Analisis Sentimen Menggunakan Algoritma Naïve Bayes Classifier Studi Kasus: Aplikasi Teman Bus Di Play Store

Asrul Syam[1], Muhammad Syahlan Natsir[2], Syafruddin Muhtamar[3], Azizah Nur Fatima Azzahra[4]
[1,2, 3,4] Jurusan Teknik Informatika, Universitas Dipa Makassar, Makassar
Jln. Perintis Kemerdekaan KM. 9 Makassar
e-mail: [1]asrulsyam12@undipa.ac.id, [2]sahlan@dipanegara.ac.id, [3]syafuddinmuhtamar@gmail.com, [4]azizah.azzahra1302@gmail.com

Abstrak

Tujuan dari penelitian ini adalah untuk mengetahui akurasi algoritma Naïve Bayes Classifier dalam menganalisa sentimen ulasan pengguna dan pelayanan aplikasi Teman Bus yang ada di Play Store. Teman Bus merupakan salah satu aplikasi inisiasi Kementrian Perhubungan Republik Indonesia yang beroperasi di bidang transportasi dan informasi. Dari segi fungsionalitas dan layanannya, Teman Bus masih memiliki kekurangan dan kelebihan yang mengundang ribuan komentar yang bersifat positif, negatif dan netral dari masyarakat. Komentar-komentar tersebut tertuang pada Play Store dengan jumlah 2256 data. Hasil akurasi yang diperoleh setelah dilakukan analisis sentimen terhadap ulasan dan layanan aplikasi Teman Bus adalah 85% tanpa teknik oversampling dan 80% dengan menggunakan teknik random oversampling. Pemvisualisasian dari kata-kata yang berpengaruh berdasarkan frekuensi kemunculan kata dari kelas positif, netral dan negatif ditampilkan dalam bentuk Word Cloud.

Kata kunci— Teman Bus, Play Store, Naïve Bayes Classifier

Abstract

The aim of this study is to determine the accuracy of the Naïve Bayes Classifier algorithm in analyzing the sentiments of user reviews and the services of the Teman Bus application on the Play Store. Teman Bus is an application initiated by the Ministry of Transportation of the Republic of Indonesia which operates in the transportation and information sector. In terms of functionality and service, Teman Bus still has strengths and weaknesses which have attracted thousands of positive, negative and neutral comments from the public. These comments are listed on the Play Store with a total of 2256 data. The accuracy results obtained after sentiment analysis of the Friends Bus application reviews and services were 85% without the oversampling technique and 80% using the random oversampling technique. The visualization of influential words based on the frequency of occurrence of words from the positive, neutral and negative classes is displayed in the form of a Word Cloud.

Keywords— *Teman Bus, Play Store, Naïve Bayes Classifier, Word Cloud*

1. Pendahuluan

Era globalisasi pada saat ini telah membuat seluruh aktivitas keseharian dapat dilakukan secara online didukung oleh berbagai macam aplikasi yang telah diciptakan sesuai dengan kebutuhan penggunanya, aplikasi-aplikasi tersebut dapat menunjang keseharian dibidang informasi, e-commerce, trading, hingga transportasi sekalipun.

Salah satu aplikasi yang ada di bidang transportasi dan informasi adalah aplikasi Teman Bus. Aplikasi Teman Bus merupakan salah satu bentuk kemudahan yang diinisiasi oleh Kementrian Perhubungan Republik Indonesia. Aplikasi Teman Bus memudahkan penumpang untuk mendapatkan informasi rute, halte dan jadwal keberangkatan bus. Namun selayaknya manusia aplikasi pun memiliki kekurangan, aplikasi Teman Bus tidak sepenuhnya mendapat respon positif dari masyarakat baik itu dari segi fungsionalitasnya maupun layanan busnya. Ulasan pengguna mengenai Aplikasi Teman Bus dan pelayanannya mendapat banyak respon yang beragam yang tertuang pada kolom komentar yang tersedia di Play Store.

