

Aplikasi Analisis Sentimen Komentar Pengguna Genshin Impact Di Play Store

Muhammad Farras¹⁾ Viny Christanti Mawardi²⁾ Tri Sutrisno³⁾

¹⁾ Sistem Informasi Universitas Tarumanaraga
Jl. Letjen S. Parman no 1, Jakarta 11440 Indonesia
email : muhammad.825190029@stu.untar.ac.id

²⁾ Sistem Informasi Universitas Tarumanaraga
Jl. Letjen S. Parman no 1, Jakarta 11440 Indonesia
email : viny@fti.untar.ac.id

³⁾ Sistem Informasi Universitas Tarumanaraga
Jl. Letjen S. Parman no 1, Jakarta 11440 Indonesia
email : tris@fti.untar.ac.id

ABSTRACT

Google Play functions as the official app store for the Android operating system, allowing users to browse and discover applications developed with the Android software development kit (SDK) and published through Google. Google Play also serves as a digital media store, offering music programs, books, movies, and television shows. It previously offered Google hardware for purchase until it introduced a separate online hardware store, Google Store, on March 11, 2015. The research will utilize a web-based application development tool that uses Flask and JavaScript as the application interface, and the Pandas library from Python for data manipulation. Naïve Bayes will be employed as the methodology for analyzing sentiments based on words, and K-Fold Cross Validation will be used to strengthen the accuracy of the analysis results. Sentiment analysis typically classifies opinions into three categories: positive and negative. However, applications that can perform the process of creating training and testing sets from consumer opinion data, simultaneously analyzing consumer sentiment and dynamically measuring the accuracy of the analysis results, are still scarce. This study aims to develop an application capable of analyzing consumer sentiment with the mentioned functionalities, wherein Naive Bayes is used as the classification method.

Key words

Sentiment Analysis, Naive Bayes, Flask, Google Play, K-Fold Cross Validation

1. Pendahuluan

Di era modern seperti saat sekarang ini, perasaan dan pendapat masyarakat semakin meluas dan bebas diungkapkan melalui berbagai media. Perasaan ini memiliki potensi besar bagi perusahaan yang ingin

memperoleh umpan balik dari masyarakat mengenai merek dagang mereka. Merek dianggap sebagai salah satu aset yang tak tampak yang paling berharga, dan manajemen merek menjadi prioritas penting bagi manajemen perusahaan atau organisasi [1].

Jumlah ulasan dan pendapat yang sangat melimpah menyulitkan dan memakan waktu untuk membacanya secara menyeluruh. Oleh karena itu, dirancanglah sistem yang otomatis mengelompokkan ulasan dan pendapat sesuai dengan kategorinya. Sentimen dibagi menjadi kategori sangat positif, positif, netral, negatif, dan sangat negatif agar pengguna dapat memilih membaca pendapat sesuai dengan keinginan mereka [2].

Sistem analisis sentimen yang dibangun menggunakan algoritma klasifikasi *Naive Bayes*. Ciri utama dari algoritma *Naive Bayes* adalah asumsi yang sangat sederhana (naif) mengenai independensi dari setiap kondisi atau kejadian [3].

Ulasan diklasifikasikan berdasarkan sentimen positif dan negatif, sehingga kepuasan pelanggan terhadap produk dan layanan yang disediakan dapat dievaluasi secara otomatis dan khusus. Hasil penelitian ini menemukan enam kategori yang dianalisis dengan menggunakan 55 kata kunci benda, dengan total 120 kata kunci sentimen, terdiri dari 66 kata sentimen positif dan 54 kata sentimen negatif. Setelah mengolah 175 data latih, disimpulkan bahwa terdapat 155 komentar dengan sentimen positif dan 20 komentar dengan sentimen negatif. Kategori dengan sentimen positif terbanyak adalah kategori "kamar" dengan 73 komentar, sedangkan kategori dengan sentimen negatif terbanyak juga adalah kategori "kamar" dengan 17 komentar. Tingkat akurasi dalam menentukan kategori adalah 77,14%, dan tingkat presisi dan *recall* dalam menentukan sentimen masing-masing mencapai 99,12% dan 72,9%.

