

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**Analisis Sentimen Masyarakat Terhadap Kenaikan Biaya Haji
Tahun 2023 Menggunakan Metode
Naïve Bayes Classifier**

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

HERTATI

NIM. 11751202147



UIN SUSKA RIAU

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2024**



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

Analisis Sentimen Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode Naïve Bayes Classifier

TUGAS AKHIR

Oleh

HERTATI
NIM. 11751202147

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 15 Januari 2024

Pembimbing I,

DR. Elin Haerani, S.T., M.Kom

NIP. 19810523 200710 2 003

UIN SUSKA RIAU



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

Analisis Sentimen Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode Naïve Bayes Classifier

Oleh

HERTATI

NIM. 11751202147

Telah dipertahankan di depan sidang dewan penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 15 Januari 2024

Mengesahkan,

Ketua Jurusan,

IWAN ISKANDAR, M.T.

NIP. 19821216 201503 1 003



DR. HARTONO, M.PD.

NIP. 19640301 199203 1 003

DEWAN PENGUJI

Ketua : Reski Mai Candra, S.T., M. Sc

Pembimbing I : DR. Elin Haerani, S.T., M. Kom.

Penguji I : Novriyanto, S.T., M. Sc

Penguji II : Fadhilah Syafria, S.T., M. Kom.



SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Hertati
 NIM : 11751202147
 Tempat, Tgl. Lahir : Sintong, 12 Agustus 1999
 Fakultas : Sains dan Teknologi
 Prodi : Teknik Informatika
 Judul Jurnal :

Analisis Sentimen Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode Naive Bayes Classifier

Menyatakan dengan sebenar-benarnya bahwa:

1. Penulisan jurnal dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu jurnal saya ini, saya nyatakan bebas dari plagiat.
4. Apabila di kemudian hari terbukti terdapat plagiat dalam penulisan jurnal saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 16 Januari 2024
Yang membuat pernyataan



Hertati
NIM. 11751202147

Analisis Sentimen Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode Naïve Bayes Classifier

Hertati*, Elin Haerani, Novriyanto, Fadhilah Syafria

Fakultas Sains Dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}11751202147@students.uin-suska.ac.id, ²elin.haerani@uin-suska.ac.id, ³novriyanto@uin-suska.ac.id,

⁴fadhilah.syafria@uin-suska.ac.id

Email Penulis Korespondensi : 11751202147@students.uin-suska.ac.id

Abstrak—Pemerintah Indonesia melalui rapat Kementerian Agama dan Komisi VIII DPR-RI menyepakati besaran Biaya Penyelenggaraan Ibadah Haji (BPIH) 1444 H/2023 M rata-rata Rp90.050.637,26 per jemaah haji reguler. Namun kebijakan ini menimbulkan berbagai respon masyarakat. Adapun respon masyarakat tentang kenaikan biaya haji tahun 2023 terdapat pada media sosial Twitter. Pada penelitian ini melakukan klasifikasi analisis sentimen terhadap Tweets untuk mengetahui opini masyarakat terhadap kenaikan biaya haji tahun 2023 menggunakan metode naïve bayes classifier karena metode ini cenderung sederhana dan mudah digunakan. Data set yang digunakan 3000 tweets dengan jumlah data positif 1866 data positif, data negatif 415. Pada penelitian ini menghasilkan nilai akurasi 81,46% pada pembagian data 70:30, pada pembagian data 80:20 yaitu 80,74% dan pada pembagian data 90:10 yaitu 79,04. Pada penelitian ini terdapat lebih banyak respon positif dari masyarakat hal ini membuktikan bahwa kenaikan biaya haji tahun 2023 dapat diterima oleh masyarakat. Akurasi tertinggi pada penelitian ini yaitu 81,46% pada pembagian data 70:30. Penelitian selanjutnya disarankan menggunakan algoritma lain untuk melihat perbandingan hasil dari algoritma yang berbeda dalam klasifikasi sentimen masyarakat terhadap kenaikan biaya haji tahun 2023.