Dalam konteks ini, penelitian berikut akan menganalisa ulasan pengguna terhadap aplikasi Teman Bus yang ada di Play Store . Dalam analisis data terdapat banyak metode pengklasifikasian. Naive bayes classifier merupakan salah satu algoritma yang terdapat pada teknik pengklasifikasian. Peneliti

akan menganalisa ulasan pengguna terhadap aplikasi Teman Bus yang ada di Play Store berdasarkan sentimennya menggunakan algoritma Naïve Bayes Classsifier, di mana penelitian sentimen analisis berdasarkan ulasan pengguna aplikasi dan layanan Teman Bus yang ada di Play Store belum pernah dilakukan. Penggunaan algoritma Naïve Bayes Classifier pada penelitian ini cocok untuk mengolah short data dan teks. Ulasan tersebut diklasifikasikan menjadi tiga kategori sentimen yaitu; positif, netral, dan negatif.

Penelitian sebelumnya pernah membahas tentang penggunaan Naïve Bayes Classifier untuk meningkatkan akurasi analisis sentimen pada optimasi normalisasi kata yang ada di Twitter mengenai respon masyarakat terhadap layanan Teman Bus menggunakan Rapid Miner dengan hasil akurasi sebesar 0,776 atau 77% .

Hasil yang diharapkan dari penelitian ini adalah untuk mengetahui bagaimana algoritma Naïve Bayes Classifier dalam menganalisa sentimen ulasan pada aplikasi Teman Bus dan diharapkan pula dapat memvisualisasikan beberapa kata-kata yang berpengaruh dari ulasan pengguna pada aplikasi Teman Bus yang ada di Play Store.

2.1 Analisis Sentimen

Analisis sentimen atau biasa disebut juga dengan opinion mining adalah sebuah studi komputasi untuk mengenali dan mengekspresikan opini, sentimen, evaluasi, sikap, emosi, subjektifitas, penilaian atau pandangan yang terdapat dalam suatu teks. Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, netral, atau negatif (Ramdhani & Rahim, 2014). Sentimen analisis dapat dibedakan berdasarkan sumber datanya yaitu sentiment analisis pada level dokumen dan analisis sentimen pada level kalimat[1].

2.2 Aplikasi Teman Bus

Teman Bus merupakan suatu implementasi program Buy the Service dari Kementerian Perhubungan Republik Indonesia untuk pengembangan angkutan umum di kawasan perkotaan berbasis jalan yang menggunakan teknologi telematika yang andal dan berbasis non tunai untuk meningkatkan keselamatan dan keamanan serta kenyamanan mobilisasi. aplikasi mobile Teman Bus diinisiasi untuk memudahkan penumpang untuk mendapatkan informasi rute, halte dan jadwal keberangkatan bus. Teman Bus diharapkan bisa menjadi angkutan dengan layanan terbaik di Indonesia. Informasi jadwal keberangkatan dan kedatangan bus secara real time dapat melalui aplikasi Teman Bus di Play Store.[2].

2.3 Play Store

Play Store merupakan layanan penyedia konten digital milik Google yang menyediakan berbagai toko produk online seperti aplikasi, game, film atau musik, dan buku dengan beragam kategori. Adanya fitur rating dan ulasan menjadi wadah sebagai tolak ukur bagi para pengguna baru suatu aplikasi apakah aplikasi tersebut recommended untuk diunduh.

2.4 Scrapping Data

Scrapping data adalah suatu teknik yang digunakan untuk mengambil, menganalisa dan memproses suatu data dari suatu sistem atau dokumen yang berbeda. Teknik scrapping biasanya digunakan untuk mengambil data dari website yang bisa disebut web scrapping [3].

2.5 Preprocessing Text

Dokumen pada umumnya mempunyai struktur yang sembarangan atau tidak terstruktur. Oleh karena itu, diperlukan suatu proses yang dapat mengubah bentuk data yang sebelumnya tidak terstruktur ke dalam bentuk data yang terstruktur. Proses pengubahan ini dikenal dengan istilah preprocessing text [4]. Beberapa tahapan yang ada pada preprocessing text adalah: Case Folding, Tokenizing, Filtering, dan Stemming.

2.6 TF-IDF

TF-IDF (Terms Frequency-Inverse Document Frequency) merupakan suatu cara untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut.

Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen [5].

