

Analisis Sentimen Publik Pada Aplikasi X Terhadap Kenaikan UKT Di Indonesia Menggunakan Algoritma Naive Bayes

Bandaharo^{*1}, Abdul Halim Hasugian²

^{1,2}Jurusan Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Indonesia
Email: ¹bandaharo08@gmail.com, ²abdulhalimhasugian@uinsu.ac.id

Abstrak

Pertumbuhan teknologi digital dan internet telah memungkinkan untuk memperoleh dan menyelesaikan berbagai tugas akademik dengan lebih cepat. Media sosial juga telah menjadi semakin penting sebagai platform untuk opini publik tentang berbagai topik, salah satunya adalah kebijakan perguruan tinggi terkait Uang Kuliah Tunggal (UKT). Kebijakan ini sering menjadi isu yang kontroversial jika itu berkaitan tentang kenaikan UKT. Permasalahan terhadap kenaikan UKT sering menjadi perhatian bagi orangtua maupun mahasiswa tersebut. Sehingga diberbagai kalangan, kenaikan UKT dapat menjadi masalah bagi mereka. Analisis Sentimen Publik adalah langkah yang tepat untuk menghitung tingkat opini positif dan negatif terkait isu kenaikan UKT. Dengan tujuan untuk mengukur sentimen publik terhadap kenaikan UKT algoritma Naive Bayes. Penelitian ini menggunakan data yang bersumber dari aplikasi X (Twitter) yang di crawling menggunakan Google Colab dengan jumlah data yaitu 1390 dan juga menggunakan metode Naive Bayes sebagai penyelesaian masalah. Hasil yang diperoleh setelah melakukan penelitian yaitu terdapat sebanyak 457 komentar negatif dan 933 komentar positif, kemudian untuk hasil evaluasi model diperoleh skor F-1 sebesar 0,82 atau 82%, akurasi sebesar 0,748 atau 74,8%, presisi sebesar 0,81 atau 81%, dan recall sebesar 0,84 atau 84%. Diharapkan kedepannya, penelitian ini mampu memberikan evaluasi kembali terhadap kementerian pendidikan apakah kenaikan UKT ini layak atau tidak untuk diperdayakan.

Kata kunci: *analisis sentimen, google colab, kenaikan UKT, naive bayes, UKT mahal*

Public Sentiment Analysis on Application X Regarding the Increase in UKT in Indonesia Using the Naive Bayes Algorithm

Abstract

The growth of digital technology and the internet has made it possible to obtain and complete various academic tasks more quickly. Social media has also become increasingly important as a platform for public opinion on various topics, one of which is the university policy regarding Single Tuition Fees (UKT). This policy often becomes a controversial issue when it involves an increase in tuition fees. The issue of the increase in UKT often becomes a concern for both parents and the students. Thus, in some circles, the increase in UKT can become a problem for them. Public Sentiment Analysis is the right step to calculate the level of positive and negative opinions related to the issue of UKT increases. With the aim of measuring public sentiment towards the increase in UKT using the Naive Bayes algorithm. This research uses data sourced from application X (Twitter) which was crawled using Google Colab with a total of 1390 data points and also employs the Naive Bayes method as a solution to the problem. The results obtained after conducting the research show that there are 457 negative comments and 933 positive comments. The model evaluation results yielded an F-1 score of 0.82 or 82%, an accuracy of 0.748 or 74.8%, a precision of 0.81 or 81%, and a recall of 0.84 or 84%. It is hoped that in the future, this research will be able to provide a reevaluation to the Ministry of Education on whether the increase in UKT is feasible or not to be implemented.

Keywords: *expensive UKT, increase UKT, google colab, naive bayes, sentiment analysis*

1. PENDAHULUAN

Kemajuan teknologi informasi telah berdampak signifikan pada berbagai aspek kehidupan, termasuk di bidang Pendidikan [1]. Dengan perkembangan internet dan teknologi digital, berbagai proses akademik kini dapat diakses dan dilakukan dengan lebih efisien. Teknologi memungkinkan penyebaran informasi yang cepat dan luas, memfasilitasi komunikasi antara institusi pendidikan dan siswa, serta menyediakan platform untuk

pembelajaran daring. Selain itu, media sosial telah menjadi alat penting untuk mengekspresikan opini publik tentang berbagai isu, termasuk kebijakan pendidikan [2].

Salah satu isu yang sering menarik perhatian di kalangan mahasiswa dan masyarakat umum adalah kenaikan Uang Kuliah Tunggal (UKT). UKT adalah komponen biaya yang harus dibayar oleh mahasiswa setiap semester dan merupakan salah satu faktor penting dalam keberlanjutan pendidikan tinggi [3]. Kenaikan UKT sering menimbulkan berbagai reaksi dari mahasiswa dan orang tua. Dalam era digital ini, sentimen publik terhadap kebijakan kenaikan biaya kuliah sering diekspresikan melalui media sosial. Oleh karena itu, analisis sentimen terhadap kebijakan tersebut dapat memberikan wawasan berharga bagi pembuat kebijakan dalam memahami opini publik dan merumuskan kebijakan yang lebih responsive [4].

Algoritma Naive Bayes adalah salah satu metode klasifikasi yang paling sederhana namun paling efektif untuk analisis sentimen. Metode ini didasarkan pada teorema Bayes dan asumsi bahwa fitur-fitur dalam dataset saling independen satu sama lain [5]. Beberapa alasan utama mengapa algoritma Naive Bayes dipilih dalam penelitian ini adalah karena algoritma ini merupakan salah satu yang paling sederhana untuk diimplementasikan dan dijalankan, Naive Bayes telah terbukti efektif dalam berbagai tugas klasifikasi teks, termasuk analisis sentimen, Naive Bayes memiliki ketahanan yang baik terhadap data yang tidak lengkap atau tidak sempurna, dan Naive Bayes sering memberikan kinerja yang sangat kompetitif, terutama ketika diterapkan pada dataset berbasis teks.

