

ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN KENAIKAN UANG KULIAH TUNGGAL (UKT) PADA MEDIA SOSIAL X MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*

Nota Krisman Jaya Gulo¹, Purwanto^{2*}

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹2011500812@student.budiluhur.ac.id, ^{2*}purwanto@budiluhur.ac.id
(* : *corresponding author*)

Abstrak-Uang Kuliah Tunggal atau disingkat dengan UKT adalah merupakan sistem pembayaran biaya pendidikan perguruan tinggi negeri yang menjadi dasar penetapan biaya pendidikan yang dibebankan kepada mahasiswa, masyarakat, dan pemerintah. Kebijakan kenaikan UKT oleh pemerintah pada tanggal 3 Mei 2024 mengakibatkan kontroversi pada masyarakat pengguna X. Analisis sentimen merupakan proses pengklasifikasikan sentimen positif atau negatif dari opini atau pendapat seseorang terhadap suatu hal. Penelitian ini, peneliti melakukan analisis sentimen pada masyarakat di media sosial X terhadap kebijakan kenaikan UKT dengan menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasikan data dan memprediksi sentimen dari masyarakat. *Naïve Bayes Classifier* adalah sebuah algoritma yang berfungsi untuk menemukan nilai probabilitas tertinggi untuk mengklasifikasikan ke dalam kategori yang sesuai. *Naïve Bayes Classifier* juga digunakan untuk klasifikasi teks, mengidentifikasi apakah suatu *tweet* bersifat negatif atau positif terkait sentimen masyarakat pada kebijakan kenaikan UKT. *Dataset* yang dianalisis bersumber dari media sosial X dengan proses *Crawling* dari tanggal 3 Mei 2024 hingga 26 Mei 2024. Hasil dari penelitian ini mendapatkan nilai sentimen positif sebesar 61 dan sentimen negatif 189 dari 250 data. Hasil pengujian menghasilkan tingkat *accuracy* sebanyak 80% atau baik, *precision* positif atau buruk 50%, *precision* negatif atau baik sebanyak 85,7%, *Recall* positif atau buruk sebanyak 40%, *Recall* negatif atau baik 90%, dengan perbandingan 80:20 untuk data latih dan data uji.

Kata Kunci: UKT, Analisis Sentimen, X, *Naive Bayes Classifier*, *Crawling*

ANALYSIS OF PUBLIC SENTIMENT TOWARDS THE POLICY OF SINGLE TUITION INCREASE (UKT) ON SOCIAL MEDIA X USING THE *NAIVE BAYES CLASSIFIER METHOD*

Abstract- *Single Tuition Fee* or abbreviated as *UKT* is a system for paying state university education fees which is the basis for determining education fees charged to students, the community and the government. The government's *UKT* increase policy on May 3 2024 resulted in controversy among the X user community. Sentiment analysis is the process of classifying positive or negative sentiment from a person's opinion or views on something. In this research, researchers conducted sentiment analysis among the public on social media X regarding the *UKT* increase policy using the *Naïve Bayes Classifier* method to classify data and predict sentiment from the public. *Naïve Bayes Classifier* is an algorithm that functions to find the highest probability value to classify it into the appropriate category. *Naïve Bayes Classifier* is also used for text classification, identifying whether a *tweet* is negative or positive regarding public sentiment regarding the *UKT* increase policy. The dataset analyzed was sourced from social media. The test results produced an accuracy level of 80% or good, positive or bad precision of 50%, negative or good precision of 85.7%, positive or bad recall of 40%, negative or good recall of 90%, with a ratio of 80:20 for the data train and test data.

Keywords: *UKT*, Sentiment Analysis, X, *Naïve Bayes Classifier*, *Crawling*

1. PENDAHULUAN

Media sosial merupakan suatu *platform* atau aplikasi yang selalu terhubung dengan internet dengan memungkinkan dapat berinteraksi satu sama lain dengan berbagi informasi seperti gambar, teks dan video. Di dalam media sosial X terdapat berbagai opini yang bisa didapatkan pada permasalahan yang sedang terjadi. Media sosial X menjadi *platform* untuk berbagi dan mendiskusikan berbagai *trend* topik, termasuk masalah kebijakan pemerintah terkait kenaikan Uang Kuliah Tunggal (UKT) yang mengakibatkan kontroversi pada masyarakat pengguna media sosial X.

