

# Analisis Sentimen dalam Penentuan Kinerja Layanan antara *In-Driver* dan Gojek berdasarkan Opini Masyarakat menggunakan Metode Naive Bayes

## *Sentiment Analysis in Determining Service Performance between In-Driver and Gojek based on Public Opinion using the Naive Bayes Method*

Agung Firmansyah\*, Rakhmat Kurniawan

Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara  
Jl. Lap. Golf, Kp. Tengah. Kec. Pancur Batu, Kab. Deli Serdang, Sumatera Utara 20353,  
Medan, Indonesia

\*e-mail: [agungfirmansyah200103@gmail.com](mailto:agungfirmansyah200103@gmail.com)

(received: 15 May 2024, revised: 26 May 2024, accepted: 28 May 2024)

### Abstrak

Sistem transportasi berbasis aplikasi yang terhubung dengan koneksi internet yang ramai dibicarakan oleh masyarakat adalah ojek online. Salah satu perusahaan ojek online terkemuka di Indonesia adalah Gojek dan In-Driver. Setiap pelanggan memiliki tingkat kepuasan berbeda terhadap layanan yang diberikan oleh Gojek dan In-Driver Indonesia, sehingga selalu ada pro dan kontra berupa saran dan keluhan. Dilihat dari permasalahan yang ada, maka diperlukan sebuah solusi berupa analisis terhadap saran maupun keluhan yang diterima oleh perusahaan. Adapun permasalahan pengklasifikasian sebuah kalimat sentimen ke dalam kelas-kelas tertentu dapat diselesaikan dengan metode Multinomial Naive Bayes Classifier. Data yang digunakan berjumlah 1000 data serta data yang digunakan yaitu sejumlah 500 data pertama dan 500 data tweet kedua sebagai nilai perbandingan. Hasil perhitungan data gojek menghasilkan nilai *accuracy* sebesar 73%, *precision* sebesar 72%, *recall* sebesar 100%, dan *f1-score* sebesar 84%. Pada hasil perhitungan data indriver diperoleh nilai *accuracy* sebesar *accuracy* sebesar 85,71%, *precision* sebesar 85,185%, *recall* sebesar 95,833%, dan *f1-score* sebesar 90,196%. Hal tersebut membuktikan bahwa algoritma klasifikasi Naive Bayes lebih cocok digunakan pada jumlah data yang lebih sedikit.

**Kata Kunci** : Sentimen ; Kinerja Layanan ; Gojek; In-Driver ; Metode Naive Bayes.

### Abstract

*An application-based transportation system connected to an internet connection that is widely discussed by the public is an online motorcycle taxi. One of the leading online motorcycle taxi companies in Indonesia is Gojek and In-Driver. Each customer has a different level of satisfaction with the services provided by Gojek and In-Driver Indonesia, so there are always pros and cons in the form of suggestions and complaints. Judging from the existing problems, a solution is needed in the form of analyzing the suggestions and complaints received by the company. The problem of classifying a sentiment sentence into certain classes can be solved by the Multinomial Naive Bayes Classifier method. The data used amounted to 1000 data and the data used were the first 500 data and the second 500 tweet data as a comparison value. The results of the gojek data calculation resulted in an accuracy value of 73%, precision of 72%, recall of 100%, and f1-score of 84%. The results of the indriver data calculation obtained an accuracy value of 85.71% accuracy, precision of 85.185%, recall of 95.833%, and f1-score of 90.196%. This proves that the Naive Bayes classification algorithm is more suitable for use on a smaller amount of data.*

<http://sistemasi.ftik.unisi.ac.id>

**Keywords:** *Sentiment; Service Performance; Gojek; In-Driver; Naive Bayes Method.*

## 1 Pendahuluan

Transportasi umum merupakan fasilitas yang digunakan masyarakat dalam aktivitas sehari-hari. Transportasi umum terbagi menjadi dua yaitu transportasi konvensional (offline) dan transportasi online. Transportasi konvensional adalah transportasi umum yang digunakan masyarakat yang tanpa perlu melakukan pemesanan menggunakan aplikasi. Sedangkan transportasi online adalah transportasi umum berbasis aplikasi sebagai media untuk melakukan pemesanan sebelum menggunakannya. Saat ini masyarakat mulai terbiasa menggunakan jasa transportasi online karena kemudahan dan efisiensinya yang membantu dalam beraktivitas. Banyak perusahaan transportasi online yang menawarkan berbagai layanan sehingga menciptakan persaingan yang ketat. Perusahaan transportasi online yang saat ini sering digunakan oleh masyarakat antara lain Gojek dan In-Driver. Gojek adalah sebuah perusahaan yang berasal dari Indonesia yang mengandung konsep teknologi yang menyediakan layanan jasa transportasi berbasis aplikasi Online. Gojek adalah sebagai pelopor transportasi online yang berdiri sejak tahun 2010 [1].

Salah satu sarana transportasi yang sering digunakan adalah angkutan umum. Mengikuti perkembangan teknologi yang semakin maju, transportasi umum berkembang dengan pesat. Sistem transportasi umum saat ini ada yang berbasis aplikasi yang terhubung dengan koneksi internet sehingga dapat disebut juga dengan transportasi online. Di Indonesia terdapat dua penyedia jasa transportasi online, yaitu *in-driver* dan Gojek. Berbagai macam layanan bahkan promosi diberikan kepada pengguna jasa transportasi online tersebut namun layanan yang disediakan juga tidak selalu sempurna dan tetap perlu adanya penyempurnaan. Oleh karena itu, tanggapan dari pengguna sangat diperlukan untuk memperbaiki layanan yang telah disediakan [2].

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan suatu informasi sentimen yang terkandung dalam suatu kalimat opini. Secara umum, analisis sentimen bertujuan untuk menentukan sikap pembicara atau penulis terhadap suatu topik atau keseluruhan polaritas kontekstual pada suatu dokumen. Sikap dapat berupa penilaian atau evaluasi, sisi emosional penulis pada saat menulis atau efek komunikasi emosional yang penulis inginkan terhadap pembacanya. X sebagai salah satu jejaring sosial yang interaktif memungkinkan penggunaannya untuk mengkritisi suatu isu maupun sebuah fasilitas pelayanan secara real time. Masyarakat yang semula membutuhkan waktu lama untuk menyampaikan aspirasinya kini dapat melakukannya dengan mudah berkat kehadiran teknologi ini [3][4].

