

Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes

Deni Wijaya, Rizki Adi Saputra, Faldy Irwiensyah*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹deniwijaya487@gmail.com, ²belajar7706@gmail.com, ^{3,*}faldy@uhamka.ac.id

Email Penulis Korespondensi: faldy@uhamka.ac.id

Abstrak—Transformasi digital telah menjadi faktor utama perubahan dalam berbagai aspek kehidupan modern, termasuk bisnis, pendidikan, dan pemerintahan. Di era transformasi digital saat ini, pemerintah berusaha meningkatkan efisiensi dan pelayanan kepada masyarakat melalui penerapan berbagai inovasi teknologi. Penerapan teknologi digital dalam layanan publik semakin meluas, termasuk di sektor pelayanan administrasi seperti Samsat Digital Nasional (SIGNAL) yang memungkinkan masyarakat dapat melakukan pembayaran pajak kendaraan secara online melalui aplikasi SIGNAL. Evaluasi pengguna terhadap aplikasi ini dapat memberikan wawasan penting bagi penyelenggara layanan. Tujuan penelitian ini untuk mengetahui sentimen pengguna dari aplikasi Samsat Digital Nasional di platform Google Playstore menggunakan algoritma Naïve Bayes. Metode ini digunakan untuk mengklasifikasikan ulasan pengguna ke dalam kategori sentimen positif dan sentimen negatif. Dari 2000 ulasan yang diambil, 1,665 ulasan dikategorikan sebagai positif dan 335 ulasan sebagai negatif setelah pelabelan manual. Proses preprocessing data menggunakan RapidMiner mencakup cleaning, transform cases, tokenizing, filter stopword, filter token by length, dan stemming. Pembobotan TF-IDF digunakan untuk memberi bobot pada setiap kata dalam dokumen. Evaluasi model Naïve Bayes menghasilkan akurasi sebesar 63.61%, dengan 307 True Positif, 74 True Negatif, 26 False Positif, dan 192 False Negatif. Precision sebesar 92.19% dan recall sebesar 61.52%. Analisis keseluruhan menunjukkan bahwa ulasan pengguna cenderung lebih positif terhadap aplikasi SIGNAL, meskipun terdapat beberapa ulasan negatif. Kesimpulan ini memberikan gambaran tentang persepsi positif pengguna terhadap aplikasi tersebut.

Kata Kunci: Samsat; Signal; Naïve Bayes; Rapidminer; Ulasan; Data; Analisis Sentimen, Google Playstore

Abstract—Digital transformation has become a major factor of change in various aspects of modern life, including business, education, and government. In the current era of digital transformation, the government is trying to improve efficiency and services to the community through the implementation of various technological innovations. The application of digital technology in public services is increasingly widespread, including in the administrative service sector such as the National Digital Samsat (SIGNAL) which allows people to make online vehicle tax payments through the SIGNAL application. User evaluations of this application can provide important insights for service providers. This research aims to analyze the sentiment of user reviews of the National Digital Samsat application on the Google Playstore platform using the Naïve Bayes algorithm. This method is used to classify user reviews into positive and negative sentiment categories. From 2000 reviews taken, 1,665 reviews were categorized as positive and 335 reviews as negative after manual labeling. Data preprocessing using RapidMiner includes cleaning, transform cases, tokenizing, stopword filter, token by length filter, and stemming. TF-IDF weighting is used to give weight to each word in the document. Evaluation of the Naïve Bayes model resulted in an accuracy of 63.61%, with 307 True Positives, 74 True Negatives, 26 False Positives, and 192 False Negatives. Precision was 92.19% and recall was 61.52%. The overall analysis shows that user reviews tend to be more positive towards the SIGNAL app, although there are some negative reviews. This conclusion gives an idea of users' positive perception of the app.

Keywords: Samsat; Signal; Naïve Bayes; Rapidminer; Reviews; Data; Sentiment Analysis, Google Playstore

1. PENDAHULUAN

Transformasi digital telah menjadi faktor utama perubahan dalam berbagai aspek kehidupan modern, termasuk bisnis, pendidikan, dan pemerintahan [1]. Dengan penerapan teknologi, cara kita berinteraksi, berkomunikasi, dan bekerja telah mengalami perubahan [2]. Di tingkat pemerintahan, transformasi digital memungkinkan penyediaan layanan publik yang lebih efisien dan transparan, serta memfasilitasi partisipasi warga dalam proses pengambilan keputusan. Pada era serba digital saat ini, pemerintah berusaha meningkatkan efisiensi dan pelayanan kepada masyarakat melalui penerapan berbagai inovasi teknologi[3].

Berdasarkan UU Lalu Lintas dan Angkutan Jalan No. 22 Tahun 2009, Sistem Administrasi Manunggal Satu Atap atau yang disebut dengan SAMSAT merupakan sebuah sistem administrasi yang diterapkan di Indonesia untuk mengatur dan mengelola administrasi perpajakan dan penerimaan negara bukan pajak atas kendaraan bermotor. Tujuan utama Samsat adalah untuk memastikan kepatuhan pemilik kendaraan terhadap kewajiban perpajakan mereka, serta memberikan pelayanan administrasi yang efisien dan terpadu. Melalui Samsat, pemerintah dapat mengawasi dan mengelola pendapatan dari pajak kendaraan bermotor untuk mendukung pembangunan infrastruktur dan layanan publik lainnya.

Untuk meningkatkan layanan administratif kepada masyarakat, SAMSAT telah memperkenalkan aplikasi digital yang disebut SAMSAT DIGITAL NASIONAL (SIGNAL). Aplikasi SIGNAL juga tersedia untuk diunduh melalui Google Play Store [4]. Berdasarkan pada penjelasan situs remi samsatdigital.id menyatakan bahwa aplikasi SIGNAL adalah suatu aplikasi yang memungkinkan masyarakat melakukan transaksi administrasi kendaraan bermotor secara digital, seperti pembayaran pajak dan perpanjangan STNK, melalui perangkat seluler dengan ujuan utamanya aplikasi ini adalah memberikan kemudahan, kecepatan, dan efektivitas dalam proses administrasi kendaraan bermotor. Berdasarkan Data yang tertera pada Google Playstore, aplikasi ini sudah lebih dari satu juta kali diunduh, dengan menerima lebih dari 113 ribu ulasan dan rating nilai sebesar 4,6.

Berdasarkan ulasan pada Google PlayStore, banyak pengguna yang merasa terbantu dengan adanya aplikasi SIGNAL ini karena pembayaran pajak dapat dilakukan kapanpun dan dimanapun. Namun, ada juga pengguna yang

mengungkapkan ketidakpuasan terhadap kinerja aplikasi ini, seperti ketidak akuratan dalam proses pembayaran pajak, kesulitan dalam verifikasi, dan proses yang memakan waktu. Hal ini menunjukkan bahwa terdapat area-area yang perlu diperbaiki agar aplikasi dapat memenuhi harapan pengguna. Oleh karena itu, kesuksesan implementasi SAMSAT Digital Nasional tidak hanya bergantung pada aspek teknis saja. Evaluasi terhadap pengalaman pengguna dan respon masyarakat terhadap aplikasi ini menjadi faktor kunci dalam menilai efektivitas dan penerimaan aplikasi tersebut di kalangan pengguna [5]. Oleh karena itu, diperlukan penelitian mendalam untuk menganalisis sentimen masyarakat terhadap aplikasi SIGNAL (SAMSAT Digital Nasional) melalui ulasan yang ditinggalkan oleh pengguna pada aplikasi Google Playstore.

