



Sentimen Analisis Pengguna Aplikasi *Grab* Menggunakan Algoritma *Naive Bayes Classifier* dan Support Vector Machine

Ipin Sugiyarto¹, Sita Anggraeni^{*2}, Umi Faddilah³, Ali Alamuddin Muzaffar⁴

^{1,2} Universitas Nusa Mandiri; Jalan Jatiwaringin Raya No 02 RT 08/13, Cipinang Melayu, Jakarta Timur, Program Studi Informatika, Fakultas Teknologi Informasi, Jakarta

³ Universitas Bina Sarana Informatika; Jalan Kramat Raya No 98 Jakarta Pusat, Program Studi Sistem Informasi Akuntansi, Fakultas Teknik dan Informatika, Jakarta

⁴ Universitas Islam Negeri Sultan Syarif Kasim Riau; Jalan HR. Soebrantas Pekanbaru, Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Riau

*Email Penulis Korespondensi: sita.sia@nusamandiri.ac.id

Abstrak

Grab Indonesia termasuk perusahaan ojek online terkemuka di Indonesia, pemrosesan terhadap saran dan keluhan dari pengguna Grab secara real time dapat disampaikan melalui akun twitter @GrabID, sehingga Grab Indonesia dapat mengetahui tanggapan secara cepat tingkat kepuasan pelanggan terhadap pelayanan yang diberikan, dibutuhkan analisis sentimen terhadap saran ataupun keluhan yang diterima oleh perusahaan Grab Indonesia. Proses dalam pengumpulan data pada penelitian ini yaitu mengambil data primer yang diperoleh secara langsung dari sumber asli. Data diambil dari twitter sebanyak 787 data. Model evaluasi juga menggunakan coding dengan hasil visual model confusion matrix berbasis Python. Pada penelitian ini tahapan yang dilakukan Pengumpulan Data, Preprocessing Data, Fitur Extraction, Klasifikasi dan Evaluasi. Maka didapatkanlah akurasi 96,02% data training dan akurasi data testing 84,17 % untuk algoritma SVM dengan perbandingan 80:20 data training dan data testing, sedangkan untuk algoritma NBC didapatkan akurasi 54,37% data training dan akurasi 66,45% data testing dengan perbandingan 80:20 data training dan data testing. Adanya penelitian diharapkan dapat memberikan pengetahuan tentang implementasi algoritma Naïve Bayes Classifier dan Support Vector untuk mengetahui respon masyarakat terhadap aplikasi atau produk tersebut apakah mempunyai respon positif atau negatif.

Kata kunci— Naive Bayes Classifier, Support Vector Machine, Analisis sentimen, Twitter, Grab

Abstract

Grab Indonesia is one of the leading online motorcycle taxi companies in Indonesia, Processing of suggestions and complaints from Grab users in real time can be submitted through the @GrabID twitter account, so Grab Indonesia can quickly find out the level of customer satisfaction with the services provided, sentiment analysis is needed for suggestions or complaints received by the Grab Indonesia company. The process of collecting data in this study is to take

primary data obtained directly from the original source. Data taken from twitter as much as 787 data. The evaluation model also uses coding with the visual results of the Python-based confusion matrix model. In this study, the stages carried out are Data Collection, Data Preprocessing, Feature Extraction, Classification and Evaluation. Then the accuracy of 96.02% training data and the accuracy of 84.17% testing data for the SVM algorithm with a ratio of 80:20 training data and testing data, while for the NBC algorithm, the accuracy of 54.37% training data and the accuracy of 66.45% testing data are obtained with a ratio of 80:20 training data and testing data. The existence of research is expected to provide knowledge about the implementation of the Naïve Bayes Classifier and Support Vector algorithms to determine the community's response to the application or product whether it has a positive or negative response.

Keywords— Naive Bayes Classifier, Support Vector Machine, Sentiment analysis, Twitter, Grab

1. PENDAHULUAN

Perkembangan teknologi yang sangat pesat membawa pengaruh yang besar bagi manusia. Salah satu pengaruh dari perkembangan teknologi adalah semakin maraknya penggunaan transportasi *online* melalui media internet. Ojek *online* atau sering disebut *ojol* merupakan transportasi dari ojek konvensional yang biasanya bertempat di pangkalan untuk menunggu pelanggan. seiring berjalannya waktu, *ojol* ini semakin diminati masyarakat apalagi dengan bantuan *smartphone*. Pelayanan yang disediakan tidak hanya transportasi angkutan penumpang saja, namun juga melayani jasa kurir untuk pemesanan makanan, jasa pembersih untuk bersih-bersih rumah dan kantor, pengiriman barang, dan lain sebagainya.