Salah satu transportasi online yang ramai dibicarakan oleh masyarakat adalah ojek online. Ojek online atau sering disebut Ojol merupakan transformasi 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, dokumen, berbelanja, dan lain sebagainya [5]. Terjadinya persaingan pada kedua perusahaan transportasi online tentu perlu saran dan kritik dari pengguna untuk meninjau kekurangan dari setiap layanan kedua transportasi online ini, sehingga dilakukan sentimen analisis untuk mengetahui tingkat kepuasan pengguna transportasi online antara 2 (dua) aplikasi tersebut.

Gojek dan In-Driver 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 Gojek dan In-Driver Indonesia, sehingga selalu ada pro dan kontra berupa saran dan keluhan. Pemrosesan terhadap saran dan keluhan kini dapat disampaikan secara real time melalui akun Twitter Gojek dan In-Driver, sehingga Gojek dan In-Driver 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 Gojek dan In-Driver yang diterima per hari [6] [7].

<http://sistemasi.ftik.unisi.ac.id>

Dilihat dari permasalahan yang ada, maka diperlukan sebuah solusi berupa analisis terhadap saran maupun keluhan yang diterima oleh perusahaan ojek online Grab Indonesia dari pelanggan yang mengirimkan tweet ke akun user Gojek dan In-Driver pada media sosial media X sehingga dapat diketahui informasi sentimen pelanggan terhadap pelayanan yang telah diberikan secara cepat. Data dapat di peroleh melalui proses crawling dengan menggunakan fasilitas Twitter API [8][9][10]. Permasalahan pengklasifikasian sebuah kalimat sentimen ke dalam kelas– kelas tertentu dapat diselesaikan dengan metode Multinomial Naive Bayes Classifier. Metode ini memiliki beberapa kelebihan antara lain, sederhana, cepat, dan berakurasi tinggi yang memiliki keunggulan dalam memproses teks. Metode MNBC untuk klasifikasi atau kategorisasi teks menggunakan atribut kata yang muncul dalam suatu dokumen sebagai dasar klasifikasinya [11][12][13][14][15].

Tujuan penelitian ini adalah mengembangkan aplikasi untuk analisis opini masyarakat terhadap keunggulan dan kekurangan dari layanan in-driver dan gojek dengan implementasi text mining menggunakan algoritma Naïve Bayes. Batasan masalah pada penelitian ini adalah analisis opini masyarakat terhadap kinerja layanan In-driver dan Gojek menggunakan klasifikasi sentimen dan ekstraksi variable yang menentukan sentimen positif dan negatif, data yang diambil melalui X berupa tweet pengguna yang berada di Indonesia, dan analisis menggunakan bahasa pemrograman Python yang kemudian diimplementasi ke dalam Jupyter Notebook.

## 2 Tinjauan Literatur

Hasil penelitian oleh Nugraha, [8] tentang Analisis Sentimen Penggunaan Aplikasi Transportasi Online pada ulasan google play store dengan metode naive bayes classifier. Berdasarkan analisis dan pengujian yang dilakukan, maka kesimpulan yang dapat diambil dari hasil implementasi algoritma NBC pada penelitian ini, data menunjukkan bahwa tingkat keakuratan tertinggi untuk aplikasi transportasi online pertama yaitu Maxim dengan nilai 93%, kedua Grab sebesar 87%, dan ketiga Gojek sebesar 86%. Penulis berasumsi bahwa aplikasi Maxim terdapat akurasi paling tinggi karena lebih banyak ulasan positif dan memiliki ulasan negatif yang lebih sedikit dibandingkan dengan aplikasi transportasi online lainnya.

Hasil penelitian oleh Yuniar, [15] tentang Analisis Sentimen Ulasan pada Gojek Menggunakan Metode Naïve Bayes. Data ulasan yang diberikan oleh pengguna aplikasi Gojek pada bulan Januari 2023 cenderung bersentimen positif. Data diklasifikasikan ke dalam sentimen positif dengan jumlah 2.802 ulasan dan sentimen negatif berjumlah 1.348 ulasan. Penelitian dengan metode *Naïve Bayes Classifier* mencapai tingkat akurasi tertinggi yaitu 89,9% dengan nilai presisi, *recall*, dan *F-measure* berturut-turut sebesar 0,828, 0,854, dan 0,841 pada pembagian 90% data *training* dan 10% data *testing*. Sejumlah 373 ulasan dari total data 415 ulasan yang diuji, dapat diklasifikasikan dengan benar. Berdasarkan asosiasi teks yang telah dilakukan, diperoleh hasil pada kelas positif antara lain gojek, aplikasi, bantu, driver, dan baik. Sedangkan pada kelas negatif sering mengeluhkan mengenai mahal, eror, gopay, lama, dan kecewa.

Hasil penelitian oleh Indarwati [16] tentang Analisis Sentimen Terhadap Kualitas Pelayanan Aplikasi Go-Jek Menggunakan Metode Naive Bayes Classifier. Berdasarkan dari hasil penelitian yang telah dilakukan mengenai analisis sentimen data ulasan pada pengguna aplikasi gojek memiliki dampak yang cukup besar dalam mencari transportasi umum. Dapat dilihat dalam tahap pembobotan kata yang dilakukan menggunakan wordcloud menampilkan perolehan kata teratas yaitu kata “mantap” dengan perolehan nilai 131.085 dan disusul dengan kata “bantu” dengan perolehan nilai 106.727. Para pengguna transportasi umum lebih memilih untuk menggunakan aplikasi secara online karena dapat digunakan secara praktis dan terpercaya, penumpang dapat mengetahui data diri. driver serta dapat mengetahui rute perjalanan yang dituju. Adapun pengguna yang lebih memilih mencari transportasi secara umum, terutama karena belum terbiasa untuk menggunakan aplikasi tersebut atau akses jaringan yang bermasalah sehingga pengguna sulit untuk memesan secara online.