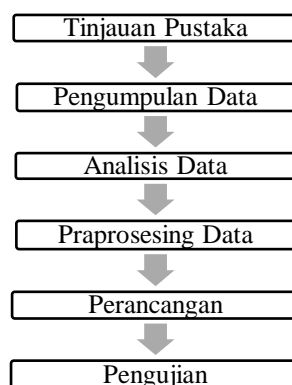
Beberapa penelitian terdahulu yang membahas tentang analisis sentimen terhadap kenaikan UKT diantaranya adalah penelitian yang dilakukan oleh [6], penelitian ini dilakukan dengan menggunakan metode VADER (Valence Aware Dictionary and Sentiment Reasoner). Dataset yang digunakan pada penelitian ini bersumber dari aplikasi X (Twitter) yang diperoleh melalui perpustakaan tweetharvest dari Node.js sebanyak 649 data. Hasil yang diperoleh pada penelitian ini berupa 243 bersifat netral, 217 bersifat positif dan 189 bersifat negatif. Penelitian berikutnya dari [7] dengan metode penelitian yang digunakan adalah SLP (Single Layer Perceptron). Penelitian ini juga melakukan analisis sentimen menggunakan SLP sebagai penyelesaian masalahnya. Dataset dari penelitian ini diperoleh melalui tools library Tweet-Harvest pada Python dengan jumlah data sebanyak 400 data. Hasil yang diperoleh pada penelitian ini berupa nilai rata-rata dari f-1 score, akurasi, presisi dan recall sebanyak 84% dari 10 kali percobaan yang telah dilakukan.

Berdasarkan penjelasan tersebut, peneliti melakukan studi berjudul "Analisis Sentimen Publik Pada Aplikasi X Terhadap Kenaikan UKT Di Indonesia Menggunakan Algoritma Naive Bayes." Dengan menggunakan algoritma Naive Bayes, penelitian ini bertujuan untuk memberikan solusi analisis sentimen yang efisien dan efektif terkait peningkatan UKT. Solusi yang diusulkan mencakup beberapa langkah seperti pengumpulan data, praproses data, pelatihan model, evaluasi model, dan analisis hasil.

2. METODE PENELITIAN

2.1. Kerangka Penelitian

Kerangka penelitian adalah struktur atau rencana dasar yang digunakan untuk membimbing penelitian dari awal hingga akhir. Kerangka ini mencakup berbagai elemen penting yang membantu peneliti dalam mengorganisir dan melaksanakan penelitian secara sistematis. Hal ini disebabkan oleh fakta bahwa penelitian selanjutnya harus disajikan dengan cara yang koheren dan teratur, sehingga memerlukan pembuatan kerangka penelitian. Oleh karena itu, kerangka penelitian harus dibuat terlebih dahulu sebelum memulai fase penelitian [8]. Adapun kerangka penelitian ini adalah sebagai berikut :



Gambar 1. Kerangka Penelitian

Selain kerangka penelitian, ada juga flowchart yang telah dibuat untuk alur dari proses penelitian yang akan dilakukan. Untuk flowchart nya dapat dilihat seperti pada gambar berikut ini :



Gambar 2. Flowchart

2.2. Pengumpulan Data

Pengumpulan data adalah proses sistematis untuk mengumpulkan informasi atau data yang relevan dan akurat guna menjawab pertanyaan penelitian, menguji hipotesis, atau mengevaluasi hasil. Metode pengumpulan data dapat bervariasi tergantung pada jenis penelitian, tujuan penelitian, dan sumber data [9]. Dalam penelitian ini, pengumpulan data dilakukan dengan mencrawling data di aplikasi Twitter (sekarang dikenal sebagai X). Proses crawling data pada aplikasi X dilakukan pada Google Colab menggunakan bahasa pemrograman Python.

2.3. Analisis Data

Dalam penelitian ini, analisis data diperlukan untuk memastikan pelaksanaannya yang sukses. Pada tahap ini, termasuk persyaratan yang digunakan dalam penelitian, terutama dataset tentang peningkatan Uang Kuliah Tunggal (UKT). Data tersebut dikumpulkan dari platform media sosial, yaitu Twitter. Kata kunci yang digunakan untuk pengumpulan data termasuk "kenaikan UKT", "UKT mahal", "biaya kuliah", dan "kebijakan UKT". Sebanyak 1397 posting teks terkait kenaikan UKT dikumpulkan selama periode 1 bulan.

2.4. Pembobotan

Pembobotan adalah suatu proses dalam menghitung nilai probabilitas untuk setiap kata yang muncul dalam data berdasarkan kategorinya. Berdasarkan pada penelitian yang dilakukan, maka perhitungan probabilitas data dilakukan untuk menghitung nilai postingan “positif” dan “negatif”. Tahapan ini sangat penting dalam algoritma Naive Bayes untuk menentukan kontribusi setiap kata terhadap klasifikasi sentimen suatu teks. Pembobotan dilakukan menggunakan metode frekuensi kata (Bag of Words) dan pendekatan probabilistik dalam algoritma Naive Bayes.

2.5. Klasifikasi Naive Bayes

Teorema Bayes digabungkan dengan "Naive", yang berarti bahwa setiap atribut atau variabel bersifat independen. Klasifikasi Naive Bayes dapat dilatih secara efisien dalam pembelajaran terawasi. Metode Pembelajaran Bayesian yang paling mudah, tercepat, dan sederhana. karena didasarkan pada teorema Bayes dan hipotesis independensi, menghasilkan pengklasifikasi statistik berdasarkan probabilitas [12]. Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada

variable kelas [13]. Algoritma ini mengasumsikan bahwa atribut objek bersifat independen. Langkah-langkah yang digunakan untuk perhitungan metode Naïve Bayes adalah sebagai berikut:

- Menentukan kriteria.
- Membuat tabel aturan.
- Menghitung probabilitas setiap nilai untuk atribut dari setiap kriteria yang sering muncul.
- Menghitung nilai dari setiap kriteria.
- Menghitung nilai probabilitas.

Probabilitas yang terlibat dalam menghasilkan estimasi akhir dihitung sebagai jumlah frekuensi dari tabel keputusan "master" [14]. Teorema Bayes memiliki rumus [15]:

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

Keterangan :

- X = Data dengan kelas yang belum diketahui
- H = Hipotesis data X merupakan suatu kelas spesifik
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)
- P(H) = Probabilitas hipotesis H (prior prob.)
- P(X|H) = Probabilitas X berdasarkan kondisi tersebut
- P(X) = Probabilitas dari X

Naïve Bayes Classifier mengadopsi ilmu statistika yaitu dengan menggunakan teori kemungkinan (Probabilitas) untuk menyelesaikan sebuah kasus Supervised Learning, artinya dalam himpunan data terdapat Label, Class atau Target sebagai acuan. Naïve bayes classifier merupakan Metode klasifikasi yang memiliki beberapa fase penyelesaian yaitu dimulai dari Training Data, Learning Algorithm, Model, Test Data dan diakhiri dengan proses Testing sehingga dihasilkan sebuah keputusan yang akurat. Model Naive Bayes mudah dibuat dan sangat berguna untuk kumpulan data yang sangat besar.