Uang Kuliah Tunggal (UKT) adalah merupakan sistem pembayaran biaya pendidikan perguruan tinggi negeri yang menjadi dasar penetapan biaya pendidikan yang dibebankan kepada mahasiswa, masyarakat, dan pemerintah. Pembayaran UKT diimplementasikan sesuai dengan pasal 88 ayat 1 sampai 5 dalam undang-undang No. 12 Tahun 2012 tentang pendidikan Tinggi. Ayat 1 menyatakan bahwa pemerintahan menetapkan standar satuan operasional

pendidikan tinggi secara berkala dengan mempertimbangkan pencapaian standar nasional pendidikan tinggi, jenis program studi, dan indeks kemahalan lokal [1].

Analisis sentimen adalah merupakan suatu pengklasifikasian sentimen terhadap sisi positif atau negatif dari berbagai hasil ekstraksi opini pada seseorang dalam peristiwa yang telah atau sedang terjadi. Analisis sentimen diberikan pada pengguna internet media sosial yang bisa memberikan suatu penilaian opini pribadi [2]. Analisis sentimen dapat digunakan untuk memahami opini, perasaan, dan pandangan publik, yang bisa membantu masyarakat dalam mengambil suatu keputusan yang lebih baik. Salah satu metode yang bisa dapat digunakan yaitu metode *Naive Bayes Classifier*.

Penelitian sebelumnya telah dilakukan oleh [3], perbedaan antara penelitian ini dengan penelitian sebelumnya adalah terletak pada pengambilan datanya. Penelitian sebelumnya menggunakan Rstudio untuk pengambilan data dari X, sedangkan penelitian ini menggunakan *crawling* data dari *tools google colab*, Penelitian ini dilakukan untuk mengamati tren kebijakan kenaikan UKT di masyarakat. Dengan demikian, diharapkan hasil dari penelitian ini dapat memberikan manfaat bagi masyarakat dalam mempertimbangkan kebijakan kenaikan UKT.

Data yang digunakan dalam penelitian ini adalah data *tweet* yang diperoleh dari X melalui data *crawling*. Data *tweet* akan diklasifikasikan menjadi 2 kategori yaitu positif dan negatif, sehingga diharapkan sistem mampu mengklasifikasikan algoritma *Naive Bayes Classifier* dapat membantu memberikan gambaran mengenai sentimen masyarakat terkait dengan kebijakan kenaikan UKT pada media sosial X.

2. METODE PENELITIAN

2.1 Data Penelitian

Data yang digunakan dalam penelitian ini berasal dari X dan terdiri dari *tweet* yang dikumpulkan antara tanggal 03 Mei 2024 hingga 26 Mei 2024. Jumlah data yang berhasil diperoleh adalah 250 baris data. Proses pengumpulan data dilakukan dengan menggunakan *tools google colab* bernama *Tweet harvest* melalui metode *crawling*. *Dataset* yang dikumpulkan berdasarkan pada parameter kata kunci ‘UKT’. Tabel 1 menunjukkan beberapa contoh hasil dari *crawling* data yang telah dikumpulkan.

Tabel 1. Data Penelitian

No	Username	Full text
1	hend_85	@BuanaJannah @MardaniAliSera @prabowo @PKSejahtera @jokowi Saya lebih suka lihat negara hancur korupsi merajalela rakyat sengsara... Dan mereka yg dungu memilihnya kemarin pada teriak kayak kejadian UKT mahasiswa yg naik...
2	CakYunus2	@tvOneNews Miris sekali Saat negara kelangkaan sarjana dengan presentasi populasi penduduk indonesia saat ini malah membuat kebijakan biaya UKT yang tinggi
3	lemon_taa	@alfan_rahadian @ardisatriawan Terjangkau kalau dapet UKT golongan 1 atau 2. Masalahnya sekarang banyak kampus negeri yang UKTnya mahal tapi fasilitas ga oke. Bahkan kampus negeri yang ga terkenal aja UKT nya mahal loh
...
250	EncepRosyidin	@MartinusButarb1 Terus kenapa ukt nya yg malah diperbesar gara2 anggota dpr pake kip. Seharusnya bikin aturannya buat cabut kip dari pejabat negara dan orang kaya bukan menaikkan ukt dan mencabut kip semuanya

2.2 Penerapan Metode

Dalam penelitian ini, terdapat beberapa tahapan yang dilakukan dalam sistem analisis sentimen. Hal ini dilakukan agar data yang digunakan dapat diteliti sesuai dengan penelitian ini. Ada beberapa tahap yang akan menjadi rancangan dari menu dalam membangun sistem, dimana rancangan ini mewakili proses tahapan awal sampai akhir. Dapat di lihat tahapan di gambar 1 berikut :



Gambar 1. Penerapan Metode

2.3 Pre-processing

Preprocessing adalah proses mengubah teks menjadi kumpulan kata. Teks adalah data yang tidak terstruktur, sehingga cukup sulit untuk diproses oleh komputer. Operasi numerik juga tidak bisa diterapkan pada data teks.

