



E-ISSN 2597-9671

ISSN 2598-2079

CYBERSPACE

Jurnal Pendidikan Teknologi Informasi

Vol. 6, No. 2, 2022

OPTIMIZATION SENTIMEN ANALYSIS USING CRISP-DM AND
NAIVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA WIRELESS NETWORK

Denny Kurniawan, Muhammad Yasir

ANALISIS SENTIMEN MENGGUNAKAN ARSITEKTUR LONG SHORT-TERM MEMORY (LSTM)
TERHADAP FENOMENA CITAYAM FASHION WEEK

Laina Farsiah, Alim Misbullah, Husaini Husaini

IMPLEMENTASI DAN ANALISIS KINERJA ANTENA WAJAN BOLIC
SEBAGAI PENERIMA SINYAL WI-FI

Firmansyah Firmansyah, Isnan Bajili, Hendri Ahmadian, Aulia Syarif Aziz

FULLY HOMOMORPHIC ENCRYPTION FOR CLOUD BASED E-GOVERNMENT DATA

Rini Deviani, Sri Azizah Nazhifah, Aulia Syarif Aziz

PENERAPAN METODE CERTAINTY FACTOR PADA
SISTEM PAKAR DIAGNOSA GANGGUAN MENTAL PADA ANAK BERBASIS WEB

Kikye Martiwi Sukiakhy, Zulfan

DETEKSI KATA TAK BAKU DAN KESALAHAN PENULISAN KATA
PADA TUGAS AKHIR MAHASISWA MENGGUNAKAN METODE DICTIONARY LOOKUP

Alim Misbullah, Viska Mutiawani, Cut Sri Mulyani

ANALISIS PEMAHAMAN LITERASI DIGITAL PADA MAHASISWA UIN ARRANIRY
TERHADAP DIGITAL SKILL DAN DIGITAL SAFETY

Ridwan M.T, Yusran Yusran, Cut Addis Maulidia

CYBERSPACE
JURNAL PENDIDIKAN TEKNOLOGI INFORMASI

Penerbit:

Center for Research and Community Service (LP2M) in cooperation with the Department of Information Technology Education, Faculty of Education and Teacher Training, Ar-Raniry State Islamic University, Banda Aceh. Fakultas Teknik Universitas Syiah Kuala

Penasehat:

Ketua LP2M Universitas Islam Negeri Ar-Raniry Banda Aceh

Penanggung Jawab:

Ketua Prodi Pendidikan Teknologi Infomasi, Fakultas Tarbiyah dan Keguruan, UIN Ar-Raniry.

Ketua Editor: Mira Maisura

Editor:

Hendri Ahmadian, (SINTA ID: 6003098) Universitas Islam Negeri Ar-Raniry
Zuhra Sofyan, Hasso-Plattner Institute, University Of Potsdam Germany, Germany
Zuhal Ozturk, Computer Science Engineering, Marmara University, Turkey, Turkey
Bustami Yusuf, (Scopus ID: 36449577800) Universitas Islam Negeri Ar-Raniry,
Ahmed Alzeitoun, Informatik, University of Bonn, Germany
Aula Andika Fikrullah Al Balad, Lehigh University, USA, Indonesia
Maya Fitria, Universitas Syiah Kuala, Indonesia
Andika Prajana, (SCOPUS ID: 57212373564) Universitas Islam Negeri Ar-Raniry,
Firmansyah, FTK Universitas Islam Negeri Ar-Raniry
Basrul Abdul Majid, Institut Agama Islam Negeri Lhokseumawe, Indonesia
Zakiul Jailani, Wageningen University and Research, Netherlands
Dien Taufan, Politeknik Aceh, Aceh, Indonesia

Redaksi Pelaksana: Aulia Syarif Aziz

Editor Tata Letak dan Publikasi: Nurul Fajri

Alamat Redaksi:

Prodi Pendidikan Teknologi Informasi, Fakultas Tarbiyah dan Keguruan (FTK) Gedung B,
Lantai 1. UIN Ar-Raniry, Darussalam, Banda Aceh, 23111
E-mail: jurnal.cyberspace@ar-raniry.ac.id



DAFTAR ISI

OPTIMIZATION SENTIMEN ANALYSIS USING CRISP-DM AND NAIVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA <i>Denni Kurniawan dan Muhammad Yasir</i>	74 – 85
ANALISIS SENTIMEN MENGGUNAKAN ARSITEKTUR LONG SHORT-TERM MEMORY (LSTM) TERHADAP FENOMENA CITAYAM FASHION WEEK <i>Laina Farsiah, Alim Misbullah, Husaini</i>	86 – 94
IMPLEMENTASI DAN ANALISIS KINERJA ANTENA WAJAN BOLIC SEBAGAI PENERIMA SINYAL WI-FI <i>Firmansyah, Isnan Bajili, Hendri Ahmadian, Aulia Syarif Aziz</i>	95 – 104
FULLY HOMOMORPHIC ENCRYPTION FOR CLOUD BASED E-GOVERNMENT DATA <i>Rini Deviani, Sri Azizah Nazhifah, Aulia Syarif Aziz</i>	105 – 118
PENERAPAN METODE CERTAINTY FACTOR PADA SISTEM PAKAR DIAGNOSA GANGGUAN MENTAL PADA ANAK BERBASIS WEB <i>Kikye Martiwi Sukiakhy, Zulfan</i>	119 – 129
DETEKSI KATA TAK BAKU DAN KESALAHAN PENULISAN KATA PADA TUGAS AKHIR MAHASISWA MENGGUNAKAN METODE DICTIONARY LOOKUP <i>Alim Misbullah, Viska Mutiawani, Cut Sri Mulyani</i>	130 – 141
ANALISIS PEMAHAMAN LITERASI DIGITAL PADA MAHASISWA UIN ARRANIRY TERHADAP DIGITAL SKILL DAN DIGITAL SAFETY <i>Ridwan M.T, Yusran Yusran, Cut Addis Maulidia</i>	142 – 148

OPTIMIZATION SENTIMENT ANALYSIS USING CRISP-DM AND NAÏVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA

Denni Kurniawan¹ dan Muhammad Yasir²

^{1,2}Magister Ilmu Komputer, Fakultas Teknologi Informasi,
Universitas Budi Luhur, Jakarta, Indonesia

E-mail: denni.kurniawan@budiluhur.ac.id, muhammad90yasir@gmail.com

Abstract

Freedom of expression on social media like Twitter will not always has positive effects, sometimes it contains negative things such as fake news, hate speech, and racism, where these kinds of tweet can be categorized as an act of Cyberbullying. This cyberbullying tends to increase every time. The aim of this study is to use the Naïve Bayes method in classifying types of sentiment on Twitter. The keyword used is Saipul Jamil, and the tweet was taken in September 2021. A total of 18,067 tweets were collected and then they will be labelled with a positive or negative value. This study also uses the CRIPS-DM method which is consist of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment stages. The results of this study obtained the value of Accuracy (85.6%), Negative Recall (82.1%), Positive Recall (90.23%), and Negative Precision (91.76%) Positive Precision (79.18%).

Keywords: Cyberbullying, CRIPS-DM, Naïve Bayes, Sentiment Analysis, Twitter

Abstrak

Kebebasan mengeluarkan pendapat di media sosial Twitter tidak selamanya bernilai positif, karena terkadang mengandung hal negatif seperti cuitan yang mengandung berita bohong, penyebaran ujaran kebencian, dan rasisme. Ramai yang berpendapat bahwa jenis cuitan ini dapat dikategorikan sebagai *cyberbulling*. Dimana tindakan *cyberbulling* ini cenderung meningkat di setiap waktunya. Penelitian ini membahas penggunaan metode Naïve Bayes dalam melakukan klasifikasi jenis sentimen pada media sosial Twitter. Adapun kata kunci yang digunakan adalah Saipul Jamil, dan cuitan tersebut diambil pada bulan September 2021. Sebanyak 18.067 cuitan berhasil dikumpulkan dan dilanjutkan dengan memberikan label positif atau negatif pada data. Penelitian ini juga menggunakan metode CRIPS-DM dengan tahapan *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*. Hasil penelitian ini mendapatkan nilai Accuracy (85,6%), Recall Negatif (82,1%), Recall Positif (90,23%), dan Precision Negatif (91,76%) Precision Positif (79,18%).

Kata Kunci: Cyberbullying, CRIPS-DM, Naïve Bayes, Analisis Sentimen; Twitter.

1. Pendahuluan

Media sosial merupakan sebuah sarana komunikasi yang paling efektif, transparan dan efisien dimasa sekarang ini. Media sosial juga turut berperan penting dalam segala jenis perubahan dan pembaharuan yang terjadi [1]. Kebebasan berekspresi di media sosial memiliki tantangan yang sangat besar, dimana penyebaran berita yang bersifat negatif menjadi sangat mudah untuk disebar. Beberapa tindakan negatif yang sering dilakukan pengguna media sosial Twitter diantaranya adalah penyebaran berita bohong, penyebaran ujaran kebencian, dan rasisme. Tindakan ini merupakan perilaku negatif yang memiliki unsur tindakan *cyberbullying*. Tindakan yang mengandung *cyberbullying* di lingkungan pendidikan maupun di sosial media memiliki kecenderungan untuk terus naik [2].

Penelitian dengan topik *cyberbullying* merupakan salah satu penelitian yang sangat menarik untuk dilakukan, dimana opini masyarakat atau disebut juga dengan sentimen analisis di media sosial dapat dilihat dengan menggunakan beberapa metode populer dalam klasifikasi. Sebagai contoh penelitian yang dilakukan oleh Maulana dan Ernawati [3], melakukan klasifikasi *cyberbullying* di media sosial Twitter terhadap akun publik figur politik di Indonesia. Metode Naïve Bayes yang ditunakan berhasil melakukan klasifikasi dengan tingkat akurasi sebesar 76%. Penelitian lainnya yang dilakukan oleh Saputril dan Zuhri [4] menggunakan metode Naïve Bayes dalam melakukan klasifikasi terhadap ujaran kebencian pada periode Pemilihan Presiden 2019. Penggunaan metode Naïve Bayes yang digunakan berhasil mendapatkan tingkat akurasi sebesar 71%. Penelitian lain yang dilakukan oleh Sulastri [5], menggunakan metode Naïve Bayes pada analisis sentimen kasus penolakan dibukanya larangan ekspor benih Lobster. Penggunaan Naïve Bayes pada penelitian ini berhasil mendapatkan nilai akurasi sebesar 72,5%.

Dari beberapa penelitian di atas, dapat dilihat bahwa dengan menggunakan metode Naïve Bayes, tingkat akurasi yang dihasilkan masih belum terlalu tinggi. Sehingga tingkat akurasi masih bisa ditingkatkan lagi dengan menggunakan fitur tambahan. Adapun tujuan dari penelitian ini adalah untuk melakukan analisis sentimen *cyberbullying* dengan menggunakan metode CRIPS-DM dan Naïve Bayes dengan tingkat akurasi yang lebih tinggi dibandingkan penelitian sebelumnya.

2. Metode Penelitian

Pada bagian ini, akan dijelaskan teori-teori dasar yang digunakan pada penelitian ini, yaitu penjelasan *cyberbullying*, metode Naïve Bayes, dan model CRISP-DM.

Cyberbullying

Ada beberapa contoh tindakan yang dilakukan pengguna media sosial dengan menggunakan perangkat teknologi yang dapat dikategorikan sebagai perbuatan *cyberbullying*, diantaranya adalah intimidasi, ujaran kebencian, memermalukan seseorang, rasisme, atau body shaming yang dilakukan oleh pengguna media sosial menggunakan perangkat teknologi [6]. Menurut Alanzi dan Alves-foss [7], terdapat juga beberapa jenis perilaku lainnya yang dianggap sebagai tindakan *cyberbullying*, diantaranya adalah:

1. *Flooding* yaitu melibatkan pelaku intimidasi berulang kali mengirimkan komentar/postingan yang tidak penting sehingga tidak memungkinkan korban target untuk berpartisipasi dalam percakapan.
2. *Masquerade* yaitu di mana pelaku intimidasi berpura-pura meniru atau menyamar sebagai korban.

3. *Flaming/Bashing* melibatkan perkelahian online di mana pelaku intimidasi mengirim atau memposting konten yang menghina, menyakitkan, dan vulgar kepada korban yang ditargetkan secara pribadi atau publik dalam grup online.
4. *Trolling* melibatkan dengan sengaja menerbitkan komentar yang tidak sesuai dengan komentar lain untuk memicu argumen atau emosi negatif meskipun komentar itu sendiri mungkin tidak vulgar atau menyakitkan.
5. Pelecehan adalah jenis percakapan di mana pelaku intimidasi sering mengirimkan pesan yang menghina dan kasar kepada korban secara pribadi.
6. Penghinaan, terjadi ketika pelaku intimidasi mengirimkan atau mempublikasikan gosip atau pernyataan palsu tentang korban untuk merusak persahabatan/reputasi korban
7. Outing terjadi ketika seorang pengganggu memposting atau mempublikasikan informasi pribadi atau memalukan di ruang obrolan atau forum publik. Jenis *cyberbullying* ini mirip dengan fitnah.
8. Pengucilan, yaitu dengan sengaja mengecualikan seseorang dari grup online (mengucilkan orang lain). Jenis *cyberbullying* ini terjadi di kalangan remaja dan remaja lebih menonjol.

Seperti yang dikemukakan oleh Abdussalam, bahwa tindakan yang mengandung *cyberbullying* di lingkungan pendidikan maupun di sosial media memiliki kecenderungan untuk terus naik [2]. Menurut penelitian yang dilakukan oleh Microsoft, pengalaman perilaku tidak sopan yang dialami oleh pengguna sosial media juga tidak dapat dianggap enteng. Seperti terlihat pada Gambar 1, terlihat sebanyak 27% responden pernah menjadi korban ujaran kebencian, 43% responden mendapatkan berita bohong, dan 13% responden pernah merasakan tindakan diskriminasi [8].



Gambar 1. Riset Microsoft tentang *cyberbullying* di Indonesia [8].

OPTIMIZATION SENTIMENT ANALYSIS USING CRISP-DM AND NAÏVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA

Metode Naïve Bayes

Penggunaan Metode Naive Bayes sering diimplementasikan pada klasifikasi di berbagai bidang. Menurut Khairati [9], bahwa hasil klasifikasi yang dilakukan dalam implementasinya menunjukkan hasil yang beragam. Meskipun pada beberapa penelitian yang menggunakan optimalisasi Naïve Bayes menunjukkan kebenaran, namun terkadang tingkat perkiraan probabilitasnya tidak terlalu tinggi. Sehingga masih memerlukan pemahaman yang lebih dalam terhadap karakteristik data yang mempengaruhi kinerja metode ini. Metode Naïve Bayes juga menyediakan mekanisme untuk menggunakan informasi dalam sampel data untuk memperkirakan probabilitas posterior $P(H|X)$ dari setiap kelas H yang diberikan objek X . Setelah memiliki perkiraan seperti itu, baru dapat menggunakannya untuk klasifikasi atau aplikasi pendukung keputusan lainnya [10]. Menurut pemaparan Bustami [11], formula yang digunakan dalam metode Naïve Bayes adalah seperti diperlihatkan pada persamaan (1) di bawah:

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

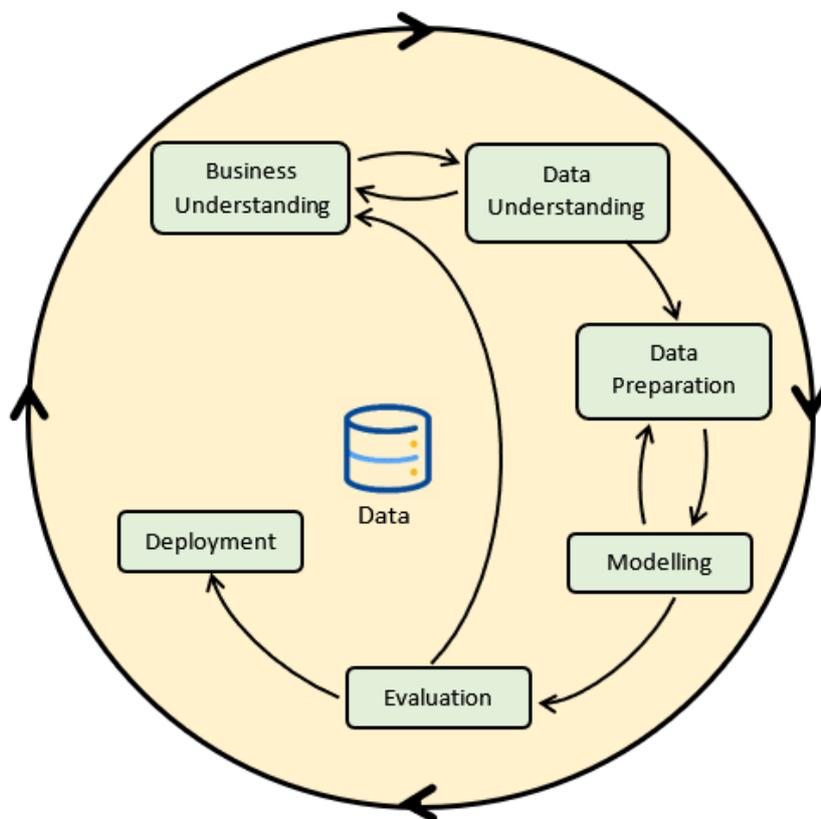
Dimana X adalah data dengan class yang belum diketahui; H adalah hipotesis data X meruapakan suatu class spesifik; $P(H|X)$ adalah probabilitas hipotesis H berdasarkan kondisi X ; $P(H)$ adalah probabilitas hipotesis H ; $P(X|H)$ adalah probabilitas X berdasarkan kondisi pada hipotesis H ; dan $P(X)$ adalah probabilitas X .

Model CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) adalah model proses industri-independen yang lazim digunakan untuk keperluan *data mining*. Metode ini terdiri dari enam fase berulang seperti terlihat pada Gambar 2. Fase pertama dimulai dari *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment* [12, 13]. Dan dalam siklus pengembangannya CRISP-DM dianggap sebagai metodologi data mining terlengkap dalam hal pemenuhan kebutuhan proyek industri, dan telah menjadi yang paling luas penggunaannya dalam proyek analitik, *data mining*, serta *data science* [14]. Tahapan penelitian menggunakan CRISP-DM terlihat pada Gambar 2, yang merupakan aliran proses pada model CRISP-DM yang biasa digunakan untuk *data mining*. Setiap proses atau fase memiliki tugas dan fungsi masing-masing. Panah pada gambar menunjukkan keterikatan antar fase, namun ada kalanya urutan fase ini tidak kaku dan juga tergantung pada hasil di setiap fase. Lingkaran luar dari setiap fase juga melambangkan siklus *data mining* itu sendiri. Model CRISP-DM juga bertujuan untuk membuat proyek penambangan data besar, lebih murah, lebih andal, lebih dapat diulang, lebih mudah dikelola, dan lebih cepat.

Fase pertama dari model CRISP-DM adalah *Business Understanding*, dimana fase ini berfokus untuk memahami tujuan dan persyaratan proyek dari perspektif bisnis. Pengetahuan ini kemudian diubah menjadi definisi masalah penambangan data dan rencana proyek awal yang dirancang untuk mencapai tujuan. Kemudian dilanjutkan dengan fase *Data Understanding* atau pemahaman data, dimana fase ini dimulai dengan proses pengumpulan data dan dilanjutkan dengan kegiatan mengidentifikasi data, mengidentifikasi masalah kualitas data, menemukan wawasan pertama ke dalam data, atau mendeteksi subset yang menarik untuk membentuk hipotesis untuk informasi

tersembunyi. Di tahap ini penulis melakukan pelabelan *dataset* manual dengan bantuan anator, ada baiknya dalam proses pelabelan ini melibatkan ahli atau pakar bahasa sesuai bahasa yang digunakan. Tujuannya adalah ahli atau pakar bahasa tersebut lebih memahami pemaknaan dari setiap teks yang tertulis [15].



Gambar 2. Ilustrasi tahapan pelaksanaan CRISP-DM [12]

Dilanjutkan dengan fase *Data Preparation* atau persiapan data, yang mencakup semua kegiatan untuk membangun dataset akhir dari data mentah awal. Tugas persiapan data kemungkinan akan dilakukan beberapa kali, tetapi tidak dalam urutan yang ditentukan. Tugas seperti tabel, catatan, dan pemilihan atribut, *preprocessing* data, konstruksi atribut baru, dan transformasi data untuk pemodelan. Tahap *preprocessing* atau pembersihan data Twitter adalah skenario standar *preprocessing* data teks. Menurut penelitian yang dilakukan oleh Giachanou dan Crestani [16], terdapat beberapa karakteristik unik tweet yang memerlukan langkah yang berbeda untuk mengatasinya. Berdasarkan penelitian sebelumnya yang dilakukan oleh Kane *et al.*, terdapat beberapa langkah yang digunakan untuk *preprocessing* Twitter yaitu *Tokenize*, *Stopword*, *Case Folding*, *Remove Punctuation*, dan *Stemming* [17]. Pada tahap tahapan *preprocessing* Twitter sebagai berikut:

1. *Tokenize*: Tahap ini merupakan proses pembagian kata. Kalimat akan dibagi menjadi beberapa bagian yang disebut token. Token dapat berupa kata, frasa, atau elemen makna lainnya. Kita bisa menggunakan *library nltk.tokenize* Kita bisa menggunakan *library nltk.tokenize* untuk membagi kata menjadi kelompok huruf.
2. *Stopword*: Tahap ini merupakan proses untuk menghilangkan kata-kata umum dan sering yang tidak memiliki pengaruh signifikan dalam sebuah kalimat. sehingga data

OPTIMIZATION SENTIMENT ANALYSIS USING CRISP-DM AND NAÏVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA

dapat diproses lebih efisien pada tahap selanjutnya. Mengimpor daftar *stopwords* bisa menggunakan dari *library nltk.corpus*.

3. *Case Folding*: Tahap ini merupakan proses mengubah kata menjadi bentuk yang sama. Pada langkah ini, mengonversi semua kata menjadi huruf kecil menggunakan metode *lower case* pada Python.
4. *Remove Punctuation*: Tahap ini merupakan proses menghapus karakter, angka, *string ASCII*, dan tanda baca. Pesan *twitter* biasanya berisi simbol, angka, dan tanda baca. Semua ini dihapus menggunakan sintaks ekspresi reguler
5. *Stemming*: Tahap ini adalah proses mendapatkan dasar atau akar kata dengan menghilangkan imbuhan dan sufiks. Penelitian ini menggunakan *library python* sastrawi untuk menghilangkan kata-kata imbuhan dalam bahasa Indonesia ke bentuk dasarnya.

Setelah fase ketiga selesai dilakukan, maka dilanjutkan dengan fase berikutnya yaitu fase *Modeling*. Dimana pada fase ini terdapat berbagai teknik pemodelan dipilih dan diterapkan. Beberapa metode memerlukan format data tertentu. Pada penelitian ini, tahap *Modeling* menggunakan metode Naïve Bayes untuk mendapatkan nilai optimal. Tahapan berikutnya adalah *Evaluation*, dimana pada tahapan ini satu atau lebih model dibangun untuk menghasilkan analisis data yang memiliki kualitas tinggi. Sebelum melanjutkan ke implementasi akhir model, penting untuk mengevaluasi model secara lebih menyeluruh, dan meninjau langkah-langkah yang diambil untuk membangun model, untuk memastikan bahwa model mencapai tujuan bisnis dengan benar. Tujuan utamanya adalah untuk menentukan apakah ada beberapa masalah yang belum dipertimbangkan secara memadai. Pada akhir fase ini, keputusan harus dibuat tentang bagaimana menggunakan hasil data mining. Fase terakhir adalah *Deployment*, dimana pada fase akhir ini pemodelan bukanlah akhir dari proyek. Secara umum, pengetahuan yang diperoleh perlu diatur dan disajikan sedemikian rupa sehingga pengguna dapat menggunakannya. Bergantung pada apa yang dibutuhkan pengguna, fase *Deployment* bisa dibuat sesederhana membuat laporan atau serumit menerapkan proses data mining berulang.

3. Hasil dan Pembahasan

Penelitian ini dilakukan dengan tahapan yang sesuai dengan pengembangan dari tahapan CRIPS-DM yang telah dijelaskan sebelumnya, yaitu terdiri dari enam tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Penjelasan secara lebih terperinci terkait dengan tahapan-tahapan tersebut adalah seperti berikut:

a) *Business Understanding*

Tahapan *Business Understanding* berfokus pada pemahaman tujuan yaitu melakukan *experiment research* bagaimana melakukan klasifikasi sentimen yang mengandung unsur *cyberbullying* pada media sosial Twitter. Kemudian data Twitter dikumpulkan dengan kata kunci “Saipul Jamil” yang diambil selama delapan hari, dimulai pada tanggal 2 sampai tanggal 9 September 2021 menggunakan API Twitter dengan menggunakan Python Google Colab.

b) *Data Understanding*

Tahapan selanjutnya yaitu tahapan *Data Understanding* yaitu proses pendataan awal terkait tweet *cyberbullying* penyanyi dangdut Saipul Jamil. Dari proses pengumpulan data

selama delapan hari ini berhasil mendapatkan sebanyak 18.0067 tweet. Di tahap ini penulis melakukan pelabelan *dataset* yang dilakukan secara manual. Proses ini juga melibatkan seorang ahli bahasa atau pakar bahasa sesuai dengan bahasa yang digunakan, yang bertujuan agar pakar bahasa tersebut dapat memberikan pelabelan tanpa kesalahan dan memiliki kemampuan pemahaman yang lebih tinggi terhadap setiap teks yang tertulis pada penelitian.

c) Data Preparation

Tahap *Data Preparation* merupakan tahap persiapan data yang mencakup semua kegiatan untuk membangun dataset akhir. Pada tahap ini, dilakukan beberapa preprocessing text yaitu tahap *Tokenize*, *Case Folding & Remove Punctuation*, *Stopword*, dan *Stemming*. Pada tahapan awal yang disebut *Tokenize*, adalah proses yang berfungsi untuk memisahkan teks menjadi potongan-potongan kata yang menyusunnya. Proses pemisahan ini dapat dilakukan dengan menggunakan `nlk.word_tokenize` pada Python. Perbedaan hasil antara sebelum dan sesudah tahapan *Tokenize* dilakukan dapat dilihat pada Tabel 1.

TABEL 1. IMPLEMENTASI *TOKENIZE*

No	Sebelum	Sesudah
1	Bebas dari Penjara, Saipul Jamil Ngaku Trauma\n@saipuljamil\n https://t.co/sgRJ4nXieH	['Bebas', 'dari', 'Penjara', ',', 'Saipul', 'Jamil', 'Ngaku', 'Trauma', '@', 'saipuljamil', 'https', ':', '//t.co/sgRJ4nXieH']
2	@saipuljamil Tuhan Maha Pemaaf dan menerima Taubat.Taubat artinya tdk melakukan perbuatan tercela yg sama lagi. Ini hanya Peringatan Bang Saipul Jamil ,Akibat perbuatan anda trauma seumur hidup diterima korban. Tdk bisa perbaiki kerusakan yg dibuat. Jadi Jgn sampai berbuat lagi ! https://t.co/iayMFyx0TT	['@', 'saipuljamil', 'Tuhan', 'Maha', 'Pemaaf', 'dan', 'menerima', 'Taubat.Taubat', 'artinya', 'tdk', 'melakukan', 'perbuatan', 'tercela', 'yg', 'sama', 'lagi', ',', 'Ini', 'hanya', 'Peringatan', 'Bang', 'Saipul', 'Jamil', ',', 'Akibat', 'perbuatan', 'anda', 'trauma', 'seumur', 'hidup', 'diterima', 'korban', ',', 'Tdk', 'bisa', 'perbaiki', 'kerusakan', 'yg', 'dibuat', ',', 'Jadi', 'Jgn', 'sampai', 'berbuat', 'lagi', '!', 'https', ':', '//t.co/iayMFyx0TT']

Tahapan selanjutnya, yaitu Tahap *Case folding* dimulai dengan proses merubah teks menjadi huruf kecil. Sedangkan *Remove Punctuation* adalah proses menghilangkan tanda baca atau simbol yang ada dalam dataset. Perbedaan hasil antara sebelum dan sesudah tahapan *Tokenize* dilakukan dapat dilihat pada Tabel 2.

TABEL 2. IMPLEMENTASI *CASE FOLDING* DAN *REMOVE PUNCTUATION*

No	Sebelum	Sesudah
1	Bebas dari Penjara, Saipul Jamil Ngaku Trauma\n@saipuljamil\n https://t.co/sgRJ4nXieH	bebas dari penjara saipul jamil ngaku trauma
2	@saipuljamil Tuhan Maha Pemaaf dan menerima Taubat.Taubat artinya tdk melakukan perbuatan tercela yg sama lagi. Ini hanya Peringatan Bang Saipul Jamil ,Akibat perbuatan anda trauma seumur hidup diterima korban. Tdk bisa perbaiki kerusakan yg dibuat. Jadi Jgn sampai berbuat lagi ! https://t.co/iayMFyx0TT	tuhan maha pemaaf dan menerima taubattaubat artinya tdk melakukan perbuatan tercela yg sama lagi ini hanya peringatan bang saipul jamil akibat perbuatan anda trauma seumur hidup diterima korban tdk bisa perbaiki kerusakan yg dibuat jadi jgn sampai berbuat lagi

Tahapan akhir yang disebut tahap *Stopword* merupakan suatu proses yang berfungsi untuk menghilangkan kata-kata umum dan sering yang tidak memiliki pengaruh

OPTIMIZATION SENTIMENT ANALYSIS USING CRISP-DM AND NAÏVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA

signifikan dalam sebuah kalimat. sehingga data dapat diproses lebih efisien pada tahap selanjutnya. Proses *import* daftar *Stopwords* dapat dilakukan dengan menggunakan *library* *nlTK.corpus*. Perbedaan hasil antara sebelum dan sesudah tahapan *Tokenize* dilakukan dapat dilihat pada Tabel 3.