Analisis sentimen adalah suatu pendekatan dalam pengolahan bahasa alami yang bertujuan untuk mengekstrak dan memahami opini ataupun perasaan yang terkandung dalam teks. Metode ini digunakan untuk menilai apakah suatu pernyataan atau dokumen bersifat positif, negatif, atau netral [6]. Dalam analisis sentimen, biasanya melibatkan penggunaan teknik pemrosesan bahasa alami, pembelajaran mesin, dan algoritma klasifikasi untuk mengidentifikasi kata-kata kunci, frasa, atau konteks yang menunjukkan sentimen tertentu [7]. Secara umum, terdapat lima tahapan dalam melakukan analisis ini, yaitu pengumpulan data, pra-pemrosesan, pemilihan fitur, klasifikasi, dan evaluasi. Terdapat metode klasifikasi data yang umum digunakan termasuk *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), dan *Naïve Bayes* [8].

Naïve Bayes, sebagai salah satu metode yang terkenal dalam domain *studi machine learning* dalam hal klasifikasi dan menonjol karena keefektifannya, terutama dalam pengolahan teks [9]. Meskipun sederhana, metode ini dapat dengan cepat menangani masalah klasifikasi dan mudah diimplementasikan [10]. Dengan menghitung probabilitas posterior dari setiap kategori atau label berdasarkan fitur-fitur yang diamati dari data pelatihan, *Naïve Bayes* dapat digunakan dalam berbagai aplikasi seperti analisis sentimen, klasifikasi teks, dan pengenalan pola [11].

Penelitian ini menggunakan metode *Naïve Bayes* untuk menganalisis sentimen ulasan aplikasi SIGNAL. Data yang dianalisis diperoleh dari ulasan pengguna di Google PlayStore melalui teknik *scraping* menggunakan Google Colab, dengan total 2000 ulasan terbaru. [12]. Setelah itu, data akan disimpan dengan format csv yang kemudian dikelompokkan menjadi kategori positif dan kategori negatif [12]. Setelah proses itu, dataset akan diproses menggunakan aplikasi RapidMiner Studio [12]. RapidMiner adalah sebuah platform perangkat lunak yang digunakan untuk analisis data, pemodelan prediktif, dan penggalian data (*data mining*). Ini menyediakan lingkungan yang kuat dan intuitif untuk mengolah data, membangun model analisis, dan menghasilkan wawasan yang berharga dari berbagai jenis data [13].

Penelitian terkait analisis sentimen juga sudah banyak dilakukan sebelumnya seperti yang dilakukan oleh Alam Sulistyio yang membahas tentang analisis sentimen terhadap aplikasi MyPertamina, mendapatkan akurasi 91%, presisi 92%, dan *recall* 100% dari 3948 data ulasan dengan menggunakan metode *Naïve Bayes* [14]. Hasibuan dan Heriyanto melakukan juga penelitian dari metode yang sama terhadap ulasan aplikasi Amazon Shopping dengan tingkat akurasi sebesar 82,15%, presisi 72%, dan *recall* 83.49% [15]. Penelitian analisis sentimen lainnya juga dilakukan oleh Nurtikasari terhadap film ngeri-ngeris sedap pada platform Twitter menggunakan algoritma *Naïve Bayes* didapat hasil akurasi mencapai 75% [16]. Sentimen analisis yang menggunakan algoritma serupa juga telah dilakukan Pratmanto, dkk terhadap aplikasi Shopee mendapat hasil 96,667% untuk tingkat akurasinya [17].

Penelitian mengenai perbandingan terhadap metode *Naïve Bayes* dengan metode lainnya seperti *K-Nearest Neighbors* (KNN) dalam melakukan analisis sentimen dilakukan oleh Ramadhani dan Wahyudin dengan akurasi sebesar 90.77% untuk metode *Naïve Bayes* dan akurasi sebesar 74.77% untuk metode KNN [18].

Perlu dilakukan penelitian terhadap aplikasi pelayanan publik untuk memahami persepsi dan kepuasan pengguna, mengidentifikasi serta memperbaiki kekurangan, dan memastikan aplikasi tersebut dapat memberikan manfaat optimal. Seperti penelitian analisis sentimen terhadap aplikasi Digital Korlantas POLRI yang dilakukan oleh Putra dan Febriawan menggunakan Algoritma *Naïve Bayes*. Hasil penelitian tersebut menunjukkan lebih banyak sentimen negatif dengan tingkat akurasi 70,33%, presisi 56,34%, dan *recall* sebesar 74,77%.

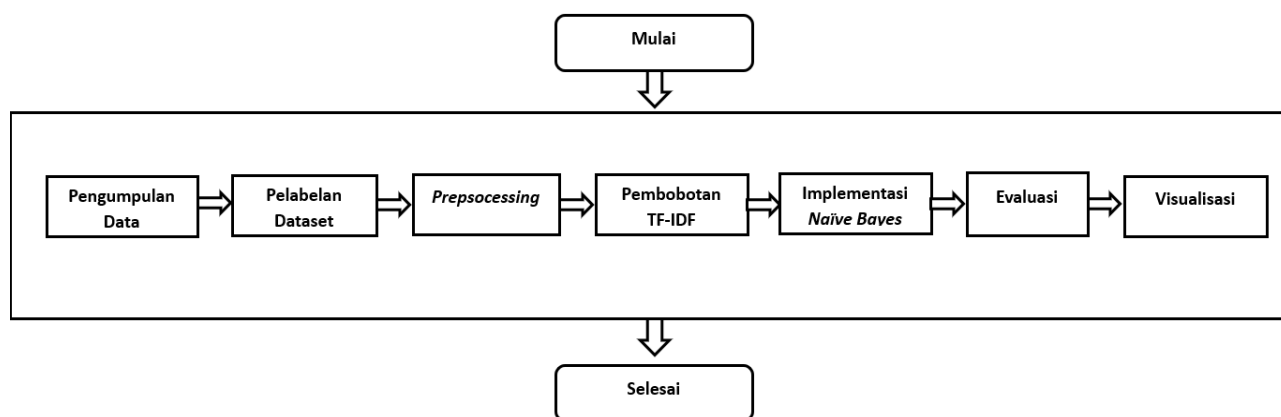
Dari permasalahan yang ada, penelitian ini dilakukan untuk mengetahui respon pengguna terhadap aplikasi pelayanan publik yaitu aplikasi SIGNAL melalui ulasan yang mereka berikan di Google Play Store. Dalam konteks ini, penelitian akan difokuskan pada analisis sentimen ulasan pengguna dengan menggunakan algoritma *Naïve Bayes*. Algoritma ini dipilih karena pada penelitian sebelumnya yang sudah dipaparkan diatas, algoritma ini memiliki akurasi yang dapat dikatakan tinggi dan bahkan lebih unggul dibandingkan algoritma klasifikasi lain seperti algoritma KNN sebagaimana penelitian yang dilakukan oleh Ramadhani dan Wahyudin sebelumnya. Melalui penggunaan algoritma ini, penelitian bertujuan untuk mengeksplorasi dan memahami pandangan pengguna terhadap aplikasi SIGNAL, termasuk aspek positif dan negatif yang dapat diidentifikasi dari ulasan mereka.