Oleh karena itu, perlu dilakukan *Preprocessing* pada teks agar data tersebut dapat diolah komputer [4]. Berikut tahap *Preprocessing* antara lain:

- Cleansing* merupakan salah satu proses untuk membersihkan sebuah kata yang tidak digunakan dari hasil *crawling* data [5].
- Case folding* adalah proses mengubah huruf yang awalnya huruf besar menjadi *lowercase* atau huruf kecil semua [5].
- slang word* yaitu mengubah sebuah kata yang baku menjadi bentuk yang sesuai ejaan bahasa Indonesia [6].
- stopword* merupakan tahapan dimana kata-kata yang umumnya sering muncul dan tidak memiliki dampak signifikan [6].
- Steeming* salah satu proses untuk mendapatkan kata dasar dari suatu kata dalam kalimat dengan cara memisahkan masing-masing kata dari kata dasar dan imbuhan dari awalan dan akhiran [6].

2.4 Labelling

labelling adalah proses untuk memberikan klasifikasi berdasarkan fitur atau karakteristik yang ada dalam sebuah kalimat atau dokumen. Pada tahapan ini akan dilakukan proses untuk memberikan klasifikasi berdasarkan fitur atau karakteristik yang ada dalam sebuah kalimat atau dokumen. Pada proses *labelling* dilakukan secara manual berdasarkan pada penelitian [6]. pada label positif untuk *tweet* yang mengungkapkan setuju kenaikan UKT atau berkomentar dengan sopan tentang kebijakan kenaikan UKT, sedangkan untuk kelas negatif akan diberikan kepada *tweet* yang cenderung tidak setuju kenaikan UKT atau berkomentar tidak sopan terhadap kenaikan UKT.

2.5 Pembagian Data

Pada proses tahapan ini, data *tweet* yang sudah diberikan label positif dan negatif akan dibagi menjadi dua bagian, yaitu data latih dan data uji, 80% data latih dan 20% data uji, Berdasarkan penelitian yang dilakukan oleh [7].

2.6 TF-IDF (Term frequency – Inverse Document Frequency)

TF-IDF adalah suatu metode yang digunakan dalam kasus penghitungan bobot kata yang paling sering digunakan pada bidang *information retrieval*. Efisiensi dari metode ini cukup dikenal selain itu metode ini juga mudah dan memiliki hasil yang akurat [8]. Cara kerja dari metode ini yaitu akan menghitung nilai *Term frequency* (TF) dan nilai *Inverse Document Frequency* (IDF) pada setiap kata (*term*) di setiap dokumen dalam korpus. Metode ini nantinya akan menghitung bobot dari setiap *term* (t) di dokumen (d) dengan rumus menggunakan persamaan 1 dan 2:

$$W_{dt} = TF_{dt} * IDF_t \quad (1)$$

Keterangan:

- d = dokumen ke-d
- t = kata ke-t dari kata kunci
- W = bobot dokumen ke-d terhadap kata ke-t
- tf = banyaknya kata yang dicari pada sebuah dokumen
- IDF = *Inversed Document Frequency*

Nilai IDF didapatkan dari:

$$IDF = \log 2 \left(\frac{D}{df} \right) \quad (2)$$

Keterangan:

- IDF = *Inversed Document Frequency*
- D = Total dokumen
- df = Banyak dokumen yang mengandung kata yang dicari