TABEL 3. IMPLEMENTASI *STOPWORD*

No	Sebelum	Sesudah
1	bebas dari penjara saipul jamil ngaku trauma	bebas penjara ngaku trauma
2	tuhan maha pemaaf dan menerima taubattaubat artinya tdk melakukan perbuatan tercela yg sama lagi ini hanya peringatan bang saipul jamil akibat perbuatan anda trauma seumur hidup diterima korban tdk bisa perbaiki kerusakan yg dibuat jadi jgn sampai berbuat lagi	tuhan maha pemaaf menerima taubattaubat tdk perbuatan tercela yg peringatan bang akibat perbuatan trauma seumur hidup diterima korban tdk perbaiki kerusakan yg jgn berbuat

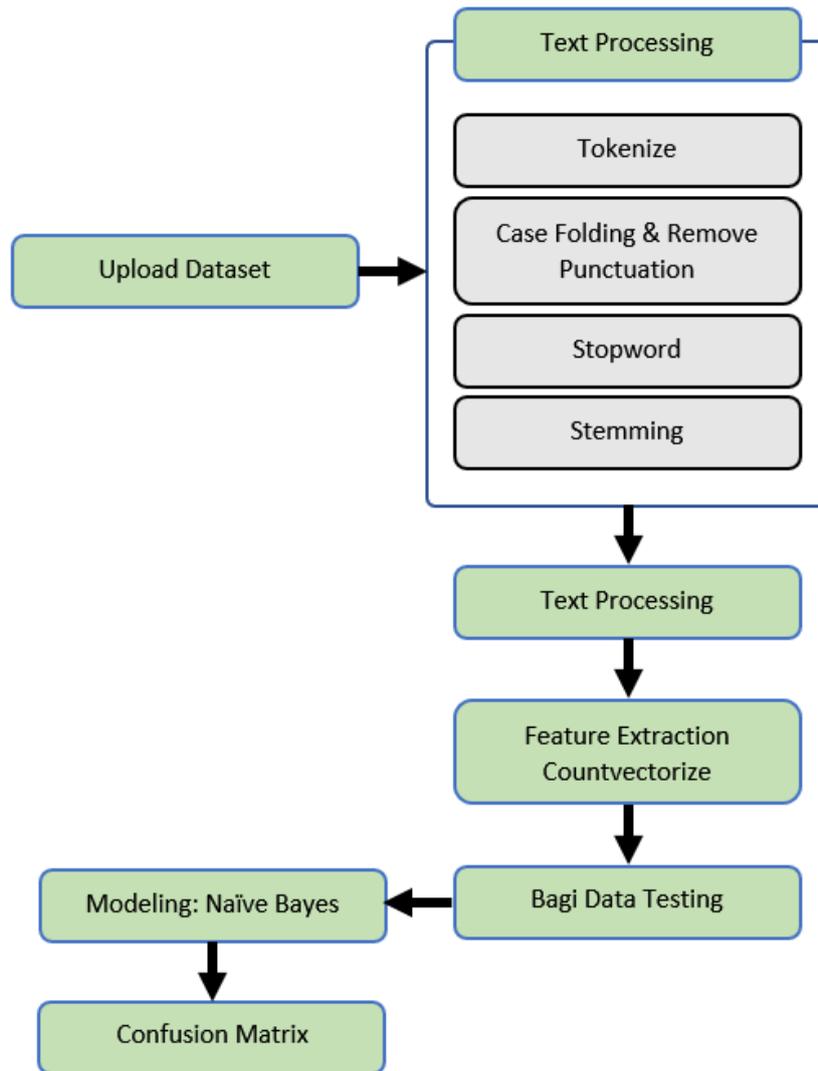
Proses mengubah kata berimbuhan menjadi kata dasar dapat dilakukan pada Tahap *Stemming*. Pada tahap ini menggunakan *library* Python Sastrawi untuk menghilangkan kata-kata imbuhan dalam bahasa Indonesia ke bentuk dasarnya. Perbedaan hasil antara sebelum dan sesudah tahapan *Tokenize* dilakukan dapat dilihat pada Tabel 4.

TABEL 4. IMPLEMENTASI *STEMMING*

No	Sebelum	Sesudah
1	bebas penjara ngaku trauma	bebas penjara ngaku trauma
2	tuhan maha pemaaf menerima taubattaubat tdk perbuatan tercela yg peringatan bang akibat perbuatan trauma seumur hidup diterima korban tdk perbaiki kerusakan yg jgn berbuat	tuhan maha maaf terima taubattaubat tdk buat cela yg ingat bang akibat buat trauma umur hidup terima korban tdk baik rusak yg jgn buat

d) Modeling

Pada tahap ini merupakan proses *mining* menggunakan model Naïve Bayes dengan alur seperti terlihat pada Gambar 3. Langkah pertama adalah melakukan *upload dataset* mentah yang sudah diberi label positif atau negatif. Kemudian dilanjutkan dengan melakukan proses *text preprocessing* menggunakan Google Collab. Selanjutnya, dilakukan *text preprocessing* data dengan melakukan *Tokenize*, *Case Folding* dan *Remove Punctuation*, *Stopword*, *Stemming*. Terdapat 18.067 data mentah dan setelah melakukan proses *text preprocessing*, hanya tersisa 6.787 data. Setelah proses pembersihan data dilakukan, maka dilanjutkan dengan melakukan perubahan informasi label tersebut menjadi angka. Fungsi ini untuk mengubah data ke dalam bentuk data integer positif (1) atau negatif (-1). Kemudian dari 6.787 data tersebut, data kembali dibagi menjadi sebanyak 2.500 yang digunakan sebagai sampel data yang didapatkan melalui perhitungan Slovin. Kemudian data ini di ubah ke dalam bentuk vektor menggunakan fitur *extraction counvectorize* untuk digunakan sebagai masukan dalam algoritma pembelajaran mesin. Dari sebanyak 2.500 data yang dimiliki, data kembali dibagi menjadi dua bagian, yang akan digunakan untuk kebutuhan *training*, dan *testing*. Pembagian data ini menggunakan perbandingan 80:20, yang bermaksud menggunakan sebanyak 80% data sebagai data *training* dan 20% data sebagai data *testing*. Setelah itu pada tahap selanjutnya dilakukan modeling menggunakan metode Naïve Bayes dan dilakukan perhitungan untuk melihat akurasi dari model. Setelah proses modeling, tahap selanjutnya adalah melakukan *Confusion Matrix* untuk mengukur performa klasifikasi.

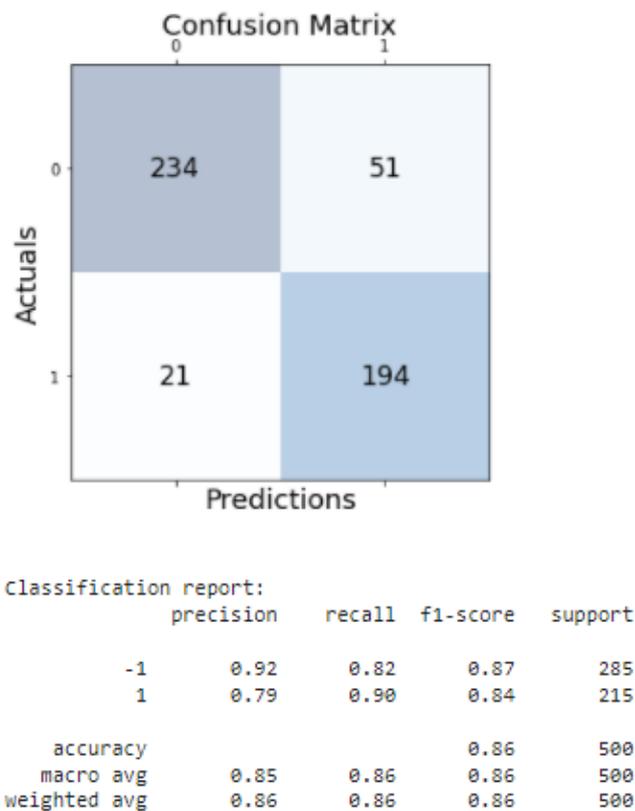


Gambar 3. Alur proses modeling

e) Evaluation

Tahapan *Evaluation* merupakan tahap laporan evaluasi model dengan menggunakan *Confusion Matrix* untuk mengukur performa klasifikasi Naïve Bayes. Perhitungan untuk menentukan *Accuracy*, *Recall Negative*, *Recall Positive*, *Precision Negative*, dan *Precision Positive* dapat dilakukan dengan menggunakan Persamaan (2), (3), (4), (5), dan (6) secara berurutan. Dimana TP adalah *True Positive*, TN adalah *True Negative*, FP adalah *False Positive*, dan FN adalah *True Negative*.

OPTIMIZATION SENTIMENT ANALYSIS USING CRISP-DM AND NAÏVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA



Gambar 4. Hasil Confusion Matrix

<i>Confusion Matrix</i>	<i>Predicted Class</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Observed Class</i>	Positive	TP = 234 FN = 21
	Negative	FP = 51 TN = 194

Gambar 5. Penjelasan pengambilan nilai TP, TN, FP, dan FN dari *Confusion Matrix*

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$Recall (neg) = \frac{TP}{TP + FP} \tag{3}$$

$$Recall (pos) = \frac{TN}{TN + FN} \tag{4}$$

$$Precision (neg) = \frac{TP}{TP + FN} \tag{5}$$

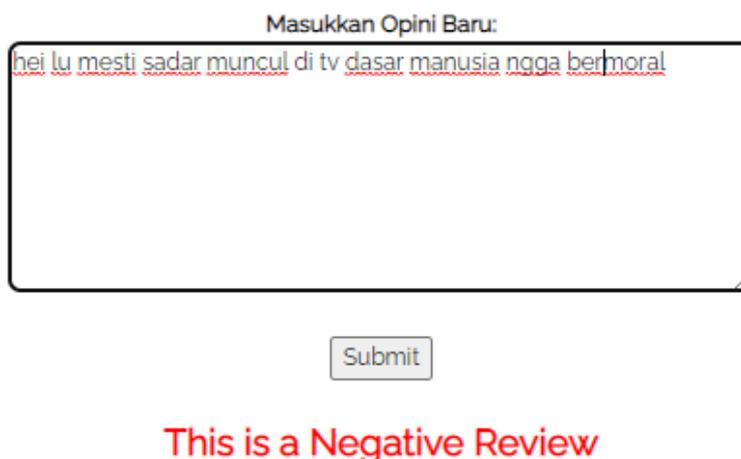
$$Precision (pos) = \frac{TN}{TN + FP} \tag{6}$$

Nilai TP, TN, FP, dan FN dapat diambil dari *Confusion Matrix* seperti pada Gambar 5. Sehingga diperoleh nilai TP, TN, FP, dan FN sebesar 234, 194, 51, dan 21, secara berurutan. Nilai ini kemudian dimasukkan ke dalam *Persamaan* (2), (3), (4), (5), dan (6)

sehingga dari hasil perhitungan diperoleh nilai *Accuracy* 85,6%; *Recall Positive* 90,23%; *Recall Negative* 82,1%; dan *Precision Positif* 79,18%.

f) Deployment

Sebuah prototipe hasil pengembangan model metode Naïve Bayes menggunakan FLASK telah dilakukan pada tahap *Deployment*, seperti diperlihatkan pada Gambar 6. Prototipe yang dibuat memiliki kemampuan untuk mendeteksi kalimat atau tweet yang memiliki unsur *cyberbullying*. Sistem ini juga mampu mengklasifikasikan kalimat opini baru tersebut dengan label *positive* atau *negative*. Prototype ini memiliki kemampuan yang sangat bagus, walaupun adanya keterbatasan penggunaan jumlah karakter yang berlaku di Twitter, yaitu maksimal 140 karakter.



Gambar 6. Implementasi Sistem

4. Kesimpulan

Berdasarkan hasil penelitian yang diperoleh, dapat disimpulkan bahwa algoritma Naïve Bayes yang digunakan untuk mengklasifikasikan tweet yang mengandung unsur *cyberbullying*. Proses pelatihan ini perlu didukung dengan data latih yang baik. Proses pemberian label pada dataset sebaiknya dilakukan dengan bantuan seorang anator atau ahli bahasa agar lebih memahami pemaknaan dari setiap teks yang tertulis. Dari hasil evaluasi menggunakan *Confusion Matrix* didapatkan hasil analisis sentimen *cyberbullying* menggunakan metode Naïve Bayes mendapatkan *Accuracy* sebesar 85,6%, *Recall Positive* sebesar 90,23%; *Recall Negative* sebesar 82,1%; *Precision Negative* sebesar 91,76%; dan *Precision Positive* sebesar 79,18%. Terlihat bahwa hasil pengujian yang dilakukan memiliki tingkat akurasi yang lebih baik dibandingkan hasil penelitian yang serupa.

Referensi

- [1] D. R. Rahadi, "Perilaku Pengguna dan Informasi Hoax di Media Sosial". *Jurnal Manajemen dan Kewirausahaan*, vol.5(1), pp. 58-70. 2017.
- [2] M. S. Abdussalam, Sejumlah kasus bullying sudah warnai catatan masalah anak di awal 2020 begini kata komisioner KPAI, <https://www.kpai.go.id/publikasi/sejumlah-kasus-bullying-sudah-warnai-catatan-masalah-anak-di-awal-2020-begini-kata-komisioner-kpai>, 2020, retrieved February 13, 2020.

OPTIMIZATION SENTIMENT ANALYSIS USING CRISP-DM AND NAÏVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA

- [3] F. A. Maulana, and I. Ernawati, “Analisa Sentimen Cyberbullying di Jejaring Sosial Twitter dengan Algoritma Naïve Bayes”, *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya*, vol. 1(2), pp. 529-538. 2020.
- [4] N. A. O. Saputril, and K. Zuhri, “Analisis sentimen masyarakat terhadap pilpres 2019 berdasarkan opini dari Twitter menggunakan metode Naïve Bayes classifier”. *Jurnal Informanika*, vol. 7(1), pp. 55-62, 2021.
- [5] K. Sulastrri, “Klasifikasi Naïve Bayes pada analisis sentimen atas penolakan dibukanya larangan ekspor benih lobster”, *Jurnal Riset Inovasi Bidang Informatika dan Pendidikan Informatika KERNEL*, vol. 1(2), pp. 68-75. 2020.
- [6] F. A. Maulana, I. Ernawati, P. Labu, and J. Selatan, “Analisa sentimen cyberbullying di jejaring sosial Twitter dengan algoritma Naïve Bayes”, *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasi (SENAMIKA)*, pp. 529–538. 2020.
- [7] I. Alanazi, and J. Alves-foss, “Cyber bullying and machine learning: A survey”, *International Journal of Computer Science Security*, vol. 18(10), pp. 1-8, 2020.
- [8] Anon, Civility, Safety & Interaction Online: Indonesia, [shttps://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4MM81](https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4MM81), 2021, retrieved November 10, 2021
- [9] A. F. Khairati, A. A. Adlina, G. F. Hertono, and B. D. Handari, “Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA”, *Prosiding Seminar Nasional Matematika*, vol. 2. pp. 161-170, 2019.
- [10] G. I. Webb. “Encyclopedia of Machine Learning and Data Mining”, *Encyclopedia of Machine Learning and Data Mining*
- [11] Bustami, “Penerapan Algoritma Naïve Bayes untuk Mengklasifikasi Data Nasabah”, *TECHSI: Jurnal Penelitian Teknik Informatika*, vol. 4, pp. 127-146. 2010.
- [12] W. Rüdiger, and J. Hipp. “CRISP-DM: Towards a Standard Process Model for Data Mining.” *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, vol. 4, pp. 29–39. 2000.
- [13] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. “CRISP-DM 1.0: Step-by-step Data Mining Guide,” 2000.
- [14] C. Schröer, F. Kruse, and J. M. Gómez, “A Systematic Literature Review on Applying CRISP-DM Process Model”, *Procedia Computer Science*, vol. 181 (2019), pp. 526–534. 2021.
- [15] M. S. Hadna, P. I. Santosa, dan W. W. Winarno, “Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter”, *Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (Sentika)*, 2016, pp. 57–64.
- [16] A. Giachanou, dan F. Crestani, “Like it or not: A survey of Twitter sentiment analysis methods”, *ACM Computing Surveys*, vol. 49 (2), 2016. pp. 1-41.
- [17] S. N. Kane, A. Mishra, dan A. K. Dutta, “Preface: International Conference on Recent Trends in Physics (ICRTP 2016)”, *Journal of Physics: Conference Series*, vol. 755 (1), 2016, pp. 3–9.

ANALISIS SENTIMEN MENGGUNAKAN ARSITEKTUR *LONG SHORT-TERM MEMORY* (LSTM) TERHADAP FENOMENA *CITAYAM FASHION WEEK*

Laina Farsiah, Alim Misbullah, dan Husaini

Jurusan Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Syiah Kuala, Darussalam, Banda Aceh, Indonesia 23111
E-mail: lainafarsiah@unsyiah.ac.id, misbullah@unsyiah.ac.id,
husaini.muhammad@usk.ac.id

Abstract

Sentiment analysis of the text aims to recognize whether a text contains positive, negative, or neutral emotions. The results of the analysis can be used as a tool for making decisions on an issue. Recently, the Citayam Fashion Week event become an issue that is extremely debated in Indonesia, especially in July 2022 on social media. The issue has motivated us to do sentiment analysis for better making decisions. In our work, the dataset is collected from Indonesian people's tweets with the keywords Citayam Fashion Week. Furthermore, each tweet will be labeled with a positive, negative, or neutral class based on the Indonesian lexical. This research produces a model based on Long Short Term Memory (LSTM) structure to predict every Indonesian tweet into the category of positive, negative, or neutral sentiment related to public views and opinions about the Citayam Fashion Week phenomenon. The model accuracy shows that the LSTM obtained good performance which is 88%.

Keywords: *sentiment analysis, deep learning, LSTM, tweet, Citayam Fashion Week*

Abstrak

Analisis sentimen pada teks bertujuan untuk melihat sebuah teks mengandung emosi positif, negatif, atau netral. Hasil analisis dapat digunakan sebagai bahan pertimbangan untuk mengambil keputusan terhadap sebuah isu. Seperti fenomena *Citayam Fashion Week* yang ramai diperdebatkan di Indonesia, khususnya pada bulan Juli 2022, sangat dibutuhkan analisis sentimen terhadap fenomena tersebut. Dataset yang digunakan berasal dari twit masyarakat Indonesia dengan kata kunci *Citayam Fashion Week*. Selanjutnya, setiap twit akan dilabeli dengan kelas positif, negatif, atau netral berdasarkan leksikal bahasa Indonesia. Penelitian ini menghasilkan model yang dapat digunakan untuk memprediksi setiap twit bahasa Indonesia ke dalam kategori sentimen positif, negatif, atau netral terkait pandangan dan pendapat masyarakat tentang fenomena *Citayam Fashion Week*. Metode membangun model yang digunakan, yaitu *Long Short Term Memory* (LSTM). Akurasi model yang dihasilkan menggunakan LSTM cukup baik, yaitu sebesar 88%.

Kata Kunci: *analisis sentimen, deep learning, LSTM, twit, Citayam Fashion Week*

ANALISIS SENTIMEN MENGGUNAKAN ARSITEKTUR *LONG SHORT-TERM MEMORY (LSTM)* TERHADAP FENOMENA *CITAYAM FASHION WEEK*

1. Pendahuluan

Fenomena *Citayam Fashion Week* pada bulan Juli 2022 sedang ramai diperbincangkan di Indonesia. *Citayam Fashion Week* merupakan aksi peragaan busana di *zebra cross* kawasan Dukuh Atas, Jakarta Pusat. Para remaja berlenggak-lenggok mengenakan busana-busana yang unik sambil menyeberangi jalan. Para remaja yang meramaikan *Citayam Fashion Week* berasal dari daerah penyangga Jakarta seperti Citayam, Bojong Gede, dan Depok.

Pro dan kontra banyak sekali terjadi terkait fenomena *Citayam Fashion Week*. Banyak yang mendukung *Citayam Fashion Week* karena mereka yakin itu akan memberikan ruang berekspresi untuk anak-anak muda dari keluarga ekonomi menengah ke bawah. Banyak juga yang memuji kreatifitas mereka dalam memadukan busana. Sebagian pengamat yakin bahwa *Citayam Fashion Week* ini juga akan berdampak positif untuk industri busana di Indonesia. Hal positif lainnya dari kegiatan tersebut, yaitu para pelaku juga dapat menjadikan hal tersebut sebagai sarana mereka untuk mencari uang, dengan cara mengunggah kegiatan mereka di aplikasi Instagram, Tiktok, dan Youtube.

Banyak juga pihak yang menentang dan mengancam fenomena *Citayam Fashion Week* itu. Diantaranya karena kehadirannya cukup mengganggu warga lokal, karena kegiatan tersebut memanfaatkan jalur trotoar dan penyeberangan jalan untuk peragaan busana. Beberapa juga berkomentar bahwa kegiatan tersebut sangat kental dengan nuansa *LGBT*. Banyak juga yang meminta agar mereka digusur karena yang dilakukan mereka sangat tidak pantas, seperti pakaian yang sangat terbuka dan pergaulan bebas. Namun ada juga yang miris karena mereka rata-rata masih berusia anak-anak dan remaja yang seharusnya masih dalam pendidikan, namun karena kondisi ekonomi terpaksa harus putus sekolah. Hal negatif lainnya yang banyak diprotes, yaitu karena sampah yang dihasilkan juga sangat banyak dari kegiatan tersebut.

Berbagai pendapat dan komentar terus bermunculan, baik dari pihak yang mendukung maupun yang menolak. Pemerintah juga dituntut untuk mengambil keputusan cepat terhadap fenomena ini. Diperlukan analisis sentimen untuk melihat tanggapan masyarakat terhadap fenomena *Citayam Fashion Week* tersebut. Analisis sentimen merupakan salah satu bidang dari *Natural Language Processing (NLP)* yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks. Kami akan mengambil data Twitter mengenai pendapat pengguna terkait fenomena *Citayam Fashion Week*. Data dari Twitter tersebut diklasifikasikan ke dalam 3 (tiga) kategori, yaitu positif, negatif, dan netral berdasarkan isi atau maksud dari postingan tersebut.

Banyak penelitian baru-baru ini yang melakukan analisis sentimen terhadap topik tertentu berdasarkan komentar masyarakat Indonesia pada media online. Pada tahun 2020, Rahmadan [1] melakukan analisis sentimen terhadap kasus bencana banjir di Jakarta menggunakan metode *Latent Dirichlet Allocation (LDA)*. Zuriel [2] menganalisis tanggapan masyarakat terhadap kebijakan pemerintah selama COVID-19, yaitu penerapan Pembatasan Sosial Berskala Besar atau yang dikenal PSBB. Widianoro [3] menganalisis ulasan pengguna terhadap layanan teknologi keuangan, khususnya layanan OVO pada tahun 2021.

Analisis sentimen terhadap BPJS Kesehatan juga pernah dilakukan oleh Dikiyanti dkk.[4] baru-baru ini berdasarkan data yang diperoleh dari Twitter. Rintyarna [5] telah melakukan analisis sentimen terhadap pandangan masyarakat Indonesia terhadap makanan organik selama pandemic COVID-19. Selama pandemic COVID-19, juga telah dilakukan analisis sentimen terhadap penyedia layanan internet [6].

Sentimen analisis tersebut sangat membantu dalam proses pengambilan

keputusan. Pengambil keputusan tidak perlu menghabiskan banyak waktu untuk mewawancarai banyak orang untuk melihat pendapat secara umum. Saat ini banyak sekali masyarakat yang mengungkapkan opininya melalui media sosial. Para peneliti dapat memperoleh data tersebut dengan mudah dan dapat menganalisis sentimen untuk mendapatkan gambaran umum.

Pada penelitian ini, kami menggunakan metode *Long Short Term Memory* (LSTM) karena metode ini sudah terbukti dapat bekerja dengan baik untuk menganalisis sentimen pada data teks [7]. Penelitian ini menghasilkan model yang dapat digunakan untuk memprediksi setiap tweet bahasa Indonesia ke dalam kategori sentimen positif, negatif, atau netral terkait pandangan dan pendapat mereka tentang fenomena *Citayam Fashion Week*. Hasil prediksi dapat digunakan untuk bahan pertimbangan mengenai langkah apa yang harus dilakukan terhadap *Citayam Fashion Week*.

2. Kajian Pustaka

Analisis sentimen pada teks bertujuan untuk melihat apakah sebuah teks mengandung emosi positif ataupun negatif. Analisis sentimen dalam bahasa Indonesia sudah banyak dilakukan sebelumnya. Analisis sentimen dilakukan untuk melihat pendapat masyarakat yang disalurkan melalui media *online* dengan topik khusus, seperti topik bencana banjir di Jakarta, kebijakan pemerintah melakukan PSBB, pelayanan OVO, BPJS Kesehatan, makanan organik, dan pelayanan penyedia internet selama COVID-19 [1]–[6].

Analisis sentimen berdasarkan topik tertentu khususnya menggunakan data dari Twitter juga sangat sering diimplementasikan baik dalam bahasa Indonesia, bahasa Inggris, maupun bahasa lainnya. Seperti yang dilakukan oleh Kim dkk. [8] pada tahun 2016, mereka menggunakan data Twitter bahasa Inggris untuk menganalisis sentimen masyarakat terhadap virus Ebola. Begitu juga dengan Suresha dkk. [9] pada tahun 2021 juga melakukan analisis sentimen pada data Twitter bahasa Inggris dengan menggunakan topik kendaraan listrik. Alayba dkk. [10] pada tahun 2017 melakukan analisis sentimen dari data Twitter bahasa Arab terhadap topik pelayanan kesehatan.

Banyak metode yang telah diimplementasikan untuk mengetahui sentimen dari sebuah data teks, contohnya metode *Latent Dirichlet Allocation* (LDA) [11]–[14]. LDA adalah sebuah metode untuk mendeteksi topik-topik yang ada pada koleksi dokumen beserta proporsi kemunculan topik tersebut, baik di koleksi maupun di dokumen tertentu. LDA juga mampu mengasosiasikan kata-kata yang ada pada dokumen dan koleksi ke topik tertentu.

Teknik pengklasifikasian sentimen ke dalam positif dan negatif yang digunakan pada penelitian-penelitian sebelumnya ada yang berupa *Machine Learning* dan ada pula yang menggunakan *Deep Learning*. Algoritma klasifikasi dengan teknik *Machine Learning* yang digunakan pada penelitian sebelumnya, yaitu algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* [2], [15].

Dalam beberapa tahun terakhir, *deep learning* telah dipakai dalam berbagai dunia industri dan akademis karena kinerjanya yang sangat baik. Metode *Long Short Term Memory* (LSTM) merupakan salah satu metode *deep learning* yang banyak diaplikasikan untuk melihat sentimen pada data teks [15]–[18]. Kami menggunakan *Long Short Term Memory* (LSTM) untuk menganalisis sentimen pada data Twitter dengan topik *Citayam Fashion Week* karena saat ini LSTM merupakan *Deep Learning* yang paling baik kinerjanya dalam menangani klasifikasi sentimen pada data teks [7].

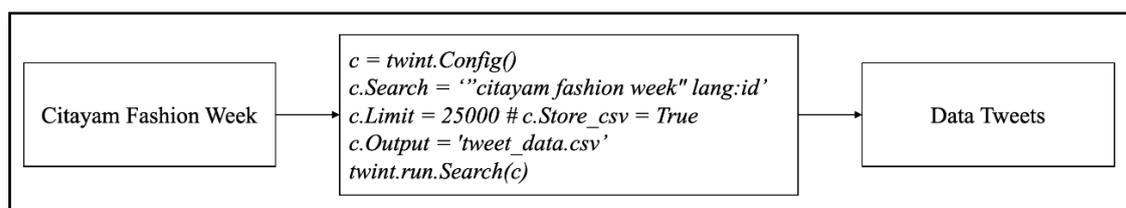
ANALISIS SENTIMEN MENGGUNAKAN ARSITEKTUR *LONG SHORT-TERM MEMORY (LSTM)* TERHADAP FENOMENA *CITAYAM FASHION WEEK*

LSTM berkerja dengan membangun model yang terdiri dari beberapa lapisan, Masing-masing lapisan mengambil *input* dari yang lapisan sebelumnya dan memberikan *output* kepada lapisan berikutnya. Lapisan pertama mengambil urutan numerik sebagai *input*, dan lapisan terakhir memberikan label prediksi sebagai *output*.

3. Metode Penelitian

A. Data Crawling

Data yang digunakan dalam penelitian ini berasal dari cuitan masyarakat Indonesia melalui sosial media Twitter. Data diambil dengan cara *crawling* menggunakan kata kunci *Citayam Fashion Week* pada bulan Juli 2022. Proses *crawling* menggunakan bantuan *Library Twint* pada Python. Data diambil dengan konfigurasi seperti ditunjukkan pada gambar 1.



Gambar 1. Konfigurasi *Twint* pada proses *crawling*

Data yang berhasil didapat dari proses *crawling* tersebut sejumlah 7.554 twit. Seluruhnya merupakan twit bahasa Indonesia yang menyebut *Citayam Fashion Week*. Gambar 2 menunjukkan beberapa contoh data twit berbahasa Indonesia yang berhasil didapatkan setelah proses *crawling*. Setelah data twit dikumpulkan, selanjutnya dilakukan proses pembersihan data twit untuk dijadikan data *training*.

Makanya udah paling pas kembalikan ke situasi awal sebelum Bonge, Jeje Dkk pas masih diwawancarai sebelum viral. Ngga perlu diseremonialisasi jadi Citayam Fashion Week sama influencer yang suka rendahkan menengah ke bawah buat konten dia, apalagi diwawancara TV One.

Bebasnya perilaku anak dan remaja tergambar dr fenomena Citayam Fashion Week. Bukan sekedar kebebasan berekspresi tapi menunjukkan generasi muda yang tak paham agamanya

Wow ... Pak #AniesBaswedan dukung fashion week Citayam...keren banget.. Jakarta itu untuk semua Jakarta itu milik Indonesia

Gambar 2. Contoh twit dengan kata kunci *Citayam Fashion Week*

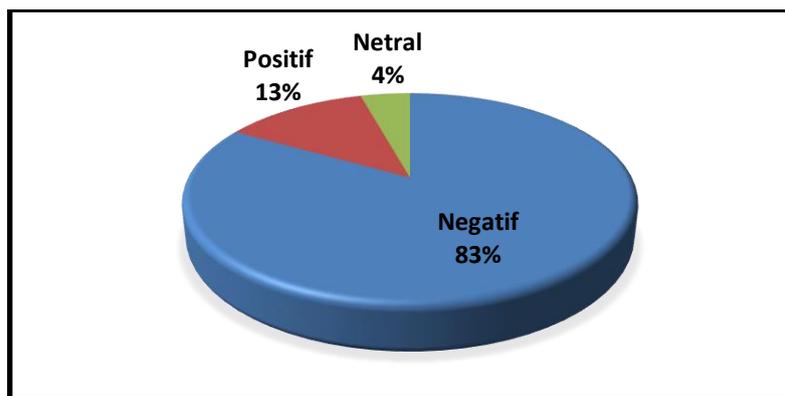
B. Persiapan Data Training

Proses persiapan data twit dimulai dengan membersihkan semua *stopword*, simbol, angka, *username*, RT, *hashtag*, tanda baca, url, dan *emoticon*. Kemudian dilakukan proses *stemming*, yaitu merubah setiap kata menjadi kata dasarnya. Setelah twit dibersihkan, setiap twit akan dilabeli dengan positif, negatif, atau netral berdasarkan kamus leksikal.

Penelitian ini menggunakan sentimen leksikal bahasa Indonesia yang dibuat oleh

Koto dkk. [19] pada tahun 2018. Leksikal tersebut terdiri dari 3.609 kata positif dan 6.609 kata negatif. Setiap kata positif dilabeli dengan skor 1 sampai 5 berdasarkan tingkat level positifnya. Begitu juga dengan kata negatif dilabeli dengan skor -1 sampai -5 berdasarkan tingkat negatifnya.

Proses penentuan setiap twit mengandung sentimen positif, negatif, atau netral, yaitu dengan cara melihat bobot pada kamus leksikal [19] dari setiap kata dalam sebuah twit. Kemudian seluruh bobot kata pada twit dijumlahkan. twit yang memiliki skor lebih besar dari nol dikategorikan ke dalam sentimen positif, twit yang memiliki skor lebih kecil dari nol dikategorikan ke dalam sentimen negatif, dan twit yang memiliki skor sama dengan nol dikategorikan ke dalam sentimen netral. Setelah proses penlabelan sentimen pada setiap twit, dataset untuk penelitian ini mengandung 940 twit positif, 6.291 twit negatif, dan 323 twit netral seperti ditunjukkan pada gambar 3.