Hasil penelitian oleh Petiwi, et al, [3] tentang Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine. Hasil kesimpulan dari penelitian dan pembahasan yang dilakukan dengan objek Gofood maka didapatkan hasil klasifikasi komentar masyarakat ke dalam tiga kelas yaitu kelas positif 5,2%, kelas netral sebesar 92,8% dan kelas negatif sebesar 2,0%. Ini menunjukkan bahwa tanggapan positif masyarakat lebih besar dari pada tanggapan

<http://sistemasi.ftik.unisi.ac.id>

negatif. Selanjutnya hasil akurasi pada penelitian ini dilakukan dengan bahasa python dan tampilan antarmuka berbasis website. Support Vector Machine menghasilkan nilai yang lebih besar dan akurat dari akurasi Naïve Bayes dalam mengelompokkan opini masyarakat berbahasa Indonesia pada media sosial Twitter. Metode Support Vector Machine menghasilkan akurasi sebesar 83% dari 5000 tweet menggunakan bahasa python dan 98,5% melalui uji coba antarmuka dengan 1000 tweet sedangkan kemampuan akurasi Naïve Bayes sebesar 74,6% melalui bahasa python dan 91,5% menggunakan uji coba antarmuka. Ini membuktikan metode SVM lebih akurat sebagai metode pengelompokan untuk proses analisis sentimen opini masyarakat berbahasa Indonesia pada Twitter dibandingkan Naïve Bayes.

Hasil penelitian oleh Mandasari, *et al*, [1] tentang Analisis Sentimen Pengguna Transportasi Online Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier. Hasil kesimpulan dari penelitian dan pembahasan yang dilakukan dengan objek Gofood maka didapatkan hasil klasifikasi komentar masyarakat kedalam tiga kelas yaitu kelas positif 5,2%, kelas netral sebesar 92,8% dan kelas negatif sebesar 2,0%. Ini menunjukkan bahwa tanggapan positif masyarakat lebih besar dari pada tanggapan negatif. Selanjutnya hasil akurasi pada penelitian ini dilakukan dengan bahasa python dan tampilan antarmuka berbasis website. Support Vector Machine menghasilkan nilai yang lebih besar dan akurat dari akurasi Naïve Bayes dalam mengelompokkan opini masyarakat berbahasa Indonesia pada media sosial Twitter. Metode Support Vector Machine menghasilkan akurasi sebesar 83% dari 5000 tweet menggunakan bahasa python dan 98,5% melalui uji coba antarmuka dengan 1000 tweet sedangkan kemampuan akurasi Naïve Bayes sebesar 74,6% melalui bahasa python dan 91,5% menggunakan uji coba antarmuka. Ini membuktikan metode SVM lebih akurat sebagai metode pengelompokan untuk proses analisis sentimen opini masyarakat berbahasa Indonesia pada Twitter dibandingkan Naïve Bayes.

Gap research dalam penelitian ini adalah Salah satu cara untuk mengetahui pengklasifikasian analisis sentimen dalam penentuan kinerja layanan antara in-driver dan gojek berdasarkan opini masyarakat di media sosial X. Kemudian untuk mengimplementasikan metode naïve bayes dalam pengklasifikasian analisis sentimen dalam penentuan kinerja layanan antara *in-driver* dan gojek berdasarkan opini masyarakat 2024 di media sosial X. Berdasarkan penelitian yang dilakukan peneliti mendapatkan kontribusi dari penelitian adalah diharapkan penelitian ini memperoleh model sentimen yang dapat melakukan praproses terhadap data tweet yang tidak terstruktur sehingga mampu mempresentasikan sebuah sentimen untuk mengetahui respon atau pandangan dari masyarakat, guna untuk mengetahui gambaran umum terkait dalam penentuan kinerja layanan antara in-driver dan gojek berdasarkan opini masyarakat. Peneliti berharap penelitian ini dapat memberikan kontribusi bagi para pembaca sebagai rekomendasi untuk memilih dalam penentuan kinerja layanan antara in-driver dan gojek berdasarkan opini masyarakat berdasarkan tweet yang sudah diklasifikasikan oleh peneliti menggunakan metode naïve bayes.

### 3 Metode Penelitian

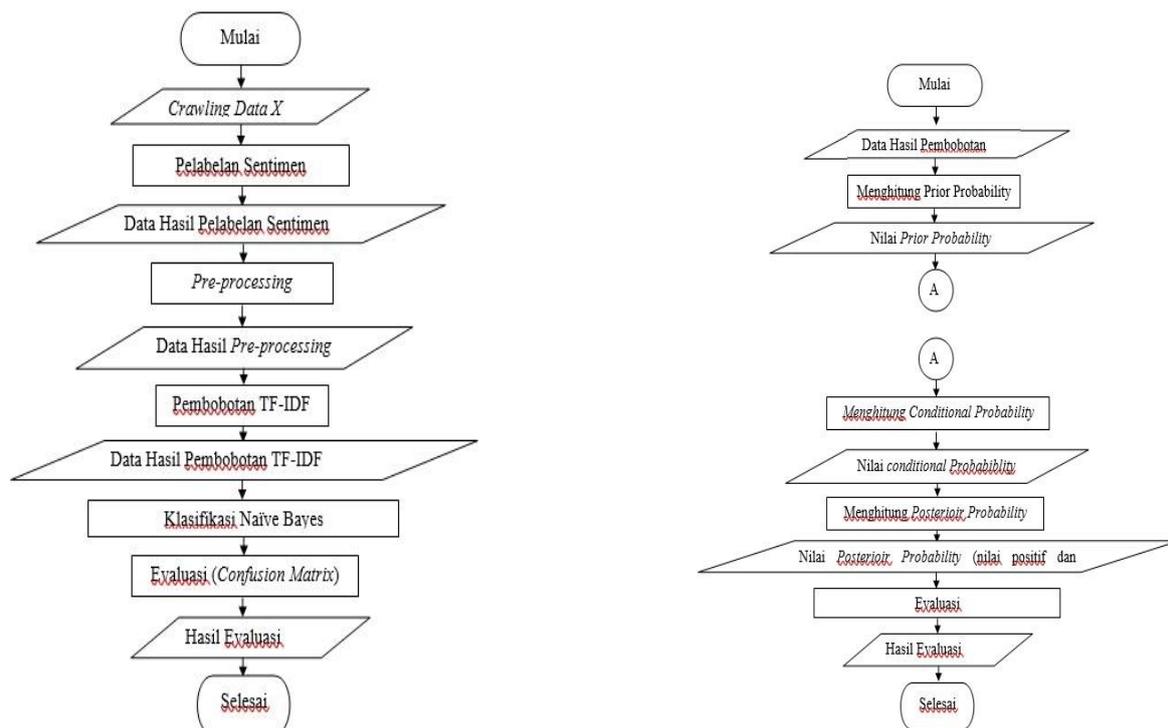
Lokasi penelitian ini mengambil data dari X sebagai tempat penelitian karna sesuai dengan topik yang diangkat peneliti mengenai analisis sentimen, selain itu peneliti juga ingin mengetahui bagaimana klasifikasi terhadap in-driver dan gojek. Sesuai dengan topik yang diangkat oleh peneliti X berperan aktif sebagai media sosial yang digunakan masyarakat dalam menyampaikan opininya sehingga peneliti dapat mengklasifikasikan pendapat masyarakat tentang in-driver dan gojek 2024. Pengujian dalam penelitian ini dilakukan untuk melihat nilai akurasi pada penentuan sentimen terhadap penentuan kinerja layanan antara in-driver dan gojek menggunakan *Recall Precision F-measure* melalui library naïve bayes, dari model training dan testing yang telah diklasifikasikan sebelumnya. Hingga menghasilkan matriks sebagai representatif kelas aktual dan kelas prediksi. Hasil dari model training dengan menggunakan data baru yang belum dilakukan training sebelumnya.

