

Analisis Sentimen Maxim dengan Perbandingan Chi Square dan MI pada Naive Bayes

Dwinda Putri Nuria¹, Ultach Enri², Yuyun Umaidah³

^{1,2,3} Program Studi Informatika, Universitas Singaperbangsa Karawang

e-mail: dwinda.putri19073@student.unsika.ac.id

Abstrak

Pesatnya perkembangan teknologi digital berdampak besar bagi manusia, terutama dengan meningkatnya popularitas layanan transportasi *online*. Saat ini penggunaan transportasi *online* masih didominasi oleh Gojek kemudian diikuti oleh Grab dan Maxim. Maxim adalah perusahaan transportasi *online* di Indonesia yang menawarkan berbagai layanan melalui aplikasinya. Meskipun menawarkan tarif yang lebih murah daripada Gojek dan Grab, persentase penggunaan Maxim masih tertinggal. Untuk mengetahui faktor penyebabnya, dilakukan analisis sentimen dari ulasan pengguna aplikasi Maxim di Google Play. Analisis sentimen ini dilakukan dengan menggunakan metodologi KDD dengan tahapan berikut: *data selection*, *preprocessing*, *transformation*, *data mining* dan *evaluation*. Pada proses penelitian digunakan algoritme *Naive Bayes* dengan seleksi fitur *Chi Square* dan *Mutual Information* untuk mengoptimalkan pengklasifikasian. Data yang digunakan merupakan data ulasan dari Google Play sebanyak 1820 data yang terdiri dari 961 data positif dan 859 data negatif. Hasil klasifikasi menggunakan algoritme *Naive Bayes* dengan seleksi fitur *Chi Square* menghasilkan tingkat akurasi terbesar yaitu 96,97%, *precision* 97%, *recall* 97%, *f1-score* 97% yang menghasilkan prediksi 978 data positif dan 842 data negatif.

Kata kunci: *Analisis Sentimen, Chi Square, Maxim, Mutual Information, Naive Bayes*

Abstract

The rapid advancement of digital technology has brought significant impacts on humanity, especially with the increasing popularity of online transportation services. Currently, the usage of online transportation is still dominated by Gojek, followed by Grab and Maxim. Maxim is an online transportation company in Indonesia that offers various services through its application. Despite providing more affordable rates compared to Gojek and Grab, Maxim's usage percentage still lags behind. To identify the contributing factors, sentiment analysis is conducted on user reviews of the Maxim application on Google Play. The sentiment analysis is carried out using the KDD methodology with the following steps: *data selection*, *preprocessing*, *transformation*, *data mining*, and *evaluation*. In the research process, the Naive Bayes algorithm with Chi Square and Mutual Information feature selection is used to optimize classification. The data used consisted of 1820 reviews from Google

Play, comprising 961 positive and 859 negative reviews. The classification results using the Naive Bayes algorithm with Chi Square feature selection yielded the highest accuracy rate of 96.97%, with a precision of 97%, recall of 97%, and an f1-score of 97%, the model predicted 978 positive reviews and 842 negative reviews.

Keywords : *Sentiment Analysis, Chi Square, Maxim, Mutual Information, Naïve Bayes*

PENDAHULUAN

Pesatnya perkembangan teknologi digital saat ini berdampak pada peningkatan popularitas layanan transportasi *online*. Aplikasi transportasi online tidak hanya memberikan layanan transportasi, tetapi telah berkembang menjadi *mobile commerce* yang menyediakan berbagai jasa lainnya, seperti memesan makanan, membeli tiket, mengirim barang, berbelanja, dan masih banyak layanan lainnya yang membantu masyarakat dalam mengefisienkan waktu (Kevin et al., 2020).

Di Indonesia, tepatnya di Kawasan Megapolitan seperti Jabodetabek (Jakarta, Bogor, Depok, Tangerang, dan Bekasi) dengan kepadatan penduduk dan aktivitas ekonomi yang tinggi membuat transportasi *online* sangat dibutuhkan karena dapat memberikan kemudahan akan mobilitas yang tinggi. Dari hasil survei, Badan Penelitian dan Pengembangan Kementerian Perhubungan menyatakan bahwa penggunaan transportasi *online* di Jabodetabek didominasi oleh Gojek dengan persentase sebesar 59,13%, Grab 32,24%, Maxim 6,93% dan 1,70% lainnya.