2.7 Algoritma Naïve Bayes Classifier (NBC)

Algoritma *Naïve Bayes Classifier* adalah metode klasifikasi yang menggunakan perhitungan probabilitas bersyarat [9]. Dasar utama dari *Naïve Bayes Classifier* berlandaskan pada teorema bayes yang pertama kali diperkenalkan oleh Thomas Bayes. Pengklasifikasi *Naïve Bayes Classifier* digunakan sebagai pengklasifikasi berbasis probabilitas. Probabilitas P didefinisikan seperti dalam persamaan 3 berikut:

$$P(A/B) = \frac{P(A)p(B \setminus A)}{P(B)} \quad (3)$$

Metode *Naïve Bayes Classifier* (NBC) adalah sebuah algoritma yang berfungsi untuk menemukan nilai probabilitas tertinggi untuk mengklasifikasikan ke dalam kategori yang sesuai. Algoritma NBC juga digunakan untuk klasifikasi teks. Kelebihan NBC terletak pada kesederhanaannya, namun tetap memberikan tingkat akurasi yang tinggi, sedangkan kelemahan NBC adalah jika probabilitas kondisional bernilai nol, maka prediksi juga akan bernilai nol, mengakibatkan ketidakakuratan tingkat akurasi. Ada dua tahapan dalam mengklasifikasikan *tweet*. Tahapan pertama adalah melibatkan algoritma menggunakan *tweet* yang kategorinya sudah diketahui. Tahapan kedua mengklasifikasikan *tweet* yang belum diketahui kategorinya. Dalam metode NBC, semua dokumen direpresentasikan sebagai pasangan atribut “ a_1, a_2, \dots, a_n ” dimana a_1 adalah kata pertama, a_2 adalah kata kedua, dan seterusnya. V adalah himpunan pengelompokan dokumen yang diuji (VMAP). Pada tahapan klasifikasi, algoritma mencari probabilitas tertinggi dari semua kelompok dokumen yang diuji (VMAP). Persamaan VMAP dapat dilihat pada persamaan 4 berikut:

$$V_{\text{map}} = \arg \max P(V_j) \prod_i P(a_i|V_j) \quad (4)$$

Nilai $P(V_j)$ dihitung dengan proses *train*, didapat dengan rumus menggunakan persamaan 5.

$$P(V_j) = \frac{|doc\ j|}{|train|} \quad (5)$$

Dimana $|doc\ j|$ adalah jumlah *tweet* yang termasuk pengelompokan j di dalam pelatihan, selanjutnya $|train|$ adalah jumlah *tweet* di dalam sampel yang digunakan untuk pelatihan. Untuk setiap probabilitas kata a_i pada setiap kategori $P(a_i|V_j)$, dihitung saat pelatihan/*train* menggunakan persamaan 6.

$$P(a_i|V_j) = \frac{n_i+1}{|n+kosakata|} \quad (6)$$

Dimana n_i merupakan jumlah kemunculan kata a_i dalam *tweet* yang termasuk dalam kategori V_j dan sedangkan n adalah total jumlah seluruh kata dalam *tweet* dengan kategori V_j dan $|kosakata|$ merupakan banyaknya kata dalam data *train*.

2.8 Confusion Matrix

Confusion Matrix adalah sebuah tabel yang menggambarkan jumlah data uji yang telah diklasifikasikan dengan benar dan jumlah data uji yang diklasifikasikan dengan salah. Pada pengukuran kinerja menggunakan *Confusion Matrix*, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Nilai *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi sebagai data positif. *True Positive* (TP) merupakan data positif yang terdeteksi benar. *False Negative* (FN) merupakan kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif [10]. Tabel 1 menunjukkan contoh *confusion matrix*.

Tabel 1. *Confusion Matrix*

Aktual/Prediksi	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Melalui 4 tahap tersebut, dilakukan dengan cara menghitung Tingkat *accuracy*, *precision*, dan *Recall* [11].

- a. *Accuracy* merupakan hasil dari skala prediksi yang akurat. Rumus *Confusion Matrix* untuk menghitung *Accuracy*, menggunakan persamaan 7 berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

- b. merupakan suatu jumlah teks ketika model prediksi positif yang memprediksi seberapa sering prediksi itu benar. Rumus *Confusion Matrix* untuk menghitung *Precision*, menggunakan persamaan 8 berikut:

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

- c. *Recall* merupakan suatu jumlah teks Ketika kelas aktualnya positif, seberapa sering model memprediksi positif. Rumus *Confusion Matrix* untuk menghitung *Recall*, menggunakan persamaan 9 berikut:

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

3. HASIL DAN PEMBAHASAN

3.1 Tahap Pengumpulan Data

data penelitian ini berasal dari media sosial X dalam bentuk *tweet*. *Dataset* ini menggunakan *tools google colab* bernama Tweet harvest, yang dikumpulkan antara tanggal 03 Mei 2024 hingga 26 Mei 2024. Jumlah data yang berhasil diperoleh adalah 250 data dengan kata kunci yang digunakan adalah “UKT”. Kemudian *dataset* yang berhasil dikumpulkan akan di *import* kedalam sistem *website* untuk dilakukan proses lebih lanjut.

