ANALISIS SENTIMEN PENERAPAN SISTEM PEMBAYARAN TOL *MULTI LANE FREE FLOW* MENGGUNAKAN *NAÏVE BAYES CLASSIFIER*

Putri Aulia Kusnadi, Tesa Nur Padilah, Betha Nurina Sari
Informatika, Universitas Singaperbangsa Karawang
Jl.HS. Ronggo Waluyo, Teluk Jambe Timur, Karawang, Indonesia
2010631170026@student.unsika.ac.id

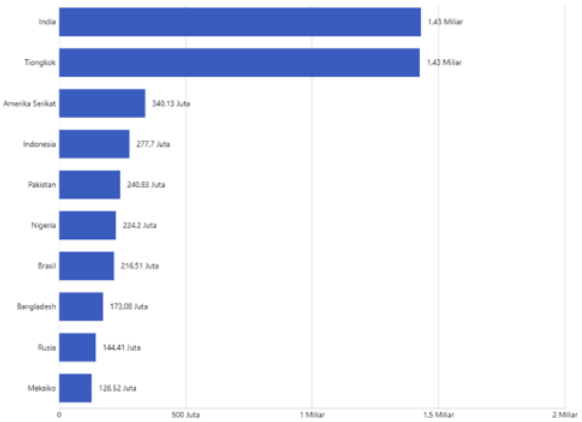
ABSTRAK

Perkembangan ilmu pengetahuan dan teknologi saat ini sedang dikembangkan alat transaksi yang disebut *Multi Lane Free Flow* (MLFF). MLFF merupakan cara pembayaran tol tanpa perlu berhenti di gerbang tol. Uji coba MLFF telah dilakukan di Tol Bali Mandara terdapat video dokumentasi uji coba sistem pembayaran tol MLFF telah disiarkan di berbagai stasiun televisi kemudian tersebar di media sosial dan menjadi topik pembicaraan di kalangan masyarakat karena terlihat beberapa kendaraan terkena palang saat hendak melewati gerbang tol. Penelitian ini bertujuan untuk menganalisis sentimen terhadap penerapan sistem pembayaran tol Multi Lane Free Flow dengan mengimplementasikan algoritma Naïve Bayes. Data tweet dihasilkan melalui teknik *Crawling data* dari media sosial Twitter sebanyak 1153 tweet dan diolah menggunakan metodologi *Knowledge Discovery in Database* (KDD) ke dalam tiga kelas yaitu positif, negatif dan netral. Dataset dilabelkan menggunakan lexicon mendapatkan hasil 194 label positif, 159 label negatif, dan 707 label netral. Kemudian pada tahap *data mining* dibagi menjadi data latih dan data uji dengan empat skenario (90:10, 80:20, 70:30, dan 60:40). Hasil penerapan algoritma *Naïve Bayes* dievaluasi menggunakan *Confusion Matrix* mendapatkan hasil skenario terbaik pada model 90:10 dengan nilai *accuracy* 67%, nilai *precision* 88%, dan nilai *recall* 45%. Analisis sentimen penerapan sistem pembayaran tol *multi lane free flow* telah memberikan gambaran tentang opini publik yang netral terhadap penerapan MLFF Indonesia

Kata kunci : *Analisis Sentimen, Multi Lane Free Flow, Knowledge Discovery in Database, Naïve Bayes Classifier*

1. PENDAHULUAN

Pertumbuhan populasi global pada tahun 2023 mencapai 8,05 miliar jiwa. Indonesia menempati peringkat keempat sebagai negara dengan total penduduk terbesar di dunia, yaitu 277,7 juta jiwa [1]. Hal tersebut dapat dilihat lebih rinci perkembangan penduduk dunia pada Gambar 1.



Gambar 1. Jumlah Penduduk Terbanyak di Dunia

Pertambahan jumlah penduduk sangat erat kaitannya dengan kemajuan sektor transportasi dan peningkatan penggunaan jalan yang tercermin dalam tingkat kemacetan karena padatnya jumlah kendaraan. Oleh karena itu, diperlukan inovasi guna menciptakan lingkungan transportasi yang aman dan nyaman bagi masyarakat [2].

Hingga saat ini, masalah kemacetan lalu lintas di jalan tol Indonesia masih belum teratasi dan menyebabkan waktu tempuh perjalanan semakin lama. Setiap kendaraan memerlukan waktu untuk melakukan pembayaran, yang menyebabkan antrian, dan beberapa pengendara masih mengalami kesulitan saat menge-tap kartu atau kehabisan saldo untuk membayar tol. [3].