Maxim merupakan layanan transportasi online yang mulai beroperasi di Indonesia pada bulan Juli 2018 di bawah naungan PT. Teknologi Perdana Indonesia. Seiring berjalannya waktu, layanan Maxim semakin berkembang dan sudah tersedia lebih dari 100 kota di Indonesia (Maxim, 2023).

Dalam beberapa waktu terakhir, Maxim menarik perhatian konsumen dengan menetapkan tarif yang cenderung murah agar dapat bersaing dengan pendahulunya, yakni Gojek dan Grab (Kapriani et al., 2021). Akan tetapi hal tersebut tidak membuat persentase penggunaan Maxim menjadi lebih tinggi dari pesaingnya, khususnya Gojek dan Grab. Hal tersebut dapat dilihat dari hasil survei Badan Penelitian dan Pengembangan Kementerian Perhubungan yang menunjukkan bahwa persentase penggunaan Maxim masih berada pada peringkat ketiga.

Untuk mengetahui faktor penyebab masih tertinggalnya persentase penggunaan Maxim dapat dilakukan dengan mengetahui analisis sentimen dari penggunanya. Analisis sentimen dapat mengklasifikasikan polaritas suatu data opini dari pengguna yang bersifat tekstual ke dalam aspek positif atau negatif (Wahyudi et al., 2021). Analisis Sentimen dilakukan untuk mengetahui dan memahami pandangan atau reaksi publik terhadap suatu produk atau layanan, individu maupun organisasi, masalah, isu, hingga suatu peristiwa (Supriyanto et al., 2023). Hasil pengolahan data analisis sentimen sangat berpengaruh bagi para calon pengguna untuk mempertimbangkan penggunaan aplikasi dan juga sebagai masukan bagi perusahaan dalam melakukan perbaikan dan pengembangan aplikasi (Tanggraeni & Sitokdana, 2022).

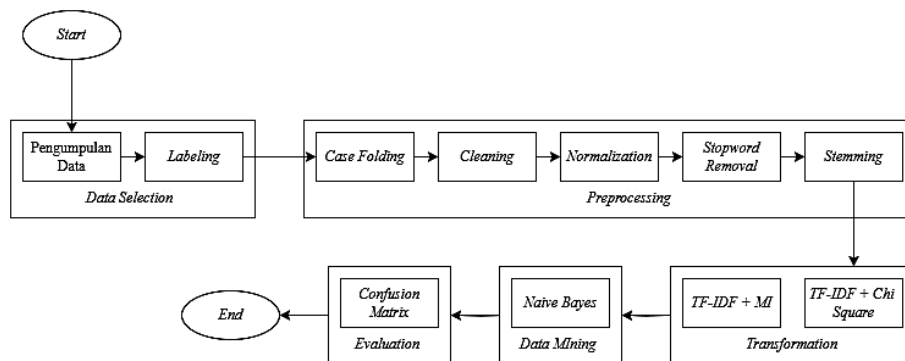
Maxim dapat diunduh dengan mudah pada *smartphone* milik pengguna melalui distributor aplikasi, salah satunya adalah Google Play. Aplikasi Maxim pada Google Play hingga bulan Maret 2023 telah mencapai 50 juta lebih unduhan dengan rating 4.8 dan 3,53 juta ulasan di Google Play. Ulasan yang diberikan mengandung opini para pengguna dalam jumlah cukup banyak dan tidak terstruktur(Fahlevvi, 2022). Data ulasan tersebut dapat dijadikan sumber dalam melakukan analisis sentimen pengguna dari Maxim.