Naive Bayes juga merupakan perhitungan yang dapat mengelompokkan variable tertentu menggunakan kemungkinan dan strategi faktual. Naïve Bayes menggunakan bagian ilmu matematika yang dikenal sebagai hipotesis kemungkinan untuk melacak kemungkinan terbaik dari pesanan potensial, dengan memeriksa pengulangan setiap karakterisasi dalam data training. Keuntungan dari penggunaan metode Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil saja untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian dan dapat bekerja jauh lebih baik dalam situasi dunia nyata yang kompleks. Naïve Bayes merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki kemampuan dan akurasi tinggi [16].

2.6. Praprosesing Data

Setelah pengumpulan data dan analisis data, tahap berikutnya adalah praproses data, yang berguna untuk menghapus data yang tidak konsisten agar memperoleh data bersih, data yang akan digunakan untuk pengelompokan dengan metode Naïve Bayes [10]. Pada tahap ini, pemilihan data dan transformasi menjadi data numerik juga dilakukan. Dari 1397 pos, 1390 pos valid diperoleh dan siap untuk analisis lebih lanjut setelah proses pembersihan dan normalisasi. Adapun tahapan-tahapan yang dilakukan selama praprosesing data pada penelitian ini yaitu:

- Cleaning; tahap ini merupakan proses pembersihan data mentah yang diperoleh untuk membuang data yang tidak layak digunakan (noise) sehingga menyisakan data yang layak digunakan.
- Tokenisasi; dalam metode membagi secara kasar serangkaian karakter dalam teks menjadi kata-kata untuk membedakan antara karakter tertentu yang mungkin atau mungkin tidak diperlakukan sebagai jeda kata.
- Penghapusan Stopword; yaitu proses filtering, pemilihan kata-kata penting dari hasil token yaitu kata-kata apa saja yang digunakan untuk mewakili dokumen.
- Stemming; merupakan sebuah proses dalam menemukan kata dasar dari sebuah kata [5].

Dari desain proses sistem pada Gambar 2, telah disesuaikan dengan tahap-tahap implementasi algoritma Naïve Bayes menggunakan Google Colab. Di mana langkah pertama yang diambil adalah memasukkan dataset yang diperoleh dari Twitter, kemudian dataset tersebut akan diklasifikasikan menjadi data pelatihan dan data pengujian. Setelah data pelatihan diproses, langkah selanjutnya di Google Colab adalah memanggil fungsi Naive Bayes. Dimaksudkan untuk menghitung probabilitas data pengujian, yang menghasilkan prediksi yang akan menjadi referensi dalam menentukan jumlah dan persentase komentar positif dan negatif terhadap kenaikan UKT.

2.7. Pelabelan

Pada tahap ini, akan dilakukan pelabelan data. Data yang dilabelkan berguna untuk mempermudah dalam proses perhitungan. Merujuk pada penelitian, pelabelan dibagi menjadi dua kategori yaitu kata positif dan kata negatif. Setiap kategori memiliki bobot dan setiap kata memiliki bobot yang berbeda-beda. Untuk kategori kata positif, bobot berada pada rentang 1 sampai 3, sedangkan pada kata negatif, bobot berada pada rentang -1 sampai -5. Sehingga saat di hitung, dapat diketahui setiap kalimat tergolong dalam kategori positif atau negatif.

2.8. Confusion Matrix

Confusion Matrix adalah metode yang digunakan untuk mengukur akurasi dalam penambangan data, dengan dua kelas: kelas positif dan kelas negatif. Terdiri dari empat sel, yaitu True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). Perhitungan akurasi dilakukan dengan membagi jumlah data yang diklasifikasikan dengan benar dengan total sampel dari data yang diuji. Berikut adalah tampilan matriks kebingungan [16]:

Tabel 1. Confusion Matrix

Classification Observed Class	Predicted Class	
	Class = Yes a (True Positive-TP)	Class = No b (True Negative-TN)
	Class = Yes c (False Positive-FP)	Class = No d (False Negative-FN)

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

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

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

$$F1 - Score = 2 \times \frac{precision \times recall}{precision+recall} \quad (5)$$

3. HASIL DAN PEMBAHASAN

3.1. Analisis Data

Informasi yang digunakan dalam penelitian ini diperoleh dari Twitter, sebuah situs media sosial yang secara rutin digunakan oleh masyarakat umum untuk menyampaikan pemikiran, keluhan, atau dukungan terhadap suatu tujuan. Masalah kenaikan biaya tunggal (UKT) adalah topik utama dari penelitian ini. Menggunakan program pihak ketiga seperti Tweepy atau Twitter API, metode scraping digunakan untuk memperoleh data. Teks dari tweet yang relevan dengan topik penelitian membentuk pengumpulan data. Kata kunci berikut digunakan untuk memastikan bahwa data yang dikumpulkan relevan dengan studi ini:

- a) "kenaikan UKT"
- b) "UKT mahal"
- c) "biaya kuliah"
- d) "kebijakan UKT"

Kata kunci dirancang untuk mencakup berbagai pendapat yang mungkin diungkapkan publik mengenai kebijakan kenaikan biaya kuliah. Proses ini menghasilkan 1397 postingan terkait dengan isu kenaikan biaya kuliah. Data dikumpulkan selama periode 1 bulan dengan langkah-langkah berikut:

- a) Penyaringan Awal; data yang dikumpulkan hanya dari akun aktif, tidak mengandung spam, dan relevan dengan topik UKT.
- b) Pengambilan Data; data dalam bentuk teks tweet, nama pengguna (tanpa identitas pribadi), tanggal, dan jumlah interaksi (suka, retweet).
- c) Pra-pemrosesan Awal; data redundan seperti tweet duplikat atau teks yang tidak relevan dihapus.

Data yang diambil dari Twitter memerlukan praproses agar dapat dianalisis dengan lebih efektif. Tahapan pra-pemrosesan meliputi:

- a) Cleaning, menghilangkan elemen-elemen yang tidak diperlukan seperti link, hastag dan mention pada dataset.
- b) Tokenisasi, yang berarti memecah kalimat menjadi kata-kata individu.
- c) Penghapusan Stopword, yang berarti menghapus kata-kata umum yang tidak memiliki nilai sentimen (misalnya, "dan", "di", "yang").
- d) Stemming, yang merupakan mengubah sebuah kata ke bentuk dasarnya menggunakan algoritma seperti Porter Stemmer atau Sastrawi.