Gambar 3. Persentase sentimen pada dataset

Berdasarkan Gambar 3 terlihat bahwa, terdapat 83,3% twit yang mengandung sentimen negatif, sedangkan twit yang mengandung sentiment positif hanya 12,4%. Ternyata lebih banyak masyarakat yang menulis twit tentang *Citayam Fashion Week* dengan sentimen negatif berdasarkan proses pelabelan dengan leksikal. Gambar 4 menunjukkan *wordcloud* dari twit dengan kata kunci *Citayam Fashion Week*.



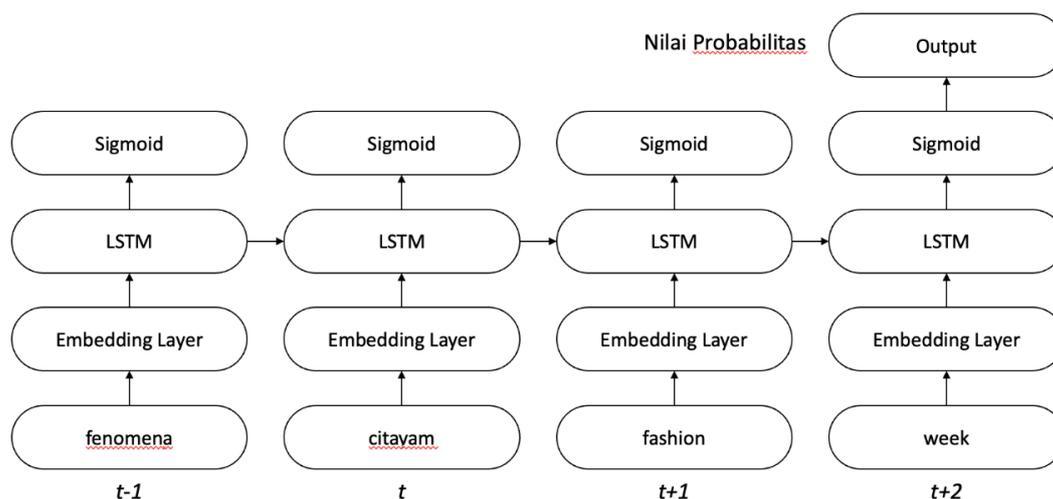
Gambar 4. *Wordcloud* dari twit-twit dengan kata kunci *Citayam Fashion Week*

ANALISIS SENTIMEN MENGGUNAKAN ARSITEKTUR *LONG SHORT-TERM MEMORY* (LSTM) TERHADAP FENOMENA *CITAYAM FASHION WEEK*

Setelah dataset yang berupa twit dilabeli dengan label sentimen positif, negatif, dan netral, selanjutnya akan dibangun model klasifikasi. Model klasifikasi yang digunakan dalam membangun model sentimen untuk kata kunci *Citayam Fashion Week* ini menggunakan metode *Long Short Term Memory* (LSTM).

C. Struktur LSTM

Pada penelitian ini, struktur *Long Short-Term Memory* (LSTM) digunakan untuk membangun model klasifikasi terhadap dataset yang telah dipersiapkan sebelumnya. Struktur LSTM terdiri dari beberapa *layer* yang saling terhubung berdasarkan langkah waktu yang ada pada setiap data yang menjadi inputan seperti yang terlihat pada Gambar 5. Selain itu, struktur LSTM mampu menyimpan informasi dari proses *training* pada tahapan sebelumnya dan menghilangkan informasi yang tidak diperlukan dari tahapan sebelumnya menggunakan gerbang kontrol, yaitu *forget gate*. Penggunaan gerbang tersebut pada struktur LSTM dapat menghindari terjadinya *vanishing/exploding gradient*.



Gambar 5. Struktur LSTM

4. Eksperimen dan Hasil

Pada penelitian ini, eksperimen dimulai dengan membagi dataset menjadi 2 (dua) bagian, yaitu *training* data dan *testing* data. Jumlah *training* data yang digunakan adalah 80% dari total dataset atau sebanyak 6.043 dan jumlah *testing* data adalah 20% dari total dataset atau sebanyak 1.511. Tahapan selanjutnya, yaitu melakukan konfigurasi *hyperparameter* yang akan digunakan. Beberapa *hyperparameter* yang akan disesuaikan diantaranya *optimizer*, *epoch*, *learning rate* dan *batch size* seperti yang ditampilkan pada Tabel 1 dan Tabel 2. Selain itu, jumlah *hidden unit* dan *embed dimension* juga disesuaikan untuk membangun model klasifikasi. Pada tahapan pertama, proses melatih model dilakukan dengan menggunakan konfigurasi tetap untuk *optimizer*, *epoch*, *dropout* dan *hidden unit* seperti yang ditampilkan pada Tabel 1. Sedangkan, nilai *batch size*, *learning rate*, *embed dimension* akan disesuaikan untuk setiap model yang dilatih. Model dilatih dengan menggunakan *framework Tensorflow 2* yang memiliki struktur LSTM seperti yang diperlihatkan pada Gambar 6. Struktur LSTM yang digunakan memiliki jumlah bobot (parameter) sebanyak 353.219 dengan *output layer* sebanyak 3 (tiga) kelas, yaitu positif, negatif, dan netral.

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 45, 64)	320000
lstm (LSTM)	(None, 64)	33024
dropout (Dropout)	(None, 64)	0
dense (Dense)	(None, 3)	195

=====
 Total params: 353,219
 Trainable params: 353,219
 Non-trainable params: 0

Gambar 6. Struktur LSTM pada *Tensorflow 2*

Pada Tabel 1, hasil akurasi terbaik adalah 88% yang diperoleh dengan menggunakan konfigurasi *hyperparameter* tidak berubah, yaitu *adam* sebagai *optimizer*; jumlah *epochs* sebanyak 50, persentase *dropout* sebesar 0,2 dan konfigurasi *hyperparameter* diperbaharui, yaitu *batch size* sebanyak 128, *learning rate* sebesar 0,001 dan *embed dimension* sebesar 64. Proses pelatihan setiap model dilakukan sebanyak 3 (tiga) kali untuk memastikan bahwa akurasi yang didapatkan adalah sama atau tidak jauh berbeda pada setiap proses. Meskipun demikian, akurasi model lain yang diperoleh juga tidak jauh berbeda dengan menggunakan konfigurasi yang sama.

Tabel 1. Konfigurasi *Batch Size*, *Learning Rate*, dan *Embed Dimension*

Konfigurasi	<i>Batch Size</i>	<i>Learning Rate</i>	<i>Embed Dimension</i>	Akurasi	
Optimizer: Adam Epoch: 50 Dropout: 0,2 Hidden Unit: 32	32	0,01	32	0,8628	
			64	0,8496	
		0,001	32	0,8711	
			64	0,8777	
		0,0001	32	0,8430	
			64	0,8430	
	64	0,01	32	0,8744	
			64	0,8529	
		0,001	32	0,8760	
			64	0,8744	
		0,0001	32	0,8628	
			64	0,8595	
		128	0,01	32	0,8628
				64	0,8678
	0,001		32	0,8744	
			64	0,8810	
0,0001	32		0,8628		
	64		0,8612		

ANALISIS SENTIMEN MENGGUNAKAN ARSITEKTUR *LONG SHORT-TERM MEMORY* (LSTM) TERHADAP FENOMENA *CITAYAM FASHION WEEK*

Selanjutnya, konfigurasi *hyperparameter* yang tetap pada proses membangun model sebelumnya akan diperbaharui, sedangkan konfigurasi *hyperparameter* yang berubah akan diatur menjadi tetap seperti yang terlihat pada Tabel 2. Pada proses melatih model ini, hasil akurasi terbaik adalah 88% yang juga diperoleh dengan konfigurasi yang sama dengan proses sebelumnya.

Tabel 2. Konfigurasi *Optimizer*, *Hidden Unit*, dan *Epoch*

Konfigurasi	<i>Optimizer</i>	<i>Hidden Unit</i>	<i>Epoch</i>	Akurasi
Batch size: 128 Learning Rate: 0,001 Embed Dimension: 64 Dropout: 0,2	Adam	16	50	0,8645
			100	0,8760
		32	50	0,8810
			100	0,8711
		64	50	0,8744
			100	0,8694
	RMSprop	16	50	0,8413
			100	0,8479
		32	50	0,8711
			100	0,8479
		64	50	0,8744
			100	0,8595

5. Kesimpulan

Pada penelitian ini, model dilatih menggunakan beberapa konfigurasi *hyperparameter* pada struktur LSTM. Model terbaik yang dihasilkan dari proses pelatihan sangat berpengaruh dengan konfigurasi dan jumlah dataset yang digunakan, Model yang dibangun dengan struktur LSTM memiliki akurasi yang cukup baik untuk dataset yang berasal dari twit dengan kata kunci *Citayam Fashion Week*. Pada penelitian kedepan, penambahan dataset untuk setiap kelas sangat diperlukan sehingga jumlah data pada setiap kelas menjadi lebih seimbang.

References

- [1] M. Choirul Rahmadan, A. Nizar Hidayanto, D. Swadani Ekasari, B. Purwandari, and Theresiawati, "Sentiment Analysis and Topic Modelling Using the LDA Method related to the Flood Disaster in Jakarta on Twitter," *Proc. - 2nd Int. Conf. Informatics, Multimedia, Cyber, Inf. Syst. ICIMCIS 2020*, pp. 126–130, Nov. 2020.
- [2] H. P. P. Zuriel and A. Fahrurrozi, "Implementasi Algoritma Klasifikasi Support Vector Machine Untuk Analisa Sentimen Pengguna Twitter Terhadap Kebijakan Psbb," *J. Ilm. Inform. Komput.*, vol. 26, no. 2, pp. 149–162, 2021.
- [3] A. D. Widiatoro, A. Wibowo, and B. Harnadi, "User Sentiment Analysis in the Fintech OVO Review Based on the Lexicon Method," *2021 6th Int. Conf. Informatics Comput. ICIC 2021*, 2021.
- [4] T. D. Dikiyanti, A. M. Rukmi, and M. I. Irawan, "Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm," *J. Phys. Conf. Ser.*,

- vol. 1821, no. 1, Mar. 2021.
- [5] B. S. Rintyarna, "Mapping acceptance of Indonesian organic food consumption under COVID-19 pandemic using sentiment analysis of Twitter dataset," *Journal of Theoretical and Applied Information Technology*, 2021. https://www.researchgate.net/publication/351990259_Mapping_acceptance_of_indonesian_organic_food_consumption_under_COVID-19_pandemic_using_sentiment_analysis_of_Twitter_dataset (accessed Jul. 27, 2022).
- [6] B. S. Rintyarna *et al.*, "Modelling Service Quality of Internet Service Providers during COVID-19: The Customer Perspective Based on Twitter Dataset," *Informatics*, vol. 9, no. 1, Mar. 2022.
- [7] Dr. G. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, and Mounika Bagadi, Mounika Belusonti, "Text based Sentiment Analysis using LSTM," *Int. J. Eng. Res.*, vol. V9, no. 05, May 2020.
- [8] E. H. J. Kim, Y. K. Jeong, Y. Kim, K. Y. Kang, and M. Song, "Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news," *J. Inf. Sci.*, vol. 42, no. 6, pp. 763–781, Dec. 2016.
- [9] H. P. Suresha and K. Kumar Tiwari, "Topic Modeling and Sentiment Analysis of Electric Vehicles of Twitter Data," *Asian J. Res. Comput. Sci.*, pp. 13–29, Oct. 2021.
- [10] A. M. Alayba, V. Palade, M. England, and R. Iqbal, "Arabic language sentiment analysis on health services," pp. 114–118, Oct. 2017.
- [11] S. Poria, I. Chaturvedi, E. Cambria, and F. Bisio, "Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2016-October, pp. 4465–4473, Oct. 2016.
- [12] F. Mazzoni, V. Marsili, S. Alvisi, and al -, "Latent Dirichlet Allocation (LDA) for Sentiment Analysis Toward Tourism Review in Indonesia," *J. Phys. Conf. Ser.*, vol. 801, no. 1, p. 012073, Jan. 2017.
- [13] I. R. Putri and R. Kusumaningrum, "Latent Dirichlet Allocation (LDA) for Sentiment Analysis Toward Tourism Review in Indonesia," *Journal of Physics: Conference Series*, 2017. <https://iopscience.iop.org/article/10.1088/1742-6596/801/1/012073/meta> (accessed Jul. 28, 2022).
- [14] J. Ye, X. Jing, and J. Li, "Sentiment Analysis Using Modified LDA," *Lect. Notes Electr. Eng.*, vol. 473, pp. 205–212, 2018.
- [15] M. Tripathi, "Sentiment Analysis of Nepali COVID19 Tweets Using NB, SVM AND LSTM," *J. Artif. Intell. Capsul. Networks*, 2021.
- [16] P. K. Jain, V. Saravanan, and R. Pamula, "A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents," *Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 5, Jul. 2021.
- [17] P. M. Sosa and C. Yang, "Twitter Sentiment Analysis using combined LSTM-CNN Models Related papers Recent Trends in Deep Learning Based Natural Language Processing," 2017.
- [18] S. Wen *et al.*, "Memristive LSTM Network for Sentiment Analysis," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 51, no. 3, pp. 1794–1804, Mar. 2021.
- [19] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, pp. 391–394, Feb. 2018.

IMPLEMENTASI DAN ANALISIS KINERJA ANTENA WAJAN BOLIC SEBAGAI PENERIMA SINYAL *WI-FI*

Firmansyah¹, Isnan Bajili², Hendri Ahmadian³, dan Aulia Syarif Aziz³

^{1,2,3,4}FTK Universitas Islam Negeri Ar-Raniry
E-mail: firmansyah.syah@ar-raniry.ac.id

Abstract

There are many free internet services available in public places. To utilize the internet service, we need a Wi-Fi receiver. However, some Wi-Fi receiver devices on the market have limitations in receiving Wi-Fi signals from signal transmitter devices such as wireless access points or Wi-Fi routers. The wajan bolic antenna is claimed able to capture Wi-Fi signals from long range area. In this study, performance testing was compared between a Wi-Fi receiver equipped with a wajan bolic antenna and without a wajan bolic antenna. Performance testing in this study was conducted to understand the effect of using a wajan bolic antenna both on signal reception strength and network quality. Analysis of signal reception is based on Received Signal Strength Indicator (RSSI) method and analysis of Wi-Fi network quality is based on Quality of Service (QoS) parameters such as packet loss, delay and jitter with certain distance limits 5 meters, 35 meters, 55 meters and 100 meters. The results of this study indicate that the implementation of the wajan bolic antenna has a better effect on increasing signal reception and network quality. The performance of the wajan bolic antenna is able to increase the performance in receiving signals up to 46%. The network quality has also increased where the packet loss value is up to 83%, the delay and jitter are each better up to 94%. At access distances above 5 meters the QoS parameter measurement increase with the RSSI measurement value, where the better the signal receiving power then the better the network quality.

Keywords: *Quality of Services, RSSI, Wireless Network*

Abstrak

Saat ini banyak tersedia layanan internet gratis di tempat umum. Untuk menikmati layanan internet tersebut dibutuhkan perangkat penerima Wi-Fi. Tetapi sebagian perangkat penerima Wi-Fi yang beredar dipasaran memiliki keterbatasan dalam menjangkau sinyal Wi-Fi dari perangkat pemancar sinyal seperti *wireless access point* or Wi-Fi router. Antena wajan bolic diklaim mampu menangkap sinyal Wi-Fi puluhan hingga ratusan meter. Pada penelitian ini dilakukan pengujian kinerja terhadap perangkat penerima Wi-Fi yang dilengkapi dengan antena wajan bolic dan tanpa antena wajan bolic. Pengujian kinerja pada penelitian ini dilakukan untuk memahami mengenai pengaruh dari penggunaan antena wajanbolic baik terhadap kekuatan daya terima sinyal dan kualitas jaringan. Analisa daya terima sinyal berdasarkan metode *Received Signal Strength Indicator* (RSSI) dan analisa kualitas jaringan Wi-Fi berdasarkan parameter *Quality of Service* (QoS) berupa *packet loss*,

IMPLEMENTASI DAN ANALISIS KINERJA ANTENA WAJAN *BOLIC* SEBAGAI PENERIMA SINYAL *WI-FI*

delay dan *jitter* dengan batasan jarak tertentu yaitu 5 meter, 35 meter, 55 meter dan 100 meter. Hasil dari penelitian ini menunjukkan bahwa implementasi antena wajan bolic berpengaruh lebih baik dalam meningkatkan daya terima sinyal maupun kualitas jaringan. Kinerja antena wajan bolic mampu meningkatkan performa dalam menerima sinyal hingga 46%. Kualitas jaringan juga meningkat dimana nilai *packetloss* hingga 83%, serta *Delay* dan *Jitter* yang masing – masing lebih baik hingga 94%. Pada jarak akses diatas 5 meter, nilai pengukuran parameter *QoS* berbanding lurus dengan nilai pengukuran *RSSI*, dimana semakin baik daya terima sinyal maka semakin baik pula kualitas jaringan.

Kata Kunci: *Quality of Services, RSSI, Wireless Network*

1 Pendahuluan

Kemajuan teknologi yang pesat mendorong penggunaan jaringan internet dengan tingkat mobilitas yang tinggi menjadi kebutuhan yang tidak dapat dihindarkan dalam keseharian kita. Hal ini terlihat dari banyaknya layanan internet gratis berupa hotspot yang tersedia di berbagai tempat publik dengan menggunakan teknologi *wireless LAN* (WLAN) [1]. Hotspot adalah tempat atau lokasi dimana kita bisa mengakses layanan internet yang disediakan umumnya berupa jaringan WLAN yang terhubung ke penyedia jasa layanan internet.

Kantor Desa Lambaro Skep merupakan salah satu gedung layanan publik yang menyediakan layanan hotspot berupa jaringan WLAN yang bisa diakses pengunjung dengan mudah dan gratis. Layanan ini tidak serta merta dapat dinikmati oleh semua orang. Masyarakat yang berada jauh dari area hotspot tidak bisa menikmati layanan tersebut sehingga beberapa orang berinisiatif untuk menggunakan antena alternatif seperti antena wajan bolic. Antena wajan bolic sendiri diklaim mampu menangkap sinyal puluhan hingga ratusan meter dan bisa dirakit sendiri menggunakan bahan – bahan yang mudah di dapat dengan harga yang murah. Masyarakat cenderung hanya mengetahui manfaat antena tersebut sebagai penguat sinyal tanpa betul - betul mengetahui bukti ilmiah yang sebenarnya baik itu dari segi performa maupun komponen yang digunakan.

Terdapat beberapa penelitian yang terkait dengan pemanfaatan antena wajan bolic. Handoko dalam penelitiannya mendapatkan hasil uji bahwa dengan reflektor bolic terdapat penguatan (*gain*) terhadap kemampuan rata-rata antena mikrostrip dalam menangkap sinyal pada jarak 5 meter, 10 meter, 15 meter dan 20 meter [2]. Rianto dalam penelitian mendapat kualitas sinyal dan *gain* yang dihasilkan oleh antena yagi dengan reflektor bolik lebih baik dari pada kualitas sinyal dan *gain* yang dihasilkan oleh antena omni pada *wireless USB Adapter* pada jarak 30 dan 100 meter [3]. Nur Huda dalam penelitiannya mendapat kualitas sinyal yang lebih baik setelah menggunakan antena wajan bolic pada jarak 50 dan 100 meter [4]. Palendra mendapat hasil yang lebih baik pada perolehan nilai parameter *QoS* seperti *throughput upload, download, latency*, dan *packetloss* saat menggunakan antena penguat sinyal *Wi-Fi* wajan bolic pada jarak 25 meter [5].

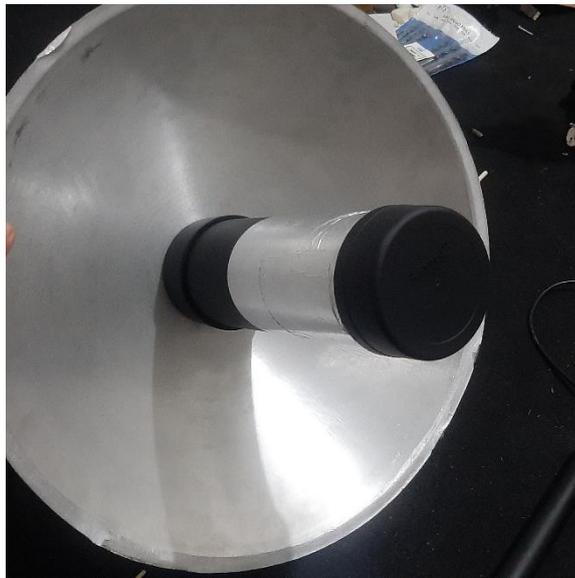
Berdasarkan permasalahan di atas pada artikel ini dilakukan penelitian yang berjudul “Implementasi dan Analisis Kinerja Antena Wajan Bolic dalam Menangkap Sinyal *Wi-Fi*” dalam upaya memahami lebih lanjut mengenai pengaruh dari penggunaan antena wajan bolic baik terhadap daya terima sinyal maupun kualitas jaringan dengan variasi jarak yang mencakup keseluruhan rentang jarak pengujian pada penelitian - penelitian sebelumnya, serta menganalisa hubungan antara kekuatan sinyal dan kualitas jaringan

dalam sebuah jaringan Wi-Fi.

2 Metode

Penelitian ini menggunakan metode eksperimental. Produk eksperimental yang dihasilkan adalah sebuah antena wajan bolic sebagai antena penerima Wi-Fi. Prosedur penelitian yang dilakukan adalah perancangan dan pembuatan antena wajan bolic sebagai penerima sinyal Wi-Fi. Selanjutnya dilakukan pengujian kinerja terhadap perangkat penerima Wi-Fi. Pengujian kinerja dilakukan dalam upaya memahami lebih lanjut mengenai pengaruh dari penggunaan antena wajan bolic terhadap kekuatan daya terima sinyal dan kualitas jaringan. Analisa daya terima sinyal berdasarkan metode *Received Signal Strength Indicator* (RSSI) dan analisa kualitas jaringan Wi-Fi berdasarkan parameter *Quality of Services* (QoS) berupa *packet loss*, *delay* dan *jitter*.

Gambar 1 merupakan hasil rancangan antena wajan bolic yang dilengkapi dengan USB Wi-Fi *adapter* TP-LINK TL WN722N dengan spesifikasi IEEE 802.11n teknologi dengan kecepatan sampai dengan 150 *Mbps*.

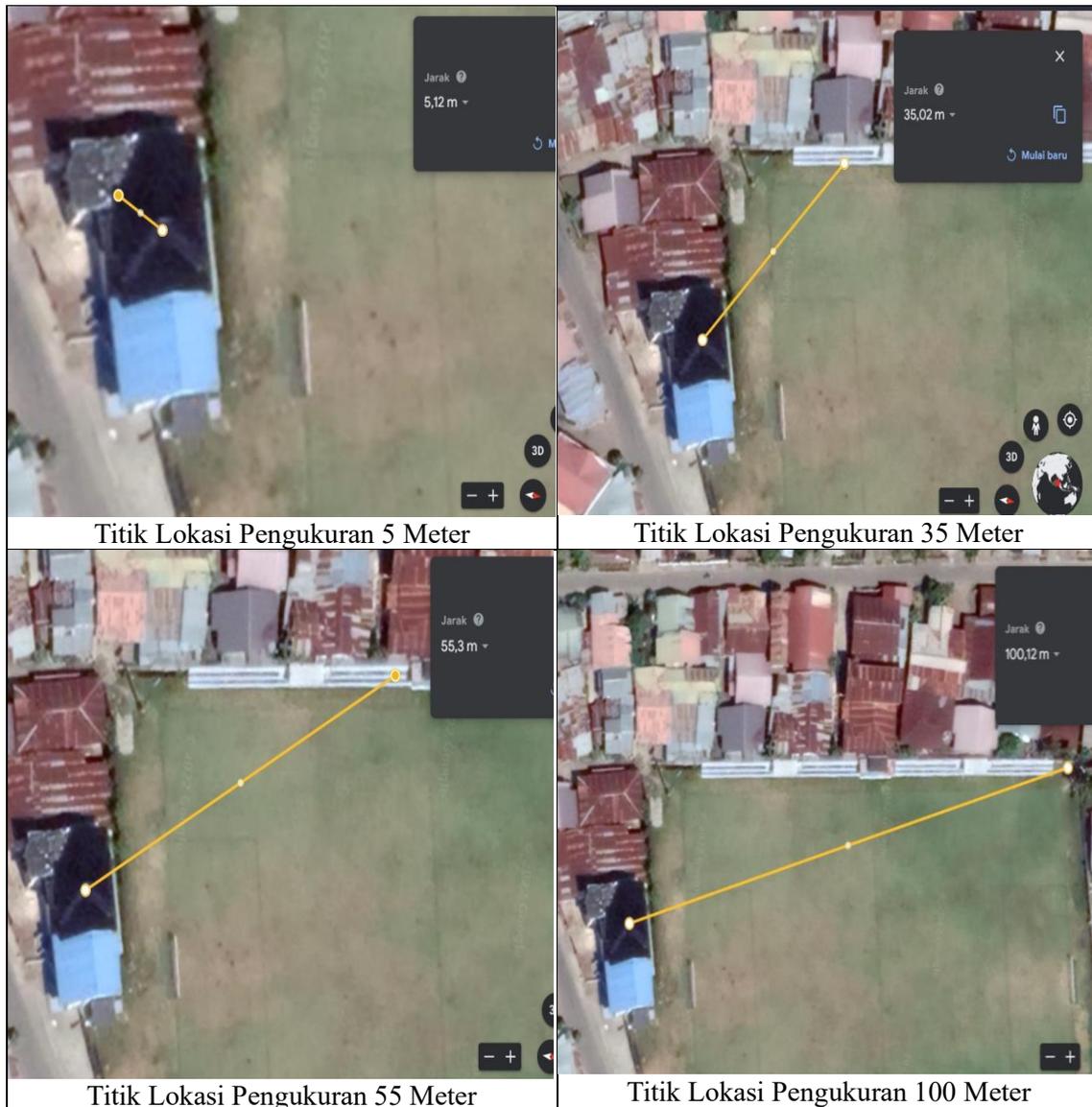


Gambar 1. Tahapan penelitian

Pengambilan data dilakukan dengan melakukan pengukuran nilai RSSI dan parameter QoS. Pengukuran dilakukan dengan dua metode, yang pertama adalah pengukuran tanpa menggunakan antena wajan bolic dan yang kedua dengan menggunakan antena wajan bolic. Antena wajan bolic memerlukan USB Wi-Fi *adapter* eksternal agar bisa digunakan pada laptop. USB Wi-Fi *adapter* digunakan sebagai perangkat eksternal yang menggantikan peran Wi-Fi adapter laptop. Pengukuran dilakukan dalam berbagai variasi jarak pengukuran. Jarak pengukuran yang dimaksud adalah jarak antara titik dimana pengukuran dilakukan dengan titik lokasi *access point* berada. Titik lokasi *access point* berada di dalam gedung kantor desa dengan ketinggian 2 meter dari permukaan tanah. Sementara itu titik lokasi pengukuran berada pada jarak masing – masing 5 meter, 35 meter, 55 meter dan 100 meter dari titik *access point* dengan ketinggian 1 meter dari permukaan tanah. Penentuan titik lokasi ini berdasarkan kombinasi jarak dari penelitian terkait yang telah dilakukan sebelumnya yang kemudian disesuaikan dengan letak geografis gedung dan lingkungan sekitarnya. Berikut adalah

IMPLEMENTASI DAN ANALISIS KINERJA ANTENA WAJAN *BOLIC* SEBAGAI PENERIMA SINYAL *WI-FI*

gambar peta titik lokasi pengukuran berdasarkan jarak.



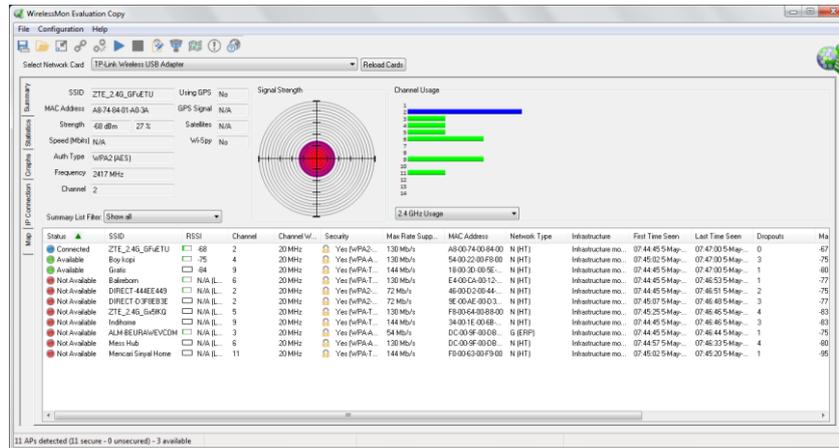
Gambar 2. Titik lokasi pengukuran

Pengukuran nilai RSSI dilakukan dengan melakukan pemantauan menggunakan aplikasi *Wirelessmon*. Laptop dihubungkan dengan *SSID* jaringan *Wi-Fi* yang tersedia pada *access point*. RSSI digunakan sebagai indeks yang menunjukkan kekuatan sinyal yang diterima oleh *receiver* dari *access point*. Satuan kekuatan sinyal *wireless* ditunjukkan dengan satuan *dBm* dengan rentang *Signal Strength* yaitu -10 *dBm* sampai -100 *dBm*. Semakin mendekati angka positif maka kualitas sinyal semakin baik [6]. Namun pemetaan langsung dari nilai RSSI yang berdasarkan jarak memiliki banyak keterbatasan karena pada dasarnya RSSI rentan terhadap *noise*, *multi-path fading*, dan lain sebagainya yang mengakibatkan fluktuasi besar dalam kekuatan yang diterima.

Tabel 1. Kategori RSSI [6]

Kategori	RSSI
Sangat Baik	> -70 dBm
Baik	-70 dBm to -85 dBm
Sedang	-86 dBm to -100 dBm
Buruk	< -100 dBm

Instrumen yang dipantau adalah nilai indikator *Strength (-dBm)* pada SSID terakses seperti yang ditunjukkan pada Gambar 3 di bawah ini.



Gambar 3. Pemantauan nilai RSSI

Quality of Service (QoS) didefinisikan sebagai suatu pengukuran tentang seberapa baik jaringan dan merupakan suatu usaha untuk mendefinisikan karakteristik dan sifat dari suatu layanan. QoS mengacu pada kemampuan jaringan untuk memberikan layanan yang lebih baik melalui berbagai teknologi yang berbeda-beda. Tujuan dari adanya QoS adalah sebagai parameter dalam penetapan standar performa untuk memenuhi kebutuhan layanan yang berbeda dalam penggunaan infrastruktur yang sama [7]. Parameter dari QoS salah satunya meliputi *packetloss*, *delay* dan *jitter* [8].