Terdapat beberapa algoritme yang dapat digunakan pada analisis sentimen, salah satunya adalah Naive Bayes yang populer dengan kemudahan dalam penggunaannya, waktu pemrosesan yang cepat, arsitektur yang cukup sederhana dan tingkat efektivitas yang tinggi(Kardian & Gustiana, 2021). Selain itu, penggunaan seleksi fitur juga dapat meningkatkan performa dari algoritme klasifikasi karena dapat mereduksi data dari fitur yang tidak diperlukan(Cahaya & Buani, 2021). Pada penelitian yang berjudul “Analisis Sentimen Aplikasi BRIMO Pada Ulasan Pengguna di Google Play Menggunakan Algoritma Naive Bayes” didapatkan hasil akurasi sebesar 84,52%(Khoirul et al., 2023). Pada penelitian selanjutnya yang berjudul “Peningkatan Performa Naive Bayes Dengan Seleksi Atribut Menggunakan Chi Square Untuk Klasifikasi Loyalitas Pelanggan Grab” menunjukkan bahwa algoritme Naive Bayes dengan penambahan seleksi fitur Chi Square memiliki akurasi yang lebih tinggi hingga 99,51% jika dibandingkan dengan algoritme Naive Bayes tanpa penambahan seleksi fitur yaitu sebesar 92,73%(Kisworini & Setiawan, 2020). Kemudian pada penelitian yang berjudul “Analisis Sentimen Twitter Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Mutual Information” menunjukkan bahwa penggunaan seleksi fitur Mutual Information pada algoritme Naive Bayes dapat meningkatkan akurasi menjadi 97,9% dari akurasi awal algoritme Naive Bayes tanpa penggunaan Mutual Information yang hanya sebesar 96,2%(Ulfa, 2018).

Berdasarkan latar belakang permasalahan dan beberapa referensi yang telah dipaparkan, penelitian ini akan melakukan analisis sentimen ulasan pengguna aplikasi Maxim dengan melakukan perbandingan akurasi antara algoritme Naive Bayes tanpa fitur seleksi dengan algoritme Naive Bayes dengan penambahan fitur seleksi Chi Square dan Mutual Information untuk mengetahui performa seleksi fitur terbaik.

METODE

Penelitian ini menggunakan metodologi *Knowledge Discovery in Database* (KDD). *Knowledge Discovery in Database* (KDD) yang merupakan sebuah metode analisis dalam penerapan *data mining* yang dapat mengidentifikasi informasi tambahan yang tidak dapat diperoleh secara manual dari sekumpulan data(Prakarsya & Prambayun, 2020). Dalam penelitian ini terdapat lima tahapan yang akan dilakukan, yang dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Data Selection

Pada tahap awal, dilakukan pengumpulan data ulasan dari pengguna aplikasi Maxim di Google Play dengan menggunakan *library google-playscraper* pada Google Colab. Pengambilan data dilakukan pada rentang waktu 5 Maret 2023 – 5 April 2023 yang didapatkan sebanyak 10500 data ulasan pengguna sesuai dengan Gambar 2.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt
0	27b1a7df-d694-4342-bfbc-63b1b883d476	Yanto Yanto	https://play-lh.googleusercontent.com/a-/ACB-R...	saya suka pakai maxsime kalo pergi naik angkutan	5	0	3.14.8	2023-04-06 08:33:57	None	NaT
1	d0fa56a7-4dc7-4dfe-8314-3c70e3fd31b	Piere Tuan	lh.googleusercontent.com/a/AGNmyx...	Tepat waktu	5	0	3.14.7	2023-04-06 08:32:09	None	NaT
2	d68c52f3-5369-4755-b671-4cf7f5e4bc00	Marcella	https://play-lh.googleusercontent.com/a-/ACB-R...	Udah lama, drivemya gak ada yg sopan.	1	0	3.14.8	2023-04-06 08:30:49	None	NaT
3	c781204a-7c83-4a56-ad74-e7b6a7d149c5	Nungki Indrasari	https://play-lh.googleusercontent.com/a-/ACB-R...	ramah sopan 🍌	5	0	3.14.8	2023-04-06 08:13:20	None	NaT
4	d000987f-9791-473f-abac-5cc4fdba98a1	Susanti aguso	lh.googleusercontent.com/a/AGNmyx...	murah dan	5	0	3.14.8	2023-04-06 08:13:08	None	NaT
...
10495	8bf906fc-749e-49d2-ad97-e4c70ab0e262	Maksum 223	https://play-lh.googleusercontent.com/a/AGNmyx...	terimakasih orang nya sangat ramah	5	1	3.14.6	2023-03-05 00:00:37	None	NaT
10496	c1f5b897-c591-4321-9e93-48b41f6a3c40	dewi angraini cahya ningrat	https://play-lh.googleusercontent.com/a/AGNmyx...	nyaman pake Maxim ^^	5	1	3.14.7	2023-03-04 23:53:31	None	NaT
10497	829af1f2-0bbe-4c23-a3c3-a29ff79bc1ca	Herpan4 Saudara	https://play-lh.googleusercontent.com/a/AGNmyx...	Ramah	5	1	3.14.7	2023-03-04 23:49:23	None	NaT
10498	344417c3-ec90-41fa-af62-3c1f6cb98ef6	Maya Renyaan	https://play-lh.googleusercontent.com/a/AGNmyx...	aplikasih bagus	5	1	3.14.7	2023-03-04 23:37:29	None	NaT
10499	7c2b9172-4230-403e-84f2-b484140549f5	Yusnifitri Yusni	https://play-lh.googleusercontent.com/a/AGNmyx...	maxim membantu sekali	5	0	3.14.7	2023-03-04 23:27:28	None	NaT