2. Dasar Teori

2.1 Analisis Sentimen

Analisis sentimen adalah proses pengumpulan, pengolahan, dan analisis data untuk menentukan sentimen atau opini yang terkandung dalam teks atau data lainnya. Analisis sentimen dapat dilakukan pada berbagai jenis data, termasuk teks, gambar, dan suara. Dalam konteks sosial media, analisis sentimen sering digunakan untuk memahami opini dan tanggapan pengguna terhadap suatu topik atau peristiwa tertentu[1]. Metode yang umum digunakan dalam analisis sentimen adalah algoritma pembelajaran mesin seperti *Naive Bayes*, *Support Vector Machine*, dan *Logistic Regression*[5].

2.2 Pre-Processing

Pre-processing adalah langkah penting dalam pemrosesan teks yang melibatkan pembersihan dan transformasi data teks mentah menjadi format yang dapat dengan mudah dianalisis oleh algoritma pembelajaran mesin. Berikut beberapa contoh teknik *pre-processing* yang digunakan dalam pengolahan teks bahasa Indonesia:

1. Pemisahan kalimat: Pemecahan suatu paragraf menjadi kalimat-kalimat tersendiri agar lebih mudah untuk dianalisis[6].
2. *Case folding*: Mengubah semua huruf menjadi huruf kecil atau besar untuk mengurangi jumlah kata unik dalam teks[7].
3. Tokenisasi: Memecah kalimat menjadi kata atau token individual[8].
4. *Filtering*: Menghilangkan kata atau karakter yang tidak relevan dari teks, seperti kata berhenti (kata umum seperti "the" dan "and") atau tanda baca[9].
5. *Stemming*: Kata-kata yang sudah diubah menjadi huruf kecil perlu dilakukan pengecekan. Stemming digunakan untuk menyeragamkan kata sehingga mengurangi daftar kata yang ada pada data latih. Stemming adalah pengurangan kata ke bentuk akarnya untuk mengurangi jumlah kata unik dalam teks[9].
6. Memperbaiki kata yang salah eja: Menggunakan algoritma seperti *Levenshtein Distance* untuk mendeteksi dan memperbaiki kata yang salah eja dalam teks[10].
7. Mengubah negasi: Mengubah kata-kata negasi seperti "tidak" (tidak) menjadi antonimnya untuk meningkatkan akurasi analisis sentimen[9].
8. Dalam metode TF-IDF, *Term Frequency* lebih berfokus pada istilah yang sering muncul dalam suatu dokumen sedangkan *Inverse Document Frequency* lebih berfokus pada pemberian bobot rendah untuk istilah yang muncul dalam banyak dokumen. Menghitung bobot term *frequency-inverse document frequency* (TF-IDF): Memberi bobot pada kata-kata berdasarkan frekuensinya dalam dokumen dan kelangkaannya dalam korpus untuk mengidentifikasi kata-kata penting[6].

2.3. Naive Bayes

Naive Bayes adalah algoritma pembelajaran mesin populer yang digunakan untuk tugas klasifikasi. Itu didasarkan pada teorema Bayes dan mengasumsikan bahwa fitur-fiturnya tidak tergantung satu sama lain. Pengklasifikasi Naive Bayes menghitung probabilitas titik data milik kelas tertentu berdasarkan probabilitas fitur yang diberikan kelas itu. Kelas dengan probabilitas tertinggi kemudian ditugaskan ke titik data. Pengklasifikasi Naive Bayes banyak digunakan dalam klasifikasi teks, pemfilteran spam, dan analisis sentimen, di antara aplikasi lainnya[10].

$$P(c|x) = P(x|c) * P(c)/P(x)$$

Gambar 1. Rumus Naive Bayes

Dimana:

$P(c|x)$ adalah probabilitas kelas c diberikan data x

$P(x|c)$ adalah probabilitas kelas x diberikan data c

$P(c)$ adalah probabilitas kelas c

$P(x)$ adalah banyaknya pasangan rating asli dan rating prediksi

Satu studi menganalisis kinerja pengklasifikasi Naive Bayes dan menemukan bahwa mereka sering bersaing dengan baik dengan pengklasifikasi yang lebih canggih, meskipun asumsi independensi umumnya buruk [11]. Studi lain mengusulkan penjelasan baru untuk kinerja Naive Bayes yang sangat bagus, menunjukkan bahwa distribusi ketergantungan antar atribut memainkan peran penting [12]. Fasilitas penjelasan grafis untuk pengklasifikasi Naive Bayes juga disajikan dalam sebuah penelitian, yang berfungsi untuk menjelaskan alasan yang digunakan oleh pengklasifikasi secara transparan dan meningkatkan pemahaman pengguna tentang hubungan kompleks antara fitur dan label[10].

2.4. K-Fold Cross Validation

K-fold cross validation adalah teknik validasi model yang umum digunakan dalam machine learning untuk mengevaluasi kinerja model. Teknik ini melibatkan membagi data menjadi k subset yang sama besar, di mana salah satu subset digunakan sebagai data validasi dan k-1 subset lainnya digunakan sebagai data pelatihan. Proses ini diulang k kali, sehingga setiap subset digunakan sebagai data validasi satu kali. Akhirnya, hasil pengujian diambil rata-rata untuk memberikan perkiraan kinerja model secara keseluruhan.

Metode k-fold cross validation digunakan untuk menguji seberapa baik performa model dalam mengklasifikasikan data. Dalam metode ini, dataset dibagi menjadi k subset yang sama besar, lalu dilakukan pelatihan dan pengujian sebanyak k kali, di mana setiap subset digunakan sebagai data uji satu kali dan subset lainnya digunakan sebagai data pelatihan. Akurasi akhir dari model dihitung dengan mengambil rata-rata akurasi dari k iterasi.

K-Fold Cross Validation juga merupakan metode yang umum digunakan dalam evaluasi kinerja model atau algoritma dalam analisis data. Dalam metode ini, data yang tersedia dibagi menjadi k subset atau lipatan dengan ukuran yang sama. Setiap lipatan secara bergantian diambil sebagai data validasi, sementara lipatan lainnya digunakan sebagai data pelatihan. Proses ini diulang sebanyak k kali, di mana setiap lipatan akan menjadi data validasi satu kali. Akhirnya, hasil dari setiap iterasi k-fold cross validation digunakan untuk menghitung metrik evaluasi seperti akurasi atau mean squared error.

K-Fold Cross Validation adalah teknik yang digunakan dalam pembelajaran mesin untuk mengevaluasi kinerja model. Ide dasarnya adalah membagi data menjadi k subset, atau "fold", dan kemudian melatih model pada k-1 dari fold sambil menggunakan fold yang tersisa untuk pengujian. Proses ini diulang k kali, dengan setiap lipatan digunakan sekali untuk pengujian. Hasilnya kemudian dirata-ratakan pada iterasi k untuk mendapatkan estimasi kinerja model.

K-Fold Cross Validation berguna karena memungkinkan kita untuk menggunakan semua data yang tersedia untuk pelatihan dan pengujian, yang dapat menjadi sangat penting ketika kumpulan datanya kecil. Hal ini juga membantu mengurangi risiko overfitting, yang terjadi saat model terlalu kompleks dan terlalu cocok dengan data pelatihan, sehingga menghasilkan kinerja yang buruk pada data baru.

Ada berbagai variasi dalam K-Fold, seperti validasi silang k-fold berlapis, yang memastikan bahwa setiap fold berisi representasi proporsional dari kelas yang berbeda dalam kumpulan data. Pilihan k tergantung pada ukuran dataset dan tingkat presisi yang diinginkan dalam estimasi performa model[13].