2.7 Naïve Bayes Classifier

Algoritma Naive Bayes Classifier merupakan metode pengklasifikasian turunan dari teorema Bayes yang sederhana, berguna untuk mencari nilai probabilitas atau peluang tertinggi untuk mengklasifikasikan data testing (uji) pada kategori yang paling tepat (Feldman, 2007). Tujuan algoritma Naïve Bayes Classifier yaitu untuk melakukan klasifikasi data pada kelas tertentu. Kelebihan Naïve Bayes Classifier adalah pengoperasiannya yang sederhana tetapi memiliki akurasi yang baik (Kusumadewi, 2009).

2.8 Multiclass Confusion Matrix

Multiclass confusion matrix merupakan bentuk confusion matrix yang memiliki kelas lebih dari dua dan menggambarkan kinerja suatu model pada serangkaian data uji yang nilai sebenarnya diketahui. Pada dasarnya confusion matrix memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. Confusion matrix dapat digunakan dalam menghitung berbagai performance metrics untuk mengukur kinerja model yang telah dibuat. Performance metrics yang sering digunakan adalah accuracy, precision, recall, dan F1-Score. Accuracy dapat menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar.

2.9 Teknik Random Oversamping

Permasalahan klasifikasi imbalanced merupakan permasalahan yang muncul, yaitu nilai kinerja klasifikasi menunjukkan nilai akurasi yang tinggi karena jumlah kelas mayor yang sangat banyak. Terkait dengan permasalahan imbalanced data, salah satu strategi yang dapat diterapkan adalah teknik sampling, baik oversampling maupun undersampling. Oversampling merupakan salah satu teknik sampling yang mampu meningkatkan akurasi dari pengklasifikasi untuk kelas minor secara random[6].

2.10 Word Cloud

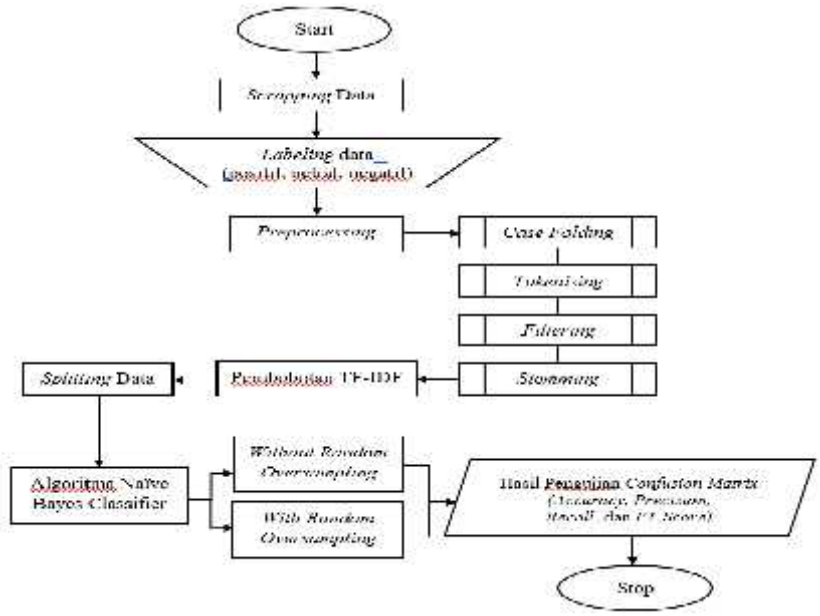
Word Cloud merupakan representasi grafis dari sebuah dokumen yang dilakukan dengan plotting kata-kata yang sering muncul pada sebuah dokumen pada ruang dua dimensi. Frekuensi dari kata yang sering muncul ditunjukkan melalui ukuran huruf kata tersebut. Semakin besar ukuran kata menunjukkan semakin besar frekuensi kata tersebut muncul dalam dokumen [7].