3.2 Tahap Preprocessing

Tahap *preprocessing* merupakan tahapan yang akan dilakukan setelah setelah tersedianya satu atau lebih *dataset* pada basis data (*database*) hasil dari tahapan pengumpulan data. Tahapan ini terdiri atas lima proses utama antara lain: *cleansing*, *case folding*, *slangword*, *stopword*, dan *stemming*. Berikut proses dari tahap *preprocessing*:

- a. *Cleansing* berfungsi untuk menghapus semua karakter yang bukan alfabet dari sebuah postingan. Berikut tabel 1:

Tabel 1. Tahap *Cleansing*

Sebelum	Sesudah
@BuanaJannah @MardaniAliSera @prabowo @PKSejahtera @jokowi Saya lebih suka lihat negara hancur korupsi merajalela rakyat sengsara... Dan mereka yg dungu memilihnya kemarin pada teriak kayak kejadian UKT mahasiswa yg naik...	Saya lebih suka lihat negara hancur korupsi merajalela rakyat sengsara Dan mereka yg dungu memilihnya kemarin pada teriak kayak kejadian UKT mahasiswa yg naik

- b. *Case Folding* berfungsi untuk mengubah huruf yang awalnya huruf besar menjadi huruf kecil. Berikut tabel 2:

Tabel 2. Tahap *Case Folding*

Sebelum	Sesudah
Miris sekali Saat negara kelangkaan sarjana dengan presentasi populasi penduduk indonesia saat ini malah membuat kebijakan biaya UKT yang tinggi.	miris sekali saat negara kelangkaan sarjana dengan presentasi populasi penduduk indonesia saat ini malah membuat kebijakan biaya UKT yang tinggi.

- c. *Slangword* dilakukan untuk mengubah *slangword* yang melibatkan sebuah kata yang baku menjadi bentuk yang sesuai ejaan Bahasa Indonesia. Berikut tabel 3.

Tabel 3. Tahap *Slangword*

Sebelum	Sesudah
saya lebih suka lihat negara hancur korupsi merajalela rakyat sengsara dan mereka yang dungu memilihnya kemarin pada teriak kayak kejadian UKT mahasiswa yang naik	saya lebih suka lihat negara hancur korupsi merajalela rakyat sengsara dan mereka yang bodoh memilihnya kemarin pada teriak kayak kejadian UKT mahasiswa yang naik

- d. *stopword* berfungsi untuk menghilangkan kata-kata yang umumnya yang sering muncul seperti “pada”, “apa”, “lagi”. Berikut tabel 4.

Tabel 4. Tahap *Stopword*

Sebelum	Sesudah
ukt 2 digit itu bapaknya pada kerja apaan ya. memelihara tuyul?	ukt digit bapaknya kerja iya memelihara tuyul

- e. *stemming* berfungsi untuk mendapatkan kata dasar dari suatu kata dalam kalimat dengan cara memisahkan masing-masing kata dari kata dasar dan imbuhan dari awalan dan akhiran. Berikut tabel 5.

Tabel 5. Tahap *Stemming*

Sebelum	Sesudah
@alfan_rahadian @ardisatriawan Terjangkau kalau dapet UKT golongan 1 atau 2. Masalahnya sekarang	jangkau ukt golongan sekarang kampus negeri uktnya mahal fasilitas oke kampus negeri kenal ukt mahal loh

banyak kampus negeri yang UKTnya mahal tapi fasilitas ga oke. Bahkan kampus negeri yang ga terkenal aja UKTnya mahal loh

3.3 Labelling

Sebelum melakukan pengujian, data perlu diberikan label yaitu positif dan negatif. Hal ini dilakukan agar memenuhi tujuan dari penelitian ini dalam menganalisis apakah suatu berita tersebut positif atau negatif. Namun dalam mengidentifikasi data termasuk dalam label positif atau label negatif, dilakukan secara manual. Berikut tabel 6.