Packetloss didefinisikan sebagai kegagalan transmisi paket dalam mencapai tujuannya. Kegagalan tersebut dapat disebabkan oleh beberapa faktor seperti *overload* trafik di dalam jaringan, tabrakan atau *congestion*, kegagalan yang terjadi pada media fisik hingga kegagalan pada sisi penerima.

Tabel 2. Kategori *Packetloss* [9]

Kategori	Packetoss (%)	Indeks
Sangat Baik	0	4
Baik	< 3	3
Sedang	< 15	2
Buruk	> 25	1

Delay adalah waktu jeda atau tunda dari suatu paket, hal ini disebabkan oleh proses transmisi paket itu sendiri dari titik awal ke titik tujuannya.

IMPLEMENTASI DAN ANALISIS KINERJA ANTENA WAJAN *BOLIC* SEBAGAI PENERIMA SINYAL *WI-FI*

Tabel 3. Kategori *Delay* [9]

Kategori	<i>Delay (ms)</i>	Indeks
Sangat Baik	< 150 <i>ms</i>	4
Baik	150 - 300 <i>ms</i>	3
Sedang	300 <i>ms</i> s/d 450 <i>ms</i>	2
Buruk	> 450 <i>ms</i>	1

Jitter adalah variasi dari *delay* antar paket yang terjadi pada sebuah jaringan. Besarnya nilai *jitter* ini dipengaruhi oleh variasi dari beban trafik dan seberapa besarnya penyumbatan yang terjadi antar paket yang ada dalam jaringan. Nilai sebuah *jitter* akan berbanding lurus dengan besaran beban trafik serta penyumbatan yang terjadi [7].

Tabel 4. Kategori *Jitter* [9]

Kategori	<i>Jitter (ms)</i>	Indeks
Sangat Baik	0	4
Baik	0 -75	3
Sedang	75-125	2
Buruk	125-225	1

3 Hasil dan Pembahasan

Pada bagian ini menunjukkan hasil dari pengukuran dan analisis dari nilai RSSI dan QOS.

3.1 RSSI

RSSI dihitung dalam satuan (*-dBm*). Rangkuman nilai pengukuran terdapat pada Tabel 5 di bawah ini.

Tabel 5. Hasil Pengukuran Nilai *RSSI*

Jarak	Tanpa Antena Wajan <i>Bolic</i>		Dengan Antena Wajan <i>Bolic</i>	
	Nilai	Kategori	Nilai	Kategori
5 meter	-39 <i>dBm</i>	Sangat Baik	-25 <i>dBm</i>	Sangat Baik
35 meter	-80 <i>dBm</i>	Baik	-62 <i>dBm</i>	Sangat Baik
55 meter	-	-	-59 <i>dBm</i>	Sangat Baik
100 meter	-	-	-72 <i>dBm</i>	Baik

3.2 Packetloss

Nilai *packetloss* di ukur dalam satuan persen (%). Semakin rendah nilai yang diperoleh maka semakin baik kualitas dari *packetloss*.

Tabel 6. Hasil Pengukuran Nilai *Packetloss*

Metode	Jarak	<i>Packetloss (%)</i>				
		Pengukuran			Rata-rata	Kategori
		1	2	3		
Tanpa Antena Wajan <i>Bolic</i>	5 meter	0,2	0,2	0,2	0,2	Sangat Baik
	35 meter	1,6	1,2	3,4	2,3	Sangat Baik
	55 meter	-	-	-	-	-
	100 meter	-	-	-	-	-
Dengan Antena Wajan <i>Bolic</i>	5 meter	0,2	0,2	0,2	0,2	Sangat Baik
	35 meter	0,3	0,5	0,3	0,4	Sangat Baik
	55 meter	0,2	0,7	0,3	0,4	Sangat Baik
	100 meter	1,2	5	2,5	2,9	Sangat Baik

Walaupun terdapat perbedaan nilai pengukuran pada pengujian dengan dan tanpa menggunakan antena wajan *bolic* pada semua jarak pengukuran, namun semua nilai pengukuran dikategorikan sangat baik. Ini artinya perbedaan nilai pengukuran tidak lah terlalu besar atau signifikan jika dinilai dari indeks parameter kualitas *packetloss*.

3.3 Delay

Nilai *delay* di ukur dalam satuan *milisecond (ms)*, semakin rendah nilai yang diperoleh maka semakin baik kualitas dari *delay*. Rangkuman hasil pengukuran nilai *delays* dapat dilihat pada Tabel 7 di bawah ini.

Tabel 7. Hasil Pengukuran Nilai *Delay*

Metode	Jarak	Delay (ms)				
		Pengukuran			Rata-rata	Kategori
		1	2	3		
Tanpa Antena Wajan <i>Bolic</i>	5 meter	1,4	1,4	1,4	1,4	Sangat Baik
	35 meter	35	35	19	29,7	Sangat Baik
	55 meter	-	-	-	-	-
	100 meter	-	-	-	-	-
Dengan Antena Wajan <i>Bolic</i>	5 meter	1,4	1,4	1,4	1,4	Sangat Baik
	35 meter	1,7	1,9	2	1,9	Sangat Baik
	55 meter	2,3	4	2,7	3	Sangat Baik
	100 meter	11,2	23,5	14,2	16,3	Sangat Baik

Walaupun terdapat perbedaan nilai pengukuran pada pengujian dengan dan tanpa menggunakan antena wajan *bolic* pada semua jarak pengukuran, namun semua nilai pengukuran dikategorikan sangat baik. Ini artinya perbedaan nilai pengukuran tidak lah terlalu besar atau signifikan jika dinilai dari indeks parameter kualitas *delay*.

3.4 Jitter

Nilai *jitter* di ukur dalam satuan *milisecond (ms)*, semakin rendah nilai yang diperoleh maka semakin baik kualitas dari *jitter*. Rangkuman hasil pengukuran nilai *jitter* dapat dilihat pada Tabel 8 di bawah ini.

Tabel 8 Hasil Pengukuran Nilai *Jitter*

Metode	Jarak	Jitter (ms)				
		Pengukuran			Rata-rata	Kategori
		1	2	3		
Tanpa Antena Wajan <i>Bolic</i>	5 meter	2,5	2,4	2,3	2,4	Baik
	35 meter	56	52	59	45,7	Baik
	55 meter	19,4	26,9	30,1	25,5	Baik
	100 meter	-	-	-	-	-
Dengan Antena Wajan <i>Bolic</i>	5 meter	2,4	2,5	2,4	2,4	Baik
	35 meter	2,7	3	3,3	3	Baik
	55 meter	3,9	6,8	4,3	5	Baik
	100 meter	17	23	23,4	26,9	Baik

Walaupun terdapat perbedaan nilai pengukuran pada pengujian dengan dan tanpa menggunakan antena wajan *bolic*, namun semua nilai pengukuran tersebut dikategorikan baik. Ini artinya perbedaan nilai pengukuran tidak terlalu besar atau signifikan jika dinilai dari indeks parameter kualitas *jitter*.

IMPLEMENTASI DAN ANALISIS KINERJA ANTENA WAJAN *BOLIC* SEBAGAI PENERIMA SINYAL *WI-FI*

3.5 Analisis Hasil Pengukuran Nilai *RSSI*

Untuk nilai *RSSI* digunakan perhitungan persentase penurunan karena semakin mendekati angka positif maka semakin baik kualitas sinyal. Artinya semakin besar nilai persentase penurunan maka semakin baik pengaruh penggunaan antena wajan *bolic* terhadap kualitas sinyal. Hasil pengukuran yang dibandingkan adalah nilai pengukuran pada jarak 5 meter, 35 meter. Sedangkan pada jarak 100 meter tidak dilakukan perhitungan nilai perbandingan dikarenakan pengujian tanpa wajan *bolic* tidak bisa dilakukan sehingga tidak menghasilkan nilai pengukuran.

Tabel 9 Analisis Perbandingan Nilai *RSSI*

Analisa Perbandingan Nilai <i>RSSI</i> (<i>dBm</i>) Sebelum dan Sesudah Menggunakan Antena Wajan <i>Bolic</i>			
Jarak	Sebelum	Sesudah	Peningkatan Performa
5 meter	-39	-25	36%
35 meter	-80	-62	46%
100 meter		-72 <i>dBm</i>	

Hasil pengukuran dengan menggunakan antena wajan *bolic* menunjukkan keunggulan pada setiap jarak pengukuran. Dimulai pada jarak 5 meter yang mampu memperoleh nilai hingga 36% lebih baik ketimbang pengukuran tanpa menggunakan antena wajan *bolic*. Juga seterusnya pada jarak 35 meter hasil 23 % lebih baik. Keunggulan paling signifikan terjadi pada jarak pengukuran 100 meter dimana pengukuran dengan menggunakan antena wajan *bolic* masih mampu menerima sinyal dari *access point*. Sementara itu tanpa menggunakan antena wajan *bolic* sinyal tidak lagi dapat diterima. Dengan begitu artinya tidak bisa dilakukan perhitungan perbandingan nilai pada jarak pengukuran ini. Kondisi ini menunjukkan bahwa tanpa antena wajan *bolic* *USB Wi-Fi Adapter* hanya mampu menerima sinyal dengan jarak dibawah 100 meter.

3.6 Analisa Hasil Pengukuran Nilai Parameter *QoS*

Analisa hasil pengukuran ini meliputi perbandingan nilai *packetloss*, *delay* dan *jitter*. Hasil pengukuran yang dibandingkan adalah nilai pengukuran pada jarak 5 meter, 35 meter. Sedangkan pada jarak 55 dan 100 meter tidak dilakukan perhitungan nilai perbandingan dikarenakan pengujian tanpa wajan *bolic* tidak menghasilkan nilai pengukuran.

Perhitungan persentase penurunan digunakan pada perbandingan nilai *packetloss*. Semakin rendah nilai yang diperoleh maka semakin baik kualitas dari *packetloss*. Artinya semakin besar nilai persentase penurunan maka semakin baik pengaruh dari penggunaan antena wajan *bolic*. Ringkasan perhitungan berdasarkan Tabel 10 di bawah ini.

Tabel 10 Analisa Perbandingan Nilai *Packetloss*

Analisa Perbandingan Nilai Rata-rata *Packetloss* (%) Sebelum dan Sesudah Menggunakan Antena Wajan *Bolic*

Jarak	Sebelum	Sesudah	Peningkatan Performa
5 meter	0,2	0,2	0%
35 meter	2,3	0,4	83%

Pada jarak 5 meter tidak terdapat perbedaan yang signifikan dimana kedua metode pengukuran memperoleh nilai pengukuran yang relatif sama. Namun pada jarak 35 meter penggunaan antena wajan *bolic* nilai pengukuran sebesar 83 % lebih baik.

Kemudian perhitungan persentase penurunan pada parameter *delay*. Semakin rendah nilai yang diperoleh maka semakin baik kualitas dari *delay*. Artinya semakin besar nilai persentase penurunan maka semakin baik pengaruh dari penggunaan antena wajan *bolic*. Ringkasan perhitungan berdasarkan Tabel 11 di bawah ini.

Tabel 11 Analisa Perbandingan Nilai *Delay*

Analisa Perbandingan Nilai Rata-rata *Delay* (ms) Sebelum dan Sesudah Menggunakan Antena Wajan *Bolic*

Jarak	Sebelum	Sesudah	Peningkatan Performa
5 meter	1,4	1,4	0%
35 meter	29,7	1,9	94%

Pada jarak 5 meter tidak terdapat perbedaan yang signifikan dimana performa pada kedua metode pengukuran relatif sama. Seterusnya pada jarak 35 meter penggunaan antena wajan *bolic* menghasilkan performa hingga 94 % lebih baik.

Selanjutnya perhitungan persentase penurunan digunakan pada perbandingan nilai *jitter*. Semakin rendah nilai yang diperoleh maka semakin baik kualitas dari *jitter*. Artinya semakin besar nilai persentase penurunan maka semakin baik pengaruh dari penggunaan antena wajan *bolic*. Ringkasan perhitungan berdasarkan Tabel 12 di bawah ini.

Tabel 12 Analisa Perbandingan Nilai *Jitter*

Analisa Perbandingan Nilai Rata-rata *Jitter* (ms) Sebelum dan Sesudah Menggunakan Antena Wajan *Bolic*

Jarak	Sebelum	Sesudah	Peningkatan Performa
5 meter	2,4	2,4	0%
35 meter	45,7	3	93%

4 Kesimpulan

Setelah dilakukan pengukuran dan analisa terhadap implementasi antena wajan *bolic* untuk mengetahui bagaimana kinerja antena wajan *bolic* menggunakan *USB Wi-Fi Adapter* terhadap kekuatan daya terima sinyal dan kualitas jaringan, maka dapat disimpulkan sebagai berikut:

1. Implementasi antena wajan *bolic* berpengaruh baik terhadap performa dalam meningkatkan daya terima sinyal maupun kualitas jaringan. Kinerja antena wajan *bolic* mampu meningkatkan performa dalam menerima sinyal hingga 46% pada jarak

IMPLEMENTASI DAN ANALISIS KINERJA ANTENA WAJAN *BOLIC* SEBAGAI PENERIMA SINYAL *WI-FI*

- 35 meter. Penggunaan wajan *bolic* juga mampu menangkap sinyal sampai 100 dengan nilai -72 dbm yang masuk dalam kategori baik.
2. Penggunaan *USB Wi-Fi Adapter* tanpa wajan *bolic* mampu menangkap sinyal maksimum jarak 35 meter sedangkan penggunaan wajan *bolic* beserta *USB Wi-Fi Adapter* mampu menangkap sinyal hingga 100 meter dengan parameter *QOS* yang baik.
 3. Kualitas jaringan juga meingkat dimana nilai *packetloss* hingga 83% pada jarak 35 meter, serta *Delay* dan *Jitter* yang masing – masing lebih baik hingga 94% dan 93% pada jarak 35 meter.

Referensi

- [1] . A. and . E., “IEEE 802.11ac sebagai Standar Pertama untuk Gigabit Wireless LAN,” *J. Rekayasa Elektr.*, vol. 11, no. 1, pp. 36–44, 2014, doi: 10.17529/jre.v11i1.1994.
- [2] F. Handoko, F. Imasnyah, and D. Suryadi, “Analisis Pengaruh Reflector Bolic Pada Antena Mikrostrip Dalam Meningkatkan Penguatan Sinyal Wifi,” pp. 6–7, 2018.
- [3] T. I. Rianto, F. Imansyah, and D. Suryadi, “Analisis Rancang Bangun Antena yagi dengan Reflektor Bolik sebagai Penguat Daya Tangkap Wireless USB Adapter dengan Frekuensi Kerja 2.4 GHz,” *Technol. Rev.*, vol. 106, no. 7, p. 81, 2003.
- [4] B. N. Huda, S. T. Umi Fadlillah, and M. Eng, “Pembuatan dan analisis perbandingan kinerja wajan *bolic* dan antena kaleng dalam menangkap sinyal wifi,” Universitas Muhammadiyah Surakarta, 2014.
- [5] R. Palendra and S. Broto, “Pengujian Quality of Service Jaringan Internet dengan Memanfaatkan Antena Wajan Bolic Sebagai Penguat Sinyal Wifi,” *J. Maest.*, vol. 2, no. 2, pp. 455–464, 2019.
- [6] T. S. J. Putra and I. R. Widiyari, “Analisis Kualitas Signal Wireless Berdasarkan Received Signal Strength Indicator (RSSI) pada Universitas Kristen Satya Wacana,” *Teknol. Informsi*, no. 672014132, 2018.
- [7] Aprianto Budiman, M. Ficky Duskarnaen, and Hamidillah Ajie, “Analisis Quality of Service (Qos) Pada Jaringan Internet Smk Negeri 7 Jakarta,” *PINTER J. Pendidik. Tek. Inform. dan Komput.*, vol. 4, no. 2, pp. 32–36, 2020, doi: 10.21009/pinter.4.2.6.
- [8] T. Hidayat, U. H. Medan, J. H. M. Joni, and C. No, “Perbandingan Quality of Service Jaringan Dan Antena Wifi Gun,” pp. 187–193, 2020.
- [9] Tiphon, “Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) General aspects of Quality of Service (QoS),” no. DTR/TIPHON-05006 (cb0010cs.PDF), 1999.

FULLY HOMOMORPHIC ENCRYPTION (FHE) PADA PENYIMPANAN DATA E-GOVERNMENT BERBASIS CLOUD

Rini Deviani¹, Sri Azizah Nazhifah¹, dan Aulia Syarif Aziz²

¹Jurusan Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Syiah Kuala, Darussalam, Banda Aceh, 23111, Indonesia

²Jurusan Pendidikan Teknologi Informasi, Fakultas Tarbiyah dan Keguruan,
UIN Ar-Raniry, Darussalam, Banda Aceh, 23111, Indonesia
E-mail: rini.deviani@unsyiah.ac.id, sriazizah07@unsyiah.ac.id,
aulia.aziz@ar-raniry.ac.id

ABSTRACT

Cloud computing is a way of providing services, networks, hardware, storage, and interfaces for the construction of E-Government infrastructure that enables the delivery of services efficiently and achieves cost savings. Cloud services enable individuals and organizations to utilize cloud providers' software and hardware resources stored remotely. The span between the client and the actual location of his data provides a barrier because this data can be obtained by a third party, risking the citizen's data privacy. We examine a method based on the Fully Homomorphic Encryption (FHE) scheme in order to facilitate the processing of sensitive information pertaining to the E-Government that does not involve the disclosure of the original data. In this paper, we consider some general data operations to evaluate the feasibility of the FHE method and show that the accuracy are similar when data operations are applied to homomorphically encrypted data. The results of the experiment highlight the potential of the various privacy-preserving data operations that can be performed under FHE approach. These methods provide results that are equivalent to those achieved by unencrypted data and models within a decent amount of time.

Keywords: *Cloud Computing, E-Government, Cryptography, Homomorphic, Encryption*

ABSTRAK

Komputasi awan adalah gabungan komponen sistem teknologi informasi seperti servis, jaringan, perangkat keras, penyimpanan, dan *interface* untuk membuat infrastruktur *E-Government* yang dapat memberikan layanan secara efektif. Layanan *cloud* menyediakan kemampuan kepada setiap pengguna untuk dapat mengakses perangkat lunak dan perangkat keras yang disimpan secara jarak jauh oleh penyedia layanan *cloud*. Pada saat proses pengiriman data atau permintaan data ke penyedia layanan *cloud*, proses ini dapat menyebabkan data tersebut diperoleh oleh pihak ketiga. Pada penelitian ini kami menguji sebuah metode skema *Fully Homomorphic Encryption* (FHE) untuk memfasilitasi perlindungan kerahasiaan dan keamanan data pribadi yang berkaitan dengan *E-Government* yang tidak melibatkan pengungkapan data asli (*plaintext*). Dalam penelitian ini, kami memperhitungkan beberapa operasi

FULLY HOMOMORPHIC ENCRYPTION (FHE) PADA PENYIMPANAN DATA E-GOVERNMENT BERBASIS CLOUD

data yang umum dilakukan dan menguji apakah hasil dari operasi data tersebut memiliki akurasi serupa, ketika operasi data diterapkan pada data yang telah dienkripsi secara homomorfik maupun pada data yang tidak dienkripsi. Hasil pengujian menunjukkan bahwa operasi data dapat dilakukan dengan metode FHE. Penerapan metode FHE pada beberapa operasi data yang dilakukan pada penelitian ini memberikan hasil yang sebanding dengan hasil yang diperoleh pada data dan model yang tidak dienkripsi. Hasil didapatkan dalam waktu komputasi yang layak.

Kata Kunci: *komputasi awan, E-Government, kriptografi, Homomorfik, enkripsi*

1. PENDAHULUAN

Berbagai jenis *framework* untuk pengembangan sistem *Electronic Government* atau biasa disebut *E-Government* sangat berkembang pesat seiring dengan perkembangan teknologi informasi dan komunikasi. Bagian utama dari serangan bisa berupa data yang tidak diautentifikasi dan adanya perubahan informasi pada *E-Government* yang drastis. Maka dari itu *framework E-Government* harus dilindungi secara kuat sehingga data dan informasinya tetap aman dari serangan apapun.

Umumnya data yang disimpan pada *cloud* akan dilakukan proses enkripsi (*encryption*). Akan tetapi, jika pengguna akan melakukan pengolahan data, maka layanan *cloud* akan dilakukan proses dekripsi (*decryption*) dan mengirimkannya kembali ke pengguna dalam bentuk yang telah didekripsi (*decrypted*). Pada saat proses *encryption-decryption* di *cloud*, kemungkinan terjadinya kebocoran dan peretasan sangatlah besar. Hal ini dapat dihindari dengan menggunakan teknik *Fully Homomorphic Encryption* (FHE).

Beberapa penelitian terkait keamanan data pada penyimpanan berbasis *cloud* menggunakan metode enkripsi homomorfik telah dilakukan pada beberapa tahun terakhir. Algoritma kriptografi homomorfik baru dan efektif yang ringan untuk tujuan keamanan data berbasis *cloud computing* dilakukan pada [1]. Adapun optimalisasi enkripsi homomorfik menggunakan *Application Programming Interface* (API) pada penyimpanan *cloud* dalam penelitian ilmu material yang melibatkan kumpulan data besar telah dilakukan pada [2]. Penelitian [3] mengusulkan skema enkripsi homomorfik praktis yang dapat memungkinkan pengguna data dalam sistem IoT untuk mengoperasikan data dengan aman melalui data terenkripsi, yang secara efektif dapat melindungi privasi data kunci dalam sistem.

Penelitian ini bertujuan untuk melakukan simulasi dan menganalisa kemampuan skema FHE dalam pengolahan data terenkripsi terhadap data yang disimpan menggunakan penyimpanan berbasis *cloud*. Penelitian ini membahas tentang hasil percobaan terhadap beberapa skema data yang umum digunakan pada aplikasi *E-Government*. Pengolahan data yang akan dilakukan percobaan diantaranya pencarian data, pengelompokan data menggunakan K-Means, dan klasifikasi berbasis *Convolutional Neural Network* (CNN).

2. ENKRIPSI HOMOMORPHIC

Enkripsi digunakan untuk melindungi semua data yang disimpan pada *cloud*. Data akan didekripsi jika pengguna akan melakukan pemrosesan di dalam *cloud* agar data tidak rentan dari *hacker*. Pada saat pengguna mengakses atau mengolah data di penyimpanan berbasis *cloud*, *hacker* akan lebih mudah meretas data tersebut, maka dari itu Enkripsi Homomorfik dikembangkan untuk menghindari peretasan data.

Rivest pertama kali mengusulkan “*privacy homomorphism*” sebagai dasar untuk *Homomorphic Encryption* (HE) pada tahun 1978. Kemudian ia menguji HE untuk keamanan *computation* pada *Big Data* [4]. HE adalah jenis metode enkripsi yang memungkinkan modifikasi langsung dari *ciphertext*. Ide dasarnya adalah untuk memungkinkan mengolah data pada data terenkripsi dengan memanfaatkan sifat matematika tertentu dari berbagai skema enkripsi. Informasi yang didapatkan tetap dalam keadaan terenkripsi dan dapat diproses lebih lanjut atau didekripsi.

Pada tahun 2009, Gentry mengusulkan teknik enkripsi homomorfik pertama berdasarkan *lattice* ideal [5]. Skema ini mampu melakukan operasi penjumlahan dan perkalian menggunakan *ciphertext* dalam beberapa kali, ini adalah versi pertama dari enkripsi homomorfik. Kemudian, perkembangan teknologi enkripsi homomorfik mengalami kemajuan yang sangat pesat. Teknologi enkripsi homomorfik dapat dipecah menjadi kategori berikut:

1. Teknik FHE pertama berbasis *lattice* ideal yang diusulkan oleh Gentry. Metode ini melibatkan konstruksi *SomeWhat Homomorphic Encryption* (SWHE), dimana operasi dilakukan pada bilangan terbatas baik operasi penjumlahan maupun perkalian.
2. Berdasarkan konsep Gentry tentang skema enkripsi homomorfik berbasis bilangan bulat [6], yang tidak memerlukan operasi berdasarkan *lattice* ideal dari *ring* polinomial ideal. Hanya integer yang digunakan dalam semua operasi. Ini juga disebut *Partially Homomorphic Encryption* (PHE), dalam jenis enkripsi ini, hanya satu operasi yang dapat dilakukan pada data terenkripsi baik dengan penambahan atau perkalian. Kriptosistem Pillar hanya melakukan operasi penjumlahan sedangkan kriptosistem RSA melakukan operasi perkalian pada data.
3. Metode terakhir adalah sistem *Fully Homomorphic Encryption* (FHE) yang didasarkan pada *Learning with Errors* (LWE) atau *Learning with Errors over Ring* (R-LWE). Sistem ini dibangun di atas pembelajaran yang toleran terhadap kesalahan, dan menghasilkan skema enkripsi yang sepenuhnya homomorfik dengan memanfaatkan non-linearisasi. Salah satu contoh skema enkripsi jenis ini adalah skema enkripsi Brakerski-Gentry-Vaikuntanathan (BGV) [7].

Karakteristik FHE yang bersifat *multi-layer* menyebabkan sistem berjalan sangat lambat. Untuk mengatasi masalah ini, banyak peneliti telah menggabungkan beberapa skema. Dalam beberapa tahun terakhir, sejumlah *library* HE yang bersifat *open source* telah berkembang. Semuanya memiliki kelebihan dan kekurangannya masing-masing sesuai algoritma enkripsi yang diterapkan [8]. *Library* yang paling sering digunakan adalah IBM HELib [9], dimana *library* tersebut mengadopsi skema yang dikembangkan oleh Brakerski-Gentry-Vaikuntanathan (BGV) [7], dan skema pada Microsoft Simple Encrypted Arithmetic Library (SEAL) [10]. Skema ini didukung oleh skema Brakerski/Fan-Vercauteren (BFV) [11] dan skema Cheon-Kim-Kim-Song (CKKS) [12].

3. DATA E-GOVERNMENT

E-Government menyediakan otomatisasi terhadap semua kegiatan pemerintah dan meningkatkan efisiensi dari sebuah organisasi dan warga yang berpartisipasi dalam pemerintahan [13]. *E-Government* yang efektif pada suatu negara dapat ditunjukkan dengan diterapkannya *E-Governance Requirements* dan *E-Governance Components*. *E-Government* diharapkan mampu meningkatkan kinerja pemerintah dan mudah berbagi informasi dengan warga kapanpun dan dimanapun. Dalam proses implementasi praktis

FULLY HOMOMORPHIC ENCRYPTION (FHE) PADA PENYIMPANAN DATA E-GOVERNMENT BERBASIS CLOUD

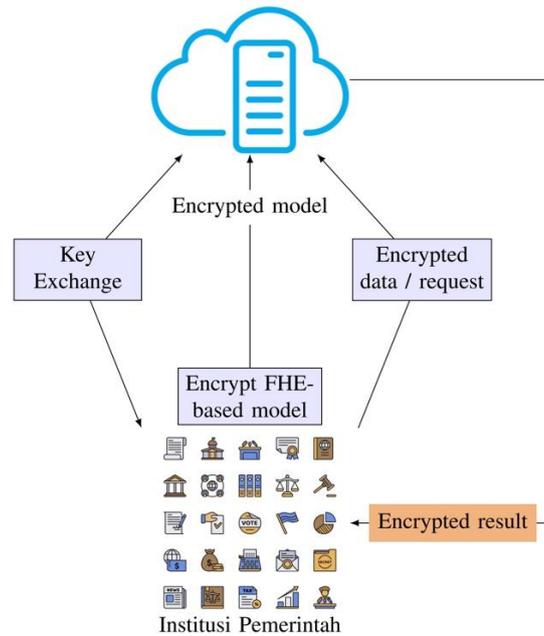
E-Government, ada beberapa hal penting untuk mengidentifikasi faktor-faktor khusus yang akan memainkan peran penting selama penerapan *E-Government*. Ada tiga aspek *E-Governance Requirements*:

1. *Government to Government*: Konektivitas antar pemerintah merupakan prasyarat untuk administrasi yang efektif, koordinasi antar pemerintah, dan akuntabilitas publik. Ini berfokus pada komunikasi antar internal pemerintahan, dan berbagai bentuk komunikasi antar lembaga pemerintahan.
2. *Government to Business*: Entitas bisnis sangat diperlukan bagi semua negara dan kontribusi secara signifikan terhadap perkembangannya. Pemerintah juga mengawasi organisasi ini untuk menjalankan kebijakan, standarisasi, dan akuntabilitas. Manajemen tender, administrasi kontrak, pembayaran pajak, dan transaksi *government to business* lainnya harus diotomatisasi.
3. *Government to Citizen*: Tugas utama setiap pemerintah adalah memberikan layanan kepada rakyatnya. Kebutuhan dasar, seperti makanan, tempat tinggal, dan pakaian, serta memberikan hak pendidikan yang sangat baik, perawatan kesehatan, dan layanan sosial lainnya, tidak dapat dicapai tanpa keterlibatan pemerintah. *E-Government* menyediakan kebutuhan warga, yang mungkin dapat dilakukan dengan sistem pemerintahan satu jendela.

4. PENERAPAN FHE PADA DATA E-GOVERNMENT

Keamanan data *E-Government* adalah salah satu masalah yang sangat penting pada skema *cloud computing*. Pengguna didorong untuk mengenkripsi data mereka sebelum menyimpannya di server untuk menjaga kerahasiaan informasi pribadi mereka. Teknik enkripsi homomorfik dapat segera mencari, menghitung, dan mengoperasikan data terenkripsi di cloud sekaligus melindungi kerahasiaan data *ciphertext* yang sedang diproses.

Pertama, pengguna masuk dan menggunakan pembuatan kunci (*key-generation*) yang disediakan server untuk menghasilkan sebuah kunci rahasia (*secret key*), yang hanya dipegang oleh pengguna. Pengguna kemudian mengenkripsi data sebelum mengirimnya ke *cloud*. Keutuhan dan keamanan data dapat dipastikan selama transmisi menggunakan teknologi kriptografi seperti tanda tangan digital. Pengguna dapat mengirimkan permintaan terenkripsi ke server *cloud* ketika ingin server melakukan pengolahan pada data terenkripsi ini (seperti pencarian). Server menjalankan proses yang diperlukan dan mengirimkan hasil terenkripsi kepada pengguna. Pengguna kemudian mendekripsi data dengan kunci rahasianya untuk menerima hasil yang benar [14]. Gambar 1 menggambarkan penerapan FHE pada data *E-Government* berbasis *cloud*.