Setelah proses pembersihan dan normalisasi, 1390 pos valid diperoleh, siap untuk analisis lebih lanjut. Data ini kemudian dibagi menjadi 2, yaitu:

- a) Data Pelatihan: 80% dari total data (1113 tweet) digunakan untuk melatih algoritma Naive Bayes.
- b) Data Pengujian: 20% dari total data (279 tweet), digunakan untuk mengukur akurasi model.

Setiap tweet diberi label dengan sentimen berdasarkan analisis manual atau bantuan anotasi data dengan ketentuan berikut:

- a) Positif, tweet yang mendukung kebijakan tersebut atau menganggap kenaikan biaya kuliah tersebut wajar.
- b) Negatif, tweet yang menentang kebijakan tersebut atau menganggap kenaikan biaya kuliah tidak adil.

Berikut adalah beberapa data pelatihan yang digunakan, termasuk tweet, label sentimen, dan hasil pra-pemrosesan:

Tabel 2. Sampel Data Latih

No	Tweet Asli	Praprosesing Tweet	Sentimen
1	"UKT makin mahal aja, nggak masuk akal!"	"ukt mahal nggak masuk akal"	Negatif
2	"Naik UKT itu wajar, asal fasilitas diperbaiki."	"naik ukt wajar fasilitas diperbaiki"	Positif
3	"Gaenak banget nagih uang spp kuliah mahal."	"spp kuliah mahal"	Positif
4	"Kenapa UKT terus naik? Kami makin berat bayar kuliah."	"ukt naik berat bayar kuliah"	Negatif
5	"Kebijakan UKT naik sangat merugikan mahasiswa."	"kebijakan ukt naik merugikan mahasiswa"	Negatif
6	"Kenaikan UKT ini adalah langkah bagus untuk kampus maju."	"kenaikan ukt langkah bagus kampus maju"	Positif
7	"Tidak ada penjelasan logis tentang kenaikan UKT ini."	"tidak ada penjelasan logis kenaikan ukt"	Negatif
8	"UKT tinggi tapi kualitas pendidikan masih kurang."	"ukt tinggi kualitas pendidikan kurang"	Negatif
9	"Fasilitas membaik, jadi kenaikan UKT itu masuk akal."	"fasilitas membaik kenaikan ukt masuk akal"	Positif
10	" halo sentuh pokok beban ukt mahal wali orang tua mahasiswa uang kuliah kuliah jangkau mahal."	"ukt mahal uang kuliah jangkau mahal"	Positif

Pada Tabel 2 diatas, dapat dilihat hasil dari praprosesing tweet dalam bentuk sampel yang asli menjadi data yang dapat dihitung. Kemudian data tersebut disimpan dalam format .xls dengan kolom untuk teks tweet asli, teks yang telah diproses sebelumnya, dan label sentimen (positif dan negatif). Berdasarkan proses anotasi manual, data pelatihan dibagi menjadi kategori-kategori berikut:

- a) Negatif: 933 post
- b) Positif: 457 post

Kata-kata unik yang muncul dalam data dikumpulkan menjadi sebuah kosakata. Kosa kata dari data di atas adalah ["ukt", "mahal", "kenaikan", "fasilitas", "kampus", "semester", "terjangkau", "wajar"].

3.2. Pembobotan

Pembobotan data adalah proses menghitung nilai probabilitas untuk setiap kata yang muncul dalam data berdasarkan kategorinya. Merujuk pada penelitian ini, maka perhitungan probabilitas data dilakukan untuk menghitung nilai postingan “positif” dan “negatif”. Tahapan ini sangat penting dalam algoritma Naive Bayes untuk menentukan kontribusi setiap kata terhadap klasifikasi sentimen suatu teks. Pembobotan dilakukan menggunakan metode frekuensi kata (*Bag of Words*) dan pendekatan probabilistik dalam algoritma Naive Bayes. Berikut frekuensi kata per katagori sentimen yang dihitung menggunakan (*Bag of Words*).

Tabel 3. Frekuensi Kata per Kategori Sentimen

Kata	Positif	Negatif
"ukt"	15	50
"mahal"	2	40
"kenaikan"	8	30
"fasilitas"	5	25
"kampus"	7	20
"semester"	2	15
"terjangkau"	10	1
"wajar"	12	0

Selanjutnya, menghitung probabilitas awal. Probabilitas awal dihitung berdasarkan proporsi jumlah tweet di setiap kategori terhadap total data. Berikut perhitungan probabilitas awal pada penelitian ini:

$$P(\text{Positive}) = \frac{\text{Number of Positive Tweet}}{\text{Tweet Total}} \quad (2)$$

$$P(\text{Negative}) = \frac{\text{Number of Negative Tweet}}{\text{Tweet Total}} \quad (3)$$

Probabilitas kondisional dalam kategori tersebut kemudian ditentukan menggunakan rumus berikut ini:

$$P(\text{Kata}|\text{Kategori}) = \frac{\text{Frekuensi Kata Dalam Kategori}+1}{\text{Total Kata Dalam Kategori}+\text{Jumlah Total}} \quad (4)$$

Setelah penemuan rumus probabilitas bersyarat, setiap nilai probabilitas kemudian dihitung menggunakan rumus yang diberikan dalam Pers 4 di atas. Hasil dari perhitungan probabilitas bersyarat secara keseluruhan adalah sebagai berikut:

Table 4. Hasil Perhitungan Probabilitas Untuk Beberapa Kata

Word	P(Positive)	P(Negative)
"ukt"	0.148	0.197
"mahal"	0.027	0.158
"kenaikan"	0.083	0.120
"fasilitas"	0.055	0.100
"kampus"	0.074	0.081
"semester"	0.074	0.062
"terjangkau"	0.101	0.007
"wajar"	0.120	0.003

3.3. Klasifikasi Naïve Bayes

Menggunakan model probabilistik untuk mengelompokkan data menurut nilai probabilitas yang ditentukan dari data pelatihan dikenal sebagai klasifikasi dengan algoritma Naive Bayes. Pendekatan ini digunakan dalam penelitian ini untuk memastikan apakah tweet yang dikumpulkan dan diproses memiliki sikap positif, negatif, atau netral. Probabilitas posterior untuk setiap kata dalam tweet dihitung untuk mengklasifikasikan menggunakan Naive Bayes. Rumus berikut digunakan untuk menentukan probabilitas posterior untuk setiap kategori sentimen:

$$P(\text{Kategori} | \text{Tweet}) = P(\text{Kategori}) \times \prod_{i=1}^n P(\text{Kategori}) \quad (5)$$