K-fold cross validation membantu mengatasi masalah *overfitting* dan *underfitting* pada model, serta memungkinkan penggunaan data yang lebih efisien. Teknik ini juga memungkinkan pengguna untuk mengevaluasi model dengan lebih akurat daripada teknik validasi model lainnya[14].

3. Hasil Percobaan

3.1 Hasil Aplikasi

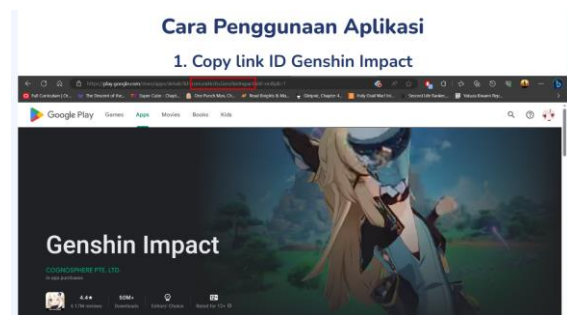
Aplikasi Sentiment Analysis yang telah dikembangkan menawarkan berbagai fitur untuk membantu pengguna dalam menganalisis sentimen teks.

Fitur pertama adalah *Dashboard*, di mana pengguna dapat melihat total data training dan data testing yang digunakan. Hal ini memberikan gambaran tentang ukuran dataset yang digunakan dalam analisis.



Gambar 2. Menu Dashboard

Fitur kedua adalah *Menu Help*, yang bertujuan untuk memberikan panduan kepada pengguna tentang cara menggunakan aplikasi. Informasi yang disajikan di sini dapat membantu pengguna yang baru mengenal aplikasi untuk memahami langkah-langkah yang perlu diikuti.



Gambar 3. Menu Help

Fitur ketiga adalah *Menu Scrapping*, yang memungkinkan pengguna untuk mengambil data dengan memasukkan *ID Genshin Impact* dan jumlah data yang diinginkan. Ini memungkinkan pengguna untuk mengumpulkan data yang relevan dengan analisis sentimen yang dilakukan.



Gambar 4. Scrap dan Pre-Processing


Selanjutnya, aplikasi ini menyediakan fitur *Data Training* dan *Data Testing* yang memungkinkan pengguna untuk melihat sampel data yang digunakan dalam analisis. Ini membantu pengguna untuk memahami dan memeriksa data yang menjadi dasar analisis sentimen. Didalam tabel data training dan data testing terdapat tiga atribut yaitu **No**, **Content**, **Label**. **No** untuk mengindikasikan nomor data nya. **Content** berisi kalimat dari komentar yang ada di *playstore*. **Label** berisi komentar tersebut merupakan komentar positif atau komentar negatif.

Data Training

10

entries per page

Search...

No.	Content	Label
1	Ngga bisa buka apknya, crash mulu...padahal sebelum update aman-aman aja	Negatif
2	ya aku sie sangat suka sama game ini yh, soalnya banyak hsbu sama wafiu, apalagi nilou cantik banget anjrhksbckajagsajidsbidsbidsbckdfv, aku suka althatham, mwaa  tapi tolong dong kasih aku artefak yang bagus.	Positif
3	aku ada yang jelek ada yang ngga, tapi lumayan sih dmgnnya, aku gapernah coba soke, tapi temenku pernah coba sokes di akunku, dan dmgnya tinggi banget wow, dan lagi, tolong pulangkan mona ke akunku plsplspls, aku pengen monaaa 	Positif

