
<b>Kesiapan Teknologi</b>			
1. Optimisme	OT1-OT4	0.858	4
2. Inovasi	INV1-INV3	0.819	3
3. Perasaan tidak nyaman saat digunakan	DI1-DI3	0.853	3
4. Merasa tidak aman	INS1-INS3	0.926	3
<b>Efek Eksternal dari Jaringan</b>			
	NE1-NE3	0.854	3
<b>Perlindungan Teknis</b>			
	TP1-TP3	0.862	3
<b>Prosedur Transaksi</b>			
	TR1-TR3	0.899	3
<b>Kesadaran Keselamatan</b>			
	PS1-PS3	0.874	3
<b>Persepsi Percaya Diri</b>			
	PT1-PT3	0.929	3
<b>Niat untuk Menggunakan e-KYC</b>			
	IU1-IU3	0.934	4

#### 3.2.2. Analisis Faktor

##### 1) *KMO dan Bartlett's Test*

Berdasarkan Tabel 10 diartikan bahwa analisis faktor tepat digunakan. Diperoleh nilai KMO sebesar 0.884 dan nilai signifikansi Bartlett's Test of Sphericity-nya adalah 0.000 sehingga dapat disimpulkan bahwa analisis faktor tepat digunakan.

**Tabel 10. Hasil *KMO and Bartlett's Test*.**

<i>KMO and Bartlett's Test</i>		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.884
Bartlett's Test of Sphericity	Approx. Chi-Square	2631.617
	df	496
	Sig.	.000

##### 2) *Measure of Sampling Adequacy (MSA)*

Pada pengukuran kecukupan sampling (MSA) untuk responden Bank SCB digunakan variabel yang sama dengan variabel untuk responden Bank Mandiri. Oleh karena itu, Tabel 11 berikut hanya menunjukkan nilai untuk 32 komponen secara urut.

**Tabel 11. *Measure of Sampling Adequacy (MSA)*.**

No	Komponen	No	Komponen	No	Komponen	No	Komponen
1	.936 <sup>a</sup>	9	.775 <sup>a</sup>	17	.860 <sup>a</sup>	25	.904 <sup>a</sup>
2	.891 <sup>a</sup>	10	.823 <sup>a</sup>	18	.829 <sup>a</sup>	26	.908 <sup>a</sup>
3	.802 <sup>a</sup>	11	.800 <sup>a</sup>	19	.868 <sup>a</sup>	27	.906 <sup>a</sup>
4	.861 <sup>a</sup>	12	.767 <sup>a</sup>	20	.923 <sup>a</sup>	28	.899 <sup>a</sup>
5	.886 <sup>a</sup>	13	.826 <sup>a</sup>	21	.889 <sup>a</sup>	29	.898 <sup>a</sup>
6	.928 <sup>a</sup>	14	.887 <sup>a</sup>	22	.928 <sup>a</sup>	30	.905 <sup>a</sup>
7	.933 <sup>a</sup>	15	.761 <sup>a</sup>	23	.923 <sup>a</sup>	31	.932 <sup>a</sup>
8	.720 <sup>a</sup>	16	.879 <sup>a</sup>	24	.934 <sup>a</sup>	32	.921 <sup>a</sup>

Berdasarkan Tabel 11 dapat dilihat bahwa 32 variabel ini mempunyai nilai MSA lebih dari 0,5. Sehingga dapat disimpulkan bahwa variabel sudah memadai untuk dianalisis lebih lanjut.



3) *Total Variance Explained*

Dari Tabel 12 diperoleh nilai Eigen yang lebih besar dari 1 adalah 1, 2, 3, 4, dan 5. Dengan kriteria ini diperoleh jumlah faktor yang digunakan adalah 5. Variabel 9-32 tidak ditampilkan.

Tabel 12. *Total Variance Explained.*

Component	Total Variance Explained					
	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	16.207	50.647	50.647	16.207	50.647	50.647
2	3.771	11.783	62.430	3.771	11.783	62.430
3	1.702	5.320	67.749	1.702	5.320	67.749
4	1.441	4.502	72.251	1.441	4.502	72.251
5	1.105	3.453	75.705	1.105	3.453	75.705
6	.920	2.874	78.578			
7	.711	2.223	80.801			
8	.681	2.127	82.928			

Dari Tabel 11 dapat dilakukan interpretasi variansi total kumulatif sampel. Jika variabel-variabel itu diringkas menjadi beberapa faktor, maka nilai total variansi adalah sebagai berikut:

- Jika ke-32 variabel diekstraksi menjadi 1 faktor, diperoleh variansi total yang dapat dijelaskan adalah  $16.207/32 \times 100\% = 50.65\%$ .
- Jika ke-32 variabel diekstraksi menjadi 2 faktor, diperoleh  $3.771/32 \times 100\% = 11.78\%$ .
- dan seterusnya.
- Jika ke-32 variabel diekstraksi menjadi 5 faktor, variansi total adalah  $1.105/32 \times 100\% = 3.45\%$ . Kumulatif untuk 5 faktor adalah 75.70%.