Akibatnya, tiga kategori—negatif, positif, dan netral—digunakan untuk menghitung probabilitas posterior. Perhitungannya adalah sebagai berikut:

Perhitungan sentimen positif:

$$P(\text{Positif}|\text{Tweet}) = 0.8 \times 0.148 \times 0.027 \times 0.083 \times 0.055 =$$

$$P(\text{Positif}|\text{Tweet}) = 0.0000145934$$

Perhitungan sentimen negatif:

$$P(\text{Negatif}|\text{Tweet}) = 0.2 \times 0.197 \times 0.158 \times 0.120 \times 0.100 =$$

$$P(\text{Negatif}|\text{Tweet}) = 0.0000747024$$

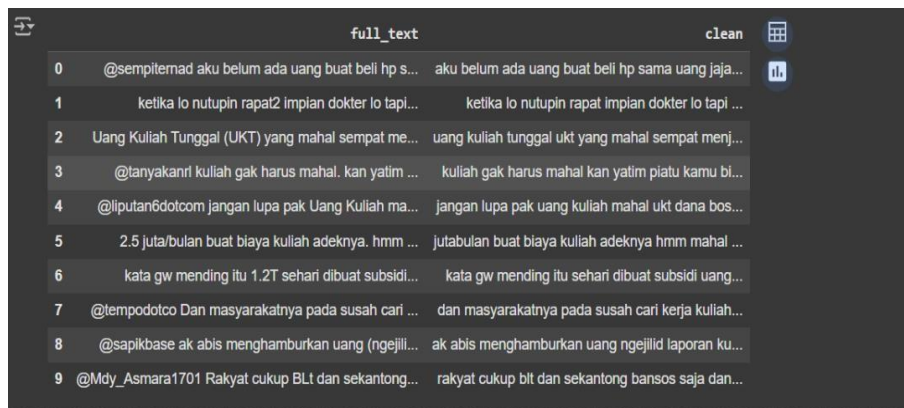
Dari hasil di atas, kategori dengan probabilitas tertinggi adalah Negatif. Oleh karena itu, tweet "UKT naik mahal, tidak masuk akal" termasuk dalam kategori negatif.

3.4. Praprosesing Data

Pada tahap ini akan dilakukan pemrosesan ulang data. Data yang telah diperoleh sebelumnya akan di lakukan pembersihan agar dapat digunakan untuk meneliti. Adapun beberapa tahapan dalam melakukan praprosesing data pada penelitian ini adalah:

a) Pembersihan (Cleaning)

Tahap ini bertujuan untuk menghilangkan elemen-elemen yang tidak diperlukan seperti mention, link dan hastag dari dataset. Dan dibawah ini merupakan hasil cleaning menggunakan Google Colab



Gambar 3. Hasil Cleaning Dataset Menggunakan Google Colab

b) Tokenizing

Setelah itu, tahap normalisasi dilakukan untuk mengubah kata-kata menjadi bentuk baku yang sesuai dengan KBBI. Hasil dari implementasi proses normalisasi di Google Colab dapat dilihat di sini.



Gambar 4. Proses Normalisasi Data Menggunakan Google Colab

c) Stopword Removal

Dalam langkah ini, kata-kata yang tidak memiliki arti dibuang. Hasil dari penggunaan metode penghapusan *stopword* di Google Colab ditunjukkan di sini.

	full_text	clean	tokenize	Normalisasi	Stopwords
0	@sompfemad aku belum ada uang buat beli hp s...	aku belum ada uang buat beli hp sama uang jajan	[aku, belum, ada, uang, buat, beli, hp, sama, k...	[saya, belum, ada, uang, buat, beli, hp, sama, k...	[uang, beli, hp, uang, jajan, slozin, banyu, k...
1	ketika lo nutupan rapat2 implan dokter lo tapx	ketika lo nutupan rapat implan dokter lo tapi ...	[ketika, lo, nutupan, rapat, implan, dokter, l...	[ketika, lo, nutupan, rapat, implan, dokter, l...	[lo, nutupan, rapat, implan, dokter, lo, kolaa...
2	Uang Kuliah Tunggal (UKT) yang mahal sempat me...	uang kuliah tunggal ukt yang mahal sempat meng...	[uang, kuliah, tunggal, ukt, yang, mahal, semp...	[uang, kuliah, tunggal, ukt, yang, mahal, semp...	[uang, kuliah, tunggal, ukt, mahal, sontan, l...
3	@tanyakani kuliah gak harus mahal kan yatin	kuliah gak harus mahal kan yatin paku koma la...	[kuliah, gak, harus, mahal, kan, yatin, paku, koma, la...	[kuliah, tidak, harus, mahal, kan, yatin, paku, koma, la...	[kuliah, mahal, yatin, paku, bantu, ipan, lo...
4	@ipufanidocom jingan lupa pak uang kuliah mahal ukt dana bos...	jinjan lupa pak uang kuliah mahal ukt dana bos...	[jinjan, lupa, pak, uang, kuliah, mahal, ukt, ...]	[jinjan, lupa, pak, uang, kuliah, mahal, ukt, ...]	[lupa, uang, kuliah, mahal, ukt, dana, bos, ua...
1393	@Nickkyandiego @tanyakani Kayaknya kpadan di video blom era ukt dh era ...	kayaknya kpadan di video blom era ukt dh era ...	[kayaknya, kpadan, di, video, blom, era, ukt, ...]	[kayaknya, kpadan, di, video, blom, era, ukt, ...]	[kayaknya, kpadan, video, blom, era, ukt, dh, ...]
1394	Masih kepikiran kalau gw masih di R wow ukt u...	masih kepikiran kalau gw masih di R wow ukt u...	[masih, kepikiran, kalau, gw, masih, di, R, w...	[masih, kepikiran, kalau, gw, masih, di, R, w...	[kepikiran, gw, R, wow, ukt, under, uang, sak...
1396	@EloKoda @ravuloka #RAYAKAN You Way @savok... ranyakanyouway pengalaman liburan pating bechi...	ranyakanyouway pengalaman liburan pating bechi...	[ranyakanyouway, pengalaman, liburan, pating, bechi, ...]	[ranyakanyouway, pengalaman, liburan, pating, bechi, ...]	[ranyakanyouway, pengalaman, liburan, bechi, ...]
1398	@tanyakani emang kuliah harus bgt pako mobil biay...	emang kuliah harus bgt pako mobil biay...	[emang, kuliah, harus, bgt, pako, mobil, biaya, ...]	[emang, kuliah, harus, bgt, pako, mobil, biaya, ...]	[emang, kuliah, harus, bgt, pako, mobil, biaya, k...
1397	@drosodrifl ipb naik bgt uang pangkalnya naik	ipb naik bgt uang pangkalnya naik lipot dri la...	[ipb, naik, bgt, uang, pangkalnya, naik, lipot, dri, la...	[ipb, naik, bgt, uang, pangkalnya, naik, lipot, dri, la...	[ipb, bngot, uang, pangkalnya, lipot, dari...