10500 rows × 10 columns

Gambar 2. Hasil Web Scraping

Penelitian ini akan dilakukan dengan menganalisis sentimen pada kolom content, sehingga data pada kolom lainnya akan dihapus. Selanjutnya dilakukan penyeleksian data dengan menghapus ulasan yang berbahasa asing, mengandung kalimat kasar, ulasan yang sama, dan ulasan yang kurang relevan dengan penelitian.

Jumlah data setelah diseleksi adalah 1820 data. Kemudian data diberikan label pada kelas positif dan negatif secara manual dengan menambahkan kolom baru yang diberi nama "label". Pelabelan dilakukan dengan memberi angka 0 sebagai label data ulasan positif dan angka 1 sebagai label data ulasan negatif. Data yang sudah diberi label divalidasi dengan

bantuan pakar Bahasa Indonesia yang menghasilkan jumlah data pada kelas positif sebanyak 961 data dan kelas negatif sebanyak 859 data. Hasil dari pelabelan dapat dilihat pada Tabel 1.

Tabel 1. Pelabelan

label	content
1	Map gak jelas
0	mas nya baik
1	Lumayan bagus,harganya juga sangat membantu,tetapi driver nya ngebut bngt kalo bawa motor,ga bisa di sesuai in sama map nya,sikap driver nya lebih ke gojek
0	Ramah
0	Driver jemput sesuai titik. Terima kasih

Preprocessing

Pada tahap ini data akan dikoreksi dan dibersihkan dari atribut yang tidak diperlukan dalam penelitian untuk mengurangi gangguan atau *noise*. Pada tahapan ini, terdapat lima proses dilalui, diantaranya adalah *case folding*, *cleaning*, *normalization*, *stopword removal*, dan *stemming*. Tabel 2 merupakan contoh perbandingan antara data awal dengan data yang sudah melewati *preprocessing*.

Tabel 2. Hasil Preprocessing

Content	Hasil preprocessing
Jadi kemarin saya pesan Maksim yang mobil pas harga awal itu masih 17k ya Tapi pas udah nyampe tujuan ko harganya naik jadi 26k Itu salah ya dimana? Kalau dari awal 26k, kenapa ngga sekalian aja tu Kan kita yang susah nantinya Untung aja saya bawa duit lebih Kalau engga gimana ?? Kok bisa harganya berubah gitu !	kemarin pesan maxim mobil pas harga tetapi pas tuju harga salah tidak susah untung bawa duit tidak harga ubah

Transformation

Pada tahap ini, fitur ekstraksi TF-IDF akan digunakan untuk menemukan fitur-fitur penting yang tersimpan di dalam data. Selanjutnya, fitur seleksi (Chi Square dan Mutual Information) mengurangi jumlah variabel dan data yang tidak terlalu diperlukan. Banyaknya dimensi fitur yang ada pada data, dimana didalamnya bisa saja berisikan data-data yang tidak relevan yang dapat mempengaruhi kinerja dari algoritme(Sugianto et al., 2018). Karena hal tersebut, penggunaan seleksi fitur dapat menjadi solusi dalam mengatasi fitur berdimensi tinggi sehingga proses klasifikasi lebih maksimal.Tahapan *transformation* ini akan dijabarkan sebagai berikut:

1. *Term Frequency Inverse Document Frequency* (TF-IDF)

TF-IDF berguna dalam mengetahui frekuensi kemunculan sebuah kata dengan memberikan nilai atau bobot ke setiap *term* dalam sebuah dokumen untuk mengetahui seberapa sering kata tersebut muncul(Pratondo, 2023). Hasil pembobotannya dapat dilihat pada Gambar 3.