4) *Rotated Component Matrix*

Berdasarkan Tabel 13, korelasi antara variabel OT1 sampai INV3 adalah dengan faktor 1, dan seterusnya. Sehingga 32 variabel bisa direduksi ke dalam lima faktor, yaitu:

- Faktor 1: OT1-OT4, INV1-INV3, dan IU1-IU4, diberikan nama Optimisme, Inovasi dan Niat untuk menggunakan e-KYC.
- Faktor 2: TP2-TP3, TR1, PS1-PS3, PT1-PT3, dan IU4, dengan nama Perlindungan Teknis, Prosedur Transaksi, Kesadaran Keselamatan, Persepsi Percaya Diri, dan Niat untuk Menggunakan e-KYC.
- Faktor 3: DI1-DI3, dan INS1-INS3, dengan nama Perasaan tidak nyaman saat digunakan, dan Merasa tidak aman.
- Faktor 4: NE1-NE2, dan TP1, diberikan nama Efek Eksternal dari Jaringan, serta Perlindungan Teknik.

5) *Component Transformation Matrix*

Matriks transformasi komponen menunjukkan bahwa nilai korelasi komponen 1 adalah .648, komponen 2, 3, dan 4 dengan nilai korelasi .942, .631, dan .645. Berdasarkan Tabel 14 disimpulkan bahwa semua komponen adalah valid. Jadi, keempat faktor tersebut adalah:

- Faktor 1: OT1-OT4, INV1-INV3, dan IU1-IU4, diberikan nama Optimisme, Inovasi dan Niat untuk menggunakan e-KYC.
- Faktor 2: TP2-TP3, TR1, PS1-PS3, PT1-PT3, dan IU4, dengan nama Perlindungan Teknis, Prosedur Transaksi, Kesadaran Keselamatan, Persepsi Percaya Diri, dan Niat untuk Menggunakan e-KYC.
- Faktor 3: DI1-DI3, dan INS1-INS3, dengan nama Perasaan tidak nyaman saat digunakan, dan Merasa tidak aman.
- Faktor 4: NE1-NE2, dan TP1, diberikan nama Efek Eksternal dari Jaringan, serta Perlindungan Teknik.

Sehingga faktor 1, 2, 3, dan 4 dapat merangkum ke-32 variabel.



Tabel 13. *Rotated Component Matrix.*

Variabel	<i>Rotated Component Matrix<sup>a</sup></i>				
	<i>Component</i>				
	1	2	3	4	5
OT1	.761	.307	.097	.188	.118
OT2	.802	.165	.037	.137	.019
OT3	.758	.167	.189	-.064	-.165
OT4	.548	.531	.223	-.055	.125
INV1	.598	.300	.274	.047	.030
INV2	.646	.289	.068	.313	.254
INV3	.714	.224	.222	.321	.049
DI1	.003	-.063	.736	.189	.210
DI2	.055	.140	.723	.215	.502
DI3	.155	.278	.817	.116	.080
INS1	.071	.120	.889	.102	-.020
INS2	.146	.162	.880	-.032	-.085
INS3	.180	.076	.854	.028	-.190
NE1	.528	.242	.221	.564	.233
NE2	.303	.131	.140	.828	.199
NE3	.665	.271	.041	.490	-.186
TP1	.081	.410	.320	.694	-.018
TP2	.165	.654	.056	.545	-.102
TP3	.343	.589	.119	.531	-.255
TR1	.495	.672	.199	.133	-.098
TR2	.649	.541	.085	.111	-.118
TR3	.668	.505	.055	.127	-.145
PS1	.370	.765	.077	.292	-.115
PS2	.433	.734	.064	.199	.030
PS3	.216	.664	.357	.202	-.141
PT1	.416	.775	.162	.103	.176
PT2	.371	.751	.152	.089	.282
PT3	.357	.692	.212	.200	.409
IU1	.660	.472	.062	.252	.238
IU2	.646	.468	.031	.178	.232
IU3	.715	.425	.004	.218	.073
IU4	.488	.640	.067	.273	.304

Tabel 14. *Component Transformation Matrix.*

<b>Component Transformation Matrix</b>						
<b>Component</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
1	.587	.585	.402	.347	.178	.587
2	.366	.264	-.796	.063	-.399	.366
3	.338	-.063	.388	-.608	-.601	.338
4	.634	-.610	-.140	-.110	.441	.634
5	.069	-.461	.187	.703	-.503	.069
6	.587	.585	.402	.347	.178	.587

### 3.3. Perbandingan Faktor yang Mempengaruhi Niat Menggunakan e-KYC Bank Mandiri Indonesia dan SCB Thailand

Berdasarkan analisis faktor yang dilakukan, ada tiga faktor: 2, 3, dan 4 yang valid dan bisa mewakili ke-32 variabel. Faktor-faktor ini adalah Perlindungan Teknis, Prosedur Transaksi, Kesadaran Keselamatan, Perasaan tidak nyaman saat digunakan, Merasa tidak aman, Optimisme, dan Inovasi yang mempengaruhi niat responden menggunakan e-KYC Bank Mandiri. Sementara itu, ada empat faktor yaitu 1, 2, 3, dan 4, yang valid dari data kuesioner responden SCB. Faktor-faktor ini adalah Optimisme, Inovasi, Efek eksternal dari jaringan, Prosedur



Transaksi, Niat untuk Menggunakan e-KYC, Perlindungan Teknis, Kesadaran Keselamatan, Persepsi Percaya Diri, Perasaan tidak nyaman saat digunakan, dan Merasa tidak aman.