Gambar 1. Penerapan FHE pada *E-Government*

Ada empat komponen utama untuk penerapan teknologi enkripsi homomorfik dalam komputasi awan [15]:

1. Pengambilan data terenkripsi dalam komputasi awan. Dengan menggunakan enkripsi yang menggunakan metode FHE, keamanan data komputasi awan dapat dipastikan. Ide dasarnya adalah bahwa data dienkripsi menggunakan enkripsi homomorfik dan disimpan di server *cloud*, yang memberikan manfaat yang besar. Kemudian, muncul masalah pada bagaimana memulihkan data terenkripsi. Pengambilan data berdasarkan teknologi enkripsi homomorfik tidak hanya dapat memperoleh data terenkripsi secara langsung, tetapi juga dapat memastikan bahwa data yang dikembalikan tidak dihitung atau diperiksa. Pada tahun 2011, Turaisingham mengakses dan mengambil data di platform *cloud* menggunakan teknologi Hive dan Hadoop [16]. Saat ini, ada beberapa perusahaan tertentu yang sudah menyediakan layanan pengambilan data terenkripsi.
2. Server *cloud* dapat memproses *ciphertext* secara langsung. Dalam komputasi awan, pemrosesan data terenkripsi sebagian besar terdiri dari pengambilan data, pengolahan data, statistik, dan analisis. Server *cloud* bekerja langsung pada data terenkripsi untuk memenuhi kebutuhan pengguna dan mengirimkan data yang telah diubah kembali ke pengguna. Setelah menerima *ciphertext*, pengguna dapat memecahkan kodenya yang mengurangi jumlah data yang telah dikirim. Ristenpart dkk menerbitkan sebuah jurnal pada tahun 2009 di mana mereka merekomendasikan penggunaan enkripsi homomorfik dalam platform layanan komputasi awan pertanian dan mengukur kemampuan beberapa teknik enkripsi [17]. Pada tahun 2013, perkembangan yang sangat pesat dibuat dalam penggunaan FHE dalam proses identifikasi biometrik. Dalam proses membandingkan dua vektor ciri biologis, jarak Hamming sering digunakan sebagai indikasi. Untuk menentukan jarak Hamming, Yasuda et al. menyajikan teknik FHE yang didasarkan pada lattice ideal [18].
3. Tempat penyimpanan data rahasia. Sejak konsep FHE pertama kali muncul, para peneliti mulai menyatakan untuk membuat gudang informasi yang dapat disimpan

FULLY HOMOMORPHIC ENCRYPTION (FHE) PADA PENYIMPANAN DATA E-GOVERNMENT BERBASIS CLOUD

secara rahasia. Data pengguna dapat disimpan di awan menggunakan komputasi awan, dan teknologi FHE dapat menjamin bahwa informasi yang diperlukan tidak diolah dan dianalisis. Teeba menyarankan pada tahun 2015 penggunaan enkripsi homomorfik hibrida untuk menjamin privasi bank data cloud [19].

4. *Multi-party cryptography* yang aman dan terhubung langsung ke enkripsi homomorfik adalah topik yang didiskusikan dalam penelitian Bendlin dkk. Dokumen [20] menunjukkan bagaimana enkripsi homomorfik dapat diterapkan untuk mengatasi tantangan *multi-party cryptography* yang aman dan luas.

5. EKSPERIMEN

Untuk memvalidasi penerapan FHE dalam berbagai jenis data *E-Government*, kami memfokuskan pada tiga jenis operasi data: pencarian data, clustering menggunakan K-Means, dan klasifikasi multikelas menggunakan *Convolutional Neural Network* (CNN). Percobaan ini dilakukan dengan menggunakan HElayers. HElayers adalah wadah Docker berbasis Linux yang dapat diimplementasikan sepenuhnya dalam perangkat lunak dan penyimpanan berbasis cloud. HElayers ditulis dalam C++ dan menyertakan API Python yang dirancang untuk memungkinkan pengembang aplikasi dan ilmuwan data dalam menerapkan teknik privasi tingkat lanjut dalam lingkungan Python terintegrasi dengan baik [21].

Tujuan dari eksperimen yang dilakukan bukanlah untuk mendapatkan metode pencarian data mutakhir, parameter pengelompokan, dan *deep learning-based outcomes* dalam penyelesaian masalahnya. Eksperimen ini dilakukan untuk melihat kelayakan dalam mempertahankan data yang bersifat privasi sekaligus dapat mengakses basis data, pengelompokan, dan model jaringan saraf untuk melakukan perhitungan dengan benar pada data dalam bentuk *ciphertext*. Penjelasan berikut ini akan membahas pengumpulan data dan metode untuk pengambilan sampel yang akan digunakan terhadap masalah yang sudah disebutkan diatas.

A. Privacy-Preserving Search

Privacy-preserving search adalah skenario umum untuk menunjukkan manfaat enkripsi homomorfik. Mampu melakukan pencarian data sambil menjaga privasi dan kerahasiaan dari parameter *query* yang memiliki banyak aplikasi di berbagai bidang industri mulai dari genomik hingga keuangan.

Contoh ini menunjukkan bahwa bagaimana seorang pengguna dapat menggunakan teknik berbasis enkripsi homomorfik untuk menghasilkan sebuah *mask*, dengan tujuan mengambil data dari database pasangan nilai kunci. Sehubungan dengan realisme data, dataset yang digunakan ialah data yang berisi nama negara-negara di dunia beserta ibukota negara tersebut dapat diakses pada laman website [22]. Dalam dunia nyata, contoh dataset ini bisa berupa informasi tentang data pelanggan atau catatan keuangan.

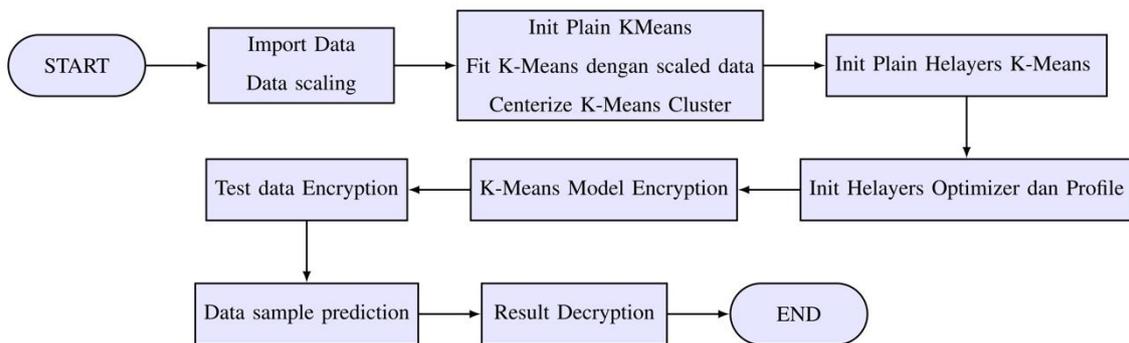
B. K-Means Clustering

K-Means Clustering adalah salah satu teknik pengelompokan data tanpa pengawasan (*unsupervised*) yang paling sederhana dan paling populer. Biasanya, algoritma tanpa pengawasan membuat kesimpulan berdasarkan kumpulan data yang hanya menggunakan vektor sebagai input tanpa mengetahui hasilnya akan seperti apa, atau diberi label. Salah satu kasus penggunaan FHE potensial yang menggunakan K-Means adalah keamanan deteksi anomali dan dapat diterapkan pada kasus penggunaan rantai pasokan di berbagai banyak industri mulai dari otomotif hingga energi serta pertahanan.

Dalam eksperimen ini, kami akan menggunakan masalah segmentasi pelanggan grosir. Data yang digunakan merupakan data dari repository *UCI Machine Learning* [23]. Tujuannya adalah untuk mengelompokkan data klien distributor grosir berdasarkan pengeluaran tahunan mereka dengan mempertimbangan kategori produk yang beragam, seperti susu, bahan makanan, wilayah, dan lain sebagainya [24]. Uraian topologi K-Means yang digunakan dapat dilihat pada Tabel 1. Langkah-langkah dalam proses penerapan FHE pada operasi klustering menggunakan K-Means ditunjukkan pada Gambar 2.

TABEL 1 PARAMETER DARI K-MEANS *CLUSTERING* PADA KASUS DATA GROSIR

Tipe Parameter	Ukuran Parameter
Dimensi	8
<i>Centroid</i>	2
<i>Batch</i>	8192
Test data	8192



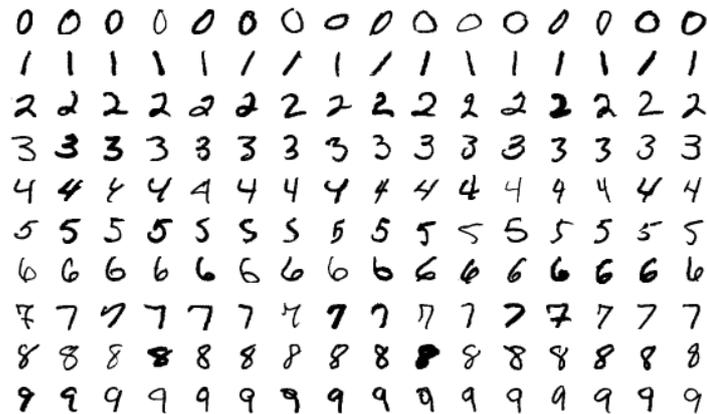
Gambar 2. Flowchart Penerapan FHE pada K-Means *Clustering* untuk Data Grosir

C. Convolutional Neural Network

Klasifikasi adalah tipe subjek yang diuji berkaitan dengan operasi jaringan saraf. Lebih khusus lagi, bidang klasifikasi citra digital berdasarkan informasi yang tergambar dalam citra digital. Sebagai patokan, untuk sistem klasifikasi gambar, basis data *Modified National Institute of Standards and Technology* (MNIST) berisi gambar digit tulisan tangan.

Dataset MNIST memiliki data training sebanyak 60,000 gambar dan data test 10,000 gambar digital dengan skala abu-abu dengan dimensi 28×28 . Setiap gambar digital diberi label dengan nomor yang diwakilinya yaitu 0 hingga 9, seperti yang ditunjukkan pada Gambar 3. Gambar hitam putih (bilevel) asli dari MNIST dinormalisasi ukurannya agar berukuran 20×20 piksel namun tetap mempertahankan rasio aspeknya. Gambar yang dihasilkan mengandung tingkat keabuan sebagai hasil dari teknik *anti-aliasing* yang digunakan oleh algoritma normalisasi. Gambar dipusatkan dalam gambar 28×28 dengan menghitung pusat massa piksel, dan menerjemahkan gambar sehingga meletakkan titiknya di tengah bidang dimensi 28×28 [25].

FULLY HOMOMORPHIC ENCRYPTION (FHE) PADA PENYIMPANAN DATA E-GOVERNMENT BERBASIS CLOUD



Gambar 3. Contoh gambar dari MNIST Dataset

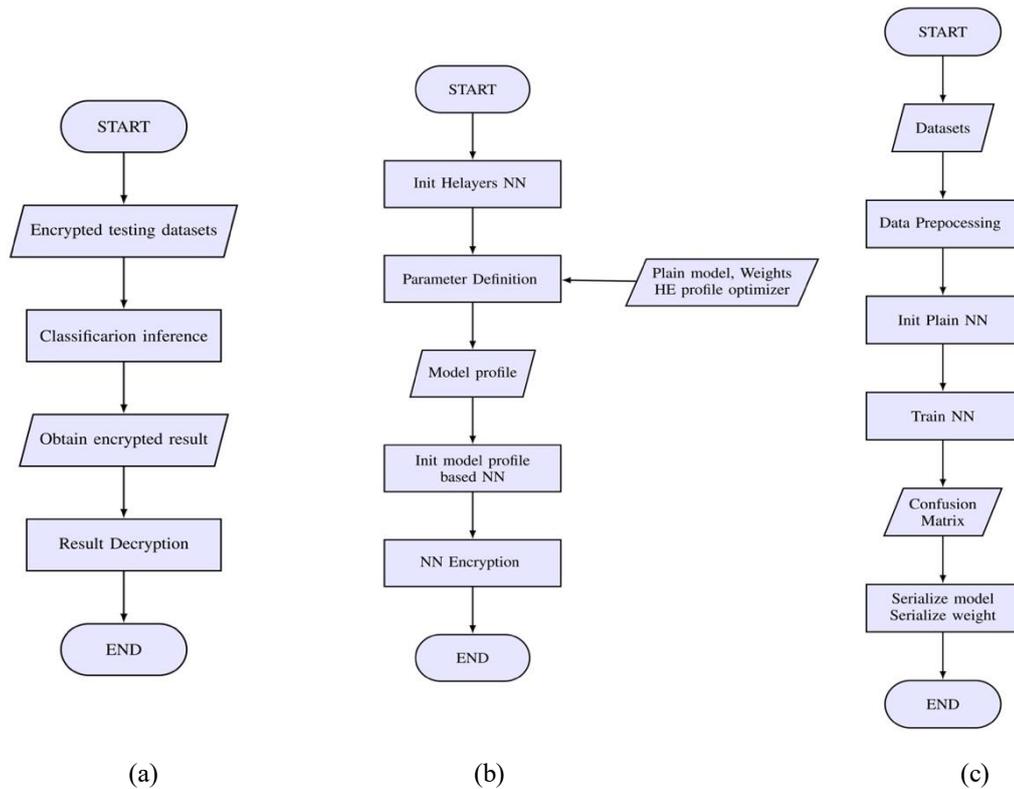
Percobaan ini dilakukan untuk mengidentifikasi satu set sampel dari dataset MNIST secara akurat menggunakan model jaringan saraf yang dikembangkan dan dilatih dengan menggunakan Python dengan *library* Keras [26]. Pada dasarnya, penerapan FHE pada *neural network* memiliki tiga alur aplikasi. Pertama, kita akan membangun model jaringan syaraf tiruan dengan mengimpornya dari *library* Python Keras. Langkah pertama ialah membuat model jaringan syaraf tiruan dalam membentuk *unencrypted (plain)* ditunjukkan pada diagram alur pada Gambar 4(a). Tabel 2 menguraikan topologi model CNN yang digunakan.

TABEL 2 PARAMETER DARI CNN UNTUK PENGENALAN TULISAN TANGAN MNIST

Layer (Tipe)	Output Shape	Jumlah Parameter
Conv2D	(13,13,5)	130
Flatten	845	0
Square Activation	845	0
Dense	100	84600
Square Activation	100	0
Dense	10	1010
Total Params		85740
Trainable Params		85740
Non-trainable Params		0
Batch size		500
Epoch		10

Langkah berikutnya adalah membangun objek jaringan saraf; yaitu versi yang terenkripsi dari jaringan dengan menggunakan perpustakaan Helayers. Seperti yang ditunjukkan pada Gambar 4(b), pengoptimal otomatis akan memeriksa jaringan dan memberikan profil HE, yang terdiri dari berbagai konfigurasi yang akan memfasilitasi eksekusi inferensi yang efisien di bawah enkripsi.

Langkah terakhir yang dilakukan ialah mengenkripsi *training dataset* menggunakan jaringan terenkripsi yang sudah dibuat untuk melakukan inferensi, dan membandingkan hasilnya dengan label yang diharapkan, seperti yang ditunjukkan pada Gambar 4(c).



Gambar 4. Diagram alur penerapan FHE pada klasifikasi digit MNIST berbasis CNN

Hasil klasifikasi dapat dilihat nilai akurasi pada setelah dilakukan klasifikasi pada data uji. Dalam hal memahami peringkat akurasi, ada aturan umum untuk menilai apakah nilai akurasi tersebut dapat dinilai baik atau tidak [27], seperti terlihat sebagai berikut:

TABEL 3 ATURAN UMUM NILAI AKURASI PADA PEMBELAJARAN MESIN

Nilai Akurasi	Kategori
> 90 %	Sangat Baik
70%-90%	Baik
60%-70%	Cukup
< 60%	Kurang

6. HASIL

Untuk menganalisis hasil dari percobaan di atas, yaitu pada bagian pencarian dengan menjaga privasi penggunaanya (*privacy-preserving search*), *unsupervised clustering* menggunakan K-Means dan klasifikasi menggunakan *CNN*, ada dua analisis yang diterapkan. Kedua analisis yang dinilai yaitu konsistensi dan kepraktisan dalam mengolah data *E-Government*. Setiap model akan diuji kinerja basis datanya dan hasil yang diperolehnya dengan menggunakan *privacy-preserving search*, *clustering* dan *deep learning models* terhadap model data *plaintext* dan *cyphertext*. Hasil yang akan didapatkan berasal dari kedua jenis data tersebut. Kami menganalisis dan mengukur kemampuan *privacy-preserving search*, *unsupervised clustering* dan klasifikasi menggunakan *CNN*. Setiap model akan dilihat akurasi dan kepraktisannya.

Untuk mendapatkan hasil yang lebih akurat, penelitian ini juga mengukur waktu (*runtime*) dari *privacy-preserving model*, *unsupervised clustering* menggunakan K-

FULLY HOMOMORPHIC ENCRYPTION (FHE) PADA PENYIMPANAN DATA E-GOVERNMENT BERBASIS CLOUD

Means dan klasifikasi menggunakan *CNN*. *Runtime* merupakan parameter yang penting dalam prosedur pengolahan data. Oleh karena itu, pada penelitian ini, pencarian data, *clustering*, dan klasifikasi baik terenkripsi maupun tidak terenkripsi, akan diuji secara detail.

A. Akurasi

Hasil data dekripsi yang diperoleh adalah sama dilihat dari segi operasi pencarian data parameter yang dipelajari modelnya, bahkan sampai level akurasi mesinnya dengan data telah dicari dan dipelajari oleh model yang tidak terenkripsi. Kinerja keseluruhan model pencarian data, pengelompokan, dan *deep-learning* pada sampel pengujian adalah identik, baik model tersebut dilatih dengan enkripsi atau tanpa enkripsi.

- 1) Pencarian data nama negara. Proses ini merupakan pencarian nama ibukota negara dari nama negara yang dimaksud. Proses ini diuji dengan melihat hasil pencarian *database* yang terenkripsi menggunakan *query* terenkripsi. Hasil pencarian data terenkripsi memiliki nilai akurasi 100% dibandingkan dengan pencarian pada *database* tidak terenkripsi.
- 2) *Clustering* data konsumen grosir. Skema enkripsi FHE mendapatkan nilai akurasi 100% dengan membandingkan *cluster* yang kami dapatkan menggunakan algoritma K-Means pada data terenkripsi (*ciphertext*) dengan *cluster* yang didapat dari data data tidak terekripsi (*plaintext*).
- 3) Pengenalan digit pada MNIST. Pengklasifikasi (*classifiers*) pada dataset MNIST dinilai berdasarkan akurasinya, dimana pengelompokan tulisan angka yang benar akan ditunjukkan dengan sebuah presentase kebenaran dari gambar angka yang diklasifikasi. Pada penelitian ini, jaringan saraf terenkripsi secara FHE memperoleh akurasi klasifikasi 98,2% pada dataset yang diuji proses dilakukan di jaringan yang terenkripsi. Pada data MNIST, distribusi kelas pada set data pengujian, 10 label digit angka dilakukan pada jumlah gambar yang seimbang. Oleh karena itu, akurasi merupakan indikator kinerja klasifikasi yang dapat dipercaya. Meskipun model pada skema FHE tidak mencapai akurasi yang diklaim 99,21% seperti yang disampaikan pada [28], pengujian pada klasifikasi dan pengenalan digit MNIST terenkripsi dengan akurasi 98,2% dianggap dapat diterima, mengingat pengenalan digit pada CNN terekripsi memiliki keamanan data pada penyimpanan *cloud* yang lebih baik.

B. Runtime

Semua *runtime* yang dilaporkan telah diukur dengan menggunakan HELayers yang berjalan di Docker Engine: 20.10.12 dengan sumber daya 8 core CPU, 8 GB memori, dan 3 GB memori *swap*. Pada percobaan ini, *runtime* diuji menggunakan sistem operasi MacOS Monterey pada Apple M1 MacBook Air dengan RAM 8 GB, CPU 8 core dan GPU 8 core.

Tabel 4 hingga tabel 6 berikut memberikan analisis yang komprehensif membandingkan lama waktu yang dibutuhkan untuk menyelesaikan setiap operasi data. Setiap proses dilakukan sebanyak 5 kali percobaan dan diambil nilai rata-rata dan dihitung standard deviasinya.

TABEL 4
 RUNTIME (RATA-RATA DAN STANDARD DEVIASI) PADA CIPHERTEXT DAN PLAINTEXT
 UNTUK PRIVACY-PRESERVING PENCARIAN DATA NAMA IBUKOTA NEGARA.

Proses	Runtime (detik) pada Ciphertext	Runtime (detik) pada Plaintext
Enkripsi database	-	0.780 ± 0.012
Enkripsi query	-	0.004 ± 0.2
Pencarian pada database terenkripsi	42.490 ± 0.01	-
Dekripsi hasil pencarian	0.008 ± 0.03	-

Tabel 4 menunjukkan perhitungan *runtime* pada pencarian nama ibukota negara. Tahap pertama yang dilakukan ialah melakukan enkripsi pada basis data yang berisi data nama negara dan nama ibukota negara tersebut. Enkripsi pada database rata-rata dapat dilakukan dalam 0.78 detik dengan standard deviasi yang cukup kecil yaitu 0.012. Hal ini menunjukkan bahwa tahap enkripsi dilakukan dengan cukup singkat, dan tidak berbeda jauh dalam setiap percobaannya. Tahap selanjutnya ialah melakukan enkripsi terhadap baris perintah (*query*) yang diperlukan untuk melakukan pencarian. Proses enkripsi baris perintah hanya memerlukan waktu sekitar 0.004 detik. Selanjutnya dilakukan pencarian data pada database dan query yang telah dienkripsi (*ciphertext*). Proses pencarian memerlukan waktu sekitar 42 detik. Hasil pencarian masih dalam bentuk terenkripsi (*ciphertext*), lalu untuk dapat mengetahui hasil pencarian tersebut, dilakukan proses dekripsi yang memerlukan waktu sekitar 0.008 detik.

TABEL 5
 RUNTIME (RATA-RATA DAN DAN STANDARD DEVIASI PADA CIPHERTEXT DAN PLAINTEXT
 K-MEANS PADA DATA KOSTUMER GROSIR.

Proses	Runtime (detik) pada Ciphertext	Runtime (detik) pada Plaintext
Bangun dan load model K-Means	-	0.670 ± 0.02
Enkripsi model	-	4.723 ± 0.03
Enkripsi data uji	-	0.356 ± 0.01
Run clustering	0.212 ± 0.01	-
Dekripsi hasil	0.071 ± 0.03	-

Tabel 5 menunjukkan perhitungan *runtime* pada proses clustering data kostumer grosir. Tahap pertama yang dilakukan ialah membuat dan membangun model K-Means. Tahap ini memerlukan waktu sekitar 0.67 detik. Selanjutnya dilakukan enkripsi terhadap model yang telah dibangun dimana membutuhkan waktu sekitar 4.723 detik. Enkripsi juga dilakukan terhadap data uji, yaitu sekitar 0.356 detik. Proses *clustering* dilakukan pada model dan data uji yang telah terenkripsi (*ciphertext*) dalam waktu 0.212 detik. Hasil dari *clustering* perlu didekrip agar dapat dibaca oleh pengguna, proses ini dilakukan dalam waktu 0.071 detik.

Tabel 6 menunjukkan proses klasifikasi gambar angka dari dataset MNIST. Tahap awal yang dilakukan adalah mendefinisikan *Neural Network* (NN) dalam waktu sekitar 1420

FULLY HOMOMORPHIC ENCRYPTION (FHE) PADA PENYIMPANAN DATA E-GOVERNMENT BERBASIS CLOUD

detik. Waktu ini merupakan training dari 10 epoch yang dilakukan, dimana waktu per epoch ialah rata-rata 141 detik. Selanjutnya NN dan data uji dienkrrip masing-masing dalam waktu 19.05 dan 2.12 detik. Proses *inference*, merupakan proses klasifikasi data dilakukan dalam 7.82 detik pada data yang telah dienkrripsi (*ciphertext*). Hasil klasifikasi kemudian didekrip dalam waktu sekitar 0.056 detik, dimana rata-rata waktu dekripsi per sampel yaitu 0.004 detik.

TABEL 6
RUNTIME (RATA-RATA DAN STANDARD DEVIASI) PADA *CIPHERTEXT* DAN *PLAINTEXT*
UNTUK KLASIFIKASI DIGIT MNIST MENGGUNAKAN CNN.

Proses	<i>Runtime</i> (detik) pada <i>Ciphertext</i>	<i>Runtime</i> (detik) pada <i>Plaintext</i>
Bangun Plain NN	-	1420 ± 0.01
Training (1 epoch)	-	141 ± 0.05
Enkripsi Plain NN	-	19.05 ± 0.03
Enkripsi Test Data	-	2.12 ± 0.01
<i>Inference</i>	7.82 ± 0.32	-
Klasifikasi per sample	0.478 ± 0.002	-
Dekripsi hasil	0.056 ± 0.009	-
Dekripsi per sample	0.004 ± 0.023	-

7. KESIMPULAN

Teknik enkripsi homomorfik memberikan pendekatan inovatif untuk masalah dalam menjaga data yang disimpan pada penyimpanan berbasis *cloud*. Ini memungkinkan penyedia layanan *cloud* untuk menawarkan kepada pelanggan dengan cara yang lebih efisien dengan menjaga integritas dan kerahasiaan data terutama pada penerapan aplikasi *e-government*. Sudah banyak penelitian tentang Enkripsi Homomorfik, karena tingkat *noise* dan kompleksitas yang terus meningkat, namun teknik FHE tidak sepenuhnya diadopsi. Dalam penelitian ini, kami telah memberikan pandangan bahwa salah satu skema homomorfik terbaru dan menyajikan studi komparatif dari skema FHE tersebut. Hasil menunjukkan bahwa skema FHE yang digunakan memberikan hasil yang sangat baik dengan *runtime* yang wajar.

REFERENSI

- [1] F. Thabit, O. Can, S. Alhomdy, G. H. Al-Gaphari, and S. Jagtap, "A Novel Effective Lightweight Homomorphic Cryptographic Algorithm for data security in cloud computing," *International Journal of Intelligent Networks*, vol. 3, pp. 16–30, 2022.
- [2] A. Jawed Khan and S. Mehfuz, "Big-data driven approaches in materials science on cloud storage API through optimal homomorphic encryption," *Mater Today Proc*, May 2021.
- [3] W. Ren *et al.*, "Privacy-preserving using homomorphic encryption in Mobile IoT systems," *Comput Commun*, vol. 165, pp. 105–111, Jan. 2021.
- [4] C. Lefebvre, "On data," *Journal of Pidgin and Creole Languages*, vol. 15, no. 2, pp. 313–337, 2000.
- [5] C. S. Gu, "Fully homomorphic encryption from approximate ideal lattices," *Ruan Jian Xue Bao/Journal of Software*, vol. 26, no. 10, pp. 2696–2719, 2015.

- [6] J. H. Cheon *et al.*, “Batch fully homomorphic encryption over the integers,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7881 LNCS, pp. 315–335, 2013.
- [7] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, “(Leveled) fully homomorphic encryption without bootstrapping,” *ACM Transactions on Computation Theory*, vol. 6, no. 3, 2014.
- [8] S. S. Sathya, P. Vepakomma, R. Raskar, R. Ramachandra, and S. Bhattacharya, “A Review of Homomorphic Encryption Libraries for Secure Computation,” pp. 1–12, 2018.
- [9] S. Halevi and V. Shoup, “Algorithms in HElib,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8616 LNCS, no. PART 1, pp. 554–571, 2014.
- [10] Microsoft Corp., “microsoft/SEAL: Microsoft SEAL is an easy-to-use and powerful homomorphic encryption library.,” 2021. <https://github.com/microsoft/SEAL>
- [11] J. Fan and F. Vercauteren, “Somewhat Practical Fully Homomorphic Encryption,” *Proceedings of the 15th international conference on Practice and Theory in Public Key Cryptography*, pp. 1–16, 2012.
- [12] J. H. Cheon, K. Han, A. Kim, M. Kim, and Y. Song, “Bootstrapping for approximate homomorphic encryption,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10820 LNCS, pp. 360–384, 2018, doi: 10.1007/978-3-319-78381-9_14.
- [13] M. Kumar, M. Shukla, and S. Agarwal, “An E Governance model using cloud computing technology for Developing Countries,” no. January, 2013.
- [14] S. Hemalatha and R. Manickachezian, “Performance of Ring Based Fully Homomorphic Encryption for securing data in Cloud Computing,” vol. 3, no. 11, pp. 8496–8500, 2014.
- [15] M. Z. E, Y. Geng, and Y. Geng, “ScienceDirect ScienceDirect ScienceDirect Homomorphic Encryption Technology for Cloud Computing Homomorphic Encryption Technology for Intangible Cloud Computing Research on the Innovation of Protecting Cultural Heritage the Plus " Era Min in E * " Internet,” *Procedia Comput Sci*, vol. 154, pp. 73–83, 2019.
- [16] B. Thuraisingham, V. Khadilkar, A. Gupta, M. Kantarcioglu, and L. Khan, “Secure Data Storage and Retrieval in the Cloud”.
- [17] S. Savage and T. C. Clouds, “Hey , you , get off of my cloud Hey , You , Get Off of My Cloud ”.
- [18] M. Yasuda, T. Shimoyama, and J. Kogure, “Packed Homomorphic Encryption Based on Ideal Lattices and Its Application to Biometrics,” pp. 55–74.
- [19] M. Tebaa, K. Zkik, and S. El Hajji, “Hybrid homomorphic encryption method for protecting the privacy of banking data in the cloud,” *International Journal of Security and its Applications*, vol. 9, no. 6, pp. 61–70, 2015.
- [20] R. Bendlin, I. Damgård, C. Orlandi, and S. Zakarias, “Semi-homomorphic encryption and multiparty computation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6632 LNCS, pp. 169–188, 2011.
- [21] “ibmcom/helayers-pylab - Docker Image | Docker Hub.”

**FULLY HOMOMORPHIC ENCRYPTION (FHE) PADA PENYIMPANAN DATA E-
GOVERNMENT BERBASIS CLOUD**

- <https://hub.docker.com/r/ibmcom/helayers-pylab>
- [22] F. Sabir ,“Countries With Capitals _ Kaggle.” [Online].Available: <https://www.kaggle.com/datasets/faransabir/countries-with-capitals..>
- [23] D. Dua and C. Graff, “UCI Machine Learning Repository: Data Sets,” *Irvine, CA: University of California, School of Information and Computer Science*. 2019.
- [24] P. Sharma, “K Means Clustering | K Means Clustering Algorithm in Python,” *Analytics Vidhya*. 2019.
- [25] Y. LeCun, C. Cortes, and C. J. C. Burges, “MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges.” 1998.
- [26] N. Dowlin *et al.*, “CryptoNets: Applying neural networks to Encrypted data with high throughput and accuracy - Microsoft research,” *Microsoft Research TechReport*, vol. 48, pp. 1–12, 2016.
- [27] Stephen Allwright, “What is a good accuracy score in machine learning?,” May 14, 2022.
- [28] F. Siddique, S. Sakib, and M. A. B. Siddique, “Recognition of handwritten digit using convolutional neural network in python with tensorflow and comparison of performance for various hidden layers,” *2019 5th International Conference on Advances in Electrical Engineering, ICAEE 2019*, pp. 541–546, 2019.

PENERAPAN METODE CERTAINTY FACTOR PADA SISTEM PAKAR DIAGNOSA GANGGUAN MENTAL PADA ANAK BERBASIS WEB

Kikye Martiwi Sukiakhy¹, Zulfan², dan Odli Aulia³

¹² Jurusan Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Syiah Kuala, Darussalam, Banda Aceh, 23111, Indonesia
E-mail: kikye.martiwi.sukiakhy@unsyiah.ac.id, zulfan.abdullah@unsyiah.ac.id

ABSTRACT

CCM (Care for Children's Mental) is a diagnostic expert system with a certainty factor method that can be done in various places and chooses some of the symptoms that occur in children so that the system can diagnose the child's mental disorders that are experienced. It is hoped that this expert system can help parents in early diagnosis of mental disorders that their children experience independently with the help of an expert based on the symptoms felt so that parents are sensitive to the child and do not take it for granted. Another benefit of this expert system is that it makes it easy to access anywhere, and it makes it easier for parents to diagnose children's disorders easily. Certainty Factor is a method of defining a measure of expert facts or rules to describe a person's beliefs about a problem. To help and make it easier for parents in the media to diagnose mental disorders in children, the certainty factor method can be used as a solution. The application of the Certainty Factor method to the expert system for diagnosing mental disorders in children with the web-based certainty factor method is built using the PHP programming language and MySQL databases and uses the CodeIgniter (CI) framework.