Untuk mencapai tujuan tersebut, diperlukan teknologi yang dapat mengatasi permasalahan ini. Salah satu aspek yang mendorong hadirnya e-tol adalah integrasi teknologi dalam pengelolaan perkotaan yang dimungkinkan oleh keberadaan *Internet of Things* (IoT), sebuah jaringan perangkat elektronik. Ini terkait dengan penerapan sistem pembayaran nontunai yang dianggap lebih praktis, cepat, aman, dan efisien. Dalam mengikuti perkembangan ilmu pengetahuan dan teknologi, saat ini sedang dikembangkan alat transaksi yang disebut *Multi Lane Free Flow* (MLFF). MLFF merupakan cara pembayaran tol tanpa perlu berhenti di gerbang tol, namun pelaksanaannya masih terkendala oleh belum adanya pelelangan teknologi yang cocok untuk diterapkan di Indonesia [4].

Dokumentasi uji coba sistem pembayaran tol nirsentuh telah disiarkan di berbagai stasiun televisi berita pada Jumat, 15 Desember 2023. Video tersebut kemudian tersebar di media sosial dan menjadi topik opini di kalangan masyarakat karena terlihat beberapa kendaraan terkena palang saat hendak melewati gerbang tol [5].

Gambar 2. Opini Positif dan Negatif di Twitter

Gambar 2 menampilkan beragam opini yang diunggah oleh masyarakat di platform media sosial Twitter. Beberapa opini positif menyatakan bahwa dengan adanya sistem pembayaran tol MLFF ini menjadi sebuah inovasi pemerintah yang sangat keren dan mempermudah pengguna tanpa harus antre di gerbang tol. Sementara opini negatif menyatakan bahwa penerapan sistem MLFF dengan menggunakan palang pintu dapat menyebabkan kerugian bagi pengguna dengan merusak kendaraan.

Kejadian tersebut menghasilkan respons dari masyarakat di platform Twitter. Twitter adalah salah satu media sosial yang digunakan untuk berbagi opini dan dilengkapi dengan fitur-fitur seperti retweet, pengambilan foto dan video, serta berbagi ke berbagai jaringan sosial lainnya. Twitter sudah menjadi bagian dari kehidupan sehari-hari. Banyak orang memanfaatkan Twitter untuk menyampaikan opini karena aksesnya yang mudah dan jumlah pengikut yang tidak terbatas. [6].

Dalam penelitian ini diterapkan metode *naïve bayes* untuk melakukan klasifikasi. *Naïve bayes* merupakan algoritma klasifikasi yang menjalankan perhitungan probabilitas dengan mengumpulkan dan menggabungkan nilai dari kumpulan data. algoritma naïve bayes mampu memberikan tingkat akurasi yang tinggi serta memproses data dalam jumlah besar [7].

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah proses otomatis untuk memahami, mengekstrak, dan memproses data teks yang tidak terstruktur dengan tujuan mendapatkan informasi yang akurat mengenai opini publik terkait topik tertentu, produk, atau layanan [8].

2.2. *Knowledge Discovery in Database* (KDD)

Knowledge Discovery in Database (KDD) merupakan suatu proses untuk mencari pengetahuan yang berguna dari data dengan menerapkan metode khusus. KDD merujuk pada rangkaian langkah untuk mengekstraksi atau mengidentifikasi pola, pengetahuan, serta informasi dari kumpulan data yang besar [9].

2.3. Algoritma *Naïve Bayes*

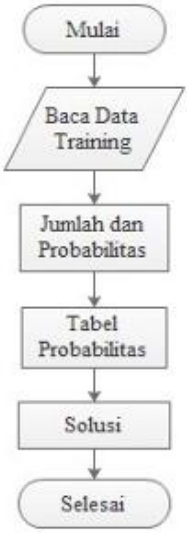
Naïve Bayes Classifier ialah sebuah algoritma dalam bidang *data mining* yang berguna untuk memprediksi probabilitas kelas dan statistik keanggotaan. Fitur utama dari algoritma *Naïve Bayes* adalah kemandiriannya (*naïve*) yang sangat kuat terhadap kondisi atau peristiwa [10]. Algoritma *naïve bayes* ini berdasarkan pada *teorema bayes* yang memeiliki bentu formula sebagai berikut.

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

Keterangan:

a. X, yaitu data dengan kelas yang belum diketahui
b. H, yaitu hipotesis data X yang merupakan suatu kelas yang spesifik
c. P(H|X), yaitu probabilitas hipotesis H berdasarkan dengan kondisi X
d. P(X|H), yaitu probabilitas hipotesis X berdasarkan dengan kondisi H
e. P(H) , yaitu probabilitas hipotesis H
f. P(X), yaitu probabilitas hipotesis X

Terdapat sebuah *flowchart* yang menggambarkan proses *Naive Bayes Classifier*.