Data yang digunakan adalah data yang diperoleh dari media sosial X. Data tersebut diperoleh dengan menggunakan teknik web crawling dan memanfaatkan sncrape. Teknik web crawling dilakukan menggunakan Google Colab dengan memanfaatkan library Bahasa pemrograman Python yaitu sncrape. Data yang diambil meliputi tweet, id tweet, tanggal tweet diunggah dan nama pengguna.

<http://sistemasi.ftik.unisi.ac.id>

Tetapi data yang akan diproses hanya data tweet. Data yang diperoleh berjumlah 1000 data serta data yang digunakan yaitu sejumlah 500 data pertama dan 500 data tweet kedua sebagai nilai perbandingan. Ini dikarenakan pada media sosial X terdapat fitur retweet dimana fitur ini memungkinkan penggunaannya untuk mengunggah kembali tweet milik pengguna lainnya yang menyebabkan duplikasi data.

Pada penelitian ini analisis data yakni penyusunan data *tweet* sebagai proses awal untuk pengolahan data, proses ini digunakan untuk menggambarkan *tweet* pengguna *X* tentang *in-driver* dan *gojek* 2024. Dengan mendisplay data, akan memudahkan untuk meneliti dan menyusun data tersebut. Kemudian akan dilakukan pelabelan kelas dengan ketentuan apabila *tweet* kurang dari 0 maka kelas tersebut negatif dan lebih dari 0 maka positif. Tahapan dalam melakukan text mining ini dilakukan dengan menggunakan metode naïve bayes dimulai dengan input data yang berupa data latih dan data uji. Data yang telah memiliki label akan digunakan untuk proses klasifikasi di tampilan dalam Gambar 1.



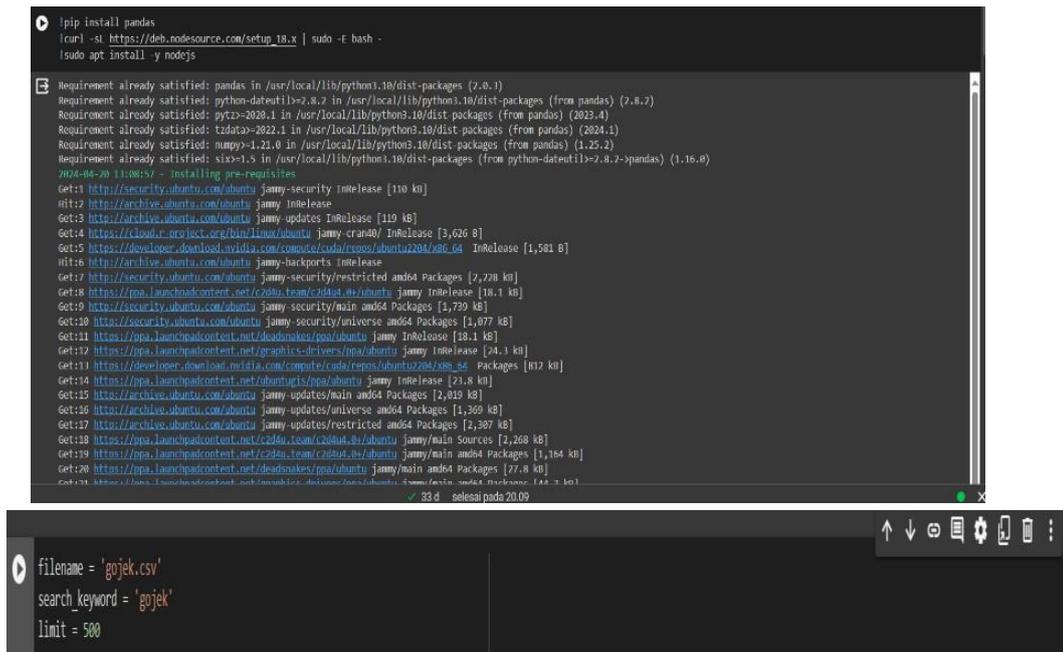
Gambar 1. Flowchart Klasifikasi Text dan Klasifikasi Naive Bayes

#### 4. Hasil dan Pembahasan

Pada bagian ini menjelaskan proses analisis data yang di dapatkan pada saat penelitian dengan analisis sebagai berikut:

##### 4.1 Analisis Data

*Dataset* pada penelitian ini berupa opini atau tanggapan masyarakat terhadap kinerja layanan antara *in-driver* dan *gojek* pada social Media X. Data berasal dari media sosial X yang diperoleh dengan teknik *crawling* dengan menggunakan *library Python* yaitu *nodejs*. Proses ini menggunakan Bahasa pemrograman *Python* dengan *tools Google Colab*. Langkah pertama dengan mengunduh dan memasang *library "Snsrape"* kemudian mendeklarasikannya ke dalam sistem. Kemudian data *tweet* dicrawling dengan kata kunci "In-drivers" dan "Gojek" ditampilkan dalam Gambar 2.