Keywords: *certainty factor, expert system, child disease, mental disorders, diagnosis*

ABSTRAK

CCM (*Care for Children's Mental*) adalah sistem pakar diagnosa dengan metode *certainty factor* yang dapat dilakukan di berbagai tempat dan memilih beberapa gejala yang terjadi pada anak agar sistem dapat mendiagnosa gangguan mental anak yang di alami. Diharapkan dengan adanya sistem pakar ini bisa membantu para orang tua dalam mendiagnosa secara dini gangguan mental yang anak mereka alami secara mandiri dengan bantuan seorang ahli dengan berdasarkan gejala-gejala yang dirasakan agar para orang tua peka terhadap anak dan tidak menganggap remeh hal tersebut. Manfaat lain dari sistem pakar ini adalah mempermudah dalam akses dimana saja, dan memudahkan orang tua untuk mendiagnosa gangguan anak dengan mudah. *Certainty Factor* merupakan suatu metode yang dapat mendefinisikan ukuran terhadap aturan atau fakta pakar untuk menggambarkan keyakinan seorang terhadap suatu masalah yang dihadapi. Guna mempermudah dan membantu para orang tua dalam mendiagnosa gangguan mental pada anak, maka metode *certainty factor* mungkin dapat dijadikan salah satu solusinya. Penerapan metode *Certainty*

PENERAPAN METODE CERTAINTY FACTOR PADA SISTEM PAKAR DIAGNOSA GANGGUAN MENTAL PADA ANAK BERBASIS WEB

Factor pada Sistem pakar diagnosa gangguan mental pada anak dengan metode *certainty factor* berbasis web ini dibangun dengan menggunakan database MySQL dan framework CodeIgniter (CI) serta bahasa pemrograman PHP.

Kata kunci: Certainty faktor, sistem pakar, penyakit anak, gangguan mental, diagnosa

1. Pendahuluan

Kesehatan mental adalah sebuah kondisi yang mana seseorang memiliki kesejahteraan yang dapat terlihat dari dirinya dalam kemampuannya menyadari potensi yang ada pada dirinya, mempunyai kemampuan dalam mengatasi dan menghadapi tekanan kehidupan normal pada berbagai situasi yang ada, dapat bekerja dengan produktif dan menghasilkan dengan menggunakan semua potensi yang ada secara maksimal, mampu berfikir secara jernih dan jauh lebih positif ketika harus menghadapi berbagai masalah dan kesulitan dalam hidup, serta dapat memberikan kontribusi yang nyata kepada komunitasnya dan lingkungan dimana seseorang itu berada. Kesehatan mental yang baik adalah dimana seseorang berada pada keadaan yang tenang serta tenang sehingga dapat menghargai orang lain dan lingkungan sekitar dan dapat menikmati kehidupan sehari-harinya. Kesehatan mental ini sering sekali diabaikan karena kebanyakan orang-orang hanya menganggap penting kesehatan fisik saja.

Kesehatan mental saat ini tidak hanya penting untuk diperhatikan bagi orang-orang dewasa saja, akan tetapi juga sangat penting untuk anak-anak dan remaja. Dalam memperhatikan kesehatan mental yang ada pada anak-anak dan juga remaja artinya perlu juga dipahami bahwa kesehatan mental pada anak dan remaja memiliki faktor-faktor yang dapat melindunginya (*protective factor*) serta faktor-faktor yang dapat membahayakan kesehatan mental tersebut (*risk factor*) [1].

Dalam bentuk apapun, kesehatan mental anak mungkin terjadi akibat dari *well-being* yang buruk, begitu pula sebaliknya gangguan apapun yang terjadi pada kesehatan mental anak mampu memberikan kontribusi serta pengaruh yang cukup besar pada keseluruhan *well-being* anak. Artinya, ketidakmampuan dan kesulitan dalam melakukan penyesuaian diri terhadap lingkungan dapat disebabkan oleh adanya faktor atau kondisi-kondisi yang mana membuat anak tertekan, contohnya anak yang tumbuh dan berkembang di lingkungan yang seringkali terjadi kekerasan, tumbuh di lingkungan keluarga yang mempunyai status dan tingkat ekonomi atau kesejahteraan yang rendah, serta memiliki pengalaman yang traumatis. Tanda awal kejahatan yang terjadi di usia remaja kemungkinan merupakan akibat dari terganggunya perkembangan kesehatan mental pada anak-anak [2].

Sistem pakar atau yang biasa juga dikenal dengan *Knowledge Based System* merupakan suatu aplikasi berbasis komputer yang mempunyai tujuan dalam membantu dalam pemecahan berbagai persoalan yang ada pada bidang yang spesifik serta membantu dalam pengambilan keputusan. Untuk dapat mendiagnosa secara dini gejala awal penyakit gangguan mental pada anak maka diperlukan adanya sebuah sistem yang dapat digunakan oleh para orang tua sehingga mereka dapat mengenali jenis gangguan perkembangan yang terjadi pada anak.

Pada penelitian ini sistem pakar digunakan guna membantu dalam diagnosa gejala awal penyakit gangguan mental pada anak-anak yaitu dengan menggunakan metode *Certainty Factor*. Metode *Certainty factor* merupakan sebuah metode yang mampu mendefinisikan tingkat kepastian suatu aturan atau fakta dalam menggambarkan

keyakinan seorang pakar terhadap masalah yang sedang dihadapi. Diharapkan dengan adanya sistem pakar ini mampu membantu mengetahui serta mendiagnosa secara lebih dini gejala awal penyakit gangguan mental yang terjadi pada anak tanpa harus datang langsung ke dokter spesialis dan hanya perlu mengaksesnya melalui internet.

CCM merupakan sebuah sistem pakar diagnosa yang menggunakan metode *certainty factor* yang bisa dilakukan di berbagai tempat dan memilih beberapa gejala yang terjadi pada anak agar sistem dapat mendiagnosa gangguan mental anak yang di alami. Harapan dari adanya sistem pakar ini adalah bisa membantu para orang tua dalam meningkatkan pengetahuan dan mendiagnosa secara lebih dini gangguan mental yang anak mereka alami secara mandiri dengan bantuan seorang ahli dengan berdasarkan gejala-gejala yang dirasakan agar para orang tua peka terhadap anak dan tidak menganggap remeh hal tersebut. Metode *certainty factor* adalah sebuah metode yang digunakan dalam pembuktian apakah suatu fakta itu tidak pasti ataupun pasti.

2. Landasan Teori

A. Sistem Pakar

Sistem pakar merupakan suatu penggabungan dari pengguna (*user*) dan komputer yang bekerja secara bersama-sama dalam pelaksanaan kegiatan analisis, manajemen, operasi, serta pengambilan keputusan terhadap penyelesaian suatu masalah yang terjadi pada sebuah organisasi guna mencapai suatu tujuan tertentu. Beberapa komponen utama yang dimiliki oleh sistem pakar antara lain yaitu antarmuka pengguna (*user interface*) yang berarti perangkat lunak komputer yang merupakan sebuah jembatan komunikasi yang digunakan antara sistem dengan pengguna. Selain itu juga terdapat basis data (*expert system database*) yang berisi berbagai informasi dan pengetahuan yang diperlukan dalam merumuskan, mengetahui dan menyelesaikan berbagai masalah berdasarkan aturan dan fakta yang didapat melalui majalah, jurnal, pakar, majalah, serta berbagai sumber pengetahuan lainnya. Lalu ada fasilitasi akuisisi pengetahuan (*knowledge acquisition facility*) atau fasilitas yang disediakan untuk memudahkan dialog antara pakar dan sistem. Dan yang terakhir adalah mekanisme inferensi (*inference mechanism*) yang merupakan sebuah perangkat lunak yang mempunyai tugas sebagai penalar untuk menghasilkan suatu kesimpulan dengan bergantung pada pengetahuan yang ada. Sistem pakar adalah sebuah sistem informasi yang berbasis pengetahuan dan metode pengambilan keputusan yang dipergunakan oleh satu atau beberapa ahli pada dibidang keahlian tertentu. Sistem pakar bekerja seperti layaknya seorang pakar pada bidang keahliannya yang berisi fakta-fakta serta aturan-aturan dalam proses pemecahan suatu masalah tertentu. Sistem pakar berkerja berdasarkan pada sistem pengetahuan, sehingga suatu sistem komputer dapat melakukan pengambilan keputusan secara tepat [3].

B. Metode Certainty Factor

Sistem pakar harus dapat bekerja dalam keadaan yang tidak pasti. Dalam menyelesaikan ketidakpastian, sejumlah teori telah ditemukan, antara lain: Probabilitas Bayes (*Bayesian probability*), Teori Hartley berdasarkan himpunan klasik (*Hartley theory based on classical sets*), Teori Dempster-Shafer (*DempsterShafer theory*), Probabilitas klasik (*classical probability*), Teori Shannon berdasarkan pada probabilitas (*Shanon theory based on probability*), Faktor kepastian (*certainty factor*) serta Teori fuzzy Zadeh (*Zadeh's fuzzy theory*) [4].

Metode *Certainty Factor* (CF) merupakan sebuah metode yang digunakan dalam pengelolaan ketidakpastian pada sistem yang berbasiskan aturan. Metode CF telah

PENERAPAN METODE CERTAINTY FACTOR PADA SISTEM PAKAR DIAGNOSA GANGGUAN MENTAL PADA ANAK BERBASIS WEB

dikembangkan oleh Shortliffe dan Buchanan (1975) pada pertengahan tahun 1970-an untuk MYCIN, yaitu sebuah sistem pakar yang digunakan dalam mengobati dan mendiagnosa penyakit infeksi darah dan meningitis. Sejak itulah pada sebuah sistem yang berbasis aturan, metode CF telah menjadi salah satu pendekatan standar dalam manajemen ketidakpastian. Berikut ini merupakan rumus umum dalam penentuan *Certainty Factor*:

$$CF[H,E] = MB[H,E] - MD[H,E] \quad (1)$$

Dimana :

CF[h,e] = faktor kepastian

MB[h,e] = ukuran kepercayaan terhadap hipotesis h, jika diberikan evidence e (antara 0 dan 1)

MD[h,e] = ukuran ketidakpercayaan terhadap evidence h, jika diberikan evidence e (antara 0 dan 1)

Pada sebuah sistem pakar cukup sering mempunyai lebih dari satu kaidah dan terdiri dari beberapa gejala atau premis yang dihubungkan dengan OR atau AND. Pengetahuan mengenai premis atau gejala juga tidak dapat dipastikan, ini karena besarnya suatu nilai (*value*) CF yang diberikan oleh pasien pada saat pasien tersebut menjawab pertanyaan yang ada pada sistem atas premis (gejala) yang dialami oleh pasien atau didapat juga dari nilai CF hipotesa. Secara umum, aturan pada Metode Certainty Factor dapat direpresentasikan ke dalam bentuk persamaan 2 berikut ini:

$$\text{IF } S1 \text{ AND } S2 \text{ AND } \dots \text{ En THEN } F \text{ (CF=CFx)} \quad (2)$$

Dimana:

S1 adalah fakta yang ada

F adalah konklusi yang dihasilkan

CF adalah tingkat keyakinan.

apabila terdapat kombinasi 2 buah aturan dengan evidence yang berbeda tetapi hipotesa maka akan dilakukan persamaan 3 berikut ini :

$$CF1 + CF2(1 - CF1) \text{ jika } CF1 > 0 \text{ dan } CF2 > 0 \quad (3) [5]$$

C. Pemrograman Web

Pemrograman diartikan sebagai suatu cara, proses, perbuatan suatu program. Sedangkan web dapat didefinisikan sebagai jaringan komputer yang terdiri dari beberapa kumpulan situs-situs internet yang menawarkan suara, grafik, teks, sumber daya animasi yang melalui HTTP (*hypertext transfer protocol*). Halaman web adalah file teks murni yang dapat dibuka, diterjemahkan dan dilihat dengan Internet Browser yang berisi sintak-sintak. Saat ini web identik dengan internet, dikarenakan web sangat dikenal merupakan standar *interface* yang ada pada berbagai layanan-layanan yang tersedia di internet, yang pada awalnya mempunyai fungsi sebagai penyedia informasi, tapi kini dipergunakan sebagai media komunikasi dari chatting hingga email, bahkan sekarang web juga dipergunakan sebagai transaksi bisnis [6].

Program komputer merupakan kumpulan dari perintah-perintah atau instruksi-instruksi yang bertugas memberitahukan komputer bagaimana mengerjakan tugas-

tugasnya. Programmer komputer atau terkadang yang dikenal juga dengan pengembang yang mempunyai tugas memodifikasi komputer dan menulis instruksi-instruksi tersebut. Dalam menciptakan atau membuat sebuah program komputer, programer biasanya membuat atau menulis kode-kode perintah dalam program dengan menggunakan bahasa pemrograman [7].

Pengertian dari bahasa pemrograman yaitu sejumlah symbol, dan kata-kata kode yang membuat seorang programmer mampu menyampaikan perintah atau instruksi kepada komputer. Dalam menciptakan sebuah program, seorang programmer menggunakan sebuah perangkat pemrograman. Dalam pembuatan sebuah aplikasi web yang dapat dikembangkan lebih lanjut, maka dibutuhkan suatu bahasa markup, bahasa pemrograman, bahasa script yang fungsinya agar web terlihat lebih dinamis dan juga terdapat bahasa stylesheet. Contohnya antara lain: HTML (*Hypertext Markup Language*), PHP *Hypertext Preprocessor* (PHP), CSS (*Cascading Style Sheet*), Javascript, dll.

D. Jaringan Komputer

Perkembangan teknologi saat ini semakin pesat serta didukung oleh kebutuhan informasi yang efisien dan efektif menyebabkan internet menjadi sangat populer di dunia. Sebagian besar masyarakat telah menggunakan aplikasi berbasis internet, seperti email, ecommerce, dan lain-lain. Internet merupakan kumpulan jaringan-jaringan komputer yang secara fisik saling berhubungan antara satu jaringan dengan jaringan yang lain dan juga memiliki kemampuan untuk menguraikan dan membaca berbagai protokol komunikasi tertentu yang sering dikenal dengan istilah Protocol (TCP) serta Internet Protocol (IP). *Transmission Control Protokol* merupakan sebuah spesifikasi yang sederhana mengenai bagaimana dua atau lebih komputer dapat saling bertukar informasi.

Internet merupakan singkatan atau kependekan dari *international network*, ini dapat diartikan sebagai suatu jaringan komputer yang sangat besar dan terdiri dari kumpulan jaringan-jaringan kecil yang saling berhubungan antara satu dengan yang lainnya. Jaringan komputer yang sangat besar ini dapat mencakup jaringan seluruh dunia. Internet saat ini adalah salah satu media pencarian informasi terpopuler melalui komputer yang dapat menjangkau hampir seluruh pelosok dunia [8].

Agar jaringan komputer dapat berjalan sesuai dengan yang diharapkan maka tiap-tiap jaringan komputer dapat memberikan layanan serta meminta dan menerima layanan. Secara umum internet dapat diartikan sebagai jaringan komputer yang tidak mempunyai batasan serta merupakan penghubung antara satu pengguna komputer dengan pengguna komputer lainnya di berbagai wilayah di seluruh dunia.

E. Database

Basis data atau *database* adalah panggalan data terintegrasi yang dikontrol serta di atur secara terpusat. Ribuan class biasanya dapat tersimpan dalam sebuah basis data. Informasi yang disimpan dalam basis data termasuk relasi antar class dan class attribute. Informasi yang bersifat deksriptif juga tersimpan dalam basis data, informasi tersebut antara lain seperti kontrol akses untuk data-data yang sensitif, nama atribut dan pemberian batasan suatu nilai. Basis data merupakan pangkalan data atau tempat penyimpanan data-data yang dibutuhkan. Kumpulan data-data tersebut tersimpan pada sistem komputer secara sistematis. [9]

Basis data mempunyai tujuan utama dalam pemeliharaan data yang telah diolah atau sebagai media penyimpanan data dan informasi dapat di akses secara cepat dan mudah. Basis data juga merupakan salah satu bagian dari rekayasa perangkat lunak yang

PENERAPAN METODE CERTAINTY FACTOR PADA SISTEM PAKAR DIAGNOSA GANGGUAN MENTAL PADA ANAK BERBASIS WEB

terkomputerisasi. Basis data adalah merupakan kumpulan data-data yang mempunyai relasi atau saling berhubungan. Maka dapat disimpulkan bahwa basis data merupakan salah satu bagian dari rekayasa perangkat lunak yang terkomputerisasi yang salah satu fungsinya adalah sebagai media penyimpanan berbagai informasi yang mempunyai relasi atau mempunyai hubungan satu dengan yang lain supaya informasi-informasi tersebut dapat di akses secara cepat dan mudah.

F. Software

Software atau yang dikenal juga dengan perangkat lunak kumpulan instruksi-instruksi atau data elektronik yang oleh sistem komputer disimpan sedemikian rupa dan data tersebut dieksekusi secara efektif dan efisien oleh mesin komputer dalam melakukan perkerjanya. Perangkat lunak komputer merupakan penghubung atau perantara antara sistem komputer dengan pengguna. Perangkat lunak komputer juga bisa diartikan sebagai penerjemah instruksi yang diberikan oleh pengguna sebagai pemberi instruksi kepada penerima instruksi yaitu sistem komputer [10]. Software yang digunakan dalam penelitian ini adalah:

1. XAMPP

XAMPP merupakan perangkat lunak bebas yang mendukung banyak sistem operasi, merupakan kompilasi dari beberapa program. XAMPP sendiri adalah singkatan dari A (*Apache*), X (*cross platform*), M (*MySQL/MariaDB*), P (*PHP*), dan P (*Perl*) dimana ini merupakan berbagai program yang tersedia di software ini.

2. Visual Studio Code

Visual Studio Code merupakan aplikasi editor yang digunakan untuk kode dan teks sumber yang kuat tetapi ringant yang berjalan pada desktop serta tersedia untuk berbagai sistem operasi, antara lain Windows, macOS, dan Linux. Muncul dengan dukungan bawaan untuk, Node.js, JavaScript dan TypeScript mempunyai ekosistem ekstensi yang kaya untuk bahasa lain (seperti C#, C++, Python, Java, Go, PHP) dan runtime (seperti Unity dan .NET). Visual Studio Code menyediakan beberapa fitur seperti penyelesaian kode, penyorotan sintaksis, merefaktor kode, kutipan kode, Git dan pengawakutuan.

3. CodeIgniter

CodeIgniter merupakan Kerangka Pengembangan Aplikasi atau toolkit yang ditujukan untuk orang yang membangun situs web menggunakan PHP. Hal ini bertujuan untuk memungkinkan dilakukannya pengembangan proyek secara lebih efektif dan efisien apabila dibandingkan jika menulis semua kode dari awal, dengan menyediakan berbagai kumpulan pustaka yang diperlukan untuk melakukan tugas-tugas secara umum, serta struktur logis dalam pengaksesan pustaka ini dan antarmuka yang cukup sederhana sederhana. CodeIgniter membuat jumlah kode yang diperlukan untuk suatu tugas tertentu dapat diminimalisir dengan secara kreatif fokus pada proyek. CodeIgniter sangat fleksibel sehingga memungkinkan untuk dapat bekerja dengan cara yang diinginkan dan tidak ada paksaan untuk bekerja dengan suatu cara tertentu.

3. Metodologi Penelitian

Metode-metode yang digunakan dalam penelitian ini:

- a. Mengumpulkan berbagai data serta jurnal-jurnal yang berkaitan dengan penyakit yang akan dimasukkan ke dalam sistem
- b. Wawancara secara langsung dengan dosen psikolog anak yang bertujuan guna mendapatkan data dan informasi mengenai penyakit anak serta apa saja gejala-gejala

yang menyertainya. Dan juga untuk memperoleh teknik inferensi guna mendapatkan suatu kesimpulan

- c. Metode yang digunakan dalam pengembangan sistem dalam penelitian ini terdiri dari beberapa tahapan, yaitu:
- Tahapan Akuisisi Pengetahuan. Dalam tahapan ini dilakukan proses perubahan data-data yang kemudian menjadi rumusan yang merupakan pengetahuan dan informasi awal tentang penyakit anak. Pada tahapan akuisisi pengetahuan ini melibatkan seorang dosen psikologi spesialis anak sebagai narasumber sekaligus pakar pada sistem.
 - Tahapan Perancangan Sistem. Dalam tahap perancangan sistem ini, rumusan-rumusan berbagai pengetahuan yang telah diperoleh sebelumnya kemudian dioleh lebih lanjut ke dalam bentuk pengetahuan dengan format JIKA MAKA. Konsekuen adalah kesimpulan yang didapat berdasarkan aturan melalui pengetahuan dasar. Interface yang juga dirancang pada tahapan ini termasuk interface untuk proses diagnosa dan interface penginputan data.
 - Tahapan Implementasi. Dalam tahapan implementasi ini rancangan yang telah diselesaikan pada tahapan perancangan sistem sebelumnya kemudian diterjemahkan ke dalam sintaks bahasa pemrograman
 - Tahapan Pengujian Sistem. Pada tahap pengujian sistem ini proses diagnosa serta data diinput dilakukan, dan keluaran sistem yang berupa hasil diagnosa kemudian dianalisis untuk dilakukan pengujian apakah kebenarannya telah sesuai dengan semua teori-teori yang telah dilibatkan dalam penelitian.

4. Hasil Dan Pembahasan

A. Tahap Analisis

Pada tahapan analisis ini dibangun tabel keputusan dasar pembuatan mesin inferensi yang didasari oleh pengetahuan dan informasi yang telah diperoleh sebelumnya dari pakar. Nilai *Certainty Factor* ditentukan terhadap setiap gejala yang melekat pada suatu gangguan dan dalam sistem pakar ini nilai *certainty factor* dapat diperbaru, berdasarkan hal tersebut maka dapat dibentuk tujuh aturan R (*rule*) Sebagai berikut :

R1: IF G001 AND G002 AND G003 AND G004 AND G005 AND G006 AND G007 THEN penyakit = Gangguan kecemasan Umum (*Generalized Anxiety Disorder*)

R2: IF G008 AND G009 AND G010 AND G011 AND G012 AND G013 THEN penyakit = Skizofrenia

R3: IF G014 AND G015 AND G016 AND G017 AND G018 THEN penyakit = Bipolar

R4: IF G019 AND G020 AND G021 AND G022 THEN penyakit = Autisme

R5: IF G023 AND G024 AND G025 AND G026 AND G027 AND G028 THEN penyakit = Defisit Perhatian atau Gangguan Hiperaktif (ADHD)

Tabel 1. Tabel Keputusan Penyakit Berdasarkan Gejala

Tabel keputusan berdasarkan gejala					
Kode gejala	Gejala	Penyakit			
		P001	P002	P003	P004 P005

**PENERAPAN METODE CERTAINTY FACTOR PADA SISTEM PAKAR
DIAGNOSA GANGGUAN MENTAL PADA ANAK BERBASIS WEB**

G001	Fobia pada hal yang spesifik atau ketakutan akan suatu benda tertentu	√	
G002	Mengalami kecemasan yang berlebihan dalam setiap situasi	√	
G003	Nyeri Otot	√	
G004	Sering merasa terkucilkan hingga berpikiran hal buruk akan terjadi	√	
G005	Sulit Tidur	√	
G006	Rasa takut dan bersalah yang hebat.	√	
G007	Mengompol di kasur pada malam hari.	√	
G008	Tidak mempunyai kepedulian terhadap kebersihan diri serta penampilan	√	
G009	Lebih memilih berdiam diri dirumah dari pada bersosialisasi dengan lingkungan luar	√	
G0010	Berbicara mendengar serta melihat hal-hal yang tidak ada	√	
G011	Memiliki ekspresi wajah yang datar walaupun sedang merasa senang ataupun sedih atau ekspresi wajah tidak sesuai dengan perasaan hatinya	√	
G012	Tidak bisa membedakan dunia nyata dan khayalan	√	
G013	Kecenderungan mengasingkan diri dari orang lain.	√	
G014	Terlihat sangat lelah atau bahkan terlihat sangat segar	√	
G015	Terlalu sering tidur atau tidak butuh tidur	√	
G016	Memiliki perilaku yang tidak sesuai dengan usianya, Terlihat sangat senang, lalu mendadak bisa menjadi sangat agresif serta sangat marah.	√	
G017	Sering adanya keluhan mengenai rasa sakit pada bagian tubuh tertentu, misalnya sakit pada kepala atau sakit pada perut dan mudah mengalami kelelahan.	√	
G018	Terus menerus mengalami dan merasa hampa, sedih atau cemas.	√	
G019	Sulit bersosialisasi		√
G020	Mengalami kesulitan dalam membaca, menulis, membaca, bahkan berbicara dan kesulitan dalam pemahaman bahasa isyarat, contohnya melambai dan menunjuk		√
G021	Kurang peduli, simpati dan empati serta kurangnya emosi terhadap suatu objek, peristiwa bahkan orang		√
G022	badan luka-luka		√
G023	Melakukan aktifitas-aktifitas yang dapat menimbulkan bahaya terhadap diri sendiri, contohnya membentur-benturkan kepala ke dinding atau menggigit tangan dengan kencang		√
G024	Tidak bisa berkonsentrasi atau fokus		√
G025	Sering tidak menyelesaikan tugas/pekerjaan		√
G026	Berbicara tanpa henti hingga mengganggu orang lain		√
G027	Berbicara berulang kali untuk diri sendiri		√
G028	memberi jawaban sebelum pertanyaan diselesaikan		√

Keterangan Penyakit: P001: Gangguan kecemasan Umum (*Generalized Anxiety Disorder*), P002: Skizofrenia, P003: Bipolar, P004: Autisme, P05: Defisit Perhatian atau Gangguan Hiperaktif (ADHD)

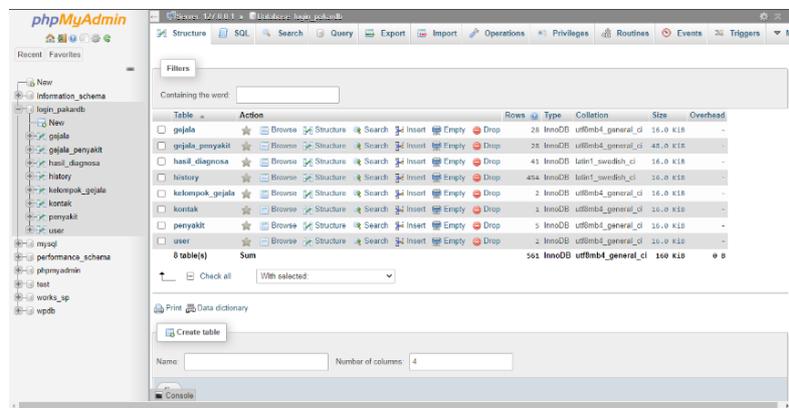
B. Tahap Perancangan

Sistem pakar diagnosa ini dapat dijalankan langsung tanpa perlu proses login maupun daftar. Kemudian user dapat langsung menginput gejala yang dialami pasien dan memproses hasil sehingga di dapat persentase gangguan yang dialami oleh anak. Sistem ini dibuat dengan sasaran orang tua yang mana dibuat untuk mendiagnosa anak, jadi sistem ini sudah dibuat sesuai kriteria umur anak menurut Departemen Kesehatan Republik Indonesia yaitu 5-11 tahun. Sehingga gejala yang disajikan umum sesuai rata-rata umur anak Indonesia.

C. Tahap Pembuatan

1. Database

Pembuatan database dengan menggunakan *localhost phpmyadmin* sebagai media untuk menyimpan data sistem dengan tabel gejala, gejala_penyakit, hasil_diagnosa, history, kelompok_gejala, kontak_penyakit, dan user. Berikut gambar 1 pembuatan *database*.



Gambar 1. Pembuatan Database

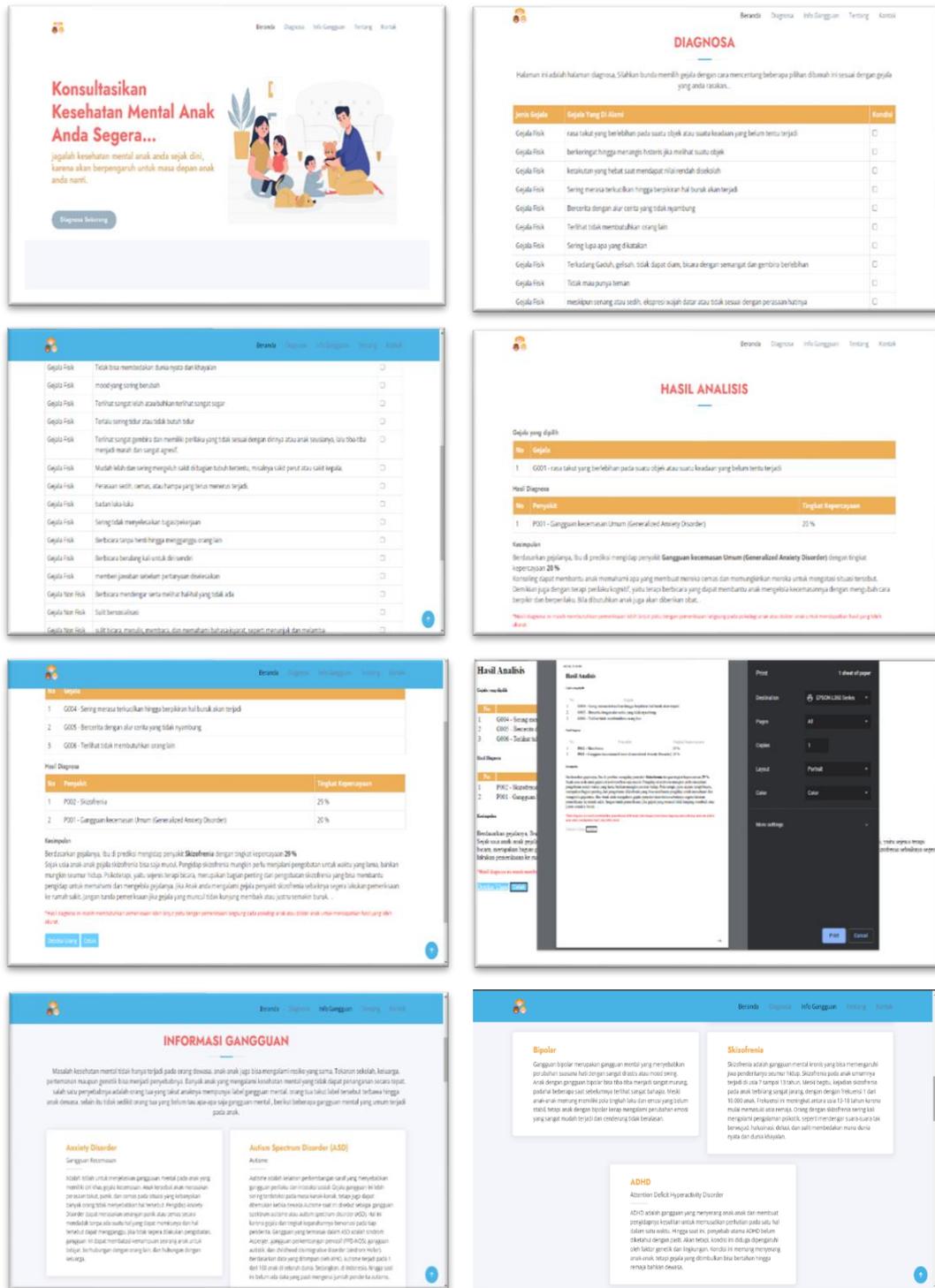
2. Pembuatan Web Aplikasi Sistem Pakar

Setelah pembuatan *database*, selanjutnya pembuatan sistem pakar diagnose berbasis *web*. Pada pembuatan *web* aplikasi ini menggunakan *framework* codeigniter 3 dengan konsep MVC (*Model View Controller*). Pada aplikasi ini memiliki beberapa halaman dan beberapa fitur. Berikut gambar-gambar pembuatan halaman dan fitur untuk *web* aplikasi sistem pakar.