Grab Indonesia sebagai salah satu perusahaan ojek *online* terkemuka di Indonesia memiliki jumlah pelanggan yang cukup besar dan mencakup hampir seluruh wilayah di Indonesia. setiap pelanggan memiliki tingkat kepuasan berbeda terhadap layanan yang diberikan oleh Grab Indonesia, sehingga selalu ada pro dan kontra berupa saran dan keluhan.

Pemrosesan terhadap saran dan keluhan kini dapat diselesaikan secara *real time* melalui akun *twitter* @GrabID, sehingga Grab Indonesia dapat mengetahui tanggapan secara cepat tingkat kepuasan pelanggan terhadap pelayanan yang diberikan. Namun mengingat banyaknya jumlah pelanggan, tidak sedikit pula saran maupun keluhan yang ditujukan ke akun @GrabID yang diterima perhari. Maka diperlukan sebuah solusi berupa analisis sentimen terhadap saran ataupun keluhan yang diterima oleh perusahaan Grab Indonesia.

Analisis sentimen disebut juga *opinion mining* merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat apakah bernilai positif atau negatif. Analisis sentimen dapat digunakan sebagai solusi dari masalah yang sudah dipaparkan sebelumnya.

Penelitian sebelumnya dari peneliti lainnya terkait sentimen mengenai [1] Komparasi Algoritma Klasifikasi Untuk Menentukan Evaluasi Kinerja Terbaik Pada Status Akreditasi Sekolah/Madrasah Kalimantan Timur Berdasarkan IASP 2020 dengan Hasil komparasi algoritma klasifikasi didapatkan bahwa algoritma SVM memiliki kinerja terbaik dibandingkan dengan algoritma klasifikasi lainnya dengan nilai *accuracy* 88,8%, *precision* 91,00%, *recall* 81,00%, dan *f1-score* 84,00%. Penelitian sebelumnya juga membahas mengenai [2] Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine dengan Hasil dari analisis menggunakan Support Vector Machine menghasilkan akurasi 85,54% dan Hasil Review positif yang paling sering diulas adalah "ovo", sedangkan review negatif yang paling sering diulas adalah "driver". Penelitian sebelumnya membahas Komparasi Algoritma Naïve Bayes, Logistic Regression dan Support Vector Machine pada Klasifikasi File Application Package Kit Android Malware [3] dengan Hasil uji akurasi menunjukkan algoritma Naive Bayes mampu mengklasifikasi keluarga malware dengan tingkat akurasi 97,75%, sedangkan algoritma Logistic Regression akurasinya 88,75% dan akurasi Support Vector Machine mencapai 96,75%. Penelitian sebelumnya menganalisis [4] tentang Analisis Perbandingan Algoritma Support Vector

Machine, *Naive Bayes*, dan Regresi Logistik untuk memprediksi Donor Darah mendapatkan hasil penelitian dengan algoritma *Naive Bayes* 89,90%, Regresi Logistik 82,59% dan SVM 94,79%. Terakhir dalam penelitian sebelumnya membahas Sentimen Analisis masyarakat mengenai COVID-19, hasil penelitian [5] menggunakan algoritma klasifikasi yang digunakan yaitu *Naive Bayes* dengan hasil akurasi 78,02% sedangkan *Support Vector Machine* menghasilkan akurasi sebesar 80,23% dan memiliki selisih akurasi 2,21%.

Dalam penelitian [6] yang menemukan optimalisasi dalam akurasi pada algoritma *Support Vector Machine* dengan nilainya sebesar 94,37%, sedangkan pada algoritma *Naive Bayes Classifier* diperoleh akurasi sebesar 86,87%. Akurasi nilai yang tinggi menunjukkan bahwa kinerja algoritma *Support Vector Machine* lebih optimal dibandingkan algoritma *Naive Bayes Classifier* dalam mengklasifikasikan kesehatan mental mahasiswa. Optimalisasi akurasi dalam penelitian lainnya mengenai [7] mengklasifikasi Keterlambatan Pembayaran Sumbangan Pengembangan Pendidikan (SPP) di Madrasah Ibtidaiyah dengan Algoritma *Support Vector Machine* (SVM) diperoleh tingkat *accuracy* sebesar 89,00%, *precison* sebesar 89,00%, *recall* sebesar 100% dan algoritma *Naive Bayes Classifier* diperoleh tingkat *accuracy* sebesar 47,00%, *precison* sebesar 27,00%, *recall* sebesar 100%.