Gambar 2. Install Library python dan Crawling Data

Berdasarkan hasil wawancara tersebut validator menyatakan bahwa kamus yang dibuat sudah sesuai dengan label kata positif dan kata negatif dengan indikator berikut di tampilkan dalam Tabel 1.

Tabel 1. Indikator Pemberian Label

Indikator	Keterangan	
	Positif	Negatif
Makna	Bersifat membangun, memotivasi dan tidak menjatuhkan atau menjelekkan sesuatu yang dibahas.	Memiliki arti yang kurang baik, dan bersifat menyimpang.

Berdasarkan kata yang sudah dilabelisasi dilakukanlah labelisasi kalimat menggunakan nilai skor sentimen (*StSc*). Berikut contoh perhitungan kalimat “santai pakai sistem bayar tempat tinggal pesan sediaan uang bayar barang mudah” dengan rumus nilai skor sentimen (*StSc*) dijelaskan dari persamaan 1.

$$StSc = \frac{\text{jumlah kata positif} - \text{jumlah kata negatif}}{\text{total jumlah kata}} = \frac{8 - 4}{10} = 0,4 \quad (1)$$

Nilai skor sentimen (*StSc*) dari data tersebut yaitu 0,4, sesuai dengan ketentuan yang telah dijabarkan sebelumnya dimana bila jika skor sentimen > 0 maka data tersebut dikategorikan sebagai sentimen **positif**. *Dataset* yang telah diperoleh disimpan dan akan diproses dalam berkas dengan format .csv. Berikut disajikan alur sistem dari proses *preprocessing data* yang meliputi *cleaning*, *case folding*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*.

#### 4.1 Pembobotan TF-IDF

Tahap ini bertujuan untuk menghitung bobot tiap kata (*term*) yang terdapat di dalam dokumen (*N*). Berikut contoh perhitungan dari proses *TF-IDF* dari dokumen yang diberikan sebelumnya di tampilkan dalam Tabel 2 dan 3.

Tabel 2. Perhitungan Nilai TF dan DF Gojek

Term	TF					DF
	D1	D2	D3	D4	D5	
sore	1	0	0	0	0	1
enak	1	0	0	0	0	1

gofoodin	1	0	0	0	0	1
...	...	...	...	...	...	...
henti	0	0	0	0	1	1

**Tabel 3. Perhitungan Nilai TF dan DF Indriver**

Term	TF					DF
	D1	D2	D3	D4	D5	
indrive	1	1	1	1	0	4
drivernya	1	0	0	0	0	1
tahi	1	0	0	0	0	1
...	...	...	...	...	...	...
indriver	0	0	0	0	1	1

Setelah tahap perhitungan *TF* dan *DF* dari setiap term selesai dilakukan, selanjutnya tahap perhitungan *IDF* menggunakan persamaan 2. Berikut disajikan tabel hasil perhitungan *IDF* melanjutkan dari tahap perhitungan sebelumnya. Berikut disajikan contoh perhitungan untuk *term* pertama yaitu “sore” yang mana telah diketahuin jumlah  $d = 5$  dan  $df = 1$  dijelaskan dari persamaan 2.

$$IDF = \ln \frac{d+1}{df+1} + 1 \tag{2}$$

Contoh Perhitungan *IDF* pada *term* pertama yaitu “sore” di bawah ini dan di tampilkan dalam Tabel 4 dan 5:

$$IDF (\text{Sore}) = \ln \frac{5 + 1}{1 + 1} + 1 = 2.09861$$

**Tabel 4. Perhitungan Nilai IDF Gojek**

Term	TF	IDF
sore	1	2.09861
enak	1	2.09861
gofoodin	1	2.09861
...	...	...
henti	1	2.09861

**Tabel 5. Perhitungan Nilai IDF Indriver**

Term	DF	IDF
indrive	4	1.182322
drivernya	1	2.098612
tahi	1	2.098612
...	...	....
indriver	1	2.098612

Setelah tahap perhitungan nilai *IDF* dari setiap term selesai dilakukan, selanjutnya tahap perhitungan nilai *TF-IDF* menggunakan persamaan 3. Berikut disajikan Tabel 6 hasil perhitungan *TF-IDF* melanjutkan dari tahap perhitungan sebelumnya. Berikut disajikan contoh perhitungan untuk term pertama yaitu “sore” yang mana telah diketahuin nilai  $tf = 1$  dan nilai  $idf = 2,098612$  dijelaskan dari persamaan 3.

$$W_{dt} = T_{f_{dt}} \times IDF \tag{3}$$

Contoh Perhitungan *TF-IDF* ( $W_{dt}$ ) pada *term* pertama yaitu “sore” di bawah ini di tampilkan dalam Tabel 6 dan 7:

$$W_{dt (\text{sore})} = 1 \times 2.09861 = 2.09861229$$



	D1	D2	D3	D4	D5
indrive	0.16052544	0.49084524	0.18456784	0.22414766	0
drivernya	0.28493151	0	0	0	0
tahi	0.28493151	0	0	0	0
...	...	...	...	...	...
indriver	0	0	0	0	0.37796

Dalam penelitian ini akan diterapkan split data dengan perbandingan 9:1 yang artinya 90% dari total dataset akan digunakan sebagai data latih dan 10% lainnya akan digunakan sebagai data uji yang merujuk pada penelitian sebelumnya.

### 4.3 Klasifikasi Naïve Bayes Classifier

Setelah tahap pembobotan kata selesai, maka akan dilakukan tahap pengklasifikasian terhadap data uji dengan metode *Naïve Bayes Classifier*. Berikut disajikan contoh perhitungan klasifikasi terhadap data uji dengan data latih yang sebelumnya sudah dilakukan tahap *preprocessing* dan pelabelan di ditampilkan dalam Tabel 10, 11,12 dan 13.