Dengan mengikutsertakan perspektif pengguna dalam penelitian ini, diharapkan dapat terwujud SAMSAT Digital Nasional (SIGNAL) yang tidak hanya bersifat teknis dan fungsional, tetapi juga responsif terhadap kebutuhan dan harapan masyarakat. Dengan demikian, diharapkan aplikasi ini bisa menjadi solusi yang lebih holistik dan terintegrasi dalam memenuhi kebutuhan administrasi kendaraan bermotor di Indonesia.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Algoritma *Naïve Bayes* digunakan dalam penelitian ini untuk menganalisis sentimen pengguna aplikasi SIGNAL berdasarkan ulasan yang ada di Google PlayStore dan menerapkannya pada *software* RapidMiner. Berikut Gambar 1 yang menjadi alur proses penelitian kali ini.



Gambar 1. Tahapan Penelitian

Pada Gambar 1, penelitian dimulai dari tahapan pengumpulan data sampai dengan visualisasi.

2.2 Pengumpulan Data

Dalam proses ini melibatkan penghimpunan informasi yang relevan untuk tujuan penelitian atau analisis [19]. Metode yang sering dipakai mencakup survei, wawancara, observasi, *web scraping* dan API (*Application Programming Interface*) [20]. Proses ini melibatkan pengumpulan sejumlah besar data dari berbagai *platform* media sosial, situs web, dan sumber online lainnya. Setelah data terkumpul, langkah berikutnya dilakukan pengolahan dan analisis untuk mendapatkan pemahaman yang mendalam [21].

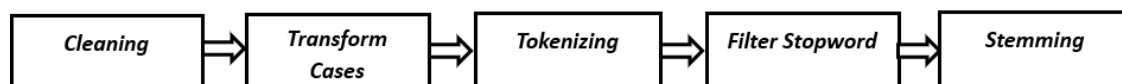
2.3 Pelabelan Data

Pelabelan data adalah proses memberikan kategori atau label pada setiap data berupa positif, negatif maupun netral, sehingga mempermudah analisis dan penggunaannya. Proses ini dapat dilakukan secara manual oleh manusia atau otomatis menggunakan algoritma. Keakuratan dan konsistensi label sangat penting dalam proses ini untuk memastikan data terstruktur dengan baik dan siap digunakan untuk tujuan tertentu. [22]

2.4 Preprocessing

Preprocessing adalah tahap awal yang mesti dilakukan sebelum menganalisis data. *Preprocessing* data bertujuan untuk merubah data yang tidak terstruktur menjadi terstruktur guna mempermudah proses analisis. Proses *preprocessing* data melibatkan beberapa tahapan seperti *cleaning*, *transform cases*, *tokenizing*, *stopword removal* dan *stemming*. [23]

- Cleaning*, Pada tahap ini, karakter selain huruf, termasuk tanda baca dan angka, akan dihapus dari data.
- Transform Cases*, Pada tahap ini, semua kata yang ada dalam dokumen nantinya akan diubah hurufnya menjadi huruf kecil semua agar teks berada dalam format standar.
- Tokenizing*, Pada tahap ini, dokumen akan dipecah menjadi bentuk tunggal dengan memotong kata-kata yang dipisahkan oleh spasi.
- Filter Stopword*, tahap ini akan menghapus semua kata yang dianggap tidak mempunyai makna, seperti kata sambung.
- Stemming*, Pada tahap ini, kata dalam dokumen akan disederhanakan menjadi bentuk dasarnya dengan menghilangkan imbuhan. Sumber [24]



Gambar 2. Tahap Preprocessing

Berikut Gambar 2 yang merupakan tahapan-tahapan yang ada dalam *preprocessing* data. Pada Gambar 2, tahap *presprocessing* dilakukan mulai dari proses *clening* sampai *stemming*.

2.5 Term Frequency Inverse Document Frequency (TF-IDF)

Pemberian bobot pada setiap kata dilakukan melalui TF-IDF. Untuk setiap term akan dilakukan perhitungan bobot secara keseluruhan pada setiap dokumen, untuk menentukan banyaknya kata yang paling sering digunakan dalam dokumen. Semakin banyak suatu kata digunakan, maka nilai atau bobot dari term tersebut akan semakin tinggi [25].

2.5 Implementasi Naïve Bayes

Penerapan Naive Bayes dalam praktik melibatkan beberapa langkah utama. Awalnya, dataset perlu disiapkan dengan baik, termasuk membagi data menjadi set pelatihan dan pengujian untuk mengevaluasi kinerja model [26]. Algoritma klasifikasi *Naïve Bayes* adalah model yang umum digunakan untuk mengklasifikasikan data dengan menggunakan perhitungan probabilitas. Dengan menghitung probabilitas untuk setiap kata dalam dokumen atau dataset, algoritma ini dapat menghasilkan dua klasifikasi, yaitu sentimen positif dan negatif [27]. Kelebihan penggunaan metode ini adalah memerlukan sedikit data training untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi, karena diasumsikan sebagai variabel yang bebas [28]. Berikut rumus perhitungan dari metode Naive Bayes.

$$P(c|x) = \frac{P(x|c) \times P(c)}{P(x)} \quad (1)$$

$P(c|x)$ adalah nilai probabilitas yang akan kita cari, $P(c)$ adalah kelas probabilitas yang berdasarkan hipotesis, $P(x|c)$ adalah prediktor probabilitas yang berdasarkan pada kelas yang diberikan dan $P(x)$ adalah prediktor probabilitas [7].

2.6 Evaluasi

Tahap evaluasi dilakukan untuk mengukur performa dari suatu model klasifikasi dengan menggunakan metode *Confusion Matrix* [29]. Saat mengukur efisiensi, terdapat empat istilah yang merepresentasikan hasil dari proses klasifikasi, yaitu TP, TN, FP, dan FN [30]. Setelah mendapatkan nilai dari keempat istilah tersebut, selanjutnya dapat dihitung nilai akurasi, nilai presisi, dan juga *recall* menggunakan rumus berikut :

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

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

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

Keterangan :

TP = True positive
TN = True Negative
FP = False Positive
FN = False Negative

2.7 Visualisasi

Visualisasi data merupakan cara untuk menggambarkan informasi dan data secara grafis, dengan tujuan untuk mempermudah pemahaman dan eksplorasi informasi [31]. Ada berbagai jenis visualisasi data, mulai dari grafik batang, diagram lingkaran, *WordCloud*, hingga visualisasi multidimensi yang lebih kompleks seperti peta panas dan diagram pencar [32]. Tujuan utama dari visualisasi data adalah untuk membantu pengguna memahami makna dan struktur data, serta mendukung pengambilan keputusan yang lebih baik dengan mengidentifikasi pola-pola yang tersembunyi [33].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan cara *scraping data* yang dilakukan menggunakan Google Colab untuk mengambil data ulasan aplikasi SIGNAL pada Google PlayStore dan kemudian disimpan dalam format *csv*. Berikut Gambar 3 adalah alur proses pengumpulan data yang diambil dari ulasan playstore menggunakan google Colab dan disimpan dalam format *csv*.