D. Implementasi

Berikut adalah implementasi dari Sistem Pakar Diagnosa Gangguan Mental pada Anak dengan Metode Certainty Factor berbasis Web.

PENERAPAN METODE CERTAINTY FACTOR PADA SISTEM PAKAR DIAGNOSA GANGGUAN MENTAL PADA ANAK BERBASIS WEB



Gambar 3. Implementasi web aplikasi sistem

5. Kesimpulan

Kesehatan mental anak sangat penting untuk masa depan anak. Kesehatan mental pada anak tidak bisa diabaikan karena merupakan faktor yang berpengaruh dalam tumbuh kembang anak. Kesehatan mental anak dapat berpengaruh terhadap keseluruhan well-being anak. Kesehatan mental anak dipengaruhi oleh berbagai faktor baik faktor internal

maupun eksternal. Dalam pemahaman akan kesehatan mental pada anak dan remaja sangat penting juga dipahami bahwa kesehatan mental memiliki faktor yang dapat melindunginya serta faktor yang dapat membahayakannya. CCM adalah suatu sistem pakar yang bertujuan melakukan diagnosa gejala awal pada penyakit gangguan mental pada anak dan remaja dengan menggunakan metode certainty faktor yang bisa dilakukan di berbagai tempat dan memilih beberapa gejala yang terjadi pada anak agar sistem dapat mendiagnosa gangguan mental anak yang di alami. Dengan adanya sistem pakar ini diharapkan dapat memberikan pengetahuan serta dapat membantu para orang tua dalam melakukan diagnosa mandiri secara dini gangguan mental yang anak mereka alami secara mandiri dengan bantuan seorang ahli dengan berdasarkan gejala-gejala yang dirasakan agar para orang tua peka terhadap anak dan tidak menganggap remeh hal tersebut.

Adapun beberapa saran yang bisa menjadi acuan untuk pengembangan sistem selanjutnya, antara lain:

1. Sistem dapat di kembangkan menjadi lebih spesifik dengan memperluas pengambilan sampel anak dengan gangguan terkait melalui jenis kelamin dan usia.
2. Penelitian melalui jurnal dapat lebih diperluas lagi agar mendapat hasil yang akurat pada setiap kemungkinan gejala yang jarang terjadi.
3. Gejala yang disajikan bisa lebih diperbanyak melalui data yang ada agar tidak terjadi gangguan yang sama melalui gejala yang sedikit berbeda.

Referensi

- [1] E. Yulindari, A. Pengertian, and K. Mental, "Kesehatan Mental Anak dan Remaja."
- [2] S. Utami and D. Hanifah, "Faktor Risiko Masalah Mental Emosional Pada Anak Prasekolah Di Kota Sukabumi," *Buletin Penelitian Sistem Kesehatan*, vol. 24, no. 3, pp. 192–201, Sep. 2021, doi: 10.22435/hsr.v24i3.4066.
- [3] Umar, Rizqi, Arnie Retno Mariana, and Oktavyani Purnamasari. "Perancangan Sistem Pakar Diagnosa Penyakit Diabetes Melitus Menggunakan Metode Forward Chaining Berbasis Web." *Jurnal Sisfotek Global* 7.1 (2017).
- [4] A. S. Sembiring *et al.*, "Implementation of Certainty Factor Method for Expert System," in *Journal of Physics: Conference Series*, Sep. 2019, vol. 1255, no. 1. doi: 10.1088/1742-6596/1255/1/012065.
- [5] Heckerman, David. "The certainty-factor model." *Encyclopedia of Artificial Intelligence, Second Edition* (1992): 131-138.
- [6] L. Oktaviani and M. Ayu, "Pengembangan Sistem Informasi Sekolah Berbasis Web Dua Bahasa SMA Muhammadiyah Gading Rejo," vol. 6, no. 2, p. 2021, 2021, doi: 10.30653/002.202162.731.
- [7] R. Puji Hastanti and B. Eka Purnama, "Sistem Penjualan Berbasis Web (E-Commerce) Pada Tata Distro Kabupaten Pacitan," *Jurnal Bianglala Informatika*, vol. 3, no. 2, 2015, [Online]. Available: <http://lppm3.bsi.ac.id/jurnal>
- [8] Sumolang, Marcelino. "Peranan Internet terhadap generasi muda di desa tounelet kecamatan Langowan barat." *ACTA DIURNA KOMUNIKASI* 2.4 (2013).
- [9] Siregar, Januari. *Pengembangan Aplikasi Pendaftaran Online Layanan Pencatatan Sipil Berbasis Web Menggunakan PHP dan Basis Data MySQL (Studi Kasus: Dispendukcapil Kabupaten Malang)*. Diss. Universitas Brawijaya, 2018.
- [10] Lubis, Aulia Rahmah. "Perangkat lunak Aulia Rahmah Lubis." (2020).

DETEKSI KATA TAK BAKU DAN KESALAHAN PENULISAN KATA PADA TUGAS AKHIR MAHASISWA MENGGUNAKAN METODE *DICTIONARY LOOKUP*

*Alim Misbullah¹, Viska Mutiawani¹, dan Cut Sri Mulyani¹

¹Jurusan Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Syiah Kuala, Darussalam, Banda Aceh, 23111, Indonesia
E-mail: misbullah@unsyiah.ac.id, viska.mw@unsyiah.ac.id,
cut.sri@s1.informatika.unsyiah.ac.id

Abstract

Dictionary lookup is a simple searching method that aimed to find a word in the dictionary. The technique is effectively implemented to find the incorrect spelling based on the lexical resource. This research will create a system to detect non-standard and misspell words by looking into Standard Indonesian Language Dictionary (KBBI) and Kateglo Dictionary. In addition, the lookup process of the system is optimized by utilizing an index in the database. To validate the system, the front-end interface is created by displaying the options for end-user to upload their documents. The system will be evaluated by using 40 documents of the students' final project. The testing result shows that the average document processing speed is 187 words per second. The average of incorrect words in each document is 2.76%, with a precision value is 28.71%. The low precision is caused by the higher value of False Positive, which was determined by non-existence words in the dictionary. However, the system has shown good performance by obtaining higher True Negative values from each document which implies the accuracy is also high.

Keywords: *dictionary lookup, word index, database, speed, accuracy*

Abstrak

Dictionary lookup merupakan sebuah metode pencarian sederhana yang bertujuan untuk menemukan kata di dalam sebuah kamus. Metode ini secara efektif dapat diimplementasikan untuk menemukan kesalahan penulisan kata berdasarkan sumber leksikal. Penelitian ini dilakukan untuk membangun sebuah sistem yang dapat mendeteksi kata yang tidak standar dan kesalahan penulisan kata dengan melakukan pencarian kata tersebut di dalam Kamus Besar Berbahasa Indonesia (KBBI) dan Kamus Kateglo. Selain itu, process pencarian kata di dalam kamus dioptimisasi dengan mengutilisasi *index* yang ada pada *database*. Validasi sistem dilakukan dengan membangun sistem antarmuka yang menampilkan opsi kepada pengguna untuk mengunggah dokumennya. Sistem dievaluasi menggunakan 40 dokumen tugas akhir mahasiswa. Hasil uji menunjukkan bahwa rata-rata dokumen dapat diproses dengan kecepatan 187 kata tiap detik. Rata-rata kesalahan kata yang terdapat di setiap dokumen adalah 2,76% dengan nilai presisi adalah 28,71%. Nilai presisi yang rendah disebabkan oleh tingginya nilai *False Positive* yang ditentukan

DETEKSI KATA TAK BAKU DAN KESALAHAN PENULISAN KATA PADA TUGAS AKHIR MAHASISWA MENGGUNAKAN METODE *DICTIONARY LOOKUP*

oleh tidak adanya kata tersebut di dalam kamus. Namun, sistem yang dibangun sudah memperoleh nilai *True Negative* yang tinggi sehingga akurasi juga menjadi tinggi.

Kata Kunci: *dictionary lookup*, *index kata*, *database*, *kecepatan*, *akurasi*

1. Pendahuluan

Bahasa Indonesia merupakan bahasa resmi yang digunakan di Indonesia. Pada penulisan karya tulis ilmiah seperti skripsi, jurnal, tesis, disertasi, dan beberapa karya sejenis lainnya, semua harus ditulis dengan Bahasa Indonesia ragam baku. Penggunaan bahasa baku pada tulisan-tulisan dimaksud akan membuat para pembaca lebih mudah memahami konteks yang dimaksud oleh penulis dan memperkecil kemungkinan salah tafsir.

Kualitas sebuah karya tulis ilmiah tidak hanya dilihat dari isi dan ide atau pemikiran penulis, tetapi juga penggunaan bahasa dalam penulisan. Salah satu penggunaan Bahasa Indonesia pada karya tulis ilmiah adalah penulisan skripsi tugas akhir yang diwajibkan oleh setiap universitas untuk jenjang sarjana. Setiap universitas menerbitkan buku panduan yang menyebutkan bahwa Bahasa Indonesia yang digunakan dalam menulis proposal atau laporan tugas akhir adalah bahasa tulis yang baku. Penggunaan ejaan juga harus berpedoman pada ejaan Bahasa Indonesia yang telah disempurnakan [1]. Semua kata yang digunakan juga harus merujuk kepada Kamus Besar Bahasa Indonesia (KBBI).

Salah satu cara untuk mengetahui apakah kata yang digunakan dalam sebuah tulisan itu baku atau tidak adalah dengan merujuk ke kata-kata yang terdapat di dalam KBBI. Jika setiap kata dari tulisan yang diperiksa terdapat di dalam kamus, maka kata itu adalah layak disebut kata baku. Sebaliknya, jika kata tersebut tidak terdapat di dalam kamus maka kata tersebut kemungkinan bukan termasuk kata baku, atau merupakan istilah asing (bukan Bahasa Indonesia), atau merupakan kata dengan kesalahan tipografi (salah tik). Salah satu kesulitan yang akan dihadapi saat melakukan pengecekan yaitu adanya kata-kata yang merupakan istilah dari bahasa asing dikarenakan tulisan-tulisan yang berasal dari tema keilmuan matematika dan pengetahuan alam, khususnya bidang informatika atau ilmu komputer.

Berdasarkan hal di atas, maka penelitian ini dilakukan untuk memeriksa kebenaran tulisan dan pengetikan dari naskah tugas akhir mahasiswa Informatika yang saat ini telah menjadi alumni. Tahapan yang akan dilakukan nanti dimulai dengan membentuk dua jenis kamus, yaitu kamus Bahasa Indonesia yang berisi kosakata Bahasa Indonesia dari semua kelas kata dan kamus istilah asing. Kosakata Bahasa Indonesia didapat dari situs KBBI daring, sedangkan kumpulan istilah asing didapat dari situs kateglo.com yang dikembangkan oleh Ivan Lanin, Romi Hardiyanto, dan Arthur Purnama. Kamus inilah yang akan digunakan nantinya dalam memeriksa tulisan tugas akhir mahasiswa.

Salah satu penelitian terkait yang dilakukan oleh Hamzah [2] pada tahun 2016 mengungkapkan bahwa kesalahan penulisan ejaan kata umumnya disebabkan oleh dua hal, yaitu galat tipografi dan ketidaktahuan penulis terhadap kosakata baku Bahasa Indonesia. Galat tipografi biasanya terjadi karena kelalaian penulis yang tidak disengaja seperti fokus yang terganggu, slip tangan atau jari akibat kedekatan tombol huruf saat mengetik, atau sebab lainnya.

Metode yang digunakan dalam pemeriksaan naskah mahasiswa untuk menemukan kesalahan penulisan adalah metode *dictionary lookup*. Metode ini telah dilakukan di beberapa penelitian terkait deteksi kesalahan kata seperti yang dikaji oleh Maghfira [3] dan Hadi [4]. Menurut Maghfira [3] dalam artikel jurnalnya, metode *dictionary lookup* dinilai efektif dalam menentukan suatu ejaan kata bernilai benar atau salah berdasarkan

Lexical Resource. Metode ini tergolong sederhana apalagi untuk menentukan non-word error. Namun penelitian tersebut belum mempertimbangkan istilah dari bahasa asing yang digunakan dalam pengujiannya. Oleh karena itu, penelitian ini selain memeriksa kesalahan penulisan dengan metode *dictionary lookup*, juga akan mempertimbangkan istilah bahasa asing dalam pengujiannya.

2. Kajian Pustaka

A. Kata Baku dalam Bahasa Indonesia

Dalam KBBI Edisi Keempat disebutkan pengertian baku adalah pokok, utama; tolok ukur yang berlaku untuk kuantitas dan kualitas yang ditetapkan berdasarkan kesepakatan; standar. Sementara menurut Kosasih dan Hermawan [5] kata baku adalah kata yang diucapkan atau ditulis oleh seseorang sesuai dengan kaidah atau pedoman yang dibakukan. Kaidah standar yang dimaksud dapat berupa pedoman ejaan yang disempurnakan (EYD), tata bahasa baku, dan kamus. Kata baku umumnya sering dipakai pada kalimat resmi atau ragam bahasa baku, baik itu melalui lisan ataupun tulisan. Kata baku dalam bahasa Indonesia ini juga memiliki beberapa ciri-ciri. Pertama, baik secara lisan maupun tulisan, kata baku digunakan dalam situasi resmi, seperti surat menyurat dinas, perundang-undangan, karangan ilmiah, laporan penelitian dan lainnya. Ragam bahasa baku tidak diwarnai atau dicampuri oleh dialek atau logat tertentu. Kedua, baik secara lisan maupun tulisan, kata baku menggunakan ketentuan-ketentuan yang berlaku dalam Pedoman Umum Ejaan Bahasa Indonesia. Ketiga, baik secara lisan maupun tulisan, ragam baku memenuhi fungsi gramatikal seperti subjek, predikat, dan objek secara eksplisit dan lengkap [6].

Setiorini [7] dalam artikel jurnalnya mengatakan bahwa penggunaan bahasa akan berubah sesuai dengan kebutuhan penuturnya. Sebagai contoh, bahasa yang digunakan saat seseorang berpidato atau berceramah dalam sebuah seminar akan berbeda dengan bahasa yang digunakannya saat mengobrol atau bercengkerama dengan keluarganya. Bahasa itu akan berubah lagi saat ia menawar atau membeli sayuran di pasar. Kesesuaian antara bahasa dan pemakaiannya ini disebut ragam bahasa. Dalam penggunaan bahasa (Indonesia) dikenal berbagai macam ragam bahasa dengan pembagiannya masing-masing, seperti ragam formal-semi formal-nonformal; ujaran-tulisan; jurnalistik; iklan; populer dan ilmiah.

Menurut Hasan Alwi [8], ragam bahasa ini memiliki dua ciri, yaitu kemantapan dinamis dan kecendekiaan. Kemantapan dinamis berarti aturan dalam ragam bahasa ini telah berlaku dengan mantap, tetapi bahasa ini tetap terbuka terhadap perubahan (terutama dalam kosakata dan istilah). Ciri kecendekiaan terlihat dalam penataan penggunaan bahasa secara teratur, logis, dan masuk akal. Ragam bahasa ini bersifat kaku dan terikat pada aturan-aturan bahasa yang berlaku. Sebagai bahasa baku, terdapat standar tertentu yang harus dipenuhi dalam penggunaan ragam bahasa ilmiah. Standar tersebut meliputi penggunaan tata bahasa dan ejaan bahasa Indonesia baku. Tata bahasa Indonesia yang baku meliputi penggunaan kata, kalimat, dan paragraf yang sesuai dengan kaidah baku. Kaidah tata bahasa Indonesia yang baku adalah kaidah tata bahasa Indonesia yang sesuai dengan aturan berbahasa yang ditetapkan oleh Pusat Bahasa Indonesia. Sementara itu, kaidah ejaan bahasa Indonesia yang baku adalah kaidah ejaan bahasa Indonesia. Sesuai dengan ragam bahasanya, aturan-aturan ini mengikat penggunaan bahasa dalam karya tulis ilmiah.

Salah satu fungsi kata baku dalam Bahasa Indonesia adalah sebagai kerangka acuan [6]. Kata baku sebagai kerangka acuan artinya kata baku menjadi patokan bagi

DETEKSI KATA TAK BAKU DAN KESALAHAN PENULISAN KATA PADA TUGAS AKHIR MAHASISWA MENGGUNAKAN METODE *DICTIONARY LOOKUP*

benar atau tidaknya pemakaian bahasa seseorang atau kelompok. Namun terkadang banyak kata tak baku yang sangat sering ditemukan pada tulisan ilmiah seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Contoh Kata Tak Baku dan Kata Baku

No.	Kata Tak Baku	Kata Baku
1	Analisa	Analisis
2	Praktek	Praktik
3	Desaign	Desain
4	Aktip	Aktif
5	Diagnosa	Diagnosis
6	Detil	Detail
7	Efektifitas	Efektivitas
8	Hirarki	Hierarki
9	Hipotesa	Hipotesis
10	Katagori	Kategori

B. *Typographical Error (Kesalahan Pengetikan)*

Ketika membuat karya tulis ilmiah, sering sekali terjadi kesalahan dalam hal pengetikan. Kesalahan tersebut dapat berupa kurangnya pengetahuan mahasiswa akan ejaan yang benar sesuai dengan kamus besar Bahasa Indonesia, kelalaian mahasiswa yang tidak disengaja, kesalahan pengaturan aplikasi yang digunakan untuk media pengetikan dan beberapa hal lain yang menyebabkan terjadinya kesalahan ejaan kata [2].

Kesalahan ejaan merupakan keadaan di mana terjadi kesalahan penulisan susunan kata. Berdasarkan sejarahnya, awalnya keadaan ini berhubungan dengan kesalahan penulisan kata secara manual, namun saat ini hal tersebut juga dapat terjadi pada proses pengetikan yang dilakukan dengan bantuan mesin ketik dan komputer. Hal tersebut dapat terjadi dikarenakan kesalahan mekanik juga tangan atau jari memeleset saat mengetik, selain itu terkadang juga disebabkan oleh ketidaktahuan seseorang tentang bagaimana pengejaan tulisan yang benar [3].

Berdasarkan jenis katanya *typographical error* dapat dibedakan menjadi 2 tipe yaitu *non-word spelling error* dan *real-word spelling error*. *Non-word spelling error* merupakan kesalahan penulisan kata di mana kata tersebut tidak dapat ditemukan dalam kamus (tidak memiliki makna). Sedangkan *real-word spelling error* merupakan kesalahan penulisan kata di mana kata tersebut dapat ditemukan dalam kamus (memiliki makna) namun bukan kata yang dimaksud dalam dokumen [3].

Tahap awal yang dilakukan untuk menghasilkan tulisan yang bebas dari salah eja adalah pendeteksian. Menurut Maghfira [3], deteksi kesalahan ejaan merupakan proses pengecekan validitas suatu kata dalam bahasa tertentu, suatu kata disebut valid jika kata tersebut dapat ditemukan dalam *lexical resource*. *Lexical resource* merupakan *database* di mana data di dalamnya dapat berupa *corpus*, *lexicon*, *word list* atau bentuk lain. Proses utama dari *error detection* adalah membandingkan kata dalam teks dengan kata yang terdapat pada *lexical resource*. Banyak metode yang dapat digunakan untuk proses deteksi kesalahan penulisan kata, namun yang umumnya digunakan untuk deteksi *non-word error* adalah deteksi menggunakan *Dictionary Lookup* dan *N-Gram Analysis*.

Metode *dictionary lookup* merupakan metode yang sering digunakan dalam menentukan *non-word error*. Proses yang dilakukan pada metode ini yaitu melakukan pengecekan apakah kata yang dimaksud terdaftar dalam kamus atau tidak, jika tidak ada maka kata ini dianggap sebagai *non-word*. Cara ini termasuk cara yang efektif untuk menentukan kata termasuk salah penulisannya atau tidak, namun jumlah kata dalam

kamus yang banyak dapat berakibat pada proses pengecekan menjadi lama, oleh karena itu dibutuhkan teknik optimasi pada teknik pencarian kata. Teknik optimasi dapat dilakukan dengan penggunaan binary search dan hash seperti pada penelitian [9].

C. Dictionary Lookup

Dictionary lookup merupakan metode yang melakukan pencarian secara sederhana untuk melihat keberadaan kata di dalam kamus atau daftar kata yang telah dibuat (Putra, Sujaini, & Safriadi, 2018). Metode dictionary lookup merupakan sebuah metode yang sering digunakan dalam menentukan non-word error, yaitu kesalahan penulisan kata yang menjadikan ia tidak dapat ditemukan dalam kamus. Proses yang dilakukan pada metode ini yaitu melakukan pengecekan apakah kata yang dimaksud terdaftar dalam kamus atau tidak, jika tidak ada maka kata ini dianggap sebagai non-word. Menurut Soleh [9], cara ini termasuk cara yang efektif untuk menentukan kata termasuk salah penulisannya atau tidak, namun jumlah kata dalam kamus yang banyak dapat berakibat pada proses pengecekan menjadi lama. Oleh karena itu dibutuhkan teknik optimasi pada teknik pencarian kata. Pada penelitian Soleh et al. (2011) teknik optimasi dilakukan dengan penggunaan binary search dan hash.

D. PHP Framework Laravel

PHP adalah singkatan perulangan untuk PHP: Hypertext Preprocessor, yaitu bahasa pemrograman yang digunakan secara luas untuk penanganan pembuatan dan pengembangan sebuah situs web. Ada banyak framework PHP yang populer, salah satunya Laravel. Framework Laravel dibuat oleh Taylor Otwell, Proyek Laravel dimulai pada April 2011. Laravel merupakan Framework PHP yang menekankan pada kesederhanaan dan fleksibilitas pada desainnya. Sama seperti framework lainnya, Laravel dibangun dengan basis MVC (Model, View, Controller). Banyak situs survei yang menulis bahwa framework Laravel ini merupakan yang terpopuler dibandingkan framework PHP lainnya. (Rohman, 2014)

3. Metode Penelitian

Penelitian ini terbagi dalam beberapa tahapan seperti yang terlihat pada Gambar 1. Tahap awal yang dilakukan pada penelitian ini yaitu mengidentifikasi masalah. Hal ini bertujuan untuk memastikan ruang lingkup penelitiannya. Pada tahap ini diidentifikasi latar belakang serta rumusan masalah yang akan diselesaikan.



DETEKSI KATA TAK BAKU DAN KESALAHAN PENULISAN KATA PADA TUGAS AKHIR MAHASISWA MENGGUNAKAN METODE *DICTIONARY LOOKUP*

Gambar 1. Alur Tahapan Penelitian

Selanjutnya, langkah yang dilakukan pada tahapan kedua yaitu mencari sumber kajian sebagai pengetahuan dasar dari berbagai macam literatur yang dapat dijadikan landasan dalam pengerjaan penelitian ini. Tujuan studi literatur dilakukan untuk memahami lebih dalam permasalahan dari penelitian ini. Sumber yang digunakan dalam studi literatur ini berasal dari beberapa artikel dan buku yang membahas topik yang relevan dengan penelitian ini.

Tahapan ketiga merupakan tahapan yang sangat penting dalam penelitian ini karena pengumpulan data merupakan komponen pendukung utama untuk melakukan penelitian ini. Data yang dikumpulkan berasal dari beberapa sumber diantaranya data kamus Bahasa Indonesia pada situs KBBI daring dan data kamus istilah asing berasal dari situs *kateglo.com*. Pengambilan data dilakukan dengan menggunakan teknik *scraping* dengan memanfaatkan fitur dari *webscraper.io*. Semua kata diambil dari situs halaman per halaman hingga tuntas, kemudian dilakukan *proprocessing* untuk membersihkan halaman tersebut sebelum disimpan dalam format *plain text*. Adapun data naskah tugas akhir mahasiswa dikumpulkan dari arsip Jurusan Informatika atau alumni secara langsung.

Tahapan implementasi program dikerjakan dengan menggunakan bahasa PHP dan *framework* Laravel. Kode skrip disusun untuk memproses tahapan dari awal hingga akhir, dimulai dari proses *preprocessing*, kemudian proses *Dictionary Lookup* pada saat melengkapi kamus dan saat mencari kata yang salah pada naskah tugas akhir. Tahapan terakhir yang dilakukan adalah analisa hasil yaitu terbentuknya kamus kota kata yang lengkap sehingga dapat digunakan untuk memeriksa dan menemukan kata-kata yang salah dan keliru dalam tulisan ilmiah tugas akhir mahasiswa Jurusan Informatika. Saat melakukan *Dictionary Lookup* pada pengecekan tugas akhir, kata-kata yang tidak terdapat di kamus akan ditandai dan dihitung jumlahnya pada setiap dokumen. Kata-kata salah yang sudah dikumpulkan dalam satu dokumen dianalisis kesalahannya. Lalu akan diuji keakuratan algoritma pada penelitian ini dengan menggunakan *confusion matrix*.

4. Hasil dan Pembahasan

Tahap awal dalam pembangunan kamus diawali dengan mengumpulkan kosakata. Kosakata yang dibutuhkan meliputi kosakata bahasa Indonesia, singkatan- singkatan, serta istilah asing. Kosakata diambil dari situs yang terdiri dari banyak halaman, sehingga digunakan *web scraper* untuk mempercepat pengumpulan kata. Setelah dilakukan proses *scraping*, diperoleh kata bahasa Indonesia dari tujuh kelas kata yaitu adjektiva, adverbial, nomina, numeralia, partikel, pronomina, dan verba dengan jumlah 77.348 kata. Singkatan-singkatan beserta istilah asing dari situs *Kateglo* didapat sebanyak 191.200 kata. Karena adanya duplikasi dan kata homonim dari seluruh halaman yang telah dilakukan *scraping*, maka kata yang sama dihapus sehingga menghasilkan kata yang unik berjumlah 130.431 kata.

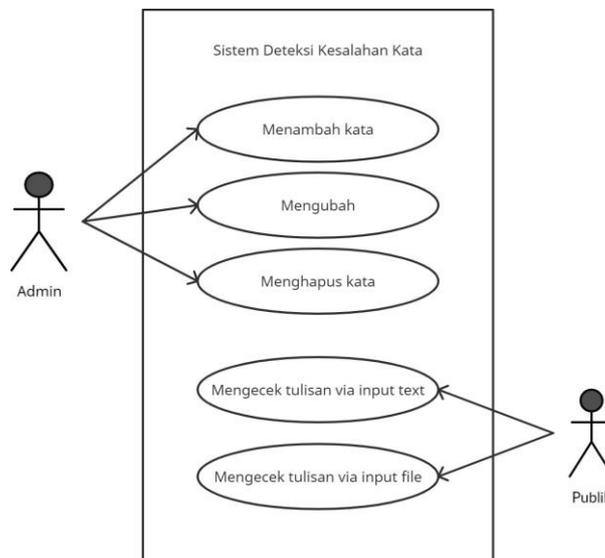
Selanjutnya, berdasarkan tujuan awal dari penelitian ini adalah mendeteksi keberadaan kata tak baku dan kesalahan penulisan kata pada naskah tugas akhir mahasiswa Jurusan Informatika Universitas Syiah Kuala. Objek data yang akan dijadikan sebagai sampel adalah bab satu dari naskah tugas akhir mahasiswa yang mewakili dari setiap tahun kelulusan. Oleh karena itu, berdasarkan data jumlah alumni Jurusan Informatika yang terlihat pada Tabel 2, maka data sampel yang akan digunakan sebagai pengujian berjumlah 40 dokumen tugas akhir. Jumlah ini dianggap memadai untuk dijadikan sampel yang optimal jika dibandingkan dengan jumlah populasi semua lulusan dari setiap tahun sebanyak 263 mahasiswa.

Tujuan dari penelitian ini adalah membentuk kamus kosakata dengan lengkap, memeriksa dan menemukan kata yang salah dan keliru dalam tulisan, serta menganalisis seberapa banyak kekeliruan yang terdapat dalam tulisan-tulisan tersebut. Tujuan-tujuan tersebut dicapai dengan implementasi *Dictionary Lookup* yang melakukan pencarian sederhana untuk melihat keberadaan kata di dalam kamus dan untuk menemukan non word error.

Tabel 2. Data jumlah lulusan Jurusan Informatika, Universitas Syiah Kuala

No.	Tahun Lulus	Jumlah Lulusan
1	2015	41 mahasiswa
2	2016	59 mahasiswa
3	2017	27 mahasiswa
4	2018	35 mahasiswa
5	2019	43 mahasiswa
6	2020	44 mahasiswa
7	2021	14 mahasiswa

Sistem pendeteksian kesalahan kata ini diimplementasikan dalam bentuk aplikasi berbasis web. Halaman web ini memiliki dua sisi pengguna, yaitu admin dan pengguna publik seperti yang digambarkan dalam diagram use case pada Gambar 2. Admin memiliki fungsi untuk mengelola kosakata dalam kamus seperti menambahkan, mengubah, dan menghapus kata. Pengguna publik hanya dapat mengakses halaman terluar yang berisi kolom input teks dan kolom input fail untuk mendeteksi kata yang salah.



Gambar 2. Diagram use case pengguna halaman deteksi kesalahan kata

Dalam implementasinya, data kamus disimpan dalam *database* dengan menggunakan *index* untuk setiap kata di dalam tabel. *Index* adalah sebuah objek dalam sistem *database* yang dapat mempercepat proses pencarian (*query*) data. *Index* merupakan objek struktur data tersendiri yang tidak bergantung kepada struktur tabel. Setiap *index* terdiri dari nilai kolom dan penunjuk ke baris yang berisi nilai tersebut. Penunjuk tersebut secara langsung menunjuk ke baris yang tepat pada tabel, sehingga menghindari terjadinya *full table-scan*. Pendeteksian kata ini membutuhkan pencarian berulang yang mengharuskan pengecekan ke kamus untuk setiap kata. Oleh sebab itu penggunaan *index* akan mempercepat kinerja sistem dalam melakukan pencarian secara berulang.

DETEKSI KATA TAK BAKU DAN KESALAHAN PENULISAN KATA PADA TUGAS AKHIR MAHASISWA MENGGUNAKAN METODE *DICTIONARY LOOKUP*

Sebanyak 40 dokumen tugas akhir mahasiswa Jurusan Informatika, Universitas Syiah Kuala digunakan sebagai sampel dalam proses pendeteksian terhadap kata-kata yang tidak baku dan salah tik. Hasil pada Tabel 3 menunjukkan jumlah kesalahan yang muncul pada setiap dokumen yang diperiksa.