Gambar 3. *Flowchart Naive Bayes Classifier*

Langkah-langkah algoritma *Naïve Bayes* adalah sebagai berikut:

a. Identifikasi atau membaca data training
b. Menghitung jumlah dan probabilitas kemunculan setiap atribut
c. Buat tabel probabilitas kemunculan setiap atribut dari semua kriteria yang ada
d. Menghitung nilai likehood ya dan likehood tidak yang diperoleh dari tabel probabilitas kemunculan masing-masing atribut.

2.4. *Confusion Matrix*

Confusion Matrix adalah suatu tabel yang memuat empat kombinasi nilai prediksi dan nilai aktual. Dalam *Confusion Matrix*, ada empat istilah yang mempresentasikan hasil dari proses klasifikasi. Istilah-istilah tersebut meliputi *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [11].

Table 1. *Confusin Matrix*

Kelas	Positif	Negatif
Prediksi Positif	TP	FN
Prediksi Negatif	FP	TN

Keterangan:

a. TP (*True Positive*), yaitu jumlah data bernilai positif dan prediksi positif
b. TN (*True Negative*), yaitu jumlah data berjumlah negatif tapi diprediksi negatif
c. FP (*False Positive*), yaitu jumlah data bernilai negatif tapi diprediksi data positif
d. FN (False Negative), yaitu jumlah data bernilai positif dan diprediksi data negatif

Pengujian dengan menggunakan *confusion matrix* menghasilkan perhitungan yang menghasilkan empat keluaran, yaitu:

a. Akurasi, digunakan untuk mengukur ketepatan suatu model dalam melakukan klasifikasi data secara keseluruhan, berikut rumus menghitung nilai akurasi.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

b. Presisi, digunakan untuk mengukur banyaknya prediksi positif yang benar terhadap total prediksi positif, serta meminimalkan *false negative.* Berikut rumus menghitung nilai presisi.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

c. *Recall*, digunakan untuk mengukur banyaknya prediksi positif yang benar terhadap total data aktual yang positif, serta meminimalkan *false negative*, berikut rumus menghitung nilai *recall.*

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

d. *F1 Score*, adalah harmonic mean dari presisi dan *recall* yang berguna untuk mengevaluasi model dnegan preisisi dan *recall* yang tinggi, berikut rumus menghitung nilai *F1 score.*

$$F\text{-}Measure = 2\ x\ \frac{Recall \times Precision}{Recall+Precision} \quad (5)$$

Keterangan :
TP : Jumlah prediksi *true pisitive*
TN : Jumlah prediksi *true negative*
FP : Jumlah prediksi *false positive*
FN : Jumlah pediksi *false negative*

2.5. *Text Mining*

Text mining yaitu jenis *data mining* yang digunakan untuk mengekstrak informasi bermanfaat dengan menemukan dan mempelajari pola yang menarik dari sekumpulan sumber data tekstual yang tidak terstruktur. Secara umum *text mining* terdiri dari beberapa tahapan yaitu *cleaning, case folding, tokenizing, stopword removal,* dan *stemming* [12].

2.6. TF-IDF

TF-IDF adalah sebuah teknik untuk menilai atau mengukur kepentingan setiap kata dalam sebuah kalimat di dalam dokumen tertentu. Seberapa sering suatu kata muncul dalam satu dokumen dapat dihitung dengan menggunakan *term frekuensi* (tf). Jika TF tinggi, itu berarti kata tersebut sering muncul [13].

2.7. Python

Bahasa pemrograman Python merupakan *interpreted programming language* yang interaktif dan berorientasi objek. Python adalah program interaktif dan berorientasi objek yang menggabungkan model, pengetikan dinamis, *exceptions*, tipe data dinamis tingkat tinggi, dan kelas. Python juga mendukung beberapa paradigma pemrograman selain pemrograman berorientasi objek, seperti pemrograman prosedural atau fungsional [14].

2.8. Twitter

Twitter adalah salah satu platform media sosial yang banyak digunakan oleh masyarakat Indonesia. Kepopulerannya membuat Twitter menjadi tempat bagi publik untuk mengekspresikan sentimen mereka mengenai isu-isu yang sedang dibicarakan. Di media sosial ini, masyarakat membahas berbagai hal yang sedang terjadi, mulai dari teknologi, pendidikan, ekonomi, politik, sosial budaya, hingga isu-isu viral di tengah masyarakat, termasuk mengenai MLFF [15].

2.9. Metode *Lexicon*

Metode *lexicon based* adalah metode klasifikasi yang menggunakan kamus kata-kata opini. Kamus *lexicon* yang akan digunakan adalah positif dan negatif *words*. Istilah "kamus" juga dapat mengacu pada daftar kosakata. Kata-kata ini biasanya terdiri dari kata sifat, yaitu kata kerja yang digunakan untuk menunjukkan kalimat opini, seperti bagus, jelek, indah, mudah, sulit, dan sebagainya. Setiap kata dalam kamus memiliki skor polaritas yang berkisar antara -1 dan +1, menunjukkan kelas negatif atau positif [16].