Gambar 3. Proses Pengumpulan Data

Dalam fase ini, dilakukannya *scraping* data menggunakan Bahasa pemrograman Python, dengan menggunakan *Google Colaboratory* dan *package google-play-scrapers* [34]. Pada proses pengambilan data ini, peneliti mengumpulkan data ulasan sebanyak 2000 data ulasan terbaru dan beberapa informasi yang diperlukan, seperti komentar ulasan, dan *rating*. Selanjutnya, data tersebut disimpan dalam format *file .csv*. Berikut Gambar 4 yang merupakan proses *scraping data*. Pada Gambar 4, proses *scraping data* dilakukan dengan menggunakan *library google-play-scrapers* pada bahasa pemrograman Python dengan ketentuan bahasa Indonesia, data yang diambil merupakan data terbaru dengan jumlah 2000 data ulasan.

```

from google_play_scraper import app

import pandas as pd

import numpy as np

#scrape jumlah ulasan yang diinginkan
from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'app.signal.id',
    lang='id', #disini kita mau men scrape data ulasan aplikasi shopee yang berada di google play store
    country='id', #kita setting bahasa nya menjadi bahasa indonesia
    sort=Sort.NEWEST, #sorting data yang terbaru
    count=2000, # disini jumlah ulasan yang mau kita ambil ada seribu
    filter_score_with=None # # kemudian di filter_score kita gunakan None untuk mengambil semua score atau rating bintang 1 sampai 5

data.to_csv("2000 Dataset Ulasan Aplikasi SIGNAL Terbaru.csv", index = False , encoding='utf-8')

```

Gambar 4. Proses Scraping Data

3.2 Pelabelan Data

Pada tahapan ini, aplikasi Microsoft excel 2019 digunakan peneliti untuk melakukan pelabelan data secara manual. Pelabelan data dikategorikan menjadi dua kategori yaitu positif dan negatif berdasarkan ulasan komentar pengguna. Kategori positif mengacu pada data ulasan yang memiliki atribut atau nilai yang dianggap menguntungkan atau diinginkan, sedangkan data negatif mengacu pada data yang memiliki atribut atau nilai yang dianggap merugikan atau tidak diinginkan [35]. Pada data ulasan aplikasi SIGNAL, terdapat 1.665 ulasan positif dan 335 ulasan negatif. Berikut Tabel 1 yang merupakan contoh hasil pelabelan data.

Tabel 1. Pelabelan Data

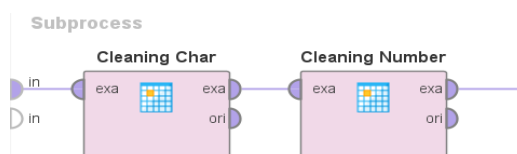
Ulasan	Label
Aplikasinya bagus sangat membantu sekali	Positif
Bagus aplikasinya bisa mempermudah buat bayar pajak,gk pake antri & langsung dikirim kerumah.	Positif
Mantap pokonya..solusi anti ribet Terimakasih Signal Terimakasih juga Polda Metro Jaya	Positif
Cs nya ancurr bot doang	Negatif
Tidak bisa melakukan transaksi pembayar gilira di komplek yg jawab boot	Negatif

Pada Tabel 1, terdapat label positif dan negatif tergantung pada isi ulasan. Apabila pengguna merasa puas maka label bernilai positif dan sebaliknya apabila pengguna merasa tidak puas maka label bernilai negatif.

3.3 Preprocessing Data

Setelah melakukan pelabelan, selanjutnya adalah tahap *preprocessing data* yang dilakukan menggunakan aplikasi *RapidMiner*. Pada penelitian kali ini terdapat beberapa tahapan dalam preprocessing, yaitu sebagai berikut ini :

- a. *Cleaning* dilakukan untuk menghapus semua karakter, terkecuali alfabet dari ulasan komentar. Termasuk tanda baca dan angka, akan dihapus dari data. Pada Gambar 5 terdapat *cleaning char* dan *cleaning number* yang digabung kedalam operator *subprocess*. *Cleaning Char* berfungsi untuk membersihkan data dari karakter atau simbol yang dianggap tidak penting seperti [-!"#\$%&'()*+,-./:;<=>?@[\\|_`{|}~]. Sedangkan *Cleaning Number* berfungsi untuk membersihkan data dari digit *number* atau angka seperti angka 0-9. Berikut Gambar 5 yang merupakan tahapan *cleaning data* dan Tabel 2 yang merupakan hasil *cleaning data*.



Gambar 5. Tahapan Cleaning Data

Pada Gambar 5, proses ini dilakukan melalui tahap *Cleaning Char* untuk membersihkan karakter dan *Cleaning Number* untuk membersihkan angka.

Tabel 2. Hasil tahapan *Cleaning Data*

Sebelum dilakukan <i>Cleaning Data</i>	Sesudah dilakukan <i>Cleaning Data</i>
puas dengan kinerja nya proses cepat..... akan terhindar dari sipat pungli kalo begini..... Alhamdulillah ini ke 4 x aku byr pajak via SIGNAL hasilnya lumayan cepat jadi dalam waktu 3 hari sdh dikirim kerumah. Tapi yg ke 4 ini kenapa Gak Ada Hologram pengganti Stempel kalau kita adh byr ya?	puas dengan kinerja nya proses cepat akan terhindar dari sipat pungli kalo begini Alhamdulillah ini ke x aku byr pajak via SIGNAL hasilnya lumayan cepat jadi dalam waktu hari sdh dikirim kerumah Tapi yg ke ini kenapa Gak Ada Hologram pengganti Stempel kalau kita adh byr ya

Pada Tabel 2, sebelum *Cleaning data* terdapat tanda titik (.) dan angka 4. Sedangkan sesudah *Cleaning Data* tanda titik (.) dan angka 4 menjadi hilang

- b. *Transform cases*. Setelah dilakukan *cleaning data*, selanjutnya tahap *transform cases*. Pada tahap ini, dilakukan perubahan semua pada huruf menjadi huruf kecil semua agar teks berada dalam format standar. Berikut Tabel 3 yang merupakan hasil implementasi dari proses *transform cases*.

Tabel 3. Hasil tahapan *Transform Cases*

Sebelum dilakuka <i>Transform Cases</i>	Sesudah dilakukan <i>Transform Cases</i>
puas dengan kinerja nya proses cepat akan terhindar dari sipat pungli kalo begini Alhamdulillah ini ke x aku byr pajak via SIGNAL hasilnya lumayan cepat jadi dalam waktu hari sdh dikirim kerumah Tapi yg ke ini kenapa Gak Ada Hologram pengganti Stempel kalau kita adh byr ya	puas dengan kinerja nya proses cepat akan terhindar dari sipat pungli kalo begini alhamdulillah ini ke x aku byr pajak via signal hasilnya lumayan cepat jadi dalam waktu hari sdh dikirim kerumah tapi yg ke ini kenapa gak ada hologram pengganti stempel kalau kita adh byr ya

Pada Tabel 3, terdapat huruf kapital sebelum dilakukan *transform cases* dan huruf kapital menjadi huruf kecil setelah dilakukannya *transform cases*., Proses ini dilakukan menggunakan operator *Transform cases* pada Rapid Miner.