Gambar 5. Hasil Proses *Stopword Removal* Menggunakan Google Colab

d) **Stemming**

Mengonversi token ke bentuk dasar dan menghapus semua imbuhan adalah langkah berikutnya dalam proses menyeragamkannya. Saat ini, Perpustakaan Sastrawi digunakan. Berikut adalah gambar yang dihasilkan dari proses stemming yang dilakukan di Google Colab dan selanjutnya proses detokenized untuk menghilangkan fitur tokenize yang dilakukan.

	full_text	clean	tokenize	Normalisasi	Stopwords	stemming	detokenized_text
0	@sompfemad aku belum ada uang buat beli hp s...	aku belum ada uang buat beli hp sama uang jajan	[aku, 'belum, 'ada, 'uang, 'buat, 'beli, 'hp, 'sama, 'u...	[saya, 'belum, 'ada, 'uang, 'buat, 'beli, ...]	[uang, 'beli, 'hp, 'uang, 'jajan, 'sana...	[uang, 'beli, 'hp, 'uang, 'jajan, 'sana...	uang beli hp uang jajan akan biaya kuliah ma...
1	ketika lo nutupan rapat2 implan dokter lo tapx	ketika lo nutupan rapat implan dokter lo tapi ...	[ketika, 'lo, 'nutupan, 'rapat, 'implan, 'dokter, 'lo, 'tapi, ...]	[ketika, 'lo, 'nutupan, 'rapat, 'implan, 'dokter, 'lo, 'tapi, ...]	[lo, 'nutupan, 'rapat, 'implan, 'dokter, 'lo, 'kolaa...	[lo, 'nutupan, 'rapat, 'implan, 'dokter, 'lo, 'kolaa...	lo nutupan rapat implan dokter lo kolaa...
2	Uang Kuliah Tunggal (UKT) yang mahal sempat me...	uang kuliah tunggal ukt yang mahal sempat meng...	[uang, 'kuliah, 'tunggal, 'ukt, 'yang, 'mahal, 'sempat, 'meng...	[uang, 'kuliah, 'tunggal, 'ukt, 'yang, 'mahal, 'sempat, 'meng...	[uang, 'kuliah, 'tunggal, 'ukt, 'mahal, 'sontan, 'l...	[uang, 'kuliah, 'tunggal, 'ukt, 'mahal, 'sontan, 'l...	uang kuliah tunggal ukt mahal sorot antar beb...
3	@tanyakani kuliah gak harus mahal kan yatin	kuliah gak harus mahal kan yatin paku koma la...	[kuliah, 'gak, 'harus, 'mahal, 'kan, 'yatin, 'paku, 'koma, 'la...	[kuliah, 'tidak, 'harus, 'mahal, 'kan, 'yatin, 'paku, 'koma, 'la...	[kuliah, 'mahal, 'yatin, 'paku, 'bantu, 'ipan, 'lo...	[kuliah, 'mahal, 'yatin, 'paku, 'bantu, 'ipan, 'lo...	kuliah mahal yatin paku bantu ipan lo kula...
4	@ipufanidocom jingan lupa pak uang kuliah mahal ukt dana bos...	jinjan lupa pak uang kuliah mahal ukt dana bos...	[jinjan, 'lupa, 'pak, 'uang, 'kuliah, 'mahal, 'ukt, 'dana, 'bos, ...]	[jinjan, 'lupa, 'pak, 'uang, 'kuliah, 'mahal, 'ukt, 'dana, 'bos, ...]	[lupa, 'uang, 'kuliah, 'mahal, 'ukt, 'dana, 'bos, 'ua...	[lupa, 'uang, 'kuliah, 'mahal, 'ukt, 'dana, 'bos, 'ua...	lupa uang kuliah mahal ukt dana bos uang sekot...
1393	@Nickkyandiego @tanyakani Kayaknya kpadan di video blom era ukt dh era ...	kayaknya kpadan di video blom era ukt dh era ...	[kayaknya, 'kpadan, 'di, 'video, 'blom, 'era, 'ukt, 'dh, 'era, ...]	[kayaknya, 'kpadan, 'di, 'video, 'blom, 'era, 'ukt, 'dh, 'era, ...]	[kayaknya, 'kpadan, 'video, 'blom, 'era, 'ukt, 'dh, 'era, ...]	[kayak, 'kpadan, 'video, 'blom, 'era, 'ukt, 'dh, 'era, ...]	kayak kpadan video blom era ukt dh era ukt ka...
1394	Masih kepikiran kalau gw masih di R wow ukt u...	masih kepikiran kalau gw masih di R wow ukt u...	[masih, 'kepikiran, 'kalau, 'gw, 'masih, 'di, 'R, 'wow, 'ukt, 'u...	[masih, 'kepikiran, 'kalau, 'gw, 'masih, 'di, 'R, 'wow, 'ukt, 'u...	[kepikiran, 'gw, 'R, 'wow, 'ukt, 'under, 'uang, 'sak...	[paku, 'gw, 'R, 'wow, 'ukt, 'under, 'uang, 'sak...	paku gw R wow ukt under uang sak...
1397	@EloKoda @ravuloka #RAYAKAN You Way @savok... ranyakanyouway pengalaman liburan pating bechi...	ranyakanyouway pengalaman liburan pating bechi...	[ranyakanyouway, 'pengalaman, 'liburan, 'pating, 'bechi, ...]	[ranyakanyouway, 'pengalaman, 'liburan, 'pating, 'bechi, ...]	[ranyakanyouway, 'pengalaman, 'liburan, 'p...	[ranyakanyouway, 'pengalaman, 'liburan, 'p...	ranyakanyouway akan libur kesian cngga seru pa...
1398	@tanyakani emang kuliah harus bgt pako mobil biay...	emang kuliah harus bgt pako mobil biay...	[emang, 'kuliah, 'harus, 'bgt, 'pako, 'mobil, 'biaya, ...]	[emang, 'kuliah, 'harus, 'bngot, 'pako, 'mobil, 'biaya, ...]	[emang, 'kuliah, 'bngot, 'pako, 'mobil, 'biaya, ...]	[emang, 'kuliah, 'bngot, 'pako, 'mobil, 'biaya, ...]	emang kuliah bngot pako mobil biaya kula...
1397	@drosodrifl ipb naik bgt uang pangkalnya naik	ipb naik bgt uang pangkalnya naik lipot dri la...	[ipb, 'naik, 'bgt, 'uang, 'pangkalnya, 'naik, 'lipot, 'dri, 'la...	[ipb, 'naik, 'bngot, 'uang, 'pangkalnya, 'naik, 'lipot, 'dri, 'la...	[ipb, 'bngot, 'uang, 'pangkalnya, 'lipot, 'dari, 'la...	[ipb, 'bngot, 'uang, 'pangkalnya, 'lipot, 'dari, 'la...	ipb bngot uang pangkalnya lipot dari lau na...