abai	abang	abis	aceh	acu ▼
0.0	0.0	0.0	0.0	0.25528459390411484
0.0	0.0	0.0	0.3570566750253328	0.0
0.2633959791954167	0.0	0.0	0.0	0.0
0.0	0.0	0.5939322504378907	0.0	0.0
0.0	0.0	0.5420334004234226	0.0	0.0
0.0	0.0	0.24688725209883866	0.0	0.0
0.0	0.9005099121841512	0.0	0.0	0.0
0.0	0.9005099121841512	0.0	0.0	0.0
0.0	0.9005099121841512	0.0	0.0	0.0
0.0	0.8478094764923308	0.0	0.0	0.0

Gambar 3. Hasil TF-IDF

Berdasarkan hasil TF-IDF yang tertera pada Gambar 3, terbentuk matriks berukuran 1820 X 1625 yang berarti terdapat 1820 dokumen juga 1625 term. Dimana setiap *term* memiliki nilai TF-IDF yang mana semakin besar nilai TF-IDF berarti semakin penting *term* tersebut dalam suatu dokumen terhadap seluruh koleksi dokumen (*corpus*) yang ada. Sedangkan nilai nol yang muncul memiliki arti bahwa term tersebut tidak ada dalam dokumen atau tidak ada dalam seluruh *corpus* dokumen yang digunakan untuk perhitungan TF-IDF.

2. Chi Square

Chi square merupakan algoritme seleksi fitur yang bertujuan untuk menguji tingkat ketergantungan suatu *term* terhadap kategorinya (Imron Maulana & Andy Soebroto, 2019). Semakin besar nilai Chi Square maka semakin penting term untuk digunakan dalam klasifikasi. Semakin kecil nilai taraf nyata (α), maka nilai kritis akan semakin besar sehingga banyak term yang akan dihilangkan (Winata et al., 2022). Pada penelitian ini proses seleksi fitur dengan algoritme Chi Square dilakukan dengan memilih 50% fitur dengan nilai Chi Square tertinggi. Jumlah fitur sebelum menggunakan Chi Square adalah sebanyak 1625 fitur dan fitur yang tersisa setelah menggunakan Chi Square adalah sebanyak 812 fitur. Tabel 3 merupakan beberapa contoh dari 5 fitur tertinggi yang terpilih.

Tabel 3. Fitur Chi Square

Nilai	Fitur
169.081859	ramah
89.923557	baik
88.308171	tidak
63.974429	peta
44.669454	cepat

3. Mutual Information

Mutual Information adalah algoritme yang menunjukkan nilai keterkaitan antara suatu fitur dengan fitur yang lainnya (Gormantara & Boli Watomakin, 2020). Pada penelitian ini proses seleksi fitur dengan algoritme Mutual Information dilakukan dengan memilih 50% fitur dengan nilai Mutual Information tertinggi. Jumlah fitur sebelum menggunakan Mutual

Information adalah sebanyak 1625 fitur dan fitur yang tersisa setelah menggunakan Mutual Information adalah sebanyak 812 fitur. Tabel 4 merupakan beberapa contoh dari 5 fitur tertinggi yang terpilih.

Tabel 4. Fitur Mutual Information

Nilai	Fitur
0.185933	ramah
0.180866	tidak
0.111836	<i>driver</i>
0.090966	baik
0.074314	peta

Data Mining

Data mining merupakan tahapan utama dalam metode KDD. Pada tahapan ini akan dilakukan penggalian dan pola dari data yang telah diproses pada tahapan sebelumnya untuk menghasilkan sebuah informasi dari dataset atau dokumen dengan menggunakan algoritme Naive Bayes. Naive Bayes merupakan salah satu algoritme klasifikasi yang menerapkan teori probabilitas dan Teorema Bayes yang mengasumsikan bahwa setiap variabel bersifat independen sehingga tidak memiliki keterkaitan dengan atribut lainnya (Nofriansyah & Nurcahyo, 2015).

Pada tahapan *data mining* ini, dilakukan perbandingan hasil akurasi antara algoritme Naive Bayes dengan penambahan Chi Square, juga Naive Bayes dengan Mutual Information. Selain itu, untuk peningkatan performa yang lebih baik, dilakukan pembagian *data train* dan *data test* menjadi 3 skenario yang dilihat pada Tabel 5.