- c. *Tokenize*. Setelah dilakukan proses *transform cases*, selanjutnya dilakukan tahap *tokenize*. Pada tahap ini, dokumen akan dipecah menjadi bentuk tunggal dengan memotong kata-kata yang dipisahkan oleh spasi. Berikut Tabel 4 yang merupakan hasil implementasi proses *tokenizing*.

Tabel 4. Hasil *Tokenize*

Sebelum proses <i>Tokenize</i>	Sesudah proses <i>Tokenize</i>
puas dengan kinerja nya proses cepat akan terhindar dari sipat pungli kalo begini alhamdulillah ini ke x aku byr pajak via signal hasilnya lumayan cepat jadi dalam waktu hari sdh dikirim kerumah tapi yg ke ini kenapa gak ada hologram pengganti stempel kalau kita adh byr ya	puas, dengan, kinerja, nya, proses, cepat, akan, terhindar, dari, sipat, pungli, kalo, begini alhamdulillah, ini, ke, x, aku, byr, pajak, via, signal, hasilnya, lumayan, cepat, jadi, dalam, waktu, hari, sdh, dikirim, kerumah, tapi, yg, ke, ini, kenapa, gak, ada, hologram, pengganti, stempel, kalau, kita, adh, byr, ya

Pada Tabel 4, sebelum dilakukan proses *Tokenize*, ulasan masih berupa kalimat dan sedangkan setelah dilakukan proses *Tokenize*, ulasan yang sebelumnya berupa paragraf berubah menjadi kata per kata yang dipisah menggunakan tanda koma seperti pada tabel 4. Pada tahapan ini dilakukan dengan operator *Tokenize* dari Rapid Miner.

- d. *Filter Stopwords*. Setelah melalui tahap *tokenizing*, selanjutnya dilakukan proses *filter stopwords*. Dalam tahap ini, langkah *filter stopwords* ditujukan untuk penghapusan beberapa kata umum yang dianggap tidak relevan dan tidak penting berdasarkan kamus *stopword* yang telah diunduh dari situs [kaggle.com](https://www.kaggle.com). Berikut Tabel 5 yang merupakan hasil dari *filter stopwords*.

Tabel 5. Hasil *Filter Stopwords*

Sebelum proses <i>Filter Stopwords</i>	Sesudah proses <i>Filter Stopwords</i>
puas dengan kinerja nya proses cepat akan terhindar dari sipat pungli kalo begini alhamdulillah ini ke x aku byr pajak via signal hasilnya lumayan cepat jadi dalam waktu hari sdh dikirim kerumah tapi yg ke ini kenapa gak ada hologram pengganti stempel kalau kita adh byr ya	puas kinerja nya proses cepat terhindar sipat pungli kalo alhamdulillah x byr pajak via signal hasilnya lumayan cepat hari sdh dikirim kerumah yg gak hologram pengganti stempel adh byr ya

Pada Tabel 5, masih terdapat kata “dengan”, “akan”, “dari” dan lainnya sebelum dilakukan *Filter stopwords*. Sedangkan setelah dilakukan *Filter Stopword*, kata tersebut akan dihilangkan. Tahapan ini dilakukan dengan penggunaan operator berupa *Filter Stopword (dictionary)* pada Rapid Miner.

- e. *Filter Token By Length*. Setelah melalui proses *filter stopwords*, selanjutnya dilakukan proses *filter token by length* yang terdapat pada Rapid Miner. Proses ini melakukan penghapusan pada kata yang kurang dari 4 karakter dan lebih dari 24 karakter. Berikut Tabel 6 yang merupakan hasil implementasi dari proses *filter token by length*.

Tabel 6. Hasil proses *Filter Token By Length*

Sebelum proses <i>Filter Token By Length</i>	Sesudah proses <i>Filter Token By Length</i>
puas kinerja nya proses cepat terhindar sipat pungli kalo alhamdulillah x byr pajak via signal hasilnya lumayan cepat hari sdh dikirim kerumah yg gak hologram pengganti stempel adh byr ya	puas kinerja proses cepat terhindar sipat pungli kalo alhamdulillah pajak signal hasilnya lumayan cepat hari dikirim kerumah yg gak hologram pengganti stempel

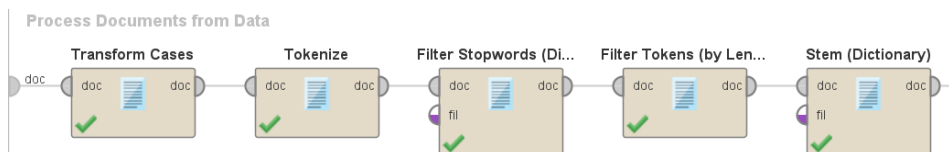
Pada Tabel 6, masih terdapat kata yang kurang dari 4 karkter seperti kata “nya”, “x”, “byr” dan lainnya sebelum dilakukan *Filter Token By Length*. Sedangkan kata tersebut sudah hilang seperti tabel 6 diatas apabila sesudah Proses *Filter Token By Length* sudah dilakukan.

- f. *Stemming*. Tahap terakhir dalam proses preprocessing dalam penelitian ini adalah melakukan proses *stemming*. Proses stemming dilakukan untuk melakukan perubahan pada kata yang memiliki imbuhan menjadi kata dasarnya berdasarkan kamus kata dasar bahasa indonesia yang didiperoleh dari situs <https://github.com/sastrawi/sastrawi/tree/master/data>. Proses ini dilakukan menggunakan operator stem dengan tambahan kamus kata dasar bahasa indonesia. Berikut Tabel 7 yang merupakan hasil dari proses *stemming*.

Tabel 7. Hasil *Stemming Data*

Sebelum proses <i>Stemming</i>	Sesudah proses <i>Stemming</i>
puas kinerja proses cepat terhindar sipat pungli kalo alhamdulillah pajak signal hasilnya lumayan cepat hari dikirim kerumah yg gak hologram pengganti stempel	puas kinerja proses cepat hindar sipat pungli kalo alhamdulillah pajak signal hasil lumayan cepat hari kirim rumah yg gak hologram ganti stempel

Pada Tabel 7, sebelum dilakukan *Stemming* masih terdapat imbuhan pada kata seperti kata “terhindar”, “hasilnya”, “dikirim”, dan lainnya. Sedangkan sesudah proses *stemming*, imbuhan pada kata akan diubah menjadi kata dasarnya seperti menjadi kata “hindar”, “hasil”, “kirim”, dan lainnya. Proses ini dilakukan menggunakan operator *stemming* pada Rapid Miner.