Gambar 6. Hasil Proses *Stemming* dan *Detokenized Text*

3.5 **Pelabelan**

Langkah selanjutnya adalah proses melabelkan data dimana dilakukan secara otomatis dengan mengimplementasikan kamus InSetLexicon. Nilai polaritas ditentukan dalam kamus InSetLexicon dengan kata-kata dan bobotnya. Untuk mengetahui nilai polaritas, Anda harus menghitung semua bobot kata dalam teks ulasan. Hasil perhitungannya adalah ulasan untuk setiap elemen dikategorikan sebagai sentimen positif atau negatif. Ulasan dianggap positif jika nilai polaritasnya lebih besar dari nol, dan negatif jika nilai polaritasnya lebih rendah dari nol. Untuk hasil pelabelan menggunakan InSetLexicon dapat dilihat seperti pada tabel dibawah ini:

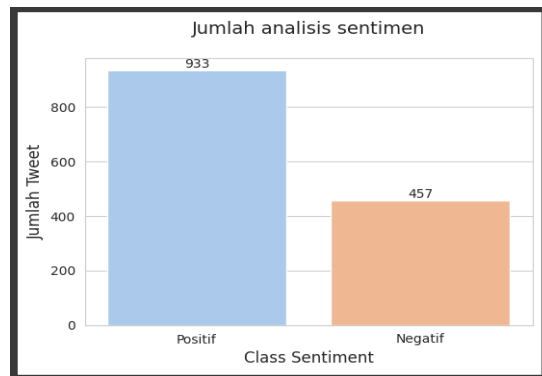
Tabel 5. Positive Lexicon

No.	Word	Weight
1	Hai	3
2	Tetap	3
3	Detail	2
4	Bagus	2
...
3610	Orisinal	3

Tabel 6. Negative Lexicon

No.	Word	Weight
1	Isak	-5
2	Sakit	-5
3	Mulu	-1
4	Gamau	-4
...
6610	Mencoreng	-4

Pada Tabel 5 dan 6 dapat dilihat hasil pelabelan positif dan negatif yang diberi nilai beragam-ragam untuk setiap kata. Setelah melakukan pelabelan dengan menetapkan setiap bobot berdasarkan kata yang ada, maka dilakukan perhitungan untuk mengetahui nilai dari setiap tweet yang ada pada data. Sehingga memperoleh hasil analisis sentimen seperti gambar dibawah in:

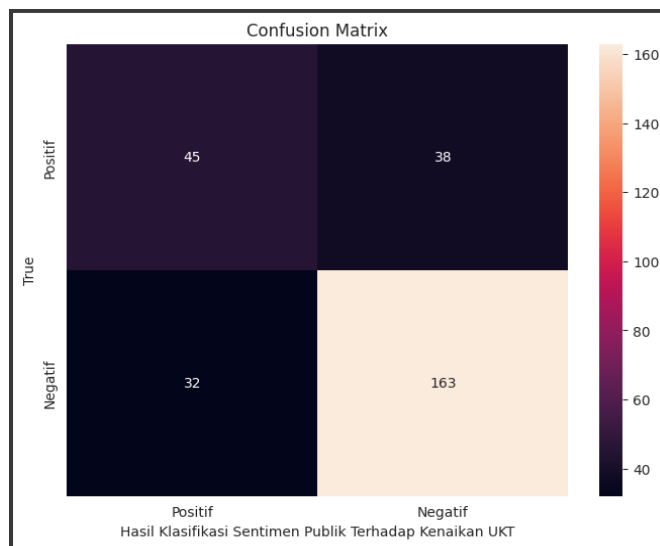


Gambar 7. Hasil Analisis Sentimen

Dari Gambar 7 diatas dapat dilihat hasil perhitungan jumlah tweet positif dan negatif menggunakan analisis sentimen melalui label-label yang telah ditetapkan. Sehingga data tweet positif pada data sebanyak 933 dan tweet negatif sebanyak 457.

3.6. Confusion Matrix

Langkah selanjutnya adalah mengevaluasi model dengan menggunakan confusion matrix untuk menghitung nilai akurasi. Selain nilai akurasi, nilai presisi, recall dan f-1 score juga akan dihitung untuk memperkuat hasil evaluasi model. Dalam proses perhitungan, nilai yang akan digunakan adalah nilai dari perhitungan tweet positif karena hasil analisis yang paling banyak diperoleh dari tweet positif yaitu berjumlah 933 data. Maka untuk hasil confusion matrix dapat dilihat seperti pada gambar berikut ini:



Gambar 8. Confusion Matrix

Setelah memperoleh nilai confusion matrix, maka langkah berikutnya yaitu menghitung nilai akurasi, presisi, recall dan F1-score menggunakan perumusan confusion matrix yang telah di jabarkan pada subbab sebelumnya yaitu rumus (2), (3), (4) dan (5). Sehingga berdasarkan perhitungan tersebut, hasil yang diperoleh untuk akurasi adalah 0.748 atau 74,8%, presisi 0.81 atau 81%, recall 0.84 atau 84% dan F1-Score yaitu 0.82 atau 82%.

```

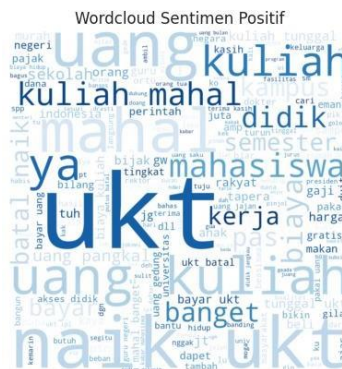
↳ Hasil Akurasi: 0.7482014388489209
      precision    recall  f1-score   support

   Negatif         0.58      0.54      0.56         83
   Positif         0.81      0.84      0.82        195

 accuracy                   0.75         278
 macro avg                 0.70         278
 weighted avg              0.74         278
    
```