**Tabel 10. Contoh Data Latih Gojek**

Dokumen	Data Opini	Label
D1	gaes sore enak gofoodin iya	Positif
D2	gojek wangi is awesome gojek helm wangi is wholesome	Positif
D3	teman teman twitter tolong bantu teledor banget akuin salah teman teman kenan bantu hp ku tinggal mobil gocar up cs gojek nya diarahin bot bot nya nyuruh call driver nya	Negatif
D4	bangsat teman st sudirman american grill sesat gojek jalan kali menit	Negatif
D5	sih ujannya iklasin gojek order gocar tunggu menit henti	Negatif

**Tabel 11. Contoh Data Latih Indriver**

Dokumen	Data Opini	Label
D1	indrive drivernya tahi aplikasi cuman nyuruh cepat bilang sop gaboleh kebut nyuruh kebut tahi	Negatif
D2	anjing kayak indrive	Negatif
D3	indrive gelar raya kartini jakarta jumat raya indrive apresiasi kemudi wanita	Positif
D4	daftar indrive motor mobil via online mudah	Positif
D5	plisss barusan indriver aa bandung wangi banget	Positif

**Tabel 12. Contoh Data Uji Gojek**

Data Tweet	Hasil <i>Preprocessing</i>	Label
Kenapa ya abang gojek kalo mas <sup>2</sup> demen banget nanyain udah nikah apa belum . Gak ada korelasinya tau gak	iya abang gojek mas suka banget nanyain nikah korelasi	?

**Tabel 13. Contoh Data Uji Indriver**

Data Tweet	Hasil <i>Preprocessing</i>	Label
@gandrasta di Bali sama juga gocar paling mahal. coba in-driver juga mas.	bali gocar mahal coba indriver mas	?

<http://sistemasi.ftik.unisi.ac.id>

Tahapan dalam proses perhitungan klasifikasi kelas pada data uji dimulai dari perhitungan nilai *prior probability*, *conditional probability*, dan *posterior probability*. Berikut disajikan tahap perhitungan metode *Naïve Bayes Classifier* terhadap data uji tersebut [17][18][19].

1. Perhitungan nilai *prior probability* Gojek

a.  $P(\text{Positif}) = \frac{2}{5} = 0,4$

b.  $P(\text{Negatif}) = \frac{3}{5} = 0,6$

2. Perhitungan nilai *conditional probability* dengan *Laplacian smoothing* Gojek

a. Positif

$$P(\text{iya}|\text{Positif}) = \frac{1+1}{100.6993161+22.16844444} = 0,016$$

$$P(\text{abang}|\text{Positif}) = \frac{0+1}{100.6993161+22.16844444} = 0,008$$

$$P(\text{gojek}|\text{Positif}) = \frac{1+1}{100.6993161+22.16844444} = 0,016$$

$$P(\text{mas}|\text{Positif}) = \frac{0+1}{100.6993161+22.16844444} = 0,008$$

$$P(\text{suka}|\text{Positif}) = \frac{0+1}{100.6993161+22.16844444} = 0,008$$

$$P(\text{banget}|\text{Positif}) = \frac{0+1}{100.6993161+22.16844444} = 0,008$$

$$P(\text{nanyain}|\text{Positif}) = \frac{0+1}{100.6993161+22.16844444} = 0,008$$

$$P(\text{nikah}|\text{Positif}) = \frac{0+1}{100.6993161+22.16844444} = 0,008$$

$$P(\text{korelasi}|\text{Positif}) = \frac{0+1}{100.6993161+22.16844444} = 0,008$$

b. Negatif

$$P(\text{iya}|\text{Negatif}) = \frac{0+1}{100.6993161+87.15727786} = 0,005$$

$$P(\text{abang}|\text{Negatif}) = \frac{0+1}{100.6993161+87.15727786} = 0,005$$

$$P(\text{gojek}|\text{Negatif}) = \frac{3+1}{100.6993161+87.15727786} = 0,021$$

$$P(\text{mas}|\text{Negatif}) = \frac{0+1}{100.6993161+87.15727786} = 0,005$$

$$P(\text{suka}|\text{Negatif}) = \frac{0+1}{100.6993161+87.15727786} = 0,005$$

$$P(\text{banget}|\text{Negatif}) = \frac{0+1}{100.6993161+87.15727786} = 0,005$$

$$P(\text{nanyain}|\text{Negatif}) = \frac{0+1}{100.6993161+87.15727786} = 0,005$$

$$P(\text{nikah}|\text{Negatif}) = \frac{0+1}{100.6993161+87.15727786} = 0,005$$

$$P(\text{korelasi}|\text{Negatif}) = \frac{0+1}{100.6993161+87.15727786} = 0,005$$

3. Perhitungan nilai *posterior probability* Gojek

$$P(\text{Opini}|\text{Positif}) = 0,4 * 0,008^7 * 0,016^2 = 8,58993459 * 10^{-20}$$

$$P(\text{Opini}|\text{Negatif}) = 0,6 * 0,005^8 * 0,021 = 4,92187500 * 10^{-21}$$

Berdasarkan contoh perhitungan didapatkan hasil dengan label yang paling tinggi adalah **Positif** dengan nilai  $8,58993459 * 10^{-20}$  Sehingga hasil klasifikasi untuk data uji adalah **Positif**.