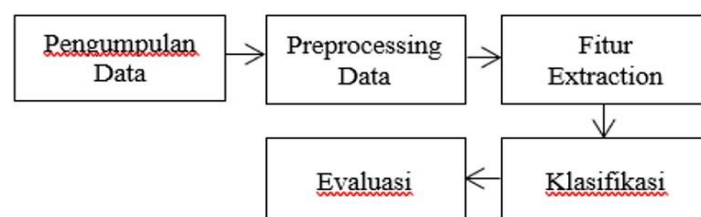
Berdasarkan uraian di atas dan beberapa penelitian terdahulu dapat disimpulkan bahwa algoritma yang menghasilkan akurasi terbaik terdapat pada *Support Vector Machine*. Pada penelitian ini penulis menggunakan dua algoritma yaitu *Naive Bayes Classifier* dan *Support Vector Machine* untuk mendapatkan akurasi terbaik dengan mengklasifikasikan opini positif dan negatif pada pelayanan aplikasi *Grab*, selain itu penelitian ini juga dilakukan untuk meningkatkan kualitas dan layanan yang ada pada aplikasi *Grab* Indonesia pada pelayanannya.

Berdasarkan latar belakang yang telah dipaparkan di atas maka dapat dirumuskan masalah sebagai berikut: Bagaimana implementasi algoritma *Naive Bayes Classifier* dan *Support Vector Machine* dalam mengklasifikasikan data *tweet* terhadap pelayanan aplikasi *Grab* Indonesia apakah data *tweet* tersebut termasuk ke dalam kelas bersentimen positif atau negatif dan manakah algoritma yang cukup baik digunakan untuk analisis sentimen aplikasi *Grab* Indonesia?

Adapun tujuan dari penelitian ini adalah untuk mendapatkan hasil akurasi yang paling baik di antara kedua algoritma yang digunakan yaitu *Naive Bayes Classifier* dan *Support Vector Machine*. Adapun manfaat dari penelitian ini adalah sebagai berikut: Penelitian ini diharapkan dapat memberi tambahan pengetahuan tentang implementasi algoritma *Naive Bayes Classifier* dan *Support Vector Machine* dan dapat memberikan gambaran mengenai bagaimana analisis sentiment terhadap aplikasi *grab* dengan media sosial *twitter*. Dan untuk mencari sebuah informasi tentang suatu layanan aplikasi atau produk untuk mengetahui respon masyarakat terhadap aplikasi atau produk tersebut apakah mempunyai respon positif atau negatif. Serta untuk mengetahui berapa keberhasilan algoritma *Naive Bayes Classifier* dan *Support Vector Machine* pada pengklasifikasian teks berbahasa Indonesia.

2. METODE PENELITIAN

Penelitian ini menganalisis sentimen aplikasi *Grab* pada media sosial *twitter* dalam Gambar 1 dimulai dengan tahapan pengumpulan data dimana dilakukan pula pelabelan dalam data yang sudah terkumpul dalam bentuk *excel*, *preprocessing* data, fitur *extraction*, klasifikasi dengan algoritma *Naive Bayes Classifier* dan *Support Vector Machine* dan tahapan terakhir yaitu evaluasi dimana model evaluasi juga menggunakan coding dengan hasil visual model *confusion matrix* berbasis *Python*.



Gambar 1 Tahapan Penelitian

2.1. Pengumpulan Data

Proses pengumpulan data pada penelitian ini yaitu data primer yang diambil dengan cara *crawling twitter* dengan bantuan *Rapid Miner*. Data yang diambil sejumlah 2000 data dengan kategori GrabID dengan review paling populer. Data yang diperoleh disimpan dalam file bertipe *xlsx*. Supaya data yang digunakan dapat terbaca dengan baik pada *read excel*. Data yang diambil dilakukan pembersihan data dengan menghilangkan kalimat atau review yang sama sehingga data tidak ada yang sama atau *double*. Hingga data yang digunakan pada penelitian ini menjadi 787 data setelah dilakukan pembersihan. Dalam Gambar 2 pengumpulan data ini sudah dilakukan pelabelan secara manual satu persatu dalam setiap hasil dari tweet yang dikumpulkan kemudian akan dimasukkan dalam tahapan selanjutnya.