Gambar 9. Hasil Perhitungan Evaluasi Model Pada Google Colab

Setelah memperoleh hasil evaluasi model, langkah selanjutnya adalah menampilkan hasil wordcloud sentimen positif dan negatif yang dilakukan pada google colab menggunakan library matplotlib. Wordcloud merupakan hasil visualisasi kumpulan kata kunci yang telah ditemukan dan disusun berdasarkan ukuran dan warna untuk membentuk suatu gambar. Sehingga untuk hasil wordcloud sentimen positif dan negatif dapat dilihat seperti pada gambar berikut ini:



Gambar 10. Wordcloud Sentimen Positif



Gambar 11. Wordcloud Sentimen Negatif

4. KESIMPULAN

Berdasarkan hasil penelitian *Analisis Sentimen Publik Terhadap Kenaikan UKT Menggunakan Algoritma Naive Bayes*, dapat disimpulkan bahwa mayoritas sentimen publik terhadap kebijakan kenaikan UKT yang dianalisis melalui platform Twitter cenderung bersifat positif dibandingkan negatif. Nilai ini menunjukkan bahwa banyak orang di Twitter atau X yang setuju atau menanggapi baik terhadap kebijakan kenaikan UKT. Jumlah yang diperoleh pada tweet positif setelah dilakukan analisis yaitu 933 sedangkan tweet negatif sebanyak 457. Namun tetap saja kenaikan UKT ini juga memberikan masalah serius bagi beberapa kalangan dikarenakan

tidak semuanya memiliki perekonomian yang baik sehingga kenaikan UKT ini menjadi permasalahan bagi beberapa pihak. Hasil ini juga memberikan evaluasi kepada pemerintah dan pemberi kebijakan untuk dapat mengatur ulang atau bertindak selayak dan seadilnya terhadap mahasiswa agar isu kenaikan UKT ini dapat di minimalisir.

Selain itu juga, berdasarkan hasil evaluasi model yang telah dilakukan untuk mengukur kelayakan model memberikan hasil berupa akurasi sebesar 74,8%, presisi sebesar 81%, recall sebesar 84% dan f1-score sebesar 82%. Berdasarkan hasil tersebut, metode naive bayes ini cukup baik dan layak untuk digunakan dalam analisis sentimen publik.

DAFTAR PUSTAKA

- [1] K. K. Nasution, I. Zufria dan M. Fakhriza, "SISTEM PENDUKUNG KEPUTUSAN PEMBERIAN KELAYAKAN KREDIT USAHA RAKYAT DENGAN MENGGUNAKAN METODE SAW & TOPSIS," *JURNAL RESPONSIF*, vol. 6, no. 1, pp. 143-154, 2024.
- [2] J. W. Iskandar dan Y. Nataliani, "Perbandingan Naïve Bayes, SVM dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1120-1126, 2021.
- [3] S. D. Prasetyo, S. S. Hilabi dan F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, vol. 10, no. 1, pp. 1-7, 2023.
- [4] N. L. P. Merawati, A. Z. Amrullah dan 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, 2021.
- [5] A. H. Hasugian, R. A. Putri dan M. A. Simatupang, "PENERAPAN ALGORITMA KLASIFIKASI NAÏVE BAYES UNTUK ANALISIS SENTIMEN TENTANG PEMINDAHAN IBU KOTA NEGARA," *Journal of Science and Social Research*, vol. 7, no. 2, pp. 635-644, 2024.
- [6] S. Z. Rozaan, M. R. Andrianto, R. S. Purnama, N. I. Ramadhan, W. P. Widyadhana dan R. B., "Analisis Sentimen terhadap Kenaikan UKT di Indonesia pasca Terpilihnya Capres 02 menggunakan VADER," *Seminar Nasional Informatika Bela Negara (SANTIKA)*, vol. 4, pp. 88-92, 2024.
- [7] A. C. N. Rosyadi, K. R. Leonida, M. Athoillah dan H. B. Rohmanto, "ANALISIS SENTIMEN DI TWITTER TERHADAP KENAIKAN UANG KULIAH TUNGGAL MENGGUNAKAN ARTIFICIAL NEURAL NETWORK (ANN)," *SEMINAR NASIONAL HASIL RISET DAN PENGABDIAN*, pp. 25-31, 2024.
- [8] A. Z. Syahputri, F. D. Fallenia dan R. Syafitri, "Kerangka Berfikir Penelitian Kuantitatif," *Tarbiyah : Jurnal Ilmu Pendidikan dan Pengajaran*, vol. 2, no. 1, pp. 160-166, 2023.
- [9] Z. Yusra, R. Zulkarnain dan Sofino, "PENGELOLAAN LKP PADA MASA PENDEMIK COVID-19," *Journal Of Lifelong Learning*, vol. 4, no. 1, pp. 15-22, 2021.
- [10] F. Alghifari dan D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *Jurnal Ilmiah Informatika (JIF)*, vol. 9, no. 2, pp. 76-81, 2021.
- [11] M. U. Albab, Y. Karuniawati dan M. N. Fawaiq, "Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic," *Jurnal TRANSFORMATIKA*, vol. 20, no. 2, pp. 1-10, 2023.
- [12] W. Wahyuni, "Analisis Sentimen terhadap Opini Feminisme Menggunakan Metode Naive Bayes," *Jurnal Informatika Ekonomi Bisnis*, vol. 4, no. 4, pp. 148-153, 2022.
- [13] E. Martantoh dan N. Yanih, "Implementasi Metode Naive Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan PHP MySQL," *JTSI*, vol. 3, no. 2, pp. 166-175, 2022.
- [14] M. I. Ghozali, W. H. Sugiharto dan A. F. Iskandar, "Analisis Sentimen Pinjaman Online Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1340-1348, 2023.
- [15] R. Saputra dan N. Hasan F, "Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes," *Jurnal Teknologi dan Sistem Informasi Bisnis (JTeksis)*, vol. 6, no. 3, pp. 411-419, 2024.
- [16] N. S. Fauzia dan R. D. Dana, "Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunung Sari," *Jurnal Blend Sains*, vol. 1, no. 4, pp. 295-305, 2023.