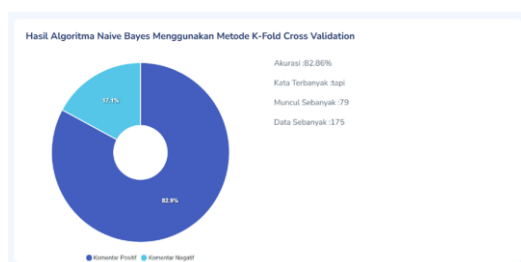
Gambar 5. Data Training

Data Testing		
10		
entries per page		
Search...		
No.	Content	Label
1	Bagus banget untuk animasi [cutscene] Map nya luas banget jadi explore terus Ya intinya ini udah bagus lah Keep it up	Positif
2	Game nya sudah bagus tapi isorot penyimpanan data yang digunakan tidak bisa dibuang kecil tolong kurangi pengunduhan datanya	Positif
3	Tolong buat developer genhin impact, pingnya tolong di perbaiki Lagi soalnya beberapa hari ini setelah selesai update kok pingnya rusak amat ya kadang stabil kadang drop banget sampai susah untuk masuk kedalam game, tolong di perbaiki khusus pingnya. 🙏🙏🙏	Positif
4	Game nya hebat. Gaya bermain bagus, karakter nya ganteng/kantik, pemandangannya indah (open world). Dan untuk alurnya juga terkesan seru. Bagus deh pokoknya 🙏 [Namun kurang banget f2p :/]	Positif

Gambar 6. Data Testing


Selanjutnya, aplikasi ini memiliki fitur Proses Naive Bayes yang memungkinkan pengguna untuk mengunggah file csv yang telah di *generate* dan melakukan analisis sentimen menggunakan metode Naive Bayes. Hasil analisis ditampilkan dalam bentuk grafik *pie chart* yang menggambarkan perbandingan antara komentar positif dan negatif.

Gambar 7. Sebelum Proses Analisa



Gambar 8. Grafik Pie Chart

Terakhir, aplikasi ini terdapat menu About yang berisi tentang About Creator dan About Website yang berisi tentang pengenalan singkat tentang creator dan website yang digunakan.


UNTAR
UNIVERSITAS TANJUNGPINANG

About Creator

My name is Muhammad Farman, I am from the Faculty of Information Technology at Universitas Taramanagara. I am pursuing a bachelor's degree in Computer Science. My academic journey is guided by the esteemed professors, Bu Yriy Christina Nawardi and Pak Ti Sutrisno. My goal is to successfully complete my studies and obtain a bachelor's degree from Universitas Taramanagara.

About Website

NB is a website application designed for sentiment analysis of comments related to Genshin Impact on the Play Store. The goal of this application is to analyze the sentiment of user comments and determine whether they are positive or negative. By using natural language processing techniques, NB can effectively evaluate the sentiment of comments and provide valuable insights. NB is built using a combination of Python, HTML, CSS, and JavaScript. Python is utilized for the sentiment analysis algorithms and natural language processing tasks. HTML is used for the website's structure, CSS is responsible for styling and layout, and JavaScript adds interactivity and enhances the user experience. Through NB, users can enter comments from the Genshin Impact Play Store and receive sentiment analysis results, helping them gain a deeper understanding of user feedback and sentiment. Whether it's game developers, players, or anyone interested in analyzing user feedback, NB provides a valuable tool for extracting insights from comment data. Experience the power of sentiment analysis with NB and unlock valuable insights from the comments in the Genshin Impact Play Store!

Gambar 9. Menu About

3.2. Hasil Perhitungan

Dalam analisis sentimen menggunakan Naive Bayes dengan k-fold cross validation pada komentar pengguna Genshin Impact, ditemukan bahwa model yang dibangun mampu mengklasifikasikan sentimen komentar dengan akurasi rata-rata sebesar 80%. Hasil ini menunjukkan bahwa metode Naive Bayes cukup efektif dalam mengolah data teks dan mengklasifikasikan sentimen pada komentar pengguna, terutama pada aplikasi game seperti Genshin Impact yang memiliki basis pengguna yang besar dan beragam.