	A	B	C
522	C Y L A • OPEN	@chilzsmile aku jual nokos nih kak buat verif aplikasi, mari lirik sini 🙏 dm @racylaaready nomor for whatsapp tw	Positif
523	Mrs. Bang	@chedtra0 @mahdinosaur emang ada opsi gocar? gimana caranya?	Positif
524	es nilai A aamiin	@koganenohikari zaman dah maju mbah, bisa pake kurir/grab/gojek	Positif
525	Grab Indonesia	@_yakalajadahl Halo Kak, mohon maaf banget ya 🙏 Aku cek laporan Kakak dinomor laporan 427472868 sdh ada	Positif
526	Izza Vithry Hayah	@tanyakanrl XL...! Beli paket kuota per bulan cuma Rp. 29.950 udah dapat 32 GB! + Unlimited WA & Line, bonus kuota	Positif
527	Bertha Darcy	RT @YeonAera1: Baru tau aplikasi grab bisa buat donasi, mana tinggal masukin code promonya simple lgi #pesanuntu	Positif
528	Fourth Avenue Cafe	! Kita baking dari cinere, pengiriman bisa dengan same day gojek/grab/paxel. Atau daerah depok jakse! bisa dengan ku	Positif
529	Sidi	@FOODFESS21. download line, biasanya dapet voucher buy1get1 atau beli via ojol2. buka aplikasi gojek/grab untuk l	Positif
530	Grab Indonesia	@manifestingyuk Maaf bgt yaa Kak jadi gak nyaman mengenai pelayanan mitra pengemudi kami 🙏 Supaya bisa al	Positif
531	klar 🙏 mt after dm	di grab nda ada diskon + mark up jauh, di shopee hari ini udh tutup 😊 padahal dekat hiks lum rejeki	Negatif
532	Silvia Julianti	Shopee diskon ongkir 2k grab 1k kaya buat ap gt	Negatif
533	Joki Tugas Unud	@aldapstrs Klo naik ojol jg, pasti klo grab ga diskon, gojek diskon, ato sebaliknya	Negatif
534	geviruw	grab ko gaada diskon grabfood lagi si :(((	Negatif
535	Raisa Adeline	Mau pesen lewat grab tidak ada diskon ya kakak, harganya jadi hampir 2x lipat bersama diskon.Tapi karena keinginan	Negatif
536	chase pacific	@tanyakanrl Tapi emg skill masak itu setidaknya harus ada, mo cewe ato cowo. Gojek, grab, dan shopee ga akan kasi l	Negatif
537	Kinan	monangis banget ini biaya layanan, ongkir, & resto buat gojek/grab/shopee gak ngotakk! beli 1 kopsus aja bisa harga	Negatif
538	Risu open!	RT @buerbato: HI HIII! dalam rangka ak akhirnya lulus (anjay), ak mw ngsih special offer nih! setiap joki di bawa	Negatif
539	ija 🙏 — OPEN! disc 15	HI HIII! dalam rangka ak akhirnya lulus (anjay), ak mw ngsih special offer nih! setiap joki di bawah 10k bakal dape	Negatif
540	open apk prem	RT @ofMagique: Help RT, please? Thank you! 🙏 Hayi. Jillie datang dengan ready stock spesial 🌟ZEROBASEONE 🌟yan	Negatif
541	Ceyen 🌟 clearanc	@KimloyG Grab banyak diskon mjrnr	Negatif
542	Babunya Anabul	@GrabID Halo min, mau tanya pesanan saya statusnya masih kayak gini tapi drivernya udh nelponin bilang udh nyam	Negatif
543	katyd 🙏	@GrabID Itu dm nya dibales, aku udah dm dari tadi loh pake keyword tersebut juga	Negatif
544	katyd 🙏	@GrabID Admin Tita tolong bales dm ku, balesnya yang bener jangan pake template ini disini malah aku yang dirugiin	Negatif

Gambar 2 Tahapan Pengumpulan Data dan Pelabelan

2.2. Preprocessing Data

Proses pada tahap *preprocessing data*, data mentah terlebih dahulu dilakukan proses *case folding*, *word normalization*, *filtering*, dan *stemming*. *Preprocessing* menghindari data yang tidak teratur, data tidak sempurna, dan data yang tidak konsisten seperti pada penelitian [8].

2.3. Fitur Extraction

Feature Extraction adalah teknik pengambilan ciri atau *feature* dari suatu bentuk yang nantinya nilai yang didapatkan akan dianalisis untuk proses selanjutnya. Dalam melakukan *feature axtraction* penulis menerapkan teknik TF-IDF seperti dalam penelitian [9] dan N-Gram (*Term Frequency-Inverse Document Frequency*) untuk memberikan bobot pada kata-kata dalam vektor fitur. Serta melakukan seleksi fitur berupa *chi-square* untuk memperoleh subset kata-kata yang paling informatif

2.4. Klasifikasi

Klasifikasi merupakan salah satu tugas utama dalam *machine learning* dan *data mining* yang termasuk ke dalam tipe *supervised learning*. Istilah klasifikasi didapat dari tujuan utama teknik ini yaitu untuk memprediksi sebuah kategori dari masukkan data [10]. Tujuan dari klasifikasi ini untuk proses perhitungan yang menerapkan algoritma data *mining* untuk mengolah

dataset dari *twitter* terhadap aplikasi *grab*. Dalam penelitian ini menggunakan dua algoritma untuk mengklasifikasi data yaitu dengan *Naïve Bayes Classifier* dan *Support Vector Machine*.