Kata kunci: Kenaikan Biaya Haji Tahun 2023; Twitter; Klasifikasi; Analisis Sentimen; Naïve Bayes Classifier

Abstract—The Indonesian government through a meeting of the Ministry of Religion and Commission IVIII of the DPR-RI agreed on the cost of organizing the Hajj pilgrimage (BPIH) 1444 iH/2023 iM, an average of IDR 90,050,637.26 per irregular pilgrimage. However, this policy gave rise to various public responses. The public's anger regarding the increase in Hajj fees in 2023 was found on the social media iTwitter. In this study, we conducted a sentiment classification analysis of Tweets to determine public opinion regarding the increase in Hajj costs in 2023 using the naïve Bayes classifier method because this method tends to be simple and easy to use. The data set used was 3000 tweets with a total of 1866 positive data, 415 negative data. This research resulted in an accuracy value of 81.46% in the 70:30 data division, in the 80:20 data division, namely 80.74% and in the data division. 90:10 which is 79.04. In this research, there were more positive responses from the public, this proves that the increase in Hajj costs in 2023 can be accepted by the public. The highest accuracy in this study was 81.46% with a 70:30 data split. It is recommended that further research use other algorithms to see a comparison of the results of different algorithms in classifying public sentiment regarding the increase in the cost of Hajj in 2023.

Keywords: Increase In Hajj Costs In 2023; Twitter; Classification; Sentiment Analysis; Naïve Bayes Classifier

1. PENDAHULUAN

Berbagai respon masyarakat muncul karena adanya kebijakan pemerintah tentang kenaikan biaya haji tahun 2023. Pemerintah melalui rapat Kementerian Agama dan Komisi VIII DPR-RI menyepakati besaran Biaya Penyelenggaraan Ibadah Haji (BPIH) 1444 H/2023 M rata-rata Rp90.050.637,26 per jemaah haji reguler. Biaya ini ditetapkan dalam Rapat Panitia Kerja (Panja) Komisi VIII DPR dan Kementerian Agama pada Rabu (15/2/2023). Nilai ini terdiri atas dua komponen, yaitu Biaya Perjalanan Ibadah Haji (Bipih) yang ditanggung jemaah dengan rata-rata Rp 49.812.700,26 (55,3%) Biaya Perjalanan tersebut digunakan untuk biaya penerbangan, biaya hidup, dan sebagian biaya paket layanan masyair haji. Kemudian sebesar Rp 40.237.937 (44,7%) digunakan untuk nilai manfaat (optimalisasi) per Jemaah yang akan dikelola Badan Pengelola Keuangan Haji (BPKH)[1]. Adapun respon masyarakat tentang kenaikan biaya haji tahun 2023 di twitter berbagai ragam, terdapat dalam bentuk positif dan juga negatif. Hal ini menjadikan Twitter menjadi sumber data yang besar berpotensi untuk dijadikan media analisis sentimen mengenai suatu topik karena banyaknya respon atau komentar dari pengguna Twitter.

Analisis sentimen merupakan suatu proses memahami, mengekstrak dan memproses data berbasis teks secara otomatis, proses ini dilakukan untuk memperoleh sebuah informasi sentimen yang tersirat dari suatu opini. Sentimen yang diberikan dapat berupa sentimen positif dan negatif [2][3][4]. Hasilnya bisa digunakan untuk menganalisis beberapa kata sentimen masyarakat terhadap kenaikan biaya haji tahun 2023 serta menghasilkan informasi berupa persentasi sentimen positif dan negatif. Adapun metode yang biasa digunakan pada klasifikasi untuk melihat tanggapan masyarakat di dalam text mining salah satunya adalah metode naïve bayes classifier.

Naive Bayes adalah sebuah pengklasifikasian atau penggolongan data yang menghitung kemungkinan dari dataset yang tersedia Naive Bayes merupakan metode klasifikasi yang menghitung probabilitas kumpulan data yang tersedia [5]. Naive Bayes memiliki waktu klasifikasi yang singkat, sehingga mempercepat proses sistem analisis sentimen. Naive Bayes terbukti menjadi algoritma yang paling terkenal untuk melakukan analisis sentimen dari sebuah teks [6]. Naive Bayes mengatasi algoritma pengklasifikasi/pembelajaran mesin lainnya seperti K-Nearest Neighbor dan Decision Tree[7]. Hal ini bisa dilihat dengan adanya beberapa riset yang terkait dengan pengklasian menggunakan metode naïve Bayes.