Tabel 5. Skenario Data Mining

Skenario	Algoritme
70:30	<i>Naive Bayes</i>
	<i>Naive Bayes dan Chi Square</i>
	<i>Naive Bayes dan Mutual Information</i>
80:20	<i>Naive Bayes</i>
	<i>Naive Bayes dan Chi Square</i>
	<i>Naive Bayes dan Mutual Information</i>
90:10	<i>Naive Bayes</i>
	<i>Naive Bayes dan Chi Square</i>
	<i>Naive Bayes dan Mutual Information</i>

Evaluation

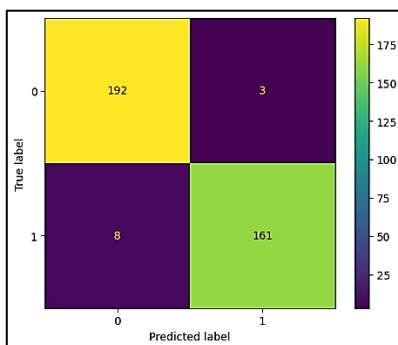
Evaluation dilakukan untuk menguji serta mengetahui performa dari model yang digunakan dalam memprediksi data menggunakan Confusion Matrix. Terdapat empat parameter yang akan dihitung dalam Confusion Matrix yaitu *accuracy*, *recall*, *precision*, dan

F1-Score(Gunawan et al., 2018). Terdapat 9 skenario percobaan yang dilakukan dalam penelitian ini. Adapun hasil evaluasi dari 9 skenario tersebut dapat dilihat pada Tabel 6.

Tabel 6. Perbandingan Hasil Akurasi

Algoritme	Skenario	Accuracy	Precision	Recall	F1-Score
Naive Bayes	70:30	96,15%	96%	96%	96%
	80:20	96,42%	97%	96%	96%
	90:10	96,15%	97%	96%	96%
Naive Bayes + Chi Square	70:30	96,70%	97%	97%	97%
	80:20	96,97%	97%	97%	97%
	90:10	96,15%	97%	96%	96%
Naive Bayes + Mutual Information	70:30	95,78%	96%	96%	96%
	80:20	95,60%	96%	95%	96%
	90:10	94,50%	96%	94%	94%

Berdasarkan Tabel 6 pemodelan dengan menggunakan algoritme Naive Bayes dan seleksi fitur Chi Square pada skenario 80:20 mendapatkan nilai akurasi terbesar yaitu 96,97%. Berikut merupakan Confusion Matrix dari pemodelan tersebut:



Gambar 4. Confusion Matrix

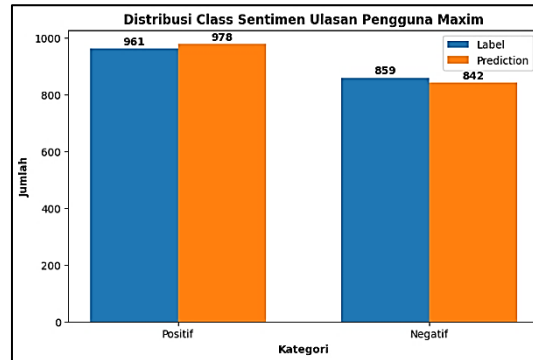
Gambar 4 menunjukkan data yang berhasil diprediksi dengan benar (*True Positive*) sebanyak 192 data, *True Negative* 161 data, *False Positive* 3 data, dan *False Negative* 8 data. Dari hasil tersebut dapat dihitung nilai dari *accuracy*, *precision*, *recall*, dan *f1-score* yang dapat dilihat pada Gambar 5.

Classification report:					
	precision	recall	f1-score	support	
0	0.96	0.98	0.97	195	
1	0.98	0.95	0.97	169	
accuracy			0.97	364	
macro avg	0.97	0.97	0.97	364	
weighted avg	0.97	0.97	0.97	364	