1. *Naïve Bayes Classifier*

Algoritma klasifikasi *Naïve Bayes* merupakan salah satu algoritma teknik klasifikasi, algoritma ini memanfaatkan metode probabilitas dan statistik. Algoritma ini merupakan salah satu metode klasifikasi yang populer dan termasuk dalam sepuluh algoritma teratas dalam data mining menurut *IEEE International Conference on Data Mining* di Hongkong [11]. Untuk rumus dasar dari *Naïve Bayes* yang digunakan, yaitu:

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

Keterangan:

X : Kriteria suatu kasus berdasarkan masukan

H_i : Kelas solusi pola ke-I, di mana I adalah jumlah label kelas

P(H_i | X) : Probabilitas kemunculan label kelas H_i dengan kriteria masukan X

P(X | H_i) : Probabilitas kriteria masukan X dengan label kelas H_i

P(H_i) : Probabilitas label kelas H_i

2. *Support Vector Machine*

Algoritma *Support Vector Machine* merupakan salah satu teknik *machine learning* yang populer untuk klasifikasi teks serta memiliki performa yang baik pada banyak domain. Kemampuan SVM dalam mengidentifikasi *hyperplane* secara terpisah diantara dua kelas berbeda sehingga termaksimalkan dan SVM menjamin untuk memaksimalkan jarak antara data yang paling dekat dengan *hyperplane* menurut uraian penelitian [12].

Support Vector Machine merupakan sebuah algoritma klasifikasi yang mampu mengenali dua kelas dalam data, sehingga termasuk dalam algoritma klasifikasi biner, karena hanya mampu mengenali dua kelas yaitu positif dan negatif dalam pengertian dalam jurnal [13]

2.5. Evaluasi

Tahap evaluasi untuk memastikan bahwa tes itu benar. Evaluasi ini adalah untuk menemukan hasil terbaik dari hasil uji. Model evaluasi juga menggunakan coding dengan hasil visual model *confusion matrix* berbasis *Python*. Pengukuran akurasi terhadap model menggunakan *confusion matrix*. *Confusion matrix* adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. *Confusion matrix* berisikan informasi mengenai hasil klasifikasi aktual dan telah diprediksi oleh sistem klasifikasi.

3. HASIL DAN PEMBAHASAN

3.1 *Pengumpulan data*

Pada Gambar 3 tahapan penelitian membahas hasil dari analisis sentimen terhadap aplikasi *grab* menggunakan Algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* dengan jumlah data yang dilabeling ada 787 data yang dibagi menjadi dua sentimen yaitu label positif sebanyak 528 data dan label negatif sebanyak 258 data. Proses dalam pengumpulan data pada penelitian ini yaitu mengambil data primer yang diperoleh secara langsung dari sumber asli. Data diambil dari *twitter* sebanyak 787 data yang diambil dengan cara *crawling* melalui aplikasi *Rapid Miner* dengan kategori *GrabID* dengan review yang paling populer. Dari 787 data dalam bentuk *excel* yang di *import* menggunakan *Google Colab* dengan bahasa pemrograman *Python* dalam gambar 3 kemudian dataset didapatkan kemudian data dilabelkan menjadi dua bagian yaitu

positif dan negatif, di mana pada nilai positif memiliki label bernilai angka 2 dan negatif memiliki label bernilai angka 1.

```

from sklearn.metrics import classification_report
%matplotlib inline

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
True

Import data

df = pd.read_excel("GrabID.xlsx")
df.head()

```

	From-User	Text tweet	Sentiment
0	Grab Indonesia	@KissTaesan Baik Kak, btw ke DM 🐼-Mamat	Positif
1	Grab Indonesia	@_ahsinar Halo Kak, Mohon maaf atas ketidak...	Positif
2	Grab Indonesia	@laciq Hi Kak Aulia, bisa dikirimkan kode pe...	Positif
3	Grab Indonesia	@dhikareani Halo juga Kak 🐼 Maaf bgt yaa Kak ...	Positif
4	VaLiA	RT @retahady: Bika selfie naik @GrabID mumpun...	Positif

```

df

```

	From-User	Text tweet	Sentiment
0	Grab Indonesia	@KissTaesan Baik Kak, btw ke DM 🐼-Mamat	Positif
1	Grab Indonesia	@_ahsinar Halo Kak, Mohon maaf atas ketidak...	Positif
2	Grab Indonesia	@laciq Hi Kak Aulia, bisa dikirimkan kode pe...	Positif
3	Grab Indonesia	@dhikareani Halo juga Kak 🐼 Maaf bgt yaa Kak ...	Positif
4	VaLiA	RT @retahady: Bika selfie naik @GrabID mumpun...	Positif
...	...	...	...
732	lalala---	@jogmfs Coba pakai kereta, lebih murah... lha...	Negatif
733	John Sutadja	Saya pantau di area mc. D dan sekitarnya, itu ...	Negatif
734	Jojon Joki Online	RT @kejO_Online: Perut orang rumah ga bisa dit...	Negatif
735	Jasa Cjek Langganan Orang	Perut orang rumah ga bisa ditunda untuk disi...	Negatif