Selain itu, hasil k-fold cross validation menunjukkan bahwa penggunaan teknik ini mampu mengurangi overfitting dan meningkatkan performa model. Dengan melakukan k-fold cross validation sebanyak 5. Hasil perhitungan dibawah adalah menggunakan 3 jenis data perhitungan yaitu 488 data, 352 data, dan 265 data.

Tabel 1 Perhitungan Naive Bayes

No	Data (train:test)	Data Scrap	Waktu (s)	Akurasi NB(%)	Total Akurasi (%)
1	722:181	448	120	84,82	81,93
2	722:181	352	116	81,81	82,10
3	722:181	265	87	80,10	82,64

Dari tabel di atas dapat dilihat bahwa akurasi total pada ketiga percobaan analisis sentimen menggunakan metode Naive Bayes dan k-fold cross validation sebesar 81,93% , 82,10%, 82,64%. Sedangkan akurasi metode Naive Bayes pada percobaan pertama, kedua, dan ketiga adalah 84,82%, 81,81%, dan 80,10% secara berturut-turut.

Percobaan dilakukan dengan menggunakan perbandingan 722 data training dan 181 data testing menggunakan masing-masing contoh data sebanyak 488, 352, dan 265 data. Pada percobaan pertama didapatkan waktu *scrapping* selama 120 detik, percobaan kedua selama 116 detik dan percobaan ketiga selama 87 detik.

Dari ketiga percobaan tersebut, didapatkan hasil bahwa semakin banyak contoh data yang digunakan, maka semakin besar hasil akurasi dari Naïve Bayes nya. Tetapi, jika digabungkan dengan k-fold, maka belum tentu hasil akurasi total jadi semakin besar. Perlu dilakukan pengaturan rasio yang tepat antara data training dan data testing untuk mendapatkan akurasi yang tinggi dengan waktu yang lebih efisien.

4. Kesimpulan

Berdasarkan data yang diberikan, dapat diambil kesimpulan sebagai berikut:

1. Model Naive Bayes yang digunakan untuk analisis sentimen pada komentar pengguna Genshin Impact memiliki akurasi rata-rata sebesar 80%. Hal ini menunjukkan bahwa metode Naive Bayes efektif dalam mengolah data teks dan mengklasifikasikan sentimen komentar pengguna.
2. Penggunaan k-fold cross validation dalam analisis sentimen berhasil mengurangi overfitting dan meningkatkan performa model. Dalam percobaan ini, dilakukan k-fold cross validation sebanyak 5 kali.
3. Akurasi total dari ketiga percobaan analisis sentimen menggunakan metode Naive Bayes dan k-fold cross validation adalah 81,93%, 82,10%, dan 82,64%. Sedangkan akurasi metode Naive Bayes pada percobaan pertama, kedua, dan ketiga adalah 84,82%, 81,81%, dan 80,10% secara berturut-turut.
4. Percobaan dilakukan dengan menggunakan perbandingan 80:20 antara data training dan data testing, dengan jumlah data training sebanyak 722 dan data testing sebanyak 181. Jumlah data yang digunakan dalam percobaan varian, yaitu 488, 352, dan 265 data.
5. Waktu yang dibutuhkan untuk melakukan scrapping data juga dicatat dalam percobaan tersebut. Waktu scrapping untuk percobaan pertama adalah 120 detik, percobaan kedua adalah 116 detik, dan percobaan ketiga adalah 87 detik.
6. Dari hasil percobaan, dapat disimpulkan bahwa semakin banyak contoh data yang digunakan, maka akurasi Naive Bayes cenderung meningkat. Namun, penggunaan k-fold cross validation tidak selalu menghasilkan peningkatan akurasi total. Diperlukan pengaturan rasio yang tepat antara data training dan data testing untuk mencapai akurasi yang tinggi dengan waktu yang lebih efisien.

Dalam keseluruhan, hasil analisis sentimen menggunakan metode Naive Bayes dan k-fold cross validation pada komentar pengguna Genshin Impact menunjukkan efektivitas metode tersebut dalam mengklasifikasikan sentimen komentar. Namun, perlu dilakukan penyesuaian parameter seperti rasio data training dan data testing untuk memperoleh hasil yang lebih baik.