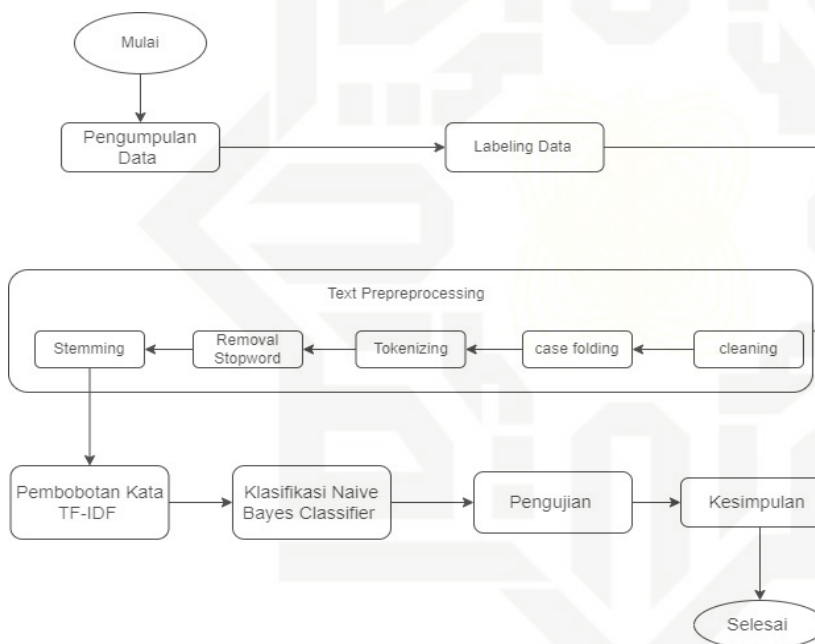


Adapun salah satu penelitian terkait yang diteliti oleh Meylan Wongkar dan Apriandy Angdresey tentang “Analisis Sentimen Untuk Twetts Terhadap Calon Presiden RI 2019, Pada penelitian ini dilakukan perbandingan menggunakan metode naïve bayes, svm dan K-Nearest Neighbor (K-NN) yang diuji menggunakan RapidMiner dengan menghasilkan nilai akurasi naïve bayes sebesar 75,58%, nilai akurasi svm sebesar 63,99% dan Nilai akurasi K-NN sebesar 73,34%”[8]. “Analisis sentimen terhadap sara, hoaks, dan radikal pada media sosial twiiter pada tahun 2020 menggunakan metode naïve bayes dengan hasil akurasi 99,62%”[9]. “Analisis sentimen terhadap provider by.u menggunakan metode naïve bayes dengan hasil akurasi 90%”[10]. “Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes dengan hasil akurasi 90.36%”[11]. “Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM tahun 2022 dan memperoleh hasil akurasi pada metode SVM 75.40% dan metode naïve bayes 83.81%”[12].

Jadi, pada penelitian ini melakukan analisis sentimen dengan menggunakan metode Naive Bayes Classifier untuk melihat pandangan masyarakat tentang kebijakan pemerintah terhadap kenaikan biaya haji tahun 2023 dengan cara meng-crawling data twitter dan membersihkan data ganda (*Double*) kemudian melabelkan data, apakah data tersebut mengandung sentimen positif atau negatif. Setelah dilakukan pelabelan, untuk mencari hasil akurasi dilakukan pengujian menggunakan *library pyhton* dan *confusion matrix* untuk melihat hasil akurasinya. Pada penelitian ini juga bermanfaat untuk mengumpulkan aspirasi dan respon dari opini masyarakat yang bertanggung jawab positif dan negatif terhadap kenaikan biaya haji tahun 2023.

2. METODOLOGI PENELITIAN

Adapun langkah-langkah yang dilakukan pada penelitian ini sbagai berikut:



Gambar 1. Desain Penelitian

2.1 Pengumpulan Data

Data pada penelitian ini dilakukan dengan mengcrawling data. Tahap crawling merupakan proses mengumpulkan tweets menggunakan Twitter API (*Application Programing Interface*) yang diambil langsung dari media sosial Twitter. Data *tweets* yang diambil menggunakan kata kunci atau *keyword* biaya haji dan kenaikan biaya haji tahun 2023.