Tabel 3. Uji pendeteksian kesalahan penulisan kata dalam dokumen

Dokumen	Jumlah Kata	Kata Yang Salah	Persentase Kata Yang Salah (%)	Waktu Eksekusi (ms)
1	517	9	1,74	3,31
2	616	6	0,97	4,04
3	830	34	4,10	4,87
4	425	19	4,47	2,77
5	655	15	2,29	4,26
6	754	20	2,65	4,04
7	1052	54	5,13	5,78
8	758	14	1,85	4,55
9	558	18	3,23	3,23
10	678	34	5,01	3,95
11	975	13	1,33	5,74
12	970	17	1,75	5,91
13	679	29	4,27	4,32
14	708	23	3,25	4,21
15	842	4	0,48	5,06
16	758	17	2,24	4,46
17	752	19	2,53	4,19
18	745	4	0,54	4,47
19	715	20	2,80	4,29
20	849	14	1,65	5,32
21	752	3	0,40	4,29
22	462	6	1,30	2,80
23	374	18	4,81	2,37
24	802	18	2,24	4,83
25	527	20	3,80	3,21
26	843	12	1,42	4,99
27	566	1	0,18	3,40
28	680	11	1,62	4,10
29	575	20	3,48	0,62
30	640	11	1,72	3,99
31	467	14	3,00	2,90
32	695	20	2,88	3,83
33	756	45	5,95	4,58
34	706	24	3,40	4,52
35	603	5	0,83	3,64
36	763	42	5,50	2,75
37	939	29	3,09	5,57
38	723	19	2,63	4,41
39	791	44	5,56	5,17
40	866	37	4,27	4,96

Berdasarkan Tabel 3 di atas, kemudian ingin dihitung rata-rata kecepatan sistem dalam proses eksekusi teks. Untuk itu, perlu dianalisis kemampuan jumlah kata per detik dalam setiap dokumen yang diproses. Hasil perhitungan jumlah kata per detik dalam setiap dokumen terlihat pada Tabel 4. Jika dihitung rata-ratanya, maka seluruh nilai pada Tabel 4 menghasilkan nilai rata-rata kecepatan eksekusi sebesar 187 kata per milidetik,

atau 187.414 kata per detik. Penggunaan metode *Dictionary Lookup* dengan memanfaatkan *index* pada DBMS dinilai efektif dari segi waktu eksekusi yang dibutuhkan.

Tabel 4. Perhitungan Kecepatan Eksekusi

Dokumen	Kecepatan (Kata/milidetik)	Kecepatan (Kata/detik)
1	156,14	156.137
2	152,39	152.392
3	170,14	170.141
4	153,09	153.087
5	153,71	153.713
6	186,62	186.624
7	181,81	181.809
8	166,58	166.582
9	172,37	172.371
10	171,64	171.637
11	169,73	169.733
12	164,09	164.093
13	157,02	157.023
14	167,84	167.844
15	166,23	166.229
16	169,94	169.944
17	179,08	179.082
18	166,32	166.321
19	166,58	166.581
20	159,43	159.428
21	175,00	174.998
22	164,58	164.577
23	157,74	157.739
24	165,86	165.864
25	164,16	164.164
26	168,64	168.644
27	166,02	166.021
28	165,60	165.603
29	913,86	913.859
30	160,07	160.072
31	160,58	160.580
32	181,17	181.169
33	164,80	164.803
34	155,94	155.943
35	165,38	165.378
36	277,43	277.434
37	168,30	168.301
38	163,60	163.601
39	152,84	152.841
40	174,38	174.375

Selanjutnya untuk mengevaluasi kinerja sistem, perlu dilakukan suatu pengujian. Pada penelitian ini kinerja sistem akan diuji dengan confusion matrix.. Penentuan TP (True Positive), FP (False Positive), TN (True Negative), dan FN (False Negative) didasari pada benar salahnya sistem mengelompokkan sebuah kata ke dalam kategori benar atau salah. Secara lebih sederhana, dua kelas yang dimiliki adalah positif dan negatif. Kelas positif adalah kelas milik tulisan berwarna merah atau kata yang terdeteksi salah tik atau tidak baku. Kelas negatif adalah kelas milik tulisan berwarna hitam atau yang dianggap benar. Agar lebih jelas, TP adalah tulisan yang dimerahkan dan ternyata

DETEKSI KATA TAK BAKU DAN KESALAHAN PENULISAN KATA PADA TUGAS AKHIR MAHASISWA MENGGUNAKAN METODE *DICTIONARY LOOKUP*

valid. TN adalah tulisan yang dihitamkan dan ternyata valid. FP adalah tulisan yang dimerahkan dan ternyata tidak valid. FN adalah tulisan yang dihitamkan dan ternyata tidak valid.

Tabel 5. Perhitungan nilai presisi pada setiap dokumen

Dokumen	Jumlah Kata	Kata Terdeteksi Salah	TP (<i>True Positive</i>)	FP (<i>False Positive</i>)	Presisi
1	517	9	4	5	44,44 %
2	616	6	2	4	33,33 %
3	830	34	18	16	52,94 %
4	425	19	8	11	42,11 %
5	655	15	0	15	0,00 %
6	754	20	0	20	0,00 %
7	1052	54	14	40	25,93 %
8	758	14	5	9	35,71 %
9	558	18	7	11	38,89 %
10	678	34	6	28	17,65 %
11	975	13	5	8	38,46 %
12	970	17	5	12	29,41 %
13	679	29	13	16	44,83 %
14	708	23	3	20	13,04 %
15	842	4	3	1	75,00 %
16	758	17	6	11	35,29 %
17	752	19	1	18	5,26 %
18	745	4	3	1	75,00 %
19	715	20	1	19	5,00 %
20	849	14	4	10	28,57 %
21	752	3	0	3	0,00 %
22	462	6	4	2	66,67 %
23	374	18	0	18	0,00 %
24	802	18	4	14	22,22 %
25	527	20	2	18	10,00 %
26	843	12	0	12	0,00 %
27	566	1	0	1	0,00 %
28	680	11	1	10	9,09 %
29	575	20	6	14	30,00 %
30	640	19	12	7	63,16 %
31	467	14	5	9	35,71 %
32	695	20	8	12	40,00 %
33	756	45	21	24	46,67 %
34	706	24	8	16	33,33 %
35	603	5	1	4	20,00 %
36	763	42	22	20	52,38 %
37	939	29	4	25	13,79 %
38	723	19	9	10	47,37 %
39	791	44	4	40	9,09 %
40	866	37	3	34	8,11 %

Nilai presisi dari setiap dokumen data sampel sangat bervariasi seperti yang terlihat pada Tabel 5. Jika dirata-ratakan, maka nilai presisi dari semua sampel adalah 28,71%. Artinya, nilai ini menjawab pertanyaan seberapa besar persentase jumlah kata yang benar-benar keliru dari seluruh kata yang dikategorikan keliru oleh sistem. Nilai presisi yang rendah ini jika dianalisis secara sederhana berdasarkan kata-kata yang dianggap salah oleh sistem, maka ditemukan bahwa penyebab utamanya adalah jumlah pada False Positive. Pada kolom FP, sebagian besar kata merupakan kata yang dianggap keliru oleh sistem karena kata tersebut adalah nama tempat, nama merek, nama orang, kutipan langsung dari bahasa lain, dan lainnya. Sedangkan sistem tetap akan mengategorikannya sebagai kata yang salah karena tidak berada dalam kamus.



Gambar 3. Hasil pendeteksian kata yang salah pada teks.

Hasil nilai presisi yang rendah pada Tabel 5 tidak menjadikan sistem ini buruk. Berdasarkan pengamatan sederhana pada hasil deteksi tulisan, sebagian besar kata yang benar tetap dikategorikan sebagai kata yang benar (dihitamkan) karena sudah tentu kata-kata tersebut berada di dalam kamus seperti terlihat di halaman sistem pada Gambar 3. Pengecualian terjadi pada kata-kata salah tik namun menghasilkan makna lain yang masih berada dalam kamus (word error). Melakukan pemindaian pada seluruh dokumen sampel untuk menemukan jumlah True Negative dan False Negative dianggap kurang efisien karena jumlah data yang sangat besar. Oleh karena itu, cukup dengan mengetahui bahwa nilai porsi TN pada semua sampel adalah sangat besar, maka kita dapat mengetahui bahwa nilai akurasi juga sangat besar.

5. Kesimpulan

Penelitian ini menghasilkan sebuah kamus kosakata dan aplikasi halaman web untuk mendeteksi kesalahan penulisan kata akibat salah tik maupun penggunaan kata tak baku. Kosa kata yang diperoleh sebanyak 131.495 yang terdiri dari kata bahasa Indonesia, singkatan, akronim, dan istilah asing. Selain itu, besar nilai rata-rata kecepatan deteksi tulisan pada percobaan sampel adalah 187.414 kata per detik. Metode *Dictionary Lookup* dinilai efektif dari segi waktu yang dibutuhkan. Hasil percobaan pendeteksian kata dari sampel yang digunakan menunjukkan bahwa rata-rata kesalahan yang ada di dalam setiap dokumen adalah sebesar 2,76%, dengan nilai presisi sebesar 28,71%. Nilai akurasi dan presisi dari hasil deteksi sangat bergantung pada tingkat kelengkapan kamus kosakata. Semakin lengkap isi kamus maka hasil yang didapat akan semakin baik.

References

- [1] U. Syiah Kuala, Panduan Tugas Akhir dan Tesis Fakultas Matematika dan Ilmu Pengetahuan Alam, Banda Aceh: FMIPA, Universitas Syiah Kuala, 2019.

**DETEKSI KATA TAK BAKU DAN KESALAHAN PENULISAN KATA PADA TUGAS
AKHIR MAHASISWA MENGGUNAKAN METODE *DICTIONARY LOOKUP***

- [2] H. R.N., "Aplikasi Perbaikan Ejaan Pada Karya Tulis Ilmiah di Program Studi Teknik Informatika dengan Menerapkan Algoritma Levenshtein Distance," Universitas Nusantara PGRI, Kediri, 2016.
- [3] T. N. Maghfira, I. Cholissodin and A. Wahyu, "Deteksi Kesalahan Ejaan dan Penentuan Rekomendasi Koreksi Kata yang Tepat Pada Dokumen Jurnal JTIK Menggunakan Dictionary Lookup dan Damerau-Levenshtein Distance," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, pp. 498-506, 2017.
- [4] A. Hadi, M. Fachrurrozi and N. Yuslianti, "Analisa Perbandingan Algoritma Damerau-Levenshtein Distance dan Soundex Similarity Pada Pengkoreksian Ejaan Kata Otomatis," Fakultas Ilmu Komputer, Universitas Sriwijaya, Palembang, 2019.
- [5] E. Kosasih and W. Hermawan, "BAHASA INDONESIA Berbasis Kepenulisan Karya Ilmiah dan Jurnal," CV. Thursina, Bandung, 2012.
- [6] S. and S. Saudah, *Buku Ajar Bahasa Indonesia Akademik*, Yogyakarta: Pustaka Pelajar, 2016.
- [7] S. R.A, *Analisis Penggunaan Tata Bahasa Indonesia dalam Penulisan Karya Tulis Ilmiah : Studi Kasus Artikel Ilmiah*, Visi Pustaka, 2010.
- [8] H. Alwi, *Tata bahasa baku bahasa Indonesia*, Jakarta: Perum Balai Pustaka, 1998.
- [9] M. Y. Soleh and A. Purwarianti, *A non word error spell checker for Indonesian using morphologically analyzer and HMM*, New Jersey, USA: IEEE, 2011.

ANALISIS PEMAHAMAN LITERASI DIGITAL PADA MAHASISWA UIN ARRANIRY TERHADAP *DIGITAL SKILL* DAN *DIGITAL SAFETY*

Ridwan¹, Yusran², dan Cut Addis Maulidia³

¹²³Pendidikan Teknologi Informasi, Universitas Islam Negeri Ar-Raniry, Banda Aceh
E-mail: ridwanmt@ar-raniry.ac.id, yusran@ar-raniry.ac.id,
cutaddis@student.arraniry.ac.id

Abstract

The development of information technology is currently developing widely, making the internet one of the needs. The rapid development of information technology allows information disclosure with various digital communication facilities or digital literacy. Digital skills and digital security are part of digital literacy that must be understood and considered when using digital devices. Based on the results of data analysis carried out using the IBM SPSS application and Microsoft Excel at UIN Ar-Raniry University, it was found that the results of the student's level of understanding in understanding digital literacy, digital skills, and digital safety were found. This study aims to analyze students' understanding in understanding digital literacy, digital skills, and digital safety. This research was conducted to see the understanding of students themselves in using digital devices and to see students' understanding of digital skills and digital safety. This research uses a quantitative methodology approach by distributing google forms to students. The sample in this study was 100 students of UIN Ar-Raniry with random sampling technique. The results showed that the results of the digital literacy variable, namely having a percentage of 86.54% with good criteria. The digital skill variable has a percentage of 83.14% with good criteria. And in the digital safety variable has a percentage of 78.95% with sufficient criteria.

Keywords: *digital literacy, digital skills, kuantitatif, internet and digital safety*

Abstrak

Perkembangan teknologi informasi saat ini berkembang secara luas sehingga menjadikan internet sebagai salah satu kebutuhan. Pesatnya perkembangan teknologi informasi memungkinkan terjadi keterbukaan informasi dengan berbagai fasilitas komunikasi digital atau literasi digital. Keterampilan digital dan keamanan digital adalah bagian dari literasi digital yang harus dipahami dan diperhatikan ketika menggunakan perangkat digital. Berdasarkan hasil analisis data yang dilakukan menggunakan aplikasi IBM SPSS dan Microsoft Excel di Perguruan Tinggi UIN Ar-Raniry, didapati hasil tingkat pemahaman mahasiswa dalam memahami literasi digital, digital *skill*, dan digital *safety*. Penelitian ini bertujuan untuk menganalisis tentang kepeahaman mahasiswa dalam memahami literasi digital, digital *skill*, dan digital *safety*. Penelitian ini dilakukan untuk melihat pemahaman mahasiswa itu

ANALISIS PEMAHAMAN LITERASI DIGITAL PADA MAHASISWA UIN ARRANIRY TERHADAP *DIGITAL SKILL* DAN *DIGITAL SAFETY*

sendiri dalam menggunakan perangkat digital dan untuk melihat pemahaman mahasiswa pada *digital skill* dan *digital safety*. Penelitian ini menggunakan pendekatan metodologi kuantitatif dengan menyebarkan google form kepada mahasiswa. Sampel pada penelitian ini adalah 100 orang mahasiswa UIN ArRaniry dengan Teknik *random sampling*. Hasil penelitian menunjukkan bahwa hasil dari variabel literasi digital yaitu memiliki persentase 86.54% dengan kriteria baik. Pada variabel *digital skill* memiliki persentase 83.14% dengan kriteria baik. Dan pada variabel *digital safety* yaitu memiliki persentase 78.95% dengan kriteria cukup.

Kata Kunci: *literasi digital, digital skill, kuantitatif, internet dan digital safety*

1. Pendahuluan

Perkembangan dunia informasi saat ini sangat berkembang luas seiring dengan kebutuhan zaman. Internet menjadi informasi teknologi terkini yang diperlukan khalayak umum, hingga dijadikan sebagai suatu alat komunikasi utama. Pesatnya teknologi informasi saat ini membuat penyebaran kebudayaan antar komunitas, bahkan antar bangsa, dimungkinkan terjadi berkat keterbukaan informasi dengan berbagai fasilitas komunikasi digital atau literasi digital.

Literasi digital ialah suatu proses berpikir, kompetensi yang dibutuhkan untuk memahami alur menyusun pengetahuan serta membangun informasi yang dapat dijadikan patokan dari berbagai sumber yang berbeda. Literasi digital dalam dimensinya terdapat *digital skill* dan *digital safety*. *Digital skill* dalam artian adalah keterampilan digital. Sedangkan *digital safety* adalah keamanan digital. Tujuan adanya *digital skill* dan *digital safety* adalah sebagai pengetahuan untuk pengguna dalam menggunakan perangkat digital.

Penelitian ini membahas mengenai pemahaman mahasiswa dalam memahami literasi digital, *digital skill* dan *digital safety*. Atas dasar ini lah penulis memilih judul “Analisis pemahaman literasi digital pada mahasiswa UIN Ar-Raniry terhadap *digital skill* dan *digital safety*”. Penelitian ini bertujuan untuk menganalisis pemahaman mahasiswa dalam memahami literasi digital, *digital skill* dan *digital safety*.

2. Studi Pustaka

A. Literasi Digital

Literasi digital adalah sebuah teori yang berfokus pada bagaimana memanfaatkan teknologi dan informasi dari platform digital secara efektif dan efisien dalam kehidupan sehari-hari. Literasi digital membutuhkan lebih dari sekedar kemampuan untuk mengoperasikan komputer atau perangkat lain, di samping itu juga membutuhkan pengetahuan tentang kapasitas kognitif, motorik, sosial dan emosional yang harus dimiliki pengguna untuk menggunakan internet secara efektif (Eshet, 2012:267).

B. *Digital Skill*

Digital Skill (keterampilan digital) adalah kemampuan dalam memahami, mengevaluasi, menggunakan, membuat, dan mengelola konten menggunakan perangkat digital seperti komputer dan smartphone. Saat ini teknologi telah menjadi bagian sentral dalam kehidupan sehari-hari, hal ini disebabkan oleh meningkatnya popularitas internet dan perkembangan komunikasi digital. Dalam keterampilan digital menggunakan media diperlukan hal yang memahami dalam memilih dan memilah data.

C. *Digital Safety*

Digital Safety (keselamatan digital) adalah pemahaman bahwa seseorang harus melindungi diri sendiri dan property digitalnya saat berada dalam lingkungan digital.

Dengan akibatnya, ini dapat digunakan untuk manajemen data digital dan banyak forum online. Selain itu juga bisa terjadi peretasan, penipuan, pencurian, pelanggaran data, dan kejahatan dunia maya lainnya terus meningkat karena perangkat digital menjadi lebih umum dan mampu digunakan untuk menipu seseorang (CFDS, 2020).

D. Cybercrime

Cybercrime (kejahatan dunia maya) adalah salah satu bentuk kejahatan kekerasan yang berkembang sebagai hasil dari pertumbuhan aplikasi internet. Kejahatan dunia maya mencakup semua bentuk kejahatan kekerasan serta metode operasinya, yang digunakan sebagai aplikasi internet negatif. Kejahatan dunia maya didefinisikan sebagai tindakan kekerasan yang dilakukan individu, sekelompok orang, atau organisasi kriminal yang menggunakan komputer atau jenis perangkat komunikasi lainnya, seperti telepon, sebagai reaksi terhadap pertumbuhan aplikasi *online* (EPTIK, 2019)

3. Metodologi Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah metode penelitian kuantitatif. Metode penelitian kuantitatif adalah metode yang digunakan untuk meneliti populasi atau sampel tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistic dengan tujuan menguji hipotesis yang ditetapkan (Sugiyono, 2017:17). Tempat penelitian dilakukan di kampus UIN Ar-Raniry Banda Aceh, Kopelma Darussalam. Waktu penelitian dilakukan pada semester genap tahun 2022/2023.

Populasi adalah keseluruhan elemen (subyek) yang akan diukur dan diteliti, kemudian dijadikan wilayah generalisasi. (Sugiyono, 2019:126). Bagian kecil dari populasi yang akan diteliti disebut sampel, yang memiliki ciri-ciri atau keadaan tertentu yang akan diteliti sehingga diharapkan dapat merepresentasi populasi dan dapat menentukan keabsahan hasil penelitian (Martono, 2016: 269). Penelitian ini memakai metode teknik *simple random sampling*, yaitu pengambilan sampel dari anggota populasi yang dilakukan secara acak tanpa memperhatikan strata yang ada dalam populasi itu.

Teknik pengumpulan data yang digunakan dalam penelitian ini yaitu Kuesioner. Kuesioner adalah Teknik pengumpulan data yang dilakukan dengan cara memberi pertanyaan atau pernyataan tertulis yang akan dibagikan kepada responden. Kuesioner bisa diberikan secara langsung atau melalui internet (Sugiyono, 2019:199)

4. Hasil dan Pembahasan

A. Deskripsi Uji Validitas

Adapun uji validitas pada hasil yang telah dilakukan ini memakai aplikasi IBM SPSS Statistic 25. rtabel yang digunakan ialah 0.195 yang ditentukan dengan melihat jumlah responden yang telah ditentukan yaitu 100 responden. Penelitian ini menggunakan uji validitas dengan Corrected Item-Total Correlation, yaitu tahap korelasi skor tiap item dengan skor totalnya. Item soal bisa dikatakan valid apabila nilai rhitung (Corrected Item-Total Correlation) > rtabel sebesar 0.195 dengan menggunakan taraf signifikan sebesar 5%, dan juga sebaliknya apabila nilai rhitung (Corrected Item Total Correlation) < rtabel maka tidak valid. Hasil uji validitas pada ketiga variabel pada literasi digital, *digital skill*, dan *digital safety* dapat ditinjau dengan tabel sebagai berikut:

Tabel 1. Hasil Uji Validitas Variabel Literasi Digital

No.	Corrected Item – Total Correlation	r tabel 5% (100)	keterangan
-----	------------------------------------	------------------	------------

**ANALISIS PEMAHAMAN LITERASI DIGITAL PADA MAHASISWA UIN ARRANIRY
TERHADAP *DIGITAL SKILL* DAN *DIGITAL SAFETY***

1.	0.439	0.195	Valid
2.	0.450	0.195	Valid
3.	0.617	0.195	Valid
4.	0.641	0.195	Valid
5.	0.647	0.195	Valid
6.	0.737	0.195	Valid
7.	0.348	0.195	Valid
8.	0.274	0.195	Valid
9.	0.607	0.195	Valid
10.	0.354	0.195	Valid

Tabel 2. Hasil Uji Validitas Variabel Digital Skill

No.	Corrected Item – Total Correlation	r tabel 5% (100)	keterangan
1.	0.592	0.195	Valid
2.	0.636	0.195	Valid
3.	0.704	0.195	Valid
4.	0.716	0.195	Valid
5.	0.696	0.195	Valid
6.	0.571	0.195	Valid
7.	0.713	0.195	Valid
8.	0.601	0.195	Valid
9.	0.569	0.195	Valid
10.	0.507	0.195	Valid

Tabel 3. Hasil Uji Validitas Variabel Digital Safety

No.	Corrected Item – Total Correlation	r tabel 5% (100)	keterangan
1.	0.623	0.195	Valid
2.	0.566	0.195	Valid
3.	0.574	0.195	Valid
4.	0.672	0.195	Valid
5.	0.750	0.195	Valid
6.	0.690	0.195	Valid
7.	0.784	0.195	Valid
8.	0.771	0.195	Valid
9.	0.761	0.195	Valid
10.	0.586	0.195	Valid
11.	0.687	0.195	Valid
12.	0.743	0.195	Valid

13.	0.621	0.195	Valid
14.	0.745	0.195	Valid
15.	0.732	0.195	Valid
16.	0.712	0.195	Valid
17.	0.786	0.195	Valid

B. Deskripsi Uji Reliabilitas

Uji reliabilitas diuji setelah dinyatakan pernyataan sudah valid. Pada penelitian ini uji reliabilitas menggunakan Cronbach's Alpha. Variabel dikatakan valid dan reliabel apabila nilai Cronbach's Alpha > dari 0,7. Sebaliknya apabila nilai Cronbach's Alpha < dari 0,7 maka variabel dikatakan tidak valid dan tidak reliabel. Uji reliabilitas pada penelitian ini dapat dilihat pada tabel sebagai berikut:

Tabel 4. Hasil uji variabel reliabilitas

No.	Corrected Item – Total Correlation	r tabel 5% (100)	keterangan
1.	Literasi digital	.797	Reliabel
2.	Digital skill	.889	Reliabel
3.	Digital safety	.946	Reliabel

C. Analisis Deskriptif

Hasil jawaban kuesioner yang telah diselesaikan oleh peneliti menggunakan Microsoft Excel untuk melihat kemampuan mahasiswa dalam memahami literasi digital, dapat diuraikan deskripsi mengenai capaian responden. Hasil analisis untuk variabel literasi digital dapat dilihat pada tabel dibawah ini.

Tabel 5. Hasil Uji Deskriptif Literasi Digital Pernyataan 1-10

Nomor Instrumen	T					T x SL					N	TCR	idx %	Kriteria TCR
	ST	TS	RR	S	SS	1	2	3	4	5				
1	0	0	8	53	39	0	0	24	212	195	100	431	86.2	Baik
2	0	1	0	47	52	0	2	0	188	260	100	450	90	Baik
3	0	0	3	53	44	0	0	9	212	220	100	441	88.2	Baik
4	0	0	3	50	47	0	0	9	200	235	100	444	88.8	Baik
5	0	1	6	57	46	0	2	18	228	230	100	478	95.6	Sangat Baik
6	0	0	6	56	38	0	0	18	224	190	100	432	86.4	Baik
7	4	15	5	45	31	4	45	15	180	155	100	399	79.8	Cukup
8	1	8	13	49	29	1	16	39	196	145	100	397	79.4	Cukup
9	0	1	22	49	22	0	2	66	196	110	100	374	74.8	Cukup
10	0	1	6	49	44	0	2	18	196	220	100	436	87.2	Baik
MEAN												428.2	85.64	Baik

Hasil jawaban kuesioner yang telah diselesaikan oleh peneliti menggunakan Microsoft Excel untuk melihat kemampuan mahasiswa dalam memahami digital skill, dapat diuraikan deskripsi mengenai capaian responden. Hasil analisis untuk variabel

**ANALISIS PEMAHAMAN LITERASI DIGITAL PADA MAHASISWA UIN ARRANIRY
TERHADAP *DIGITAL SKILL* DAN *DIGITAL SAFETY***

digital skill dapat dilihat pada tabel dibawah ini.

Tabel 6. Hasil Uji Deskriptif Digital Skill Pernyataan 11-20

Nomor Instrumen Penelitian (Pernyataan)	T					T x SL					N	TCR	idx	Kriteria TCR
	STS	TS	RR	S	SS	1	2	3	4	5			%	
11	0	2	26	45	27	0	4	78	180	135	100	397	79.4	Cukup
12	0	3	28	42	27	0	6	84	168	135	100	393	78.6	Cukup
13	0	3	21	46	30	0	6	63	184	150	100	403	80.6	Cukup
14	0	3	20	57	26	0	6	60	228	130	100	424	84.8	Baik
15	0	2	22	50	26	0	4	66	200	130	100	400	80	Cukup
16	0	1	5	59	35	0	2	15	236	175	100	428	85.6	Baik
17	0	1	3	57	39	0	2	9	228	195	100	434	86.8	Baik
18	1	1	9	55	34	1	2	27	220	170	100	420	84	Baik
19	0	0	3	64	33	0	0	9	256	165	100	430	86	Baik
20	0	2	7	52	39	0	4	21	208	195	100	428	85.6	Baik
MEAN												415.7	83.14	Baik

Hasil jawaban kuesioner yang telah diselesaikan oleh peneliti menggunakan Microsoft Excel untuk melihat kemampuan mahasiswa dalam memahami digital, dapat diuraikan deskripsi mengenai tingkat capaian responden. Hasil analisis untuk variabel digital safety dapat dilihat pada tabel dibawah ini.

Tabel 7. Hasil Uji Deskriptif Digital Safety Pernyataan 21-37

Nomor Instrumen Penelitian (Pernyataan)	T					T x SL					N	TCR	idx	Kriteria TCR
	STS	TS	RR	S	SS	1	2	3	4	5			%	
21	0	0	3	63	34	0	0	9	252	170	100	431	86.2	Baik
22	0	1	2	54	43	0	2	6	216	215	100	439	87.8	Baik
23	0	1	5	54	40	0	2	15	216	200	100	433	86.6	Baik
24	0	3	24	49	24	0	6	72	196	120	100	390	78	Cukup
25	0	7	42	31	20	0	14	126	124	100	100	364	72.8	Cukup
26	0	0	30	42	28	0	0	90	168	140	100	398	79.6	Cukup
27	0	3	25	52	20	0	6	75	208	100	100	389	77.8	Cukup
28	0	4	34	41	21	0	8	102	164	105	100	379	75.8	Cukup
29	0	6	37	40	17	0	0	111	160	85	100	356	71.2	Cukup
30	2	14	37	32	15	2	28	111	128	75	100	342	68.4	Cukup
31	0	5	21	53	21	0	10	63	212	105	100	390	78	Cukup
32	0	2	16	54	28	0	4	48	216	140	100	408	81.6	Baik
33	0	3	8	56	33	0	6	24	224	165	100	419	83.8	Baik
34	1	2	20	47	30	1	4	60	188	150	100	402	80.4	Baik
35	0	2	22	47	29	0	4	66	188	145	100	403	80.6	Baik
36	0	4	24	51	21	0	8	72	204	105	100	389	77.8	Cukup
37	1	7	28	39	25	1	14	84	156	125	100	379	75.8	Cukup
MEAN												394.76471	78.952941	Cukup

5. Kesimpulan

Berdasarkan hasil dari penelitian dan hasil analisis data dapat disimpulkan sebagai berikut: Pada variabel literasi digital, digital skill, dan digital safety penulis mengukur hasil yang ingin diketahui dengan menggunakan Aplikasi IBM SPSS dan juga Microsoft Excel. Pada hasil dari variabel literasi digital memiliki presentase 85.64% dengan kriteria baik. Dengan demikian dapat dikatakan bahwa pemahaman mahasiswa pada literasi digital memiliki kriteria yang baik. Pada hasil variabel digital skill memiliki presentase 83.14% dengan kriteria baik. Dengan demikian dapat dikatakan bahwa pemahaman mahasiswa pada digital skill memiliki kriteria yang baik. Dan pada hasil variabel digital safety memiliki presentase 78.95% dengan kriteria cukup. Dengan demikian dapat dikatakan bahwa pemahaman mahasiswa pada digital safety memiliki kriteria yang cukup.

Referensi

- [1] Eshet, Y. (2012). Thinking in the digital Era: A Revised Model For Digital Literacy. *Issues in Informing Science and Information Technology*, 9, 267-276.
- [2] CFDS, “Kajian Peningkatan Kompetensi Keamanan Digital di Indonesia: Analisis Fenomena Penipuan dengan Teknik Rekayasa Sosial”, UGM, 2020 “Etika Profesi Teknologi Informasi dan Komunikasi (Cyber Crime dan Cyber Law)”
- [3] Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi Dalam Sistem Elektronik.
- [4] Sugiyono. (2017&2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- [5] Martono, N. (2016). *Metode Penelitian Sosial: Konsep-konsep Kunci*. Jakarta: PT Raja Grafindo Persada.
- [6] Basuki. (2021). *Pengantar Metode Penelitian Kuantitatif*. Bandung: CV Media Sains Indonesia.



Prodi Pendidikan Teknologi Informasi
Gedung B Lantai 1
Fakultas Tarbiyah dan Keguruan (FTK)
UIN Ar-Raniry Banda Aceh