2.10. *Word Cloud*

Word cloud adalah metode analisis teks yang menampilkan grafik frekuensi kata untuk membantu mengidentifikasi kata-kata yang sering muncul dalam teks sumber. Semakin besar ukuran kata dalam gambar, semakin sering kata tersebut muncul dalam data penelitian [17].

2.11. Sistem Pembayaran Tol

Ada serangkaian perubahan dan evolusi yang telah terjadi berkaitan dengan transaksi jalan tol di Indonesia, yaitu [4]:

a. Pembayaran Tunai (*Cash*)
   Metode pembayaran tunai (*cash*) merupakan cara pembayaran di gerbang tol yang menggunakan uang kontan. Pengenalan metode ini dimulai pada tahun 1987 ketika jalan tol pertama kali diperkenalkan di Indonesia. Penggunaan metode ini dapat memakan waktu yang cukup lama, terutama jika pengguna jalan tidak memiliki uang dengan jumlah yang pas.
b. Kartu Elektronik (*Electronic Card*)
   Sejak dari tanggal 31 Oktober 2017, pengguna kartu *e-money* menjadi wajib di Indonesia untuk melakukan pembayaran tol. Mereka hanya perlu

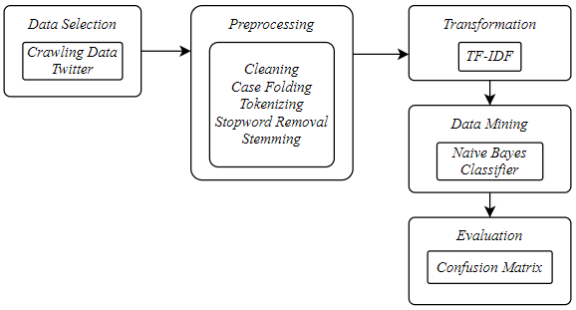
menggunakan kartu elektronik yang sudah diisi saldo untuk masuk ke jalan tol.

c. *Single Lane Free Flow* (SLFF)
Single Lane Free Flow (SLFF) adalah sistem pembayaran di mana kendaraan tidak perlu berhenti di setiap jalur transaksi untuk melakukan transaksi dengan tapping. Namun, hanya satu jalur yang menggunakan sistem *free flow.*

d. *Multi Lane Free Flow* (MLFF)
Multi Lane Free Flow (MLFF) adalah suatu teknologi terbaru dalam pembayaran tol yang memungkinkan kendaraan untuk melewati gerbang tol tanpa menghentikan atau melambatkan laju kendaraan. Proses pembayaran menggunakan sistem MLFF dirancang untuk mendeteksi kendaraan secara otomatis dan melakukan pembayaran tanpa memerlukan kendaraan untuk berhenti. Tujuannya adalah untuk meningkatkan efektivitas dan efisiensi, serta mengurangi kemacetan lalu lintas di jalan tol, meningkatkan kelancaran arus lalu lintas, memberikan dampak positif bagi lingkungan, dan memberikan kenyamanan lebih bagi pengemudi.

3. METODE PENELITIAN

Penelitian ini akan dilakukan sesuai dengan tahapan metode KDD. Gambar dibawah ini merupakan rincian dari rancangan penelitian yang akan dilaksanakan.



Gambar 4. Alur Penelitian

3.1. *Data Selection*

Penelitian ini menggunakan teknik crawling pada Twitter dan bahasa pemrograman python di google colaboratory untuk mengumpulkan informasi yang relevan mengenai penerapan sistem pembayaran tol *multi lane free flow.* Data yang telah terkumpul selanjutnya akan diproses dengan menghapus kolom yang tidak diperlukan dan hanya menggunakan kolom yang relevan.

3.2. *Preprocessing*

preprocesing dilakukan untuk memperbaiki penulisan dalam data dan membersihkan data melalui 5 tahapan yaitu *cleaning, case folding, tokenizing, stopword removal,* dan *stemming*

3.3. *Transformation*

Dalam melakukan *transformation data,* metode *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk membantu proses ini. TF-IDF akan menghitung bobot untuk setiap kata dengan menghasilkan skor frekuensi. Kata-kata yang dianggap penting adalah kata-kata yang sering muncul dalam data tweet dan memiliki tingkat frekuensi yang tinggi.

3.4. *Data Mining*

Pada tahap ini merupakan proses mengubah data awal menjadi informasi yang bermanfaat. Proses ini melibatkan eksplorasi dan analisis untuk mengidentifikasi pola yang berguna dan informatif. Dalam klasifikasi sentimen, metode digunakan untuk mengelompokkan kalimat menjadi kelas positif, negatif, dan netral. Algoritma *Naïve Bayes* diterapkan dengan merujuk pada Teorema Bayes.