Gambar 3 Import Data

3.2. Preprocessing Data

Setelah melakukan analisis data, proses selanjutnya adalah melakukan pembersihan data dalam keterangan Tabel 1 melalui proses *preprocessing*. Tahap *preprocessing* data bertujuan untuk mengubah data mentah menjadi data yang siap diolah seperti dalam penjelasan penelitian. Berikut adalah tahapan dari *preprocessing* data:

1. *Case folding* adalah tahapan teknik pembersihan yang dilakukan untuk menyisakan kata dengan karakter huruf kecil. Oleh karena itu, dilakukan penghilangan angka, tanda baca, emotikon, dan lain sebagainya serta pengubahan karakter ke huruf kecil.
2. *Word normalization* adalah mengubah kata-kata menjadi bentuk dasarnya
3. *Filtering* adalah tahapan untuk menghilangkan kata yang sering muncul namun tidak memiliki relevansi dengan teks atau dapat dikatakan kurang bermakna. Untuk kamus kata dan *stopwords* yang digunakan merupakan kombinasi dari list *stopwords* yang disediakan oleh *library Sastrawi* dan media internet [14].

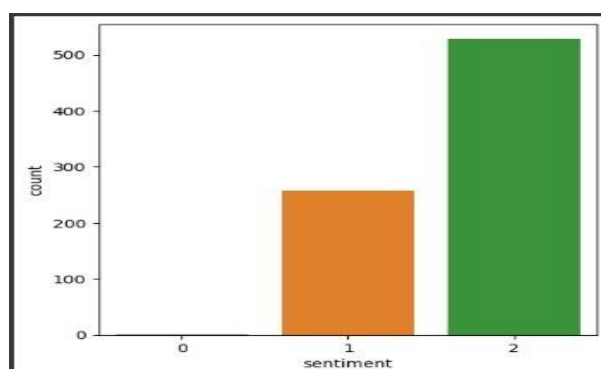
4. *Stemming* adalah tahapan untuk mengubah kata menjadi bentuk dasar dengan menghilangkan semua imbuhan. Sebagai contoh imbuhan yaitu: *meng-*, *me-*, *kan-*, *di-*, *pe-*, *peng-* dan lain sebagainya.

Tabel 1 Preprocessing Data

Proses	Input	Output
Case Folding	@Grabid Halo min aku mau infoin nih, klo barang yg tertinggal sudah ada di aku sekarang. Terimakasih gerakan cepatnya min, jadi sudah tak khawatir lagi. Dan mau pake Grab terus, Thanks Min □ □ https://t.co/x2JGdpxq9o	grabid halo min aku mau infoin nih klo barang yg tertinggal sudah ada di aku sekarang terimakasih gerakan cepatnya min jadi sudah tak khawatir lagi dan mau pake grab terus thanks min
Filtering	@Grabid Halo min aku mau infoin nih, klo barang yg tertinggal sudah ada di aku sekarang. Terimakasih gerakan cepatnya min, jadi sudah tak khawatir lagi. Dan mau pake Grab terus, Thanks Min □ □ https://t.co/x2JGdpxq9o	grabid halo min mau infoin nih klo barang yg tertinggal terimakasih gerakan cepatnya min sudah tak khawatir pake grab terus thanks min
Stemming	@Grabid Halo min aku mau infoin nih, klo barang yg tertinggal sudah ada di aku sekarang. Terimakasih g erakan cepatnya min, jadi sudah tak khawatir lagi. Dan mau pake Grab terus, Thanks Min □ □ https://t.co/x2JGdpxq9o	grabid halo min infoin nih klo barang yg tinggal terimakasih gerak cepat min sudah tak khawatir pake grab thanks min

3.3. Klasifikasi

Setelah proses *preprocessing* selesai, selanjutnya yaitu melakukan klasifikasi. Proses ini dilakukan menggunakan *Google Colab* dengan bahasa pemrograman *Python* serta menggunakan algoritma *Naïve Bayes Classifier* dan *Support Vector Machine*. Pada hasil proses klasifikasi data pada Gambar 4 yang diambil sebanyak 787 data yang dilabelkan menjadi label positif yang bernilai 2 dan negatif yang bernilai 1.



Gambar 4 Klasifikasi Sentimen Positif dan Negatif

Gambar 4 menjelaskan bahwa label positif mendapatkan jumlah sebanyak 528 data sedangkan label negatif sebanyak 258 data. Pada proses klasifikasi ini data yang diambil sebanyak 787 data kemudian data dilabelkan menjadi 2 kelas yaitu Positif dan Negatif. Didapatkanlah hasil

akurasi 96,0% data *training* dan akurasi data *testing* 84,17 % untuk algoritma *Support Vector Machine* dengan perbandingan 80:20 data *training* dan data *testing*, sedangkan untuk algoritma *Naïve Bayes* didapatkan akurasi 54.37% data *training* dan akurasi 66.45% data *testing* dengan perbandingan 80:20 data *training* dan data *testing*.