Tabel 6. Labelling

No	Data Asli	Label
1	Maaf marah Pagi-pagi. Akhir bulan harus bayar UKT lagi soalnya jadi kepikiran	Negatif
2	Semoga Janji Ini Akan Terealisasi dan Tidak Perlu Ada Lagi UKT	Positif
...
250	UKT mahal tapi fasilitas tidak memadai. Kampus jahat banget asli	Negatif

3.4 Pembagian Data

Setelah melakukan pelabelan, maka data akan di bagi menjadi 2 yaitu 80% data latih dan 20% data uji (perbandingan 80:20). Berikut tabel 7.

Tabel 7. Pembagian Data

Jenis Data	Jumlah
Data Latih	200
Data Uji	50
Dataset	250

3.5 Tahap TF-IDF

Pada tahapan ini, akan dilakukan pembobotan kata dengan TF-IDF. Maksud dari tahapan ini adalah untuk melakukan pembobotan pada setiap kata dalam dokumen. Berikut tabel 8.

Tabel 8. Dataset Training

Doc	Data asli	label
D1	moga janji realisasi ukt	Positif
D2	ukt seram banget gila udahlah nasib miskin	Negatif

Selanjutnya akan dilakukan perhitungan pembobotan kata menggunakan TF-IDF, menggunakan persamaan 1. Berikut tabel 9.

Tabel 9. Perhitungan TF-IDF

Term	D1	D2	DF	IDF	TF x IDF	
					D1	D2
moga	1/4	0	1	0.602	0.151	0
Janji	1/4	0	1	0.602	0.151	0
Realisasi	1/4	0	1	0.602	0.151	0
ukt	1/4	1/7	2	0.301	0.075	0.043
Seram	0	1/7	1	0.602	0	0.086
banget	0	1/7	1	0.602	0	0.086
gila	0	1/7	1	0.602	0	0.086
udahlah	0	1/7	1	0.602	0	0.086
nasib	0	1/7	1	0.602	0	0.086
miskin	0	1/7	2	0.301	0	0.086

Setelah melakukan perhitungan TF-IDF kemudian membuat tabel untuk menghitung frekuensi kata positif dan kata negatif. Berikut tabel 10.

Tabel 10. Frekuensi Kata Positif dan Kata Negatif

Positif	DF	Negatif	DF
moga	1	moga	0
Janji	1	Janji	0
Realisasi	1	Realisasi	0
ukt	1	ukt	1
Seram	0	Seram	1
banget	0	banget	1
gila	0	gila	1
udahnya	0	udahnya	1
nasib	0	nasib	1
miskin	0	miskin	1

3.6 Klasifikasi Naïve Bayes Classifier

Setelah melakukan pembobotan kata TF-IDF, kemudian akan dilakukan perhitungan klasifikasi sentimen dengan Algoritma NBC dari contoh *tweet* pada klasifikasi dari data *Training* sistem yang telah dibuat. Langkah awal yang dilakukan yaitu menghitung probabilitas pada setiap kelas sentimen menggunakan persamaan 5 yaitu:

$$P(V_1) = \frac{|doc1|}{|training|} = \frac{1}{2} = 0,5$$

$$P(V_2) = \frac{|doc1|}{|training|} = \frac{1}{2} = 0,5$$

Setelah mendapatkan nilai probabilitas pada setiap kelas sentimen, kemudian dilakukan perhitungan probabilitas kemunculan setiap kata pada masing-masing kategori, menggunakan persamaan 6.

$$P(moga | v1) = (1+1)/(4+11) = 0,13$$

$$P(moga | v2) = (0+1)/(7+11) = 0,05$$

$$P(janji | v1) = (1+1)/(4+11) = 0,13$$

$$P(janji | v2) = (0+1)/(7+11) = 0,05$$

$$P(realisasi | v1) = (1+1)/(4+11) = 0,13$$

$$P(realisasi | v2) = (0+1)/(7+11) = 0,05$$

$$P(ukt | v1) = (1+1)/(4+11) = 0,13$$

$$P(ukt | v2) = (1+1)/(7+11) = 0,11$$

$$P(seram | v1) = (0+1)/(4+11) = 0,06$$

$$P(seram | v2) = (1+1)/(7+11) = 0,11$$

$$P(banget | v1) = (0+1)/(4+11) = 0,06$$

$$P(banget | v2) = (1+1)/(7+11) = 0,11$$

$$P(gila | v1) = (0+1)/(4+11) = 0,06$$

$$P(bila | v2) = (1+1)/(7+11) = 0,11$$

$$P(udahnya | v1) = (0+1)/(4+11) = 0,06$$

$$P(udahnya | v2) = (1+1)/(7+11) = 0,11$$

$$P(nasib | v1) = (0+1)/(4+11) = 0,06$$

$$P(nasib | v2) = (1+1)/(7+11) = 0,11$$

$$P(miskin | v1) = (0+1)/(4+11) = 0,06$$

$$P(miskin | v2) = (1+1)/(7+11) = 0,11$$

Setelah memperoleh probabilitas setiap kata dalam sentimen positif dan negatif, selanjutnya melakukan pengujian pada data *testing*. Berikut tabel 11.