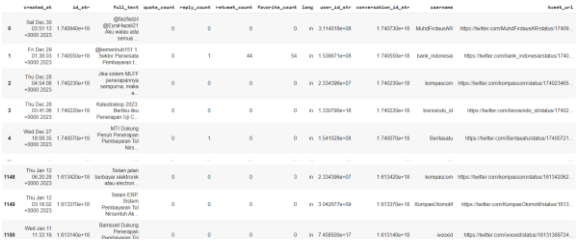
3.5. *Evaluation*

Tahap *evaluation* dilakukan untuk menilai kinerja algoritma yang digunakan. Pada tahap ini, *Confusion Matrix* diterapkan untuk memvisualisasikan hasil prediksi dari model yang sudah dibentuk dan kondisi sebenarnya dari algoritma yang diaplikasikan. Hasil evaluasi yang diperoleh yaitu berupa nilai *accuracy*, *precision, recall*, dan *f-measure*. Dari nilai-nilai ini, akan dievaluasi seberapa baik dan tepat klasifikasi yang telah dilakukan.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Data selection merupakan tahap awal dari penelitian. Pengambilan dataset dilakukan dengan teknik *crawling data* melalui Twitter harvest menggunakan kata kunci ‘mlff’, ‘tol nirsentuh, dan ‘mlff indonesia’. Data yang diperoleh ialah cuitan dengan rentang waktu 1 Januari sampai 31 Desember 2023. Peneliti mengambil data dengan batas jumlah rekord 1000 data dan kemudian menghasilkan 1153 data awal.



Gambar 5. Hasil Crawling Data

4.2. Preprocessing

Hasil *crawling data* yang telah ditunjukkan pada proses pengumpulan dataset, jenis dataset yang seharusnya digunakan untuk proses *data mining* masih belum ideal untuk dilanjutkan ke tahap selanjutnya. Oleh karena itu, diharapkan bahwa tahan

preprocessing akan menghilangkan kata atau karakter yang tidak diperlukan.

a. *Cleaning*

Proses *cleaning* dilakukan dengan menghilangkan emotikon, hashtag, URL, nama pengguna, tanda baca atau simbol lainnya yang tidak dapat diklasifikasikan. Hasil penerapan tahap *cleaning*, dapat diketahui pada Tabel 2.

Tabel 2. Hasil *Cleaning*

Sebelum	Sesudah
Uji coba sistem bayar tol tanpa berhenti atau Multi Lane Free Flow (MLFF) di Tol Bali-Mandara disebut gagal. #bisnisupdate #update #bisnis #text https://t.co/OIGXPBXrYb	Uji coba sistem bayar tol tanpa berhenti atau Multi Lane Free Flow MLFF di Tol Bali-Mandara disebut gagal.

b. *Case Folding*

Tahap *case folding* mengubah semua huruf kapital dalam kumpulan data menjadi huruf kecil.

Tabel 3. Hasil *Case Folding*

Sebelum	Sesudah
Uji coba sistem bayar tol tanpa berhenti atau Multi Lane Free Flow MLFF di Tol Bali-Mandara disebut gagal	uji coba sistem bayar tol tanpa berhenti atau multi lane free flow mlff di tol bali-mandara disebut gagal.

c. *Tokenizing*

Tahap *tokenizing* ialah proses memecah kalimat dalam data komentar menjadi kata-kata individual. Proses ini memudahkan tahap transformasi selanjutnya, karena yang diproses adalah kata per kata dari kalimat tersebut, bukan keseluruhan kalimat.

Tabel 4. Hasil *Tokenizing*

Sebelum	Sesudah
uji coba sistem bayar tol tanpa berhenti atau multi lane free flow mlff di tol bali-mandara disebut gagal.	['uji', 'coba', 'sistem', 'bayar', 'tol', 'tanpa', 'berhenti', 'atau', 'multi', 'lane', 'free', 'flow', 'mlff', 'di', 'tol', 'bali', 'mandara', 'disebut', 'gagal']

d. *Stopword Removal*

Kemudian akan dilakukan tahap *stopword removal*. Dalam proses ini, hendak digunakan dokumen yang berisi kata-kata *stopword* untuk menyaring kata-kata penghubung dan kata-kata yang tidak relevan. Contoh kata penghubung adalah "dari", "ke", "yang", "di", "nya", dan sebagainya.

Tabel 5. Hasil *Stopword Removal*

Sebelum	Sesudah
['uji', 'coba', 'sistem', 'bayar', 'tol', 'tanpa', 'berhenti', 'atau', 'multi', 'lane', 'free', 'flow', 'mlff', 'di', 'tol', 'bali', 'mandara', 'disebut', 'gagal']	['uji', 'coba', 'sistem', 'bayar', 'tol', 'berhenti', 'multi', 'lane', 'free', 'flow', 'mlff', 'tol', 'bali', 'mandara', 'gagal']

e. *Stemming*

Tahap terakhir dari *preprocessing* adalah *stemming*, yang melibatkan penghapusan imbuhan pada awal, akhir, atau kombinasi keduanya dalam sebuah kata.