Gambar 5 Classification Report

HASIL DAN PEMBAHASAN

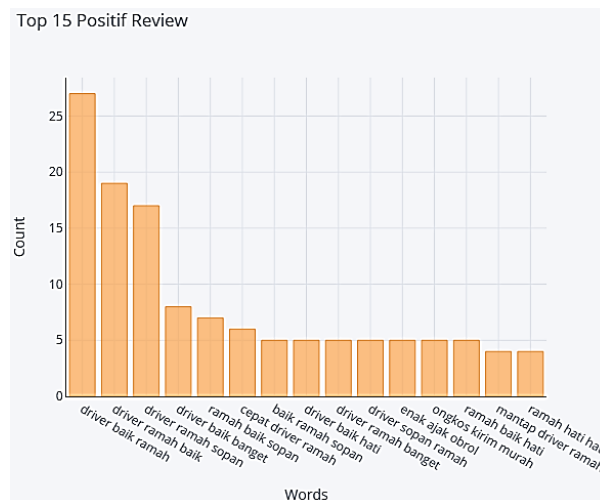
Setelah melakukan pemodelan, dapat diketahui bahwa akurasi paling tinggi diperoleh dari pemodelan Naive Bayes dengan seleksi fitur Chi Square pada skenario 80:20 mendapatkan nilai akurasi terbesar yaitu 96,97%. Ketika model tersebut digunakan untuk memprediksi keseluruhan dataset, berikut merupakan perbandingan antara data aktual dengan data yang diprediksi:



Gambar 6. Perbandingan Kelas Hasil Prediksi

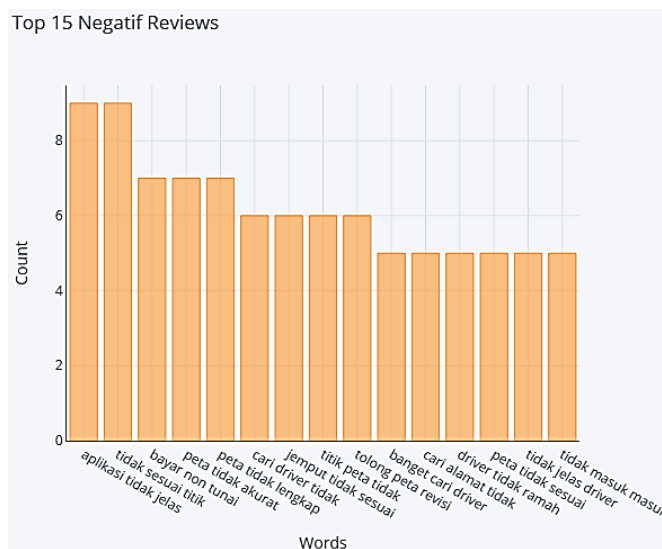
Dari Gambar 6 dapat dilihat bahwa persebaran label data yang sebenarnya (warna biru) memiliki sebanyak 961 data kelas positif dan 859 data kelas negatif. Sedangkan persebaran label data dari prediksi model (warna oranye) memiliki sebanyak 978 data kelas positif dan 842 data kelas negatif.

Adapun pengklasifikasian ulasan pengguna aplikasi Maxim dapat divisualisasikan menggunakan bar plot dengan tujuan untuk memberikan informasi umum mengenai data ulasan yang diberikan oleh pengguna aplikasi Maxim pada Google Play.



Gambar 7. Visualisasi Ulasan Positif

Berdasarkan Gambar 7 dapat dilihat frekuensi kemunculan 15 kalimat yang paling tinggi pada kelas positif. Dari visualisasi tersebut, dapat ditarik kesimpulan bahwa beberapa pengguna puas karena driver yang baik, ramah dan sopan, serta harga yang cenderung murah.



Gambar 8. Visualisasi Ulasan Negatif

Sedangkan pada visualisasi negatif pada Gambar 8, dapat ditarik kesimpulan bahwa beberapa pengguna mengalami masalah pada aplikasi yang tidak jelas, titik jemput yang tidak sesuai, pembayaran nontunai, dan peta yang tidak akurat.

SIMPULAN

Berdasarkan hasil akurasi yang didapatkan pada setiap skenario, dapat disimpulkan bahwa penggunaan seleksi fitur pada Naive Bayes dalam penelitian ini tidak berpengaruh. Penerapan Chi Square pada algoritme Naive Bayes mendapatkan nilai *accuracy* sebesar 96.97%, *precision* 97%, *recall* 97%, dan *f1-score* 97%, dimana hasil tersebut tidak menunjukkan peningkatan akurasi yang signifikan dari penggunaan Naive Bayes tanpa seleksi fitur yang mendapatkan nilai *accuracy* 96.42%, *precision* 97%, *recall* 96%, dan *f1-score* 96%. Bahkan pada penggunaan algoritme Naive Bayes dengan Mutual Information menghasilkan *accuracy* lebih kecil disbanding ketiganya yaitu sebesar 95.78%, *precision* 96%, *recall* 96%, dan *f1-score* 96%.