2. Metode Penelitian

2.1 Jenis dan Variabel Penelitian

Jenis penelitian yang digunakan adalah penelitian kualitatif yang di mana data ulasan pengguna pada aplikasi Teman Bus di analisis secara detail dan bersifat deskriptif. Pengklasifikasian data akan menggunakan pendekatan kuantitatif secara statistik yang dijelaskan dengan hasil perhitungan angka, tabel, atau diagram. Sedangkan variabel pada penelitian ini adalah variabel independen atau bebas yang di mana ulasan pengguna pada aplikasi Teman Bus yang terdapat di Play Store menjadi pengaruh pada sentimen analisisnya. Kemudian yang menjadi variabel terikat pada penelitian ini adalah kelas sentimen (positif, netral, dan negatif).

2.2 Perancangan Solusi



Gambar 1 *Flowchart* Perancangan Solusi

Adapun perancangan solusi pada gambar diatas dapat dijelaskan dengan diawali pengambilan data ulasan terhadap aplikasi dan layanan Teman Bus yang ada di Play Store dengan teknik *scrapping* menggunakan *software* Google Colab. Atribut yang diambil terdiri dari *username, rating* dan ulasan pengguna, data komentar tersebut berkisar pada tahun 2020 sampai dengan 2023. Kemudian dilakukan pelabelan data secara manual oleh satu orang aktor untuk mengetahui kelas dari tiap komentar yang telah di *scrapping,* kelas sentimen tersebut dibagi menjadi kelas positif, netral dan negatif. Tahap pengolahan data yang pertama dilakukan adalah *Preprocessing Text* atau pembersihan teks diantaranya yaitu *Case Folding, Tokenizing, Filtering,* dan *Stemming.* Setelah melewati rangkaian tahap *Preprocessing Text*, dataset tersebut kemudian dilakukan ekstraksi fitur atau pembobotan dengan *Term-Frequency – Inverse Document Frequency* untuk menentukan bobot masing-masing kata yang ada pada dataset. Selanjutnya data diolah dan diuji menggunakan metode *Naïve Bayes Classifier*. Pengujian dilakukan dengan dua teknik berbeda yaitu tanpa menggunakan teknik *random oversampling* dan dengan menggunakan teknik *random oversampling* untuk mengetahui penggunaan teknik yang terbaik dilihat dari hasil pengujian yang tertuang dalam bentuk *confusion matrix* berupa nilai *Accuracy, Precision, Recall,* dan *F1-Score.*

2.3 Pengumpulan Data

Data yang digunakan adalah dataset hasil dari scrapping menggunakan Google Collab mengenai ulasan pada aplikasi Teman Bus yang terdapat di Play Store. Data yang di Scrapping adalah data komentar dari tahun 2020 hingga 2023 sebanyak 2256 data.

2.4 Pelabelan Data

Proses pelabelan data dilakukan oleh satu aktor secara manual yaitu sang peneliti. Peneliti melakukan pelabelan secara manual dengan cara menentukan secara pribadi suatu data masuk kedalam kelas positif, kelas netral atau kelas negatif. Pelabelan data dan pengelompokan kelas dilakukan berdasarkan batasan yang ditentukan oleh aktor di antaranya :

1. Kelas Positif : dilihat dari isi ulasan yang mengandung kata yang bermakna positif atau pujian.
2. Kelas Netral : kelas yang berisi ulasan yang tidak mengarah ke positif maupun negatif, ulasan yang diluar konteks, ulasan yang bersifat saran, dan kalimat tanya.
3. Kelas Negatif : berisi ulasan yang mengandung kata yang bermakna negatif, celaan, kalimat yang berbentuk protes ataupun menyampaikan keluhan terhadap layanan dan aplikasi Teman Bus.

Saat proses pelabelan data dilakukan pula penghapusan beberapa data yang kiranya tidak dapat di proses oleh sistem, data yang dihapus berupa ulasan atau komentar yang hanya menggunakan *emoticon.* Sehingga dataset yang awalnya berjumlah 2370, telah tersisa 2256 data.