Tabel 6. Hasil *Stemming*

Sebelum	Sesudah
['uji', 'coba', 'sistem', 'bayar', 'tol', 'berhenti', 'multi', 'lane', 'free', 'flow', 'mlff', 'tol', 'bali', 'mandara', 'gagal']	uji coba sistem bayar tol henti multi lane free flow mlff tol bali-mandara gagal.

4.3. *Transformation Data*

Tahap Pelabelan akan dilakukan untuk memproses data yang telah dikumpulkan. Pelabelan data dilakukan dengan metode *lexicon*, yang menggunakan kamus kata=kata positif dan negatif. Sampel hasil pelabelan data dapat dilihat pada Gamabr berikut ini.

	clean_text	polarity	label
0	rfid smart tag amp tng lorong tng kosong prefe...	-2	negatif
1	sektor pariwisata bayar tiket wisata inap amp ...	1	positif
2	sistem mlff terap sempurna efektif kurang mace...	1	positif
3	kaleidoskop berlikuliku terap uji coba bayar t...	0	netral
4	mti dukung penuh terap bayar tol nirsentuh mlff	0	netral

Gambar 6. Hasil Pelabelan

Hasil pelabelan data mendapatkan kelas positif berjumlah 194 data, kelas negatif 159 data dan kelas netral 707 data. Selanjutnya, *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk mengubah dataset berupa data teks menjadi data numerik.

	abah	abbas	abdul	abu	abuabu	action	acu	ada	adab	adakan	...	yg	ynr	you	youtube	ytl	yuk
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1055	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
1056	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
1057	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
1058	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
1059	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0

1060 rows × 2199 columns

Gambar 7. Hasil TF-IDF

Gambar 6. Menunjukkan bahwa hasil dari proses TF-IDF adalah matriks 1000 x 2199 dengan 2199 *term* atau fitur dan seribu dokumen. Pada tahap *data mining*, nilai pembobotan terhadap kemunculan kata dalam setiap ulasan kemudian digunakan untuk membantu proses klasifikasi.

4.4. *Data Mining*

Data mining dimulai setelah tahap *transformation data* selesai. Proses pengolahan data menggunakan model *Multinomial Naïve Bayes* yang dilakukan dengan memanfaatkan *library sklearn.naive_bayes* yang dapat diakses dalam python.

Proses penelitian ini menggunakan algortima *Naïve Bayes* untuk memproses data dengan melakukan *split data* dalam empat skenario yang terdiri dari 90:10, 80:20, 70:30 dan 60:40. pembagian data training dan data testing bisa dilihat pada Tabel 7.

Tabel 7. *Split Data*

Presentasi Data		Jumlah Data	
Training	**Testing**	**Training**	**Testing**
90%	10%	954	106
80%	20%	848	212
70%	30%	742	318
60%	40%	636	424
Total		**1060**	

Pada tabel 7 merupakan pembagian dataset berdasarkan skenario yang telah dibuat. Pembagian data tersebut akan dievaluasi menggunakan *confusion matrix*. *Confusion matrix* akan menghasilkan nilai akurasi, presisi, *recall*, dan *f1-score.*.

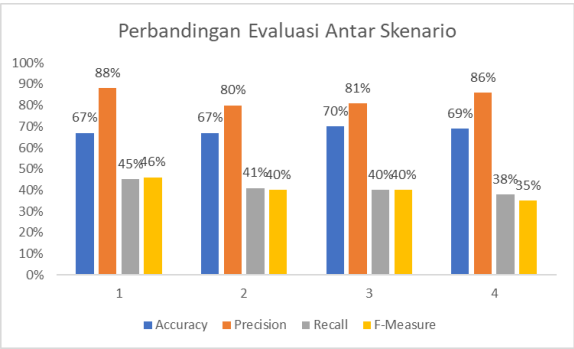
4.5. *Evaluation*

Hasil dari tahap ini berupa perhitungan evaluasi menggunakan *confusion matrix* untuk setiap model skenario yang diuji mencakup nilai *accuracy*, *precision*, *recall*, dan *f-measure*. Berikut hasil perhitungan evaluasi setiap model skenario yang diuji.