Tabel 11. Data *Training* dan Data *Testing*

Doc	Data <i>Training</i>	label
D1	moga janji realisasi ukt	Positif
D2	ukt seram banget gila udahnya nasib miskin	Negatif
Data <i>Testing</i>		
D3	iya allah adek ukt gol moga ringan	?

Setelah mendapatkan data *training* dan data *testing*, data *testing* akan dihitung probabilitasnya dan mencari probabilitas tertinggi menggunakan persamaan 4.

$$P(V_1) \prod_i P(a_i | V_1) = (0,5) (p(iya|V_1) \times p(allah | V_1) \times p(adek | V_1) \times p(ukt | V_1) \times p(gol | V_1) \times p(moga | V_1) \times p(ringan | V_1))$$

$$= (0,5) (0,25 \times 0,25 \times 0,25 \times 0,5 \times 0,25 \times 0,5 \times 0,25)$$

$$= 0,00006103515625$$

$$\begin{aligned}
 P(V_2) \prod_i P(a_i | V_2) &= (0,5) (p(\text{iya} | V_2) \times (p(\text{allah} | V_2) \times (p(\text{adek} | V_2) \times (p(\text{ukt} | V_2) \times (p(\text{gol} | V_2) \times (p(\text{moga} | V_2) \times \\
 &\quad (p(\text{ringan} | V_2) \\
 &= (0,5) (0,14 \times 0,14 \times 0,14 \times 0,28 \times 0,14 \times 0,14 \times 0,14) \\
 &= 0,0000010541350
 \end{aligned}$$

Berdasarkan hasil yang diperoleh, nilai probabilitas positif lebih besar dari pada nilai probabilitas negatif dapat disimpulkan klasifikasi *tweet* “sentimen positif”.

3.7 Pengujian

Pada tahapan ini untuk mendapatkan hasil tingkat akurasi, diperlukan perhitungan *confusion matrix* untuk menghitung tingkat *accuracy*, *precision* dan *recall*, dalam memprediksi label data uji dengan mengimplementasikan algoritma *Naïve Bayes Classifier*. Berikut tabel 12.

Tabel 12. Prediksi Dari Data Uji

No	Data	Label	Prediksi
1	moga janji realisasi ukt	Positif	Positif
2	selesai heboh ukt variasi mahasiswa ppn naik injak rem kas konsumtif paylater karya	Negatif	Negatif
3	pusing ukt iya semangat teman teman	Positif	Negatif
...
50	ukt banget sehat sehat camaba kampus merdeka kayak neraka	Negatif	Positif

Pada tabel 12 didapatkan label prediksi dari sistem yang dibuat. Label merupakan pelabelan yang dilakukan secara manual, sedangkan prediksi merupakan label yang dihasilkan dari proses klasifikasi menggunakan *Naive Bayes Classifier*. Berikut tabel 13.

Tabel 13. Hasil *Confusion Matrix*

Aktual/Prediksi	Positif	Negatif
Positif	36	6
Negatif	4	4

Berdasarkan hasil *confusion matrix* maka dilakukan perhitungan pengujian menggunakan persamaan 7, 8, dan 9 dalam tabel 14 pengujian dapat dilihat pada tabel 14 berikut ini.

Tabel 14. Pengujian

Pengujian			
<i>Accuracy</i>	$\frac{(TP + TN)}{(TP + TN + FP + FN)}$	$\frac{(36 + 4)}{(36 + 4 + 6 + 4)}$	80%
<i>Precision (Positive)</i>	$\frac{TP}{(TP + FP)}$	$\frac{4}{(4 + 4)}$	50%
<i>Precision (Negative)</i>	$\frac{TN}{(TN + FN)}$	$\frac{36}{(36 + 6)}$	85,7%
<i>Recall (positive)</i>	$\frac{TP}{(TP + FN)}$	$\frac{4}{(4 + 6)}$	40%
<i>Recal (Negative)</i>	$\frac{TN}{(TN + FP)}$	$\frac{36}{(36 + 4)}$	90%

Berdasarkan hasil pengujian pada tabel 14 didapatkan *accuracy* sebanyak 80% atau baik, *precision* positif atau buruk 50%, *precision* negatif atau baik sebanyak 85,7%, *recall* positif atau buruk sebanyak 40%, *recall* negatif atau baik 90%.