2.5 Preprocessing Text

1. *Case Folding*

Case folding dibutuhkan dalam mengkonversi keseluruhan teks yang ada dalam dataset menjadi bentuk standarnya. Proses ini akan menyamaratakan semua huruf yang ada di dataset menjadi huruf kecil(*Lowercase)*. Proses ini juga menghapus tanda baca dan karakter kosong(*Whitepace)* yang ada pada dataset ulasan aplikasi dan layanan aplikasi Teman Bus. Pada tabel 2 ditunjukkan contoh hasil *case folding* terhadap data *training.*

Tabel 1 Data *Training*

No.	Komentar	Kelas
1.	Sangat nyaman dan aman sebagai teman bepergian..	positif
2.	Aplikasi sering error, terus sering tidak akurat juga.	negatif
3.	Sudah bagus,tapi kendalanya adalah titik bus sering tidak sesuai....	netral

Tabel 2 Proses *Case Folding*

Komentar	hasil *case folding*
Sangat nyaman dan aman sebagai teman bepergian..	sangat nyaman dan aman sebagai teman bepergian
Aplikasi sering error, terus sering tidak akurat juga.	aplikasi sering error terus sering tidak akurat juga
Sudah bagus,tapi kendalanya adalah titik bus sering tidak sesuai....	sudah bagus tapi kendalanya adalah titik bus sering tidak sesuai

2. *Tokenizing*

Untuk proses analisis lebih lanjut maka proses *tokenizing* diperlukan untuk melakukan pemisahan kata dalam suatu kalimat. Teks yang telah menjadi potongan-potongan kata disebut sebagai token. Tabel 3 memperlihatkan hasil dari proses *tokenizing.*

Tabel 3 Proses *Tokenizing*

sebelum *tokenizing*	setelah *tokenizing*
sangat nyaman dan aman sebagai teman bepergian	[sangat, nyaman, dan, aman, sebagai, teman, bepergian,]
aplikasi sering error terus sering tidak akurat juga	[aplikasi, sering, error, terus, sering, tidak, akurat, juga]
sudah bagus tapi kendalanya adalah titik bus sering tidak sesuai	[sudah, bagus, tapi, kendalanya, adalah, titik, bus, sering, tidak, sesuai]

3. *Filtering*

Proses *filtering* atau *stopword* dalam penelitian ini merupakan tahap mengambil kata-kata yang penting dan menghilangkan atau membuang kata-kata yang tidak penting dalam kalimat tanpa merubah makna dari kalimat tersebut. Pada tabel 4 memperlihatkan hasil dari proses *filtering.*

Tabel 4 Proses *Filtering*

sebelum *filtering*	setelah *filtering*
[sangat, nyaman, dan, aman, sebagai, teman, bepergian,]	[sangat, nyaman, aman, sebagai, teman, pergi]
[aplikasi, sering, error, terus, sering, tidak, akurat, juga]	[aplikasi, error, tidak, akurat]
[sudah, bagus, tapi, kendalanya, adalah, titik, bus, sering, tidak, sesuai]	[bagus, tapi, kendalanya, titik, bus, tidak, sesuai]

4. Stemming

Setelah melalui proses *filtering*, selanjutnya adalah proses *stemming* yaitu mengembalikan suatu kata ke akar katanya. Dalam hal hal ini proses *stemming* dilakukan untuk menghilangkan imbuhan pada suatu kata. Gambar 4.6 memperlihatkan hasil dari proses *stemming*.

Tabel 5 Proses *Stemming*

sebelum *stemming*	setelah *stemming*
[sangat, nyaman, aman, sebagai, teman, pergi]	[sangat, nyaman, aman, sebagai, teman, pergi]
[aplikasi, error, tidak, akurat]	[aplikasi, error, tidak, akurat]
[sudah, bagus, tapi, kendalanya, titik, bus, sering, tidak, sesuai]	[bagus, tapi, kendala, titik, bus, tidak, sesuai]