4. Perhitungan nilai *prior probability* Indriver

a.  $P(\text{Positif}) = \frac{3}{5} = 0,6$

b.  $P(\text{Negatif}) = \frac{2}{5} = 0,4$

c. Perhitungan nilai *conditional probability* dengan *Laplacian smoothing* Indriver

1. Positif

$$P(\text{Bali}|\text{Positif}) = \frac{0+1}{74.63375166+48.53411346} = 0,008$$

$$P(\text{Gocar}|\text{Positif}) = \frac{0+1}{74.63375166+48.53411346} = 0,008$$

$$P(\text{Mahal}|\text{Positif}) = \frac{0+1}{74.63375166+48.53411346} = 0,008$$

$$P(\text{Coba}|\text{Positif}) = \frac{0+1}{74.63375166+48.53411346} = 0,008$$

$$P(\text{Indriver}|\text{Positif}) = \frac{1+1}{74.63375166+48.53411346} = 0,016$$

$$P(\text{Mas}|\text{Positif}) = \frac{0+1}{74.63375166+48.53411346} = 0,008$$

2. Negatif

$$P(\text{Bali}|\text{Negatif}) = \frac{0+1}{74.63375166+29.64660287} = 0,009$$

$$P(\text{Gocar}|\text{Negatif}) = \frac{0+1}{74.63375166+29.64660287} = 0,009$$

$$P(\text{Mahal}|\text{Negatif}) = \frac{0+1}{74.63375166+29.64660287} = 0,009$$

$$P(\text{Coba}|\text{Negatif}) = \frac{0+1}{74.63375166+29.64660287} = 0,009$$

$$P(\text{Indriver}|\text{Negatif}) = \frac{0+1}{74.63375166+29.64660287} = 0,009$$

$$P(\text{Mas}|\text{Negatif}) = \frac{0+1}{74.63375166+29.64660287} = 0,009$$

3. Perhitungan nilai *posterior probability* Indriver

$$P(\text{Opini}|\text{Positif}) = 0,6 * 0,008^5 * 0,016 = 3,14572800 * 10^{-13}$$

$$P(\text{Opini}|\text{Negatif}) = 0,4 * 0,009^6 = 2,12576400 * 10^{-13}$$

Berdasarkan contoh perhitungan didapatkan hasil dengan label yang paling tinggi adalah **Positif** dengan nilai  $3,14572800 * 10^{-13}$  Sehingga hasil klasifikasi untuk data uji adalah **Positif**.

4.3 Implementasi

Semua tahapan yang telah dilakukan pada saat analisis data akan di implementasikan kedalam suatu sistem dengan tujuan untuk melakukan analisis sentimen. Pada proses implementasi sistem analisis sentimen menggunakan bahasa pemrograman Python dan tools text editor yang digunakan Jupyter Notebook ditampilkan dalam Gambar 3 dan 4.

[2]:	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	lang	quote_count	reply_cc
0	1.780000e+18	Wed Apr 24 05:59:08 +0000 2024	816	gojek ketika kita ga buru2: gojek ketika ...	1.780000e+18		in	23	
1	1.780000e+18	Thu Apr 25 09:20:30 +0000 2024	697	vid ngabarin ke ayang klo otw pake gojek https...	1.780000e+18	https://pbs.twimg.com/ext_tw_video_thumb/17834...	in	1	
2	1.780000e+18	Sat Apr 27 11:33:26 +0000 2024	21	gojek yg gue pesen kenapa seneng bgt ya melawa...	1.780000e+18	https://pbs.twimg.com/media/GMKx8-pbAAAwRqC.jpg	in	1	
3	1.780000e+18	Sun Apr 21 07:47:57 +0000 2024	1796	AOWKOWOWKOWOWOK LUCU BGT GOJEK KETUKER https://...	1.780000e+18	https://pbs.twimg.com/media/GLuEzZrwbAAVtSx.jpg	in	16	
4	1.780000e+18	Tue Apr 23 10:59:40 +0000 2024	1730	Semangaaaat latihannya sampe naik gojek ngejar...	1.780000e+18	https://pbs.twimg.com/media/GL2D3oUaMAAbdns.jpg	in	29	
...	...	...	...	...	...	...	...	...	...

Gambar 3. Tampilan DataSet Gojek

[61]:	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	lo
0	1.783410e+18	Thu Apr 25 08:01:55 +0000 2024	2	romanticizing bandung with aa indrive yg outfi...	1.783410e+18	https://pbs.twimg.com/media/GL_uXH5bsAAMVoE.jpg	NaN	in	Ba
1	1.781410e+18	Fri Apr 19 19:56:16 +0000 2024	10	Ndamuwona ari muGD6 yeBlack ine masticker emus...	1.781410e+18	https://pbs.twimg.com/media/GUjYUxMXKAaegtj.jpg	NaN	in	I
2	1.783070e+18	Wed Apr 24 10:05:48 +0000 2024	71	Ndekupi kunohayiswia maFerrari nemaLambo chii c...	1.783070e+18	NaN	NaN	in	
3	1.781690e+18	Sat Apr 20 14:22:36 +0000 2024	2	NR nya bareng Abang indrive dulu bareng kamu k...	1.781690e+18	https://pbs.twimg.com/media/GLnVjKEbEAAe6Uz.jpg	NaN	in	Bar
4	1.782070e+18	Sun Apr 21 15:36:48 +0000 2024	0	jalan jalan bersama @inDrive https://t.co/nrFL...	1.782070e+18	https://pbs.twimg.com/media/GLswG_5bYAAzWKp.jpg	NaN	in	
...	...	...	...	...	...	...	...	...	...

Gambar 4. Tampilan DataSet Indriver

Pada proses ini akan digunakan data training yang sudah memiliki kelas (nilai) untuk melatih model sistem klasifikasi yang dibangun untuk melakukan prediksi dan klasifikasi pada data testing, sehingga pada saat proses testing berlangsung, sistem dapat mengetahui kelas dari setiap opini yang terdapat pada data testing. Berikut ini adalah tahapan untuk melakukan prediksi klasifikasi menggunakan metode Naïve Bayes Classifier di tampilan dalam Gambar 5 dan 6.

```
[54]: # objek naive bayes classifier
mnb = MultinomialNB()
mnb.fit(x_train, y_train)
mnb_pred = mnb.predict(x_test)
mnb_acc = accuracy_score(y_test, mnb_pred)
print("Test accuracy: {:.2f}%".format(mnb_acc*100))

Test accuracy: 73.33%
```

Gambar 5. Proses Klasifikasi Dataset Gojek

```
mnbIn = MultinomialNB()
mnbIn.fit(x_trainIn, y_trainIn)
mnb_predIn = mnbIn.predict(x_testIn)
mnb_accIn = accuracy_score(y_testIn, mnb_predIn)
print("Test accuracy: {:.2f}%".format(mnb_accIn*100))

Test accuracy: 85.71%
```