3.4. Pengujian dan Evaluasi

Pengujian dilakukan untuk mencari akurasi terbaik menggunakan *confusion matrix* melalui *library* NBC dan *library* SVM, dari model *training* dan *testing* berbanding 80:20 yang telah diklasifikasikan sebelumnya. Proses ini dilakukan menggunakan *Google Colab* dengan bahasa pemrograman *Python* serta menggunakan algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* dalam Gambar 5. Hingga menghasilkan matriks dengan ukuran 3x3 sebagai representatif kelas aktual dan kelas prediksi. Pengujian ini menghasilkan akurasi *recall*, *precision* serta *F1-Score* pada *consol spyder*. Berikut adalah hasil pengujian akurasi dengan algoritma NBC yang dapat dilihat pada gambar 5 berikut:

Model Evaluation NBC

```
[ ] # Perhitungan Performa Model untuk Data Latih
CM = confusion_matrix(y_train, predicted_train)

print(CM, '\n')
```

```
[[ 65 153]
 [134 277]]
```

```
print("ACCURACY\t:", accuracy_score(y_train, predicted_train)*100, "%\n")
```

```
print(classification_report(y_train, predicted_train))
```

```
ACCURACY      : 54.372019077901435 %
```

	precision	recall	f1-score	support
1	0.33	0.30	0.31	218
2	0.64	0.67	0.66	411
accuracy			0.54	629
macro avg	0.49	0.49	0.49	629
weighted avg	0.53	0.54	0.54	629

Gambar 5 Akurasi Data Training Algoritma NBC

```
# Perhitungan Performa Model untuk Data Uji
CM = confusion_matrix(y_test, predicted_test)

print(CM, '\n')
```

```
[[ 0  0  1]
 [ 0 11 29]
 [ 0 23 94]]
```

```
print("ACCURACY\t:", accuracy_score(y_test, predicted_test)*100, "%\n")
```

```
print(classification_report(y_test, predicted_test))
```

```
ACCURACY      : 66.45569620253164 %
```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.32	0.28	0.30	40
2	0.76	0.80	0.78	117
accuracy			0.66	158
macro avg	0.36	0.36	0.36	158
weighted avg	0.64	0.66	0.65	158

Gambar 6 Akurasi Data Testing Algoritma NBC

Gambar 5 menunjukkan bahwa hasil pengujian untuk algoritma *Naive Bayes Classifier* mendapatkan akurasi sebesar 54,37% untuk data *training*. Gambar 6 menunjukkan bahwa hasil

pengujian untuk algoritma *Naive Bayes Classifier* mendapatkan akurasi sebesar 66,45% untuk *data testing*. Pengujian ini menghasilkan akurasi *recall*, *precision* serta *F1-Score* pada *console* *spyder*. Gambar 7 menunjukkan bahwa hasil pengujian untuk algoritma *Support Vector Machine* mendapatkan akurasi sebesar 96.02% untuk data training. Gambar 8 menunjukkan bahwa hasil pengujian untuk algoritma *Support Vector Machine* mendapatkan akurasi sebesar 84.17% untuk data testing.

ACCURACY : 96.02543720190779 %					
	precision	recall	f1-score	support	
1	0.96	0.92	0.94	218	
2	0.96	0.98	0.97	411	
accuracy			0.96	629	
macro avg	0.96	0.95	0.96	629	
weighted avg	0.96	0.96	0.96	629	

Gambar 7 Akurasi Data Training Algoritma SVM

```
[ ] # Perhitungan Performa Model untuk Data Uji
CM = confusion_matrix(y_test, predicted_test)

print(CM, '\n')

print("ACCURACY\t:", accuracy_score(y_test, predicted_test)*100, "%\n")

print(classification_report(y_test, predicted_test))
```

[[0 0 1]					
[0 30 20]					
[0 4 103]]					

ACCURACY : 84.17721518987342 %					
	precision	recall	f1-score	support	
0	0.00	0.00	0.00	1	
1	0.88	0.60	0.71	50	
2	0.83	0.96	0.89	107	
accuracy			0.84	158	
macro avg	0.57	0.52	0.54	158	
weighted avg	0.84	0.84	0.83	158	

Gambar 8 Akurasi Data Testing Algoritma SVM

4. KESIMPULAN

Kesimpulan dari penelitian telah menunjukkan bahwa algoritma *Naive Bayes Classifier* dan *Support Vector Machine* dapat digunakan dalam melakukan klasifikasi data analisis sentimen terhadap aplikasi Grab. Dari hasil akurasi yang didapatkan dari algoritma *Naive Bayes* yaitu 54,37% untuk data *training* dan akurasi 66,45% untuk data *testing* dengan perbandingan 80:20 data *training* dan data *testing*. Dan hasil akurasi dari algoritma *Support Vector Machine* yaitu 96,02% untuk data *training* dan 84,17 % untuk data *testing* dengan perbandingan 80:20 data *training* dan data *testing*.