2.2 Pelabelan Data

Selanjutnya tahap *labeling* data berupa lebel atau kelas positif dan negatif yang di lakukan oleh tenaga pakar yaitu guru bahasa indonesia. Label positif adalah Teks yang mengandung pujian, kepuasan, atau persetujuan terhadap subjek atau topik yang dibahas. Label negatif adalah Teks yang biasanya mengandung kritik, ketidakpuasan, atau ketidaksetujuan terhadap subjek atau topik yang dibahas[13].

2.3 Text Pre-Processing

Text Preprocessing proses mengubah data awal yang belum berurutan menjadi data yang berurutan atau terstruktur untuk dianalisis dalam penelitian[14]. *Text Preprocessing* dapat dibagi menjadi beberapa proses yaitu:



- a. *Cleaning* yaitu proses membuang atribut yang tidak diperlukan seperti membersihkan kata, menghapuskan simbol, angka dan emotikon
- b. *Case Folding* proses mengubah semua huruf dalam teks menjadi huruf kecil atau huruf besar
- c. *Tokenizing* yaitu proses Membagi atau memecahkan teks menjadi "token" atau unit-unit kecil, seperti kata-kata atau frasa.
- d. *Stopwordremoval* yaitu proses menghapus atau membuang kata-kata tidak penting atau relevan dari hasil *tokenizing*
- e. *Stemming* yaitu proses menghapus akhiran kata untuk mengembalikan kata ke bentuk dasarnya

2.4 Pembobotan TF-IDF

TF-IDF, singkatan dari *Term Frequency-Inverse Document Frequency*, merupakan suatu proses mengkonversi sekumpulan teks menjadi data numerik agar dapat memberikan bobot pada setiap kata atau fitur. TF-IDF adalah metrik statistik yang digunakan untuk menilai tingkat signifikansi sebuah kata dalam suatu dokumen. Faktor "TF" (*Term Frequency*) mengukur seberapa sering kata tersebut muncul dalam dokumen tertentu, mencerminkan pentingnya kata tersebut dalam dokumen tersebut. Sementara itu, faktor "DF" (*Document Frequency*) mengindikasikan seberapa umumnya kata tersebut muncul dalam seluruh koleksi dokumen. [15]. Metode TF-IDF dirumuskan sebagai berikut :

$$IDF = \log \frac{D}{DF} \tag{1}$$

$$TF-IDF = tf * idf \tag{2}$$

Catatan:

- D : Jumlah dokumen dalam kumpulan data
- DF : Jumlah dokumen yang mengandung kata tersebut
- TF : Frekuensi Kemunculan Kata (Term Frequency)
- Idf : Inves Frekuensi Dokumen

2.5 Naïve Bayes Classifier

Menurut seorang ilmuwan inggris Thomas Bayes Igoritma Naïve Bayes merupakan salah satu metode klasifikasi yang menggunakan metode probabilitas dan metode statistik [16]. Naïve bayes adalah metode yang menggunakan *teorema Bayes* dan berasumsi bahwa nilai antar variabel pada suatu nilai output[17]. Klasifikasi Naïve Bayes diakui memiliki kinerja yang sangat baik jika dibandingkan dengan metode klasifikasi lainnya, terutama dalam hal akurasi, yang cenderung lebih tinggi daripada metode klasifikasi alternatif.[18]. Rumus algoritma naïve bayes classifier dapat dilihat pada persamaan berikut:

$$P(X) = \frac{P(C)P(c)}{P(x)} \tag{3}$$

- X : Data yang memiliki kelas yang belum diketahui
- C : Hipotesis data yang merujuk pada kategori tertentu
- P(C|X) : Probabilitas kelas (C) yang diberikan atribut atau fitur (X).
- P(c) : Probabilitas awal (prior) dari kelas C
- P(x|c) : Probabilitas kemungkinan (likelihood) fitur X dalam kelasC.
- P(x) : Probabilitas awal (prior) dari fitur X

2.6 Confusion matrix

Confusion matrix merupakan suatu tabel yang menggambarkan perincian klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah[19]. Tabel dalam confusion matrix mencantumkan jumlah data uji yang telah diklasifikasikan dengan benar dan jumlah data uji