Gambar 6. Proses Klasifikasi Dataset Indriver

## 5. Kesimpulan

Berdasarkan hasil dari penelitian analisis sentimen masyarakat tentang kinerja layanan antara in-driver dan gojek dengan menggunakan Naïve Bayes Classifier yang telah dilakukan, dapat ditarik kesimpulan adalah Penerapan algoritma klasifikasi Naïve Bayes pada penelitian ini adalah menggunakan metode probabilitas untuk mengklasifikasikan dan melakukan prediksi kepada data latih gojek sebanyak 446 data dan data uji sebanyak 45 data pada data gojek, serta 346 data latih Indriver dan 35 data uji pada data indriver dengan Jupyter Notebook. Pada penelitian ini digunakan dataset yang diperoleh dengan teknik crawling sebanyak 792 data. Setelah dilakukan penyaringan secara manual dengan cara menghilangkan kalimat sarkasme, sindiran, serta data duplikat maka diperoleh 446 data

<http://sistemasi.ftik.unisi.ac.id>

tweet pertama dan 346 data tweet kedua dengan penyeragaman kata kunci “gojek” dan “indriner”. Pelabelan data dalam penelitian ini menggunakan Teknik lexicon-based dengan kamus berbahasa Indonesia. Teknik ini akan menghitung seberapa banyak kata yang terdeteksi dalam kamus terhadap dokumen (opini) dan menghitung skor sentimennya. Dengan menerapkan Teknik ini didapat rincian data dalam kelas positif sebanyak 208 data, dan negatif sebanyak 292 data. Hasil perhitungan data gojek menghasilkan nilai *accuracy* sebesar 73%, *precision* sebesar 72%, *recall* sebesar 100%, dan *f1-score* sebesar 84%. Pada hasil perhitungan data indriver diperoleh nilai *accuracy* sebesar 85,71%, *precision* sebesar 85,185%, *recall* sebesar 95,833%, dan *f1-score* sebesar 90,196%. Hal tersebut membuktikan bahwa algoritma klasifikasi Naïve Bayes lebih cocok digunakan pada jumlah data yang lebih sedikit.

## Referensi

- [1] S. Mandasari, B. H. Hayadi, and R. Gunawan, “Analisis Sentimen Terhadap Kualitas Pelayanan Aplikasi Gojek Menggunakan Metode Naive Bayes Classifier”, *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, Volume, vol. 5, pp. 118–126, 2022, [Online]. Available: <https://ojs.trigunadharna.ac.id/index.php/jsk/index>
- [2] N. M. A. J. Astari, Dewa Gede Hendra Divayana, and Gede Indrawan, “Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier,” *Jurnal Sistem dan Informatika (JSI)*, vol. 15, no. 1, pp. 27–29, Nov. 2020, doi: 10.30864/jsi.v15i1.332.
- [3] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, “Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, p. 542, Jan. 2022, doi: 10.30865/mib.v6i1.3530.
- [4] G. T. Santoso, “Analisis Sentimen Pada Tweet Dengan Tagar #Bpjsrasarentenir Menggunakan Metode Support Vectore Machine (Svm),” pp. 12–13, 2021.
- [5] Fitri Wulandari, Elin Haerani, Muhammad Fikry, and Elvia Budianita, “Analisis sentimen larangan penggunaan obat sirup menggunakan algoritma naive bayes classifier,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 88–96, May 2023, doi: 10.37859/coscitech.v4i1.4781.
- [6] A. Syakuro, “Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG) Halaman Judul Skripsi Oleh : Abdan Syakuro,” *Analisis sentimen masyarakat terhadap e-commerce pada media sosial menggunakan metode naive bayes classifier (NBC) dengan seleksi fitur information gain (IG)*, pp. 1–89, 2017.
- [7] M. Furqan, S. Sriani, and S. M. Sari, “Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia,” *Techno.Com*, vol. 21, no. 1, pp. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.
- [8] D. Nugraha and D. Gustian, “Analisis Sentimen Penggunaan Aplikasi Transportasi Online Pada Ulasan Google Play Store dengan Metode Naive Bayes Classifier,” 2024.
- [9] R. Amaliyyah, “Title,” no. February, p. 6, 2021.
- [10] S. Nurul, J. Fitriyyah, N. Safriadi, E. Esyudha, and P. #3, “JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes,” (*Jurnal Edukasi dan Penelitian Informatika*), vol. 5, no. 3, pp. 279–285, 2019.
- [11] H. Irsyad, A. Farisi, and M. R. Pribadi, “Klasifikasi Opini Masyarakat Terhadap Jasa ISP MyRepublic dengan Naïve Bayes,” 2019. [Online]. Available: <https://t.co/Q3btIa6MoF>
- [12] D. Garbian Nugroho, Y. Herry Chrisnanto, A. Wahana Jurusan Informatika, and F. Matematika dan Ilmu Pengetahuan Alam Universitas Jenderal Achmad Yani Jalan Terusan Jenderal Sudirman,

- Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode Naïve Bayes. International Journal of Science, Technology & Management*, vol. 4, no. 5, pp. 1132-1138, Sep. 2023.
- [13] M. Syarifuddin, “Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan KNN,” *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, Aug. 2020, doi: 10.33480/inti.v15i1.1347.
- [14] F. V. Sari and A. Wibowo, “Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,” *Jurnal SIMETRIS*, vol. 10, no. 2, 2019.
- [15] P. Yuniar and Kismiantini, “Analisis Sentimen Ulasan pada Gojek Menggunakan Metode Naive Bayes,” *Statistika*, vol. 23, no. 2, pp. 164–175, Dec. 2023, doi: 10.29313/statistika.v23i2.2353.
- [16] K. D. Indarwati and H. Februiyanti, “Analisis Sentimen Terhadap Kualitas Pelayanan Aplikasi Gojek Menggunakan Metode Naive Bayes Classifier,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 1, 2023, doi: 10.35957/jatisi.v10i1.2643.
- [17] R. T. Aldisa and P. Maulana, “Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM,” *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, pp. 106–109, Jun. 2022, doi: 10.47065/bits.v4i1.1581.
- [18] L. Oktasari *et al.*, *Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naïve Bayes Classifier*.
- [19] W. Yulita *et al.*, “Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier,” *JDMSI*, vol. 2, no. 2, pp. 1–9, 2021.