DAFTAR PUSTAKA

- Cahya, D., & Buani, P. (2021). Penerapan Algoritma Naive Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung. *Jurnal Sains Dan Manajemen*, 9(2).
- Fahlevvi, M. R. (2022). Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google

- Playstore Menggunakan Metode Support Vector Machine. *Jurnal Teknologi Dan Komunikasi Pemerintahan*, 4(1), 1–13. <http://ejournal.ipdn.ac.id/JTKP>,
- Gormantara, A., & Boli Watomakin, D. (2020). *Klasifikasi Kategori dan Pelabelan Berita Bahasa Indonesia Menggunakan Mutual Information Dan K-Nearest Neighbor*.
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika*, 4(2), 17–29. www.femaledaily.com
- Imron Maulana, M., & Andy Soebroto, A. (2019). *Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter menggunakan Metode Improved k-Nearest Neighbor dan Seleksi Fitur Chi-square* (Vol. 3, Issue 7). <http://j-ptiik.ub.ac.id>
- Kapriani, Asmawiyah, Thaha, S., & Hariyanti. (2021). Analisis Pengaruh Harga dan Kualitas. *Al-Buhuts*, 17(1), 142–161.
- Kardian, A. R., & Gustiana, D. (2021). Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan Naive Bayes Classifier. *Jurnal Ilmiah Komputasi*, 20(1). <https://doi.org/10.32409/jikstik.20.1.401>
- Kevin, V., Que, S., Analisis, :, Transportasi, S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization). In *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* | (Vol. 9, Issue 2). www.tripadvisor.com,
- Khoirul, M., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi BRIMO Pada Ulasan Pengguna di Google Play Menggunakan Algoritma Naive Bayes. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1).
- Kisworini, R. Y., & Setiawan, M. A. (2020). Peningkatan Performa Naive Bayes Dengan Seleksi Atribut Menggunakan Chi Square Untuk Klasifikasi Loyalitas Pelanggan Grab. *Journal of Informatics, Information System, Software Engineering and Applications*, 2(2), 69–075. <https://doi.org/10.20895/INISTA.V2I2>
- Maxim. (2023). *Maxim Indonesia*. <https://id.taximaxim.com/>
- Nofriansyah, D., & Nurcahyo, G. W. (2015). *Algoritma Data Mining dan Pengujian* (1st ed.). Deepublish.
- Prakarsya, A., & Prambayun, A. (2020). Implementasi Data Mining Untuk Prediksi Penyebaran Virus HIV/AIDS di Bandar Lampung dengan Teknik Decision Tree. *Jurnal Siskomti*, 3(2). <http://www.ejournal.lembahdempo.ac.id>
- Pratondo, D. A. (2023). *Pengembangan Sistem Rekomendasi Berbasis Content-based Filtering pada Data Dinamis*.
- Sugianto, N. A., Cholissodin, I., & Widodo, A. W. (2018). *Klasifikasi Keminatan Menggunakan Algoritme Extreme Learning Machine dan Particle Swarm Optimization untuk Seleksi Fitur (Studi Kasus: Program Studi Teknik Informatika FILKOM UB)* (Vol. 2, Issue 5). <http://j-ptiik.ub.ac.id>
- Supriyanto, J., Alita, D., & Rahman Isnain, A. (2023). Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring. *JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)*, 4, 74–80. <https://doi.org/10.33365/jatika.v4i1.2468>

- Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Informatika Dan Sistem Informasi*, 9(2), 785–795.
- Ulfa, M. A. (2018). *Analisis Sentimen Twitter Menggunakan Naïve Bayes Classifier dengan Seleksi Fitur Mutual Information* [Tugas Akhir, Universitas Mataram]. <http://eprints.unram.ac.id/5961/1/Jurnal%20Untuk%20UPT.pdf>
- Wahyudi, R., Kusumawardhana, G., Purwokerto, A., Letjend, J., Soemarto, P., Purwanegara, K., Purwokerto, T., & Banyumas, K. (2021). Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *JURNAL INFORMATIKA*, 8(2). <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- Winata, W., Zaidiah, A., & Chamidah, N. (2022). *Analisis Sentimen pada Ulasan Produk Masker di Marketplace Shopee Menggunakan Support Vector Machine dan Seleksi Fitur Chi Square*.