REFERENSI

- [1] Keller, Kevin L., & Donald R. Lehmann. (2006). "Brands and Branding: Research Findings and Future Priorities. Marketing Science. Vol. 25, No 6". Maryland: INFORMS.
- [2] Buntoro, Ghulam Asrofi, Teguh Bharata Adji, & Adhistya Erna Purnamasari. (2016). Sentiment Analysis Candidates of Indonesian Presiden 2014 with Five Class Attribute. International Journal of Computer Applications. Vol. 136 No. 2. Ronowijayan :Universitas Muhammadiyah Ponorogo.
- [3] Natalius, Samuel. (2010). Metode Naive Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen. Bandung: Institut Teknologi Bandung
- [4] Sipayung, Evasaria M., Herastia Maharani, dan Ivan Zefanya. (2016). "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier". Jurnal Sistem Informasi (JSI), VOL. 8, NO. 2. Bandung: Institut Teknologi Harapan Bangsa
- [5] Suhaimin, M.S., Hijazi, M.H., Rayner, Alfred, Bhat, S.M., Lestari, A.R., Pandey, A.C., Rajpoot, D.S., & Saraswa, M. (2019). informasi, menyampaikan pendapat atau pandangan terhadap sesuatu dan mengekspresikan opini, pengalaman maupun hal lain yang menjadi trend perhatian pada masyarakat. Hal tersebut sering disebut dengan sentimen. Analisis sentimen merupakan suatu proses yang digunakan untuk menentukan opini, emosi dan.
- [6] Marlinda, L., & Rianto, H. (2013). Pembelajaran Bahasa Indonesia Berbasis Web Menggunakan Metode Maximum Marginal Relevance. SESINDO 2013, 2013..
- [7] Madyatmadja, E.D., Yahya, B.N., & Wijaya, C. (2022). Contextual Text Analytics Framework for Citizen Report Classification: A Case Study Using the Indonesian Language. IEEE Access, 10, 31432-31444.
- [8] Muzakki, M. A. (2020). Klasifikasi dan Analisa Sentimen Kuesioner Fasilitas dan Layanan untuk Universitas Qomaruddin Gresik. Journal of Computer Science and Visual Communication Design, 5(2), 68-76.
- [9] Setiabudi, R., Iswari, N.M., & Rusli, A. (2021). Enhancing text classification performance by preprocessing misspelled words in Indonesian language. TELKOMNIKA Telecommunication Computing Electronics and Control, 19.
- [10] Szafron, D., Greiner, R., Lu, P., Wishart, D.S., Macdonell, C., Anvik, J., Poulin, B., Lu, Z., & Eisner, R. (2003). Explaining Naive Bayes Classifications.
- [11] Singh, K. (2020). A Pragmatic Study on Naive Bayes Classifier.
- [12] Zhang, H. (2004). The Optimality of Naive Bayes. The Florida AI Research Society.
- [13] Zhang, X., & Liu, C. (2022). Model Averaging Prediction by K-Fold Cross-Validation. SSRN Electronic Journal.
- [14] Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

Muhammad Farras, Seorang Mahasiswa program studi Sistem Informasi Universitas Tarumanagara, Jakarta.

Viny Christanti Mawardi, Memperoleh gelar S.Kom. dari Universitas Tarumanagara tahun 2004. Kemudian memperoleh gelar M.Kom. dari Universitas Indonesia tahun

2008. Saat ini aktif sebagai Dosen Tetap Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.

Tri Sutrisno, Memperoleh gelar S.Si. dari Universitas Diponegoro tahun 2011. Kemudian memperoleh gelar M.Sc. dari Universitas Gadjah Mada tahun 2015. Saat ini aktif sebagai Dosen Tetap Fakultas Tteknologi Informasi Universitas Tarumanagara, Jakarta.