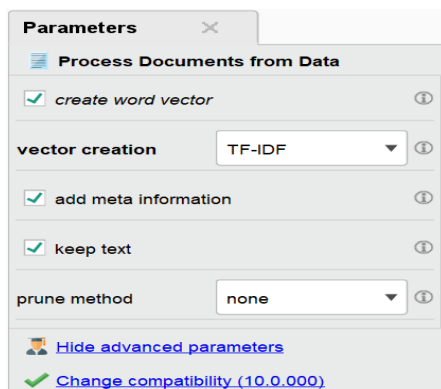


Gambar 6. Proses *Preprocessing Data*

Gambar 6 merupakan tahapan preprocessing yang diantaranya terdapat *transform cases*, *tokenize*, *filter stopwords*, *filter token by length* dan *Stemming*. Beberapa proses yang tersebut di gabungan kedalam operator process document from data yang ada pada aplikasi Rapidminer.

3.4 Pembobotan TF-IDF

pada tahap pembobotan kata ini, peneliti melakukan pembobotan pada setiap kata yang ada pada dokumen data ulasan menggunakan metode *Term Frequency Inverse Document Frequency* (TF-IDF). Besar kecilnya bobot kata tergantung pada banyaknya jumlah kata yang terdapat pada data ulasan pengguna aplikasi SIGNAL. Berikut Gambar 7 yang merupakan proses penggunaan metode TF-IDF yang dilakukan terhadap *parameters* yang ada pada *Process Documents from Data*.

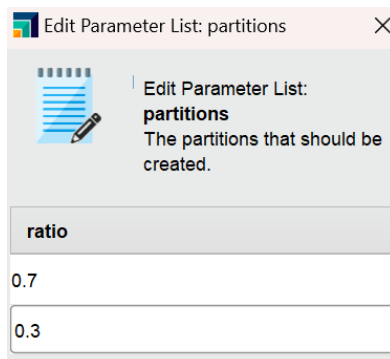


Gambar 7. Proses TF-IDF

Gambar 7 menunjukkan parameter dari operator "Process Documents from Data". Salah satu opsi yang tersedia adalah "create word vector", yang telah dicentang. Untuk proses vector creation, pilihan yang dipilih adalah TF-IDF.

3.5 Pembagian Data

Pembagian data dilakukan sebelum melakukan implementasi Naïve Bayes. Pada tahap ini, peneliti membagi data menjadi data latih dan data uji. Rasio pembagian data dilakukan sebesar 70 berbanding 30. Dimana 70% untuk data latih dan 30% untuk data uji. Dengan perbandingan rasio tersebut, didapat 1.400 untuk data latih dan 600 untuk data uji. Berikut Gambar 8 yang merupakan proses pemberian ratio yang terdapat pada parameter dari operator split data.

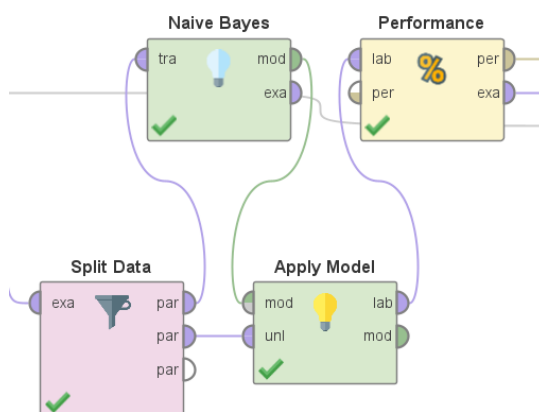


Gambar 8. Proses Pembagian Data

Gambar 8 menggambarkan parameter yang digunakan untuk membagi data dengan operator Split Data. Dalam proses ini, data dibagi dengan rasio 0.7:0.3 yang berarti 70% dari data akan digunakan sebagai data latih dan 30% sisanya sebagai data uji.

3.6 Implementasi Naïve Bayes

Setelah dilakukan pembagian data, tahap selanjutnya adalah melakukan klasifikasi menggunakan metode naïve bayes terhadap data uji untuk mengetahui sentimen positif dan negatif pengguna terhadap aplikasi SIGNAL. Dalam implementasi Naïve Bayes, peneliti menggunakan beberapa operator yang ada pada Rapidminer, diantaranya adalah operator Naïve Bayes, Apply model dan performance.



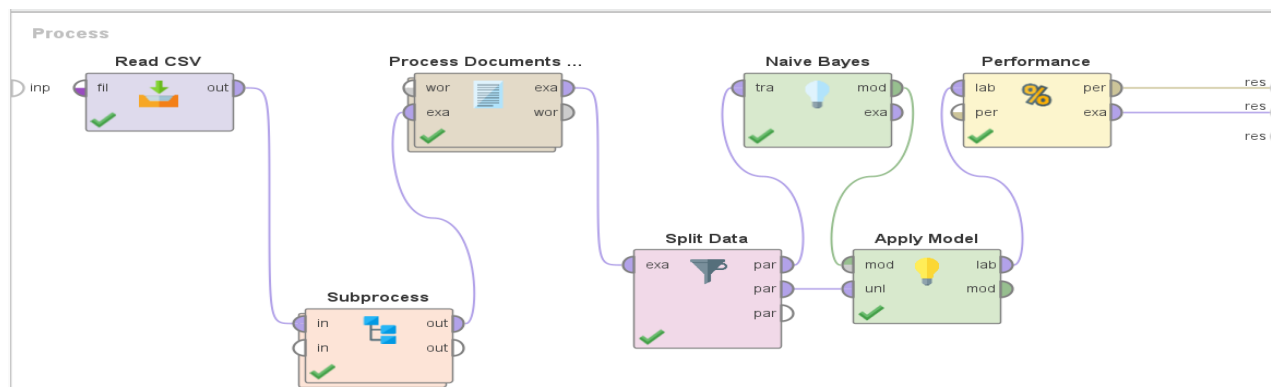
Gambar 9. Proses Implementasi Naïve Bayes

Gambar 9 diatas merupakan proses penerapan algoritma *Naïve Bayes*. Pada tahap ini, data yang sudah melalui tahap pembagian data yang menggunakan *operator split data* akan disambungkan dengan operator *naïve bayes* untuk dilakukan proses klasifikasi. Dari operator *Naïve Bayes*, data latih yang sebelumnya sudah di kalsifikasi kemudian akan disambungkan kembali pada operator *Apply Model* untuk mengetahui hasil prediksi yang kemudian akan diujikan terhadap data uji. Untuk mengetahui performa dari klasifikasi *naïve bayes*, selanjutnya operator model akan disambungkan dengan operator *performance*. Berikut gambar 10 yang merupakan hasil implentasi *Naïve Bayes* tarhadap data latih. Metode *Naïve Bayes* memberikan prediksi label secara terhadap data latih seperti dibawah ini.

prediction(Label)	text
Positif	respon cepat thanks polisi pelayanannya
Negatif	membantu
Positif	mantap cuman pengiriman
Negatif	bagus cepatklu cepat diprosesnyadan dicuekin email masuk pes...
Positif	membantu aplikasi capek antri bayar pajak lgsg kirim
Positif	mudah cepat
Negatif	update resi doank lambat selebihy
Negatif	bayar online kurir
Negatif	membantu samsat repot antri mantap dikembangkan pelayanannya
Negatif	terima kasih aplikasinya membantu bayar pajak kendaraan samsat
Positif	membantu gamau ribet kesamsat ngantri

Gambar 10. Hasil Proses implementasi *Naïve Bayes*

Gambar 10 merupakan hasil pelabelan manual dari data latih menggunakan operator *Naïve Bayes* setelah melewati tahap *preprocessing data*. Pelabelan ini terdapat dua kategori yaitu positif dan negatif.