5. SARAN

Dalam penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* lebih baik dari *Naive Bayes* untuk kasus analisis sentimen terhadap aplikasi *Grab*. Jumlah data yang

digunakan sebanyak 787 data dari *crawling twitter* yang lebih banyak bersentimen positif yaitu 528 data positif. Diharapkan penulis lainnya dapat mengkomparasi dengan beberapa metode seperti *K-Nearest Neighbors* (KNN) seperti dalam penelitian perbandingan ketiga metode tersebut [15], metode *Deep Learning* dan lain sebagainya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim Redaksi Jurnal Teknik Politeknik Negeri Sriwijaya yang telah memberi kesempatan, sehingga artikel ilmiah ini dapat diterbitkan.

DAFTAR PUSTAKA

- [1] N. A. V. Taghfirul Azhima Yoga Siswa, "Komparasi Algoritma Klasifikasi untuk menentukan Evaluasi Kinerja Terbaik pada Status Akreditasi Sekolah/Madrasah Kalimantan Timur Berdasarkan IASP 2020," *JINTEKS (Jurnal Informatika Teknologi dan Sains)*, vol. 4, no. 3, Aug. 2022.
- [2] R. Wahyudi *et al.*, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *JURNAL INFORMATIKA*, vol. 8, no. 2, 2021, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [3] Hendriyana, "Komparasi Algoritma Naïve Bayes, Logistic Regression Dan Support Vector Machine pada Klasifikasi File Application Package Kit Android Malware," *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 11, no. 1, pp. 109–122, Apr. 2022.
- [4] I. M. K. ,Sri D. Hendriyana, "Analisis Perbandingan Algoritma Support Vector Machine, Naive Bayes, dan Regresi Logistik untuk Memprediksi Donor Darah," *Jurnal Teknologi Terpadu*, vol. 8, no. 2, pp. 121–126, Dec. 2022.
- [5] N. Hafidz, S. Anggraeni, W. Gata, "Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes," *Jurnal Jupiter*, vol. 12, no. 2, pp. 1–11, 2020.
- [6] H. Dwi Putra, L. Khairani, D. Hastari, P. Studi Sistem Informasi, F. Sains dan Teknologi, and C. Author, "Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa," in *Prosiding Seminas*, Institut Riset dan Publikasi Indonesia (IRPI), Aug. 2023, pp. 120–125. [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [7] D. Kristianti and Ma. Hariyadi, "Support Vector Machine (SVM) dan Algoritma Naïve Bayes (NB) Untuk Mengklasifikasi Keterlambatan Pembayaran Sumbangan Pendidikan di Madrasah Ibtidaiyah," *Jurnal Pendidikan Tambusai*, vol. 6, no. 2, 2022.
- [8] I. P. Rahayu, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, p. 296, Dec. 2022, doi: 10.30865/json.v4i2.5381.
- [9] D. Alita and R. A. Shodiqin, "Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 1–12, Feb. 2023, doi: 10.58602/jaiti.v1i1.20.
- [10] A. Supriyatna, W. Prima Mustika, A, "Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil," 2018. [Online]. Available: <http://tunasbangsa.ac.id/ejurnal/index.php/jsakti>
- [11] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *Jurnal Teknoinfo*, vol. 14, no. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.
- [12] D. Gunawan, D. Riana, D. Ardiansyah, F. Akbar, and S. Alfarizi, "Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023," *Jurnal Teknik Komputer AMIK BSI*, vol. 6, no. 1, Aug. 2020, doi: 10.31294/jtk.v4i2.

-
- [13] C. Aulia Sari, A. Sukmawati, R. Putri Aprilli, P. Sarah Kayaningtias, and N. Yudistira, "Perbandingan Metode Naive Bayes, Support Vector Machine Dan Decision Tree Dalam Klasifikasi Konsumsi Obat," vol. 3, pp. 33–41, May 2022, doi: 10.51402/jle.v3i1.47.
 - [14] B. Liu, "Sentiment Analysis and Opinion Mining," in *Graeme Hirst, Series Editor*, Morgan & Claypool Publishers, 2012.
 - [15] A. Putri, C. Syaficha Hardiana, E. Novfuja, F. Try Puspa Siregar, Y. Fatma, and R. Wahyuni, "Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *Institut Riset dan Publikasi Indonesia (IRPI) MALCOM: Indonesian Journal of Machine Learning and Computer Science Journal Homepage*, vol. 3, no. 1, pp. 20–26, 2023.