4. KESIMPULAN

Penelitian ini melakukan analisis sentimen di media sosial X terhadap kebijakan kenaikan Uang Kuliah Tunggal (UKT), dengan *crawling* data menggunakan metode *Naïve Bayes Classifier*. Dari 250 *tweet* dari hasil

crawling data di X, mengenai sentimen masyarakat terhadap kebijakan kenaikan UKT, dimana sentimen positif 61 data (24,4%) dan sentimen negatif 189 data (75,6%) sejak 3 Mei 2024 hingga 26 Mei 2024. Pada penelitian ini dilakukan 6 tahapan penting yaitu: *Crawling* data, *Preprocessing*, *Labelling*, pembagian data, pembobotan TF-IDF dan klasifikasi *Naïve Bayes Classifier*. Berdasarkan hasil pengujian didapatkan *accuracy* sebanyak 80% atau baik, *precision* positif atau buruk 50%, *precision* negatif atau baik sebanyak 85,7%, *recall* positif atau buruk sebanyak 40%, *Recall* negatif atau baik 90%.

DAFTAR PUSTAKA

- [1] H. Hasanuddin, I. Ibrahim, and S. Syamsul, "Persepsi Mahasiswa terhadap Implementasi Uang Kuliah Tunggal di Fakultas Ilmu Sosial Universitas Negeri Makassar," *Doctoral dissertation*, Universitas Negeri Makassar, 2019.
- [2] Y. Yadi, A. Asminah, M. Purba, and I. R. Padya, "Analisis Sentimen Marketplace di Era Society 5.0 Menggunakan Algoritma Naive Bayes," *JUSIM (Jurnal Sistem Informasi Musirawas)*, vol. 8, no. 1, pp. 16-23, 2023.
- [3] A. H. Ruger, "Sentimen Analisis Pelanggan Shopee di Twitter menggunakan Algoritma Naive Bayes," *JIFOTECH (Journal of Information Technology)*, vol. 1, no. 2, 2021.
- [4] A. Sabrani, I. G. P. W. Wedashwara, and F. Bimantoro, "Metode Multinomial Naïve Bayes untuk Klasifikasi Artikel Online Tentang Gempa di Indonesia (Multinomial Naïve Bayes Method for Classification of Online Articles About Earthquake in Indonesia)," *Jurnal Teknologi Informasi*, vol. 8, no. 2, pp. 123-131, 2020.
- [5] T. Ramadhan, D. Wahiddin, and E. Awal, "Klasifikasi Sentimen terhadap Pinjaman Online (Pinjol) Menggunakan Algoritma Naive Bayes," *Scientific Student Journal for Information, Technology and Science*, vol. 4, no. 1, pp. 82-87, 2023.
- [6] M. Rifki and I. Imelda, "Analisis sentimen wacana kenaikan harga tiket Candi Borobudur menggunakan Multinomial Naïve Bayes," *JIKO (Jurnal Informatika dan Komputer)*, vol. 5, no. 2, pp. 156-163, 2022.
- [7] S. Fatimah, E. Purwanto, and H. Permatasari, "Analisis Sentimen Twitter Tentang Wisata di Kota Solo," *Techno. Com*, vol. 22, no. 4, pp. 854-869, 2023.
- [8] N. S. Marga, "Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes," *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, vol. 2, no. 4, pp. 453-463, 2021.
- [9] T. Ridwansyah, "Implementasi text mining terhadap analisis sentimen masyarakat dunia di twitter terhadap Kota Medan menggunakan k-fold cross validation dan Naïve Bayes Classifier," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178-185, 2022.
- [10] K. Karsito and S. Susanti, "Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia," *Jurnal SIGMA*, vol. 9, no. 3, pp. 43-48, 2019.
- [11] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier dan Confusion Matrix pada Analisis Sentimen Berbasis Teks pada Twitter," *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 5, no. 2, pp. 697-711, 2021.