Tabel 8. Hasil Evaluasi pada Setiap Model

Model	*Accuracy*	*Precision*	*Recall*	*F-Measure*
90:10	67%	88%	45%	46%
80:20	67%	80%	41%	40%
70:30	70%	81%	40%	40%
60:40	69%	86%	38%	35%

Pada Tabel 8 merupakan hasil dari seluruh skenario melihat *confusion matrix* dengan pemodelan menggunakan algoritma *Naïve Bayes Classifier*. Bisa dilihat dari tabel diatas menjelaskan bahwa hasil pengujian *confusion matrix* dilakukan sebanyak 4 kali, kemudian didapatkan hasil akurasi tertinggi yaitu 70% pada percobaan ke-3, presisi tertinggi yaitu 88% pada percobaan ke-1, *recall* tertinggi yaitu 45% pada percobaan ke-1, dan *f-measure* tertinggi yaitu 46% pada percobaan ke-1. Hasil evaluasi juga ditampilkan dalam bentuk grafik perbandingan antar skenario.

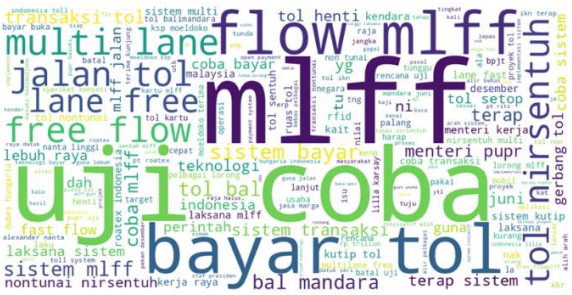


Gambar 8. Perbandingan Hasil *Evaluation*

Dalam gambar 8 menampilkan grafik persentase hasil evaluasi antar skenario yang diuji, dimana dalam semua pengujian yang dilakukan pada semua skenario menghasilkan persentase nilai yang berbeda. Semua skenario mempunyai perbedaan persentase yang tidak lerlalu banyak, hanya perbedaan nilai yang kecil saja yang membedakan.

Berdasarkan perbandingan hasil pengujian evaluasi, skenario terbaik untuk model klasifikasi *Naïve Bayes* adalah skenario pertama (90% data training dan 10% data testing) dengan nilai *accuracy* sebesar 67%, nilai *precision* sebesar 88%, nilai *recall* sebesar 45%, dan nilai *f-measure* sebesar 46%. Model ini dinilai baik karena memiliki persentase nilai tertinggi pada empat pengujian, yaitu *accuracy* walau tidak mempunyai nilai tertinggi tetapi nilai *precision*, *recall* dan f-*measure* mempunyai persentase yang baik dibandingkan skenario lainnya.

Hasil klasifikasi analisis sentimen penerapan sistem pembayaran tol *multi lane free flow* menggunakan *naïve bayes classifier* dapat divisualisasikan dengan *word cloud* untuk mengetahui gambaran atau informasi umum mengenai cuitan tentang MLFF di twitter.



Gambar 9. Hasil *Wordcloud*

Gambar 9 menujukkan *Wordcloud* yaitu kata yang sering digunakan pada dataset. Kata yang sering muncul yakni kata "uji coba", "mlff", "bayar tol", "flow", "tol nirsentuh". Semakin besar ukuran kata di *wordcloud*, maka semakin tinggi pula pengulangan kata tersebut, artinya kata tersebut sering digunakan oleh individu untuk mengirim tweet di Twitter.

5. KESIMPULAN DAN SARAN

Analisis sentimen terhadap penerapan sistem pembayaran tol *multi lane free flow* (MLFF) menggunakan total dataset sebanyak 1060 data menunjukkan bahwa terdapat 194 opini positif, 159 opini negatif, dan 707 opini netral. Hasil ini menunjukkan bahwa sentimen terhadap penerapan sistem pembayaran tol MLFF didominasi oleh sentimen netral. Analisis ini dilakukan dengan menggunakan metode lexicon pada kamus kata positif dan negatif. Kata yang sering muncul dalam opini masyarakat di Twitter tentang MLFF adalah "uji coba", "MLFF", "bayar tol", dan "flow". Dari sini, dapat disimpulkan bahwa mayoritas masyarakat tidak menunjukkan ekspresi yang kuat dalam mendukung

atau menentang penerapan sistem pembayaran tol *multi lane free flow*.

Evaluasi dengan algoritma *Naïve Bayes Classifier* dalam mengklasifikasikan sentimen opini masyarakat terhadap MLFF dilakukan menggunakan confusion matrix dengan empat skenario percobaan yaitu 90:10, 80:20, 70:30, dan 60:40. Hasil terbaik diperoleh pada skenario pertama (90:10) dengan nilai akurasi 67%, presisi 88%, *recal*l 45%, dan *f-measure* 46%. Untuk penelitian selanjutnya, disarankan untuk menggunakan algoritma klasifikasi lainnya seperti *Support Vector Machine* atau *K-nearest Neighbor* untuk menemukan nilai evaluasi yang lebih baik. Selain itu, penggunaan dataset yang lebih besar dan penambahan fitur seleksi seperti *chi-square* dapat meningkatkan akurasi analisis.