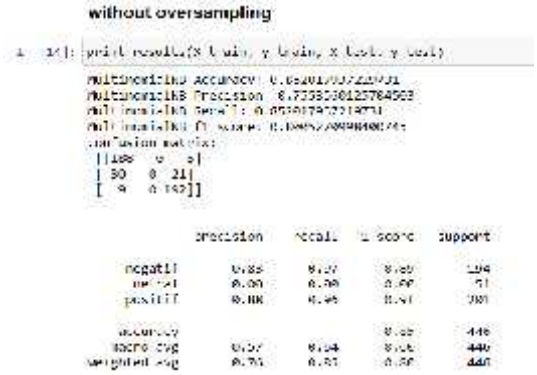
3. Hasil dan Pembahasan

3.1 Pengujian Algoritma Naïve Bayes Classifier

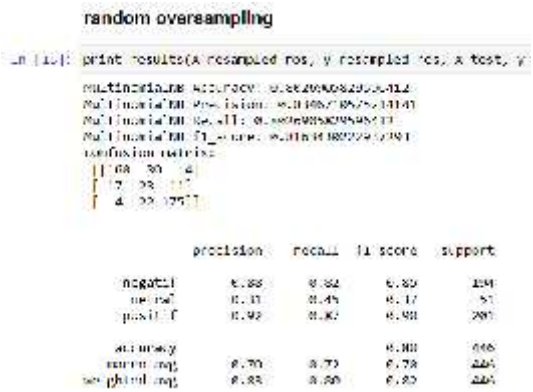
Data bersih yang didapatkan dari hasil preprocessing kemudian dilakukan ekstraksi fitur menggunakan TF-IDF*(Term Frequency-Inverse Document Frequency)* untuk menghitung bobot setiap kata yang umum digunakan dalam suatu dokumen, dalam hal ini adalah dataset ulasan pengguna aplikasi dan layanan Teman Bus yang ada di Play Store. Pembobotan kata menggunakan TF-IDF ini menjadi salah satu faktor yang mempengaruhi tingkat akurasi pada suatu algoritma. Pada penelitian ini, hasil dari TF-IDF yang dilakukan pada ulasan aplikasi dan layanan Teman Bus adalah sparse matrix yang artinya sebagian besar elemen dari dataset ulasan aplikasi dan layanan Teman Bus memiliki nilai nol.

Pada tahap pengujian terlebih dahulu dilakukan splitting data 80% 20% secara acak oleh sistem pada 2230 data yang telah diberi label sentimen. Dalam penelitian ini 80% dataset ulasan digunakan sebagai data *training*, dan 20% dataset ulasan digunakan sebagai data *testing*. Oleh algoritma *Naive Bayes Classifier*, data *training* digunakan untuk membentuk tabel probabilitas, dan data *testing* digunakan untuk menguji tabel probabilitas yang telah terbentuk atau dapat diasumsikan bahwa data *training* adalah data yang digunakan sebagai acuan untuk membangun model klasifikasi, sedangkan data *testing* adalah data yang digunakan untuk menguji performa dari model klasifikasi tersebut[8].

Proses evaluasi dari penelitian ini dilakukan dengan menggunakan *confusion matrix.* Saat dilakukan proses evaluasi, peneliti mencoba membandingkan proses evaluasi tanpa teknik *oversampling* dan dengan yang menggunakan teknik *random oversampling.* Hasil dari keduanya berbeda, seperti ditampilkan oleh gambar 3.1 berikut.



Gambar 2 *Confusion Matrix* tanpa *Random Oversampling*



Gambar 3 *Confusion Matrix* dengan *Random Oversampling*

Hasil penelitian berdasarkan pengujian oleh *confusion matrix* memperlihatkan hasil akurasi yang berbeda dari keduanya, Seperti yang ditunjukkan pada Gambar 2 hasil akurasi tertinggi terdapat pada *confusion matrix* tanpa *random oversampling* dengan nilai akurasi sebesar 0.852017937219731 atau 85%. Namun pada teknik tanpa *random oversampling* ada ketimpangan data yang terjadi yaitu dataset kelas netral hanya memperlihatkan angka 0 pada *confusion matrix*nya. Hal ini terjadi dikarenakan dalam dataset ulasan aplikasi dan pelayanan Teman Bus hasil klasifikasi data sentimen netral terlalu sedikit jika dibandingkan dengan data sentimen yang dimiliki oleh kelas positif dan kelas negatif.

Pada gambar 3 dilakukan teknik *random oversampling* yaitu membangkitkan data minoritas sebanyak data mayoritas untuk menyeimbangkan dataset yang digunakan. Hasil yang diperlihatkan adalah akurasi yang lebih rendah dibandingkan dengan tanpa menggunakan teknik *random oversampling* yaitu 80%. Nilai *confusion matrix* data kelas netral tertampil pada teknik ini, yang berarti dataset tersebut telah diseimbangkan.