#### 4. KESIMPULAN

Disimpulkan bahwa kesediaan untuk menggunakan layanan verifikasi identitas elektronik dipengaruhi oleh semua faktor yang digunakan di dalam penelitian ini. Bank Mandiri Indonesia sangat mempengaruhi dalam penggunaan e-KYC dengan mendapatkan nilai validasi KMO sebesar 0.744, Bank Commercial Thailand dengan nilai validasi KMO sebesar 0.884.

#### DAFTAR PUSTAKA

- Abrazhevich, D. (2004). *Electronic Payment Systems: a User-Centered Perspective and Interaction Design*. Technische Universiteit Eindhoven.
- Alwi, I. (2015). Kriteria Empirik dalam Menentukan Ukuran Sampel Pada Pengujian Hipotesis Statistika dan Analisis Butir. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, 2(2). <https://doi.org/10.30998/formatif.v2i2.95>
- Aoki, K., & Downes, E. J. (2003). An analysis of young people's use of and attitudes toward cell phones. *Telematics and Informatics*, 20(4), 349–364. [https://doi.org/10.1016/S0736-5853\(03\)00018-2](https://doi.org/10.1016/S0736-5853(03)00018-2)
- Ardiani, F. (2020). Online Public Access Catalogue: Factors Affecting Use E-Catalog. *IJID (International Journal on Informatics for Development)*, 9(2), 94–99. <https://doi.org/10.14421/ijid.2020.09206>
- Ardiansyah, H., & Fatwanto, A. (2020). Application Design for Registration of Civil Appeals with Intuitive District Courts. *IJID (International Journal on Informatics for Development)*, 9(1), 45. <https://doi.org/10.14421/ijid.2020.09107>
- Katz, M. L., & Shapiro, C. (1994). Systems Competition and Network Effects. *Journal of Economic Perspectives*, 8(2), 93–115. <https://doi.org/10.1257/jep.8.2.93>
- Kim, C., Tao, W., Shin, N., & Kim, K.-S. S. (2010). An empirical study of customers' perceptions of security and trust in e-payment systems. *Electronic Commerce Research and Applications*, 9(1), 84–95. <https://doi.org/10.1016/j.elerap.2009.04.014>
- Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems*, 44(2), 544–564. <https://doi.org/10.1016/j.dss.2007.07.001>
- Lin, C.-H., Shih, H.-Y., Sher, P. J., & Wang, Y.-L. (2005). Consumer adoption of e-service: integrating technology readiness with the technology acceptance model. *A Unifying Discipline for Melting the Boundaries Technology Management*, 2005, 483–488. <https://doi.org/10.1109/PICMET.2005.1509728>
- Linck, K., Pousttchi, K., & Wiedemann, D. G. (2006). Security Issues in Mobile Payment from the Customer Viewpoint. *14th European Conference on Information Systems (ECIS)*, 1–11.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709. <https://doi.org/10.2307/258792>
- Parasuraman, A., & Colby, C. L. (2015). An Updated and Streamlined Technology Readiness Index. *Journal of Service Research*, 18(1), 59–74. <https://doi.org/10.1177/1094670514539730>
- Setiawan, N. (2005). *Teknik Sampling* (pp. 1–10). Universitas Padjadjaran.
- Shin, D. (2010). Analysis of online social networks: a cross-national study. *Online Information Review*, 34(3), 473–495. <https://doi.org/10.1108/14684521011054080>
- Siregar, M. U., Amalia, A., & Sugiantoro, B. (2020). ANALISIS KESIAPAN PENGIMPLEMENTASIAN EDUROAM DI UIN SUNAN KALIJAGA. *JURNAL TEKNIK INFORMATIKA*, 13(1), 11–20. <https://doi.org/10.15408/jti.v13i1.11889>
- Thangrattanasuwan, P. (2017). *ความพึงพอใจของผู้ใช้บริการชำระเงินของการประชาสัมพันธ์ สาขาทาดใหญ่ (ชั้นพิเศษ)*. Hatyai University.
- Wijtmoko, T. E., & Siregar, M. U. (2020). Evaluation of e-Government Service Quality Using e-GovQual Dimensions Case Study Regional Office Ministry of Law and Human Rights DIY. *IJID (International Journal on Informatics for Development)*, 8(2), 55. <https://doi.org/10.14421/ijid.2019.08202>





9 772527 583007

LABORATORIUM AGAMA  
MASJID SUNAN KALIJAGA