DAFTAR PUSTAKA

[1] C. M. Annur, "10 Negara dengan Jumlah Penduduk Terbanyak di Dunia Pertengahan 2023". Databoks, 28 Juli 2023, [Online]. Tersedia: https://databoks.katadata.co.id/datapublish/2023/07/28/10-negara-dengan-jumlah-penduduk-terbanyak-di-dunia-pertengahan-2023

[2] E. P. Astutik and G. Gunartin, "Analisis Kota Jakarta Sebagai Smart City dan Penggunaan Teknologi Informasi dan Komunikasi Menuju Masyarakat Madani," *Inovasi: Jurnal Ilmiah Ilmu Manajemen*, vol. 6, no. 2, pp. 41-58, Desember 2019.

[3] N. Rahmadhiratri, F. Fuad, and Suartini, "Studi Perbandingan Hukum Sistem Transaksi Tol Nontunai Niersentuh Antara Negara Taiwan dan Turki Dengan di Indonesia," *Jurnal Bedah Hukum*, vol. 7, no. 2, pp. 170–185, 2023

[4] A. Budiharjo and S. R. Margarani, "Kajian Penerapan Multi Lane Fee Flow (Mlff) di Jalan Tol Indonesia," *Jurnal Keselamatan Transportasi Jalan (Indonesian Journal of Road Safety)*, vol. 6, no. 2, pp. 1–14, 2019

[5] A. Asmaaysi, "Uji coba dinilai gagal, proyek mlff rp 4,6 triliun di ujung tanduk", Ekonomi Bisnis, 19 Desember 2023, [Online]. Tersedia: https://ekonomi.bisnis.com/read/20231219/45/1725086/uji-coba-dinilai-gagal-proyek-mlff-rp46-triliun-di-ujung-tanduk

[6] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indoensia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *Jurnal Riset Komputer*, vol. 10, no. 2, pp. 479-487, April 2023.

[7] R. A. Nugroho and I. Cholissodin, "Implementasi Naïve Bayes Classifier Untuk Klasifikasi Emosi Tweet Berbahasa Indonesia pada Spark," *Jurnal Pengembangan Teknolohi Informasi dan Ilmu Komputer*, vol. 5, no. 1, pp. 301–310, Januari 2021.

[8] S. S. Salim and J. Mayary, "Analisis Sentimen Pengguna Twitter Terhadap Dompet Elektronik Dengan Metode Lexicon Based dan K-Nearest Neighbor," *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 1, pp. 1–17, Januari 2020.

[9] E. Bulolo, *Data Mining untuk Perguruan Tinggi.* Yogyakarta: Deepublish, 2020

[10] M. A. Permana & S. Widiastuti, "Analisis Sentimen Penggunaan Aplikasi Video Conference Pada Ulasan Google Play Store Menggunakan Metode Nbc (Naive Bayes Classifier)," *Jurnal Riset Sistem Informasi dan Teknologi Informasi (JURSISTEKNI)*, vol. 5, no. 1, pp. 178-191, Januari 2023.

[11] E. Fauziningrum and E. I. Suryaningsih, "Evaluasi dan prediksi penguasaan bahasa inggris maritim menggunakan metode decision tree dan confusion matrix (studi kasus di universitas maritim amni)," *Prosiding Kemaritiman*, 217-231, 2021.

[12] S. W. Handani, D. I. S. Saputra, Hasirun, and R. M. Arino, "Sentiment analysis for go-jek on google play store," *Journal of Physics: Conference Series*, vol 1196, no. 1, Maret 2019.

[13] S. Fide, S. Supardi, and S. Sudarno "Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (Svm) Dan Asosiasi," *Jurnal Gaussian*, vol. 10, no. 3, pp. 346–358, Desember 2021

[14] Python.org, "General Python Faq," Python, 4 April 2022, [Online]. Tersedia: https://docs.python.org/3/faq/general.html#what-is-python

[15] T. N. Wijaya, R. Indriati, and M. N. Muzaki, "Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter," *Jambura J. Electr. Electron. Eng.*, vol. 3, no. 2, pp. 78–83, 2021, doi: 10.37905/jjeee.v3i2.10885.

[16] Y. Fauziah, B. Yuwono, dan A. S. Aribowo, "Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review," *RSF Conference Series: Engineering and Technology*, vol. 1, no. 1, hlm. 363–367, Des 2021, doi: 10.31098/cset.v1i1.397.

[17] H. Andriana, S. S. Hilab, and A. Hananto, "Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter terhadap KTT G20 Indoensia," *Jurnal Riset Komputer.*, vol. 10, no. 1, pp. 60–67, 2019, doi: 10.30865/jurikom.v10i1.5427.