Untuk lebih jelasnya perbandingan hasil pengujian algoritma *Naïve Bayes Classifier* tanpa teknik *random oversampling* dan dengan menggunakan teknik *random oversampling* dapat disajikan dalam tabel berikut.

Tabel 6 Perbandingan *Confusion Matrix*

Kelas	*Precision*	*recall*	*f1-score*	*Accuracy*	
Negatif	0.83	0.97	0.89	0.85	Tanpa *oversampling*
Netral	0.00	0.00	0.00		
Positif	0.88	0.96	0.91		
Negatif	0.88	0.82	0.85	0.80	Dengan *oversampling*
Netral	0.31	0.45	0.37		
Positif	0.92	0.87	0.90		

Pada tanpa *oversampling* algoritma *Naïve Bayes Classifier* memiliki niali akurasi sebesar 85% dengan nilai presisi pada kelas negatif 83%, kelas netral 0.00%, kelas positif 88%, nilai *recall* pada kelas negatif 97%, kelas netral 0.00%, kelas positif 0.96%, nilai *f1-score* pada kelas negatif 89%, kelas netral 0.00%, kelas positif 0.91%. Dan dengan teknik *oversampling* algoritma *Naïve Bayes Classifier* memiliki niali akurasi sebesar 80% dengan nilai presisi pada kelas negatif 88%, kelas netral 31%, kelas positif 92%, nilai *recall* pada kelas negatif 82%, kelas netral 45%, kelas positif 87%, nilai *f1-score* pada kelas negatif 85%, kelas netral 37%, kelas positif 90%.

3.2 Visualisasi Data

Gambar 4 Wordcloud *pada kelas positif*

Setelah melakukan pengujian terhadap penggunaan algoritma *Naïve Bayes Classifier* pada analisa sentimen ulasan aplikasi dan layanan Teman Bus dan mendapatkan hasil akurasinya, kemudian dilakukan pembentukan *wordcloud* untuk mengetahui frekuensi kemunculan kata-kata yang berpengaruh pada objek penelitian ini. Semakin besar suatu kata ditampilkan pada *wordcloud* menunjukkan bahwa semakin sering pula kata tersebut muncul pada suatu dataset. Seperti yang dapat dilihat pada gambar *Wordcloud* pada kelas positif menunjukkan kata "bantu", "bagus", "nyaman", dan "mantap" memiliki ukuran yang lebih besar dibandingkan kata lainnya yang berarti bahwa keempat kata tersebut memiliki frekuensi paling besar pada kelas positif dari dataset sentimen ulasan pada aplikasi Teman Bus di Play Store.

Gambar 5 Wordcloud *pada kelas Netral*

Seperti yang dapat dilihat juga pada gambar 5 *Wordcloud* pada kelas netral menunjukkan kata "tolong", "aplikasi", "moga", "mohon", "bingung" yang menunjukkan bahwa kata tersebut memiliki frekuensi yang cukup besar pada kelas netral dari dataset sentimen ulasan pada aplikasi Teman Bus di Play Store.

Gambar 6 Wordcloud *pada kelas Negatif*

Seperti yang dapat dilihat juga pada gambar 6 *Wordcloud* pada kelas negatif menunjukkan kata "aplikasi", "*error*", "*force close*", "*crash*", "ribet" yang menunjukkan bahwa kata tersebut memiliki frekuensi yang cukup besar pada kelas negatif dari dataset sentiment ulasan pada aplikasi Teman Bus di Play Store

4. Analisa Hasil Penelitian

Berdasarkan hasil eksprimen yang dialukukan pada data sinyal wicara hasil rekaman ternyata suatu sinyal wicara memiliki suatu cirri yang istimewa pada suatu kawasan tertentu. Bahwa sinyal wicara merupakan suatu fungsi yang bergantung pada waktu (time invariant).