Gambar 11. Proses Keseluruhan Analisis Sentimen *Naïve Bayes*

Gambar 11 merupakan proses keseluruhan dari tahapan proses penelitian analisis sentimen pada aplikasi Samsat Digital Nasional (SIGNAL) berdasarkan ulasan pengguna menggunakan metode naïve bayes. Proses ini dimulai dari operator *Read csv* sampai dengan operator *performance*.

3.7 Evaluasi

Setelah melakukan implementasi *Naïve Bayes*, selanjutnya dilakukan proses evaluasi untuk mengetahui performa dari algoritma *Naïve Bayes*.

accuracy: 63.61%

	true Negatif	true Positif	class precision
pred. Negatif	74	192	27.82%
pred. Positif	26	307	92.19%
class recall	74.00%	61.52%	

Gambar 12. Hasil Evaluasi Pada Rapidminer

Pada Gambar 12 merupakan Hasil pengujian evaluasi menggunakan menggunakan operator *performance* pada Rapidminer dan diperoleh nilai *accuracy sebesar 63.61%* dengan nilai TP 307, TN 74, FP 26 FN 192.

Berikut proses evaluasi dengan perhitungan menggunakan rumus *confusion matrix* untuk mencari nilai *accuracy*, *recall* dan *precision*.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} = \frac{(307+74)}{(307+74+26+192)} = \frac{381}{599} = 0,6361$$

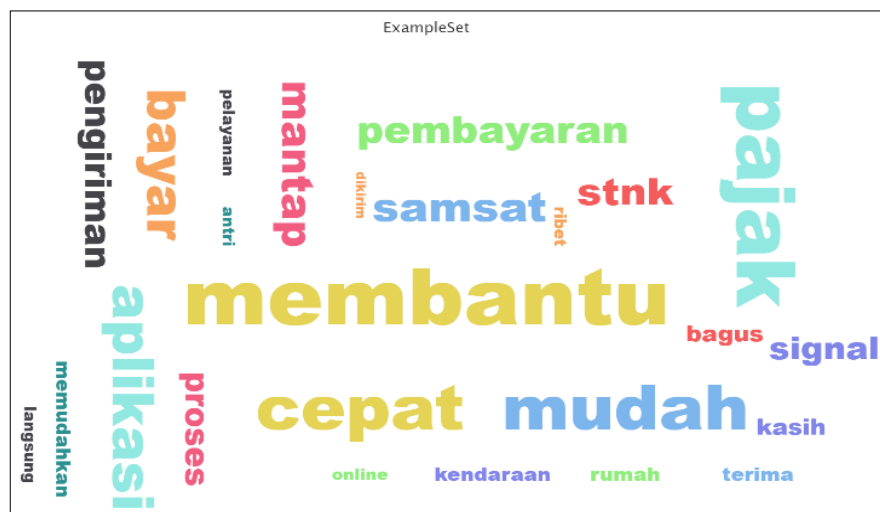
$$Precision = \frac{TP}{(TP+FP)} = \frac{307}{(307+26)} = \frac{307}{333} = 0,9219$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{307}{(307+192)} = \frac{307}{499} = 0,6165$$

Dari hasil evaluasi menggunakan *confusion matrix* diperoleh nilai *Accuracy* sebesar 63,61%, nilai *precision* sebesar 92,19% dan nilai *recall* sebesar 61,52%.

3.8 Visualisasi

Pada tahap ini, visualisasi hasil klasifikasi dipresentasikan dalam bentuk *Wordcloud*. Visualisasi word cloud adalah representasi grafis dari sejumlah kata dalam teks, di mana kata-kata yang paling sering muncul ditampilkan dengan ukuran yang lebih besar, sementara kata-kata yang muncul lebih jarang memiliki ukuran yang lebih kecil. Berikut Gambar 13 merupakan hasil visualisasi *wordcloud* dari hasil penelitian kali ini.



Gambar 13. Hasil Visualisasi *Wordcloud*

Pada Gambar 13 adalah 25 kata teratas yang sering muncul pada ulasan aplikasi SIGNAL. Kata “membantu” merupakan kata yang paling banyak digunakan dengan total mencapai 380 kali. Semakin banyak kata digunakan maka semakin besar ukuran katanya.

4. KESIMPULAN

Penelitian ini menggunakan metode Naïve Bayes untuk menganalisis sentimen pengguna terhadap aplikasi SIGNAL berdasarkan ulasan di Google PlayStore. Proses pengumpulan data melibatkan *scraping* data dengan Google Colab dan pengolahan menggunakan RapidMiner. Dari 2000 ulasan yang dikumpulkan, 1,665 ulasan dikategorikan sebagai positif, sementara 335 ulasan dikategorikan sebagai negatif setelah pelabelan manual. Tahap *preprocessing* mencakup *cleaning*, *transform cases*, *tokenizing*, *filter stopword*, *filter token by length*, dan *stemming*. Pembobotan TF-IDF digunakan untuk memberi bobot pada setiap kata. Pembagian data dilakukan menjadi 70% untuk data latih dan 30% untuk data uji. Setelah implementasi Naïve Bayes, hasil evaluasi menunjukkan akurasi sebesar 63.61%, dengan 307 True Positif, 74 True Negatif, 26 False Positif, dan 192 False Negatif. Dengan tingkat *precision* sebesar 92.19% dan *recall* sebesar 61.52%, dapat disimpulkan bahwa ulasan pengguna cenderung lebih positif terhadap aplikasi SIGNAL. Visualisasi *Wordcloud* menyoroti kata "membantu" sebagai yang paling dominan. Meskipun terdapat ulasan negatif, namun proporsinya lebih kecil dibanding ulasan positif, menunjukkan bahwa secara umum, aplikasi ini mendapatkan respon positif dari pengguna.