Hasil pengujian yang telah dilakukan pada database sinyal wicara dari perekaman yang dilakukan di dalam ruangan kedap suara dengan proses algoritma CCA dengan bantuan *software* Matlab dapat diterapkan untuk mendeteksi , menghilangkan atau meminimalisis noise pada sinyal wicara. Implementasinya juga dapat membagi sinyal menjadi beberapa bagian untuk komputasi pada daerah domain frekuensi. Sekaligus juga untuk mencari komponen frekuensi dari sinyal yang bercampur dan tersembunyi oleh noise dalam sinyal waktu domain. Algoritma CCA yang identic dengan algoritma *Autocorrelation* merupakan salah satu metode untuk mentransformasi sinyal wicara (*speech*) dalam domain waktu menjadi sinyal domain frekuensi yang artinya proses perekaman suara disimpan dalan bentuk digital berupa gelombang spektrum suara yang berbasis frekuensi sehingga lebih mudah dalam menganalisa spektrum. Sedangkan secara kompleks setelah diproses dengan domain frekuensi tersebut, maka dapat digunakan untuk menentukan nilai *avarerange* Energi untuk proses pemrosesan selanjutnya atau melakukan riset selanjutnya.

4. Kesimpulan

Kesimpulan akhir dari penelitian yang telah dilakukan mengenai analisis sentimen ulasan pada aplikasi Teman Bus di Play Store menggunakan algoritma Naïve Bayes Classifier adalah hasil akurasi yang menjadi patokan seberapa baik algoritma Naïve Bayes Classifier dalam menganalisis sentimen ulasan pada aplikasi Teman Bus di Play Store menggunakan software Jupyter Notebook. Hasil akurasi yang diperoleh adalah 85%, dan jika dilakukan teknik random oversampling pada dataset maka diperoleh hasil akurasi 80%. kedua hasil akurasi tersebut menunjukkan bahwa penggunaan algoritma Naïve Bayes Classifier dalam menganalisis sentimen ulasan pada aplikasi Teman Bus di Play Store menggunakan software Jupyter Notebook sudah cukup baik. Penggunaan Word cloud dalam penelitian ini menampilkan beberapa kata-kata yang berpengaruh terhadap frekuensi kemunculan kata pada masing-masing kelasnya positif, netral, dan negatif dalam dataset analisa sentimen Teman Bus.

5. Saran

Algoritma yang digunakan pada penelitian ini termasuk dalam metode pengklasifikasian, diharapkan akan adanya penelitian lebih lanjut mengenai objek dari penelitian ini menggunakan salah satu metode klasterisasi yang ada sehingga dapat menjadi pembanding algoritma seperti apa yang dapat menganalisa sentimen ulasan pada aplikasi Teman Bus di Play Store dengan lebih baik.

Daftar Pustaka

[1] D. Ali Imron, 'Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier', 2019.
[2] Teman Bus, 'Teman Bus #KamiAdaUntukAnda', 2020. https://temanbus.com/ (accessed Feb. 19, 2023).
[3] M. A. Adli and L. Firgia, 'Rancang Bangun Web Scraping Pada Media Online Berita Nasional', Agustus 2018.
[4] F. S. Jumeilah, 'Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian', J. RESTI Rekayasa Sist. Dan Teknol. Inf., vol. 1, no. 1, pp. 19–25, Jul. 2017, doi: 10.29207/resti.v1i1.11.
[5] M. Nurjannah and I. F. Astuti, 'Penerapan Algoritma Term Frequency-Invers Document Frequency (TF-IDF) Untuk Text Mining', 2013.
[6] A. Y. Triyanto and R. Kusumaningrum, 'Implementasi Teknik Sampling untuk Mengatasi Imbalanced Data pada Penentuan Status Gizi Balita dengan Menggunakan Learning Vector Quantization', J. IPTEKKOM J. Ilmu Pengetah. Teknol. Inf., vol. 19, no. 1, p. 39, Jul. 2017, doi: 10.33164/iptekkom.19.1.2017.39-50.
[7] T. Kurniawan, 'Implementasi text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naive Bayes Classifier dadn Support Vector Machine', 2017.
[8] D. Darwis, N. Siskawati, and Z. Abidin, 'PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL', J. Tekno Kompak, vol. 15, no. 1, p. 131, Feb. 2021, doi: 10.33365/jtk.v15i1.744.