REFERENCES

- [1] V. Lauwrenza and W. Agustiniingsih, "Pengaruh Pengetahuan Wajib Pajak, Sosialisasi Pajak, Dan Penerapan Aplikasi Samsat Digital Nasional (Signal) Terhadap Kepatuhan Wajib Pajak Di Kabupaten Tangerang," *JURNAL PAJAK INDONESIA (Indonesian Tax Review)*, vol. 7, no. 1, pp. 37–44, 2023.
- [2] S. H. W. Putra and D. Febriawan, "Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naïve Bayes pada Google Play Store," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 4, pp. 1962–1971, 2024.
- [3] S. Devaranti, H. A. Murodi, and M. Machrunnisa, "INOVASI PELAYANAN PAJAK KENDARAAN BERMOTOR MELALUI APLIKASI SIGNAL (SAMSAT DIGITAL NASIONAL) DI UPTD KANTOR SAMSAT BALARAJA KABUPATEN TANGERANG," *Jurnal Administrasi Negara*, vol. 29, no. 2, pp. 127–146, Aug. 2023, doi: 10.33509/jan.v29i2.2637.
- [4] Siska Rafitanuri, Nur Arsyida, and Rizky Gunawan, "ANALISIS TINGKAT KEPUASAN MASYARAKAT TERHADAP PELAYANAN PEMBAYARAN PAJAK KENDARAAN BERMOTOR BERBASIS APLIKASI SIGNAL DI KANTOR SAMSAT KOTA TANJUNGPINANG," *JURNAL HUKUM, POLITIK DAN ILMU SOSIAL*, vol. 1, no. 3, pp. 92–103, Sep. 2022, doi: 10.55606/jhps.v1i3.537.
- [5] M. R. Fahlevvi, "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*, vol. 4, no. 1, pp. 1–13, Jun. 2022, doi: 10.33701/jtkp.v4i1.2701.
- [6] Rina Noviana and Isram Rasal, "PENERAPAN ALGORITMA NAIVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN BOY BAND BTS PADA MEDIA SOSIAL TWITTER," *Jurnal Teknik dan Science*, vol. 2, no. 2, pp. 51–60, Jun. 2023, doi: 10.56127/jts.v2i2.791.
- [7] V. Alviani, S. Alam, and I. Kurniawan, "ANALISIS SENTIMEN REVIEW APLIKASI WETV PADA PLATFORM TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 143–149, Aug. 2023, doi: 10.55123/storage.v2i3.2351.
- [8] R. Maulana, A. Voutama, and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC," *Jurnal Teknologi Terpadu*, vol. 9, no. 1, pp. 42–48, Jul. 2023, doi: 10.54914/jtt.v9i1.609.
- [9] Merinda Lestandy, Abdurrahim Abdurrahim, and Lailis Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 802–808, Aug. 2021, doi: 10.29207/resti.v5i4.3308.
- [10] H. Z. Muflih, A. R. Abdillah, and F. N. Hasan, "Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 3, pp. 1613–1621, 2023.
- [11] J. Jefriyanto, N. Ainun, and M. A. Al Ardha, "Application of Naïve Bayes Classification to Analyze Performance Using Stopwords," *Journal of Information System, Technology and Engineering*, vol. 1, no. 2, pp. 49–53, Jun. 2023, doi: 10.61487/jiste.v1i2.15.
- [12] E. P. Sutrisno and S. Amini, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR PADA ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI DIGITAL KORLANTAS POLRI," in *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, vol. 2, no. 2, pp. 687–695, 2023.
- [13] N. L. Lavenia and R. Permatasari, "Sentiment Analysis on Twitter Social Media Regarding Depression Disorder Using the Naive Bayes Method," *CoreID Journal*, vol. 1, no. 2, pp. 66–74, Jul. 2023, doi: 10.60005/coreid.v1i2.14.
- [14] S. Alam and M. I. Sulisty, "Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes," *Storage: Jurnal Ilmiah Teknik Dan Ilmu Komputer*, vol. 2, no. 3, pp. 100–108, 2023.
- [15] E. Hasibuan and E. A. Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *Jurnal Teknik dan Science*, vol. 1, no. 3, pp. 13–24, 2022.
- [16] Y. Nurtikasari, S. Alam, and T. I. Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 1, no. 4, pp. 411–423, 2022.
- [17] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm," *J Phys Conf Ser*, vol. 1641, no. 1, p. 012043, Nov. 2020, doi: 10.1088/1742-6596/1641/1/012043.
- [18] S. H. Ramadhani and M. I. Wahyudin, "Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 6, no. 4, pp. 526–534, 2022.
- [19] W. A. Prabowo and F. Azizah, "Sentiment Analysis for Detecting Cyberbullying Using TF-IDF and SVM," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 6, Dec. 2020, doi: 10.29207/resti.v4i6.2753.
- [20] A. Saputra and F. Noor Hasan, "ANALISIS SENTIMEN TERHADAP APLIKASI COFFEE MEETS BAGEL DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER," *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, vol. 2, no. 2, pp. 465–474, Jan. 2023, doi: 10.54443/sibatik.v2i2.579.
- [21] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 123–131, Feb. 2021, doi: 10.29207/resti.v5i1.2587.
- [22] Ericha Apriliyani and Y. Salim, "Analisis performa metode klasifikasi Naïve Bayes Classifier pada Unbalanced Dataset," *Indonesian Journal of Data and Science*, vol. 3, no. 2, pp. 47–54, Jul. 2022, doi: 10.56705/ijodas.v3i2.45.
- [23] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/resti.v5i4.3146.
- [24] N. Riyanah and F. Fatmawati, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 2, no. 4, pp. 206–213, Feb. 2021, doi: 10.35746/jtim.v2i4.117.
- [25] D.-H. Vu, "Privacy-preserving Naive Bayes classification in semi-fully distributed data model," *Comput Secur*, vol. 115, p. 102630, Apr. 2022, doi: 10.1016/j.cose.2022.102630.

- [26] A. R. Damanik, S. Sumijan, and G. W. Nurcahyo, "Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma Naïve Bayes," *Jurnal Sistem Informasi dan Teknologi*, pp. 88–94, Aug. 2021, doi: 10.37034/jsisfotek.v3i3.49.
- [27] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. A. Abo-Elhoud, "Accurate detection of COVID-19 patients based on distance biased Naïve Bayes (DBNB) classification strategy," *Pattern Recognit*, vol. 119, p. 108110, Nov. 2021, doi: 10.1016/j.patcog.2021.108110.
- [28] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews," in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, IEEE, Feb. 2020, pp. 217–220. doi: 10.1109/IC3A48958.2020.233300.
- [29] Y. Choi, G. Farnadi, B. Babaki, and G. Van den Broeck, "Learning Fair Naive Bayes Classifiers by Discovering and Eliminating Discrimination Patterns," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 06, pp. 10077–10084, Apr. 2020, doi: 10.1609/aaai.v34i06.6565.
- [30] Pristiyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm," *IOP Conf Ser Mater Sci Eng*, vol. 1088, no. 1, p. 012045, Feb. 2021, doi: 10.1088/1757-899X/1088/1/012045.
- [31] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, "SENTIMENT ANALYSIS ON E-SPORTS FOR EDUCATION CURRICULUM USING NAIVE BAYES AND SUPPORT VECTOR MACHINE," *Jurnal Ilmu Komputer dan Informasi*, vol. 13, no. 2, pp. 109–122, Jul. 2020, doi: 10.21609/jiki.v13i2.885.
- [32] J. Gu and S. Lu, "An effective intrusion detection approach using SVM with naïve Bayes feature embedding," *Comput Secur*, vol. 103, p. 102158, Apr. 2021, doi: 10.1016/j.cose.2020.102158.
- [33] Alvina Felicia Watratan, Arwini Puspita. B, and Dikwan Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *Journal of Applied Computer Science and Technology*, vol. 1, no. 1, pp. 7–14, Jul. 2020, doi: 10.52158/jacost.v1i1.9.
- [34] F. I. Adiba, T. Islam, M. S. Kaiser, M. Mahmud, and M. A. Rahman, "Effect of Corpora on Classification of Fake News using Naive Bayes Classifier," *International Journal of Automation, Artificial Intelligence and Machine Learning*, Oct. 2020, doi: 10.61797/ijaauml.v1i1.45.
- [35] M. D. Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 5, no. 1, pp. 1–7, Apr. 2022, doi: 10.31539/intecom.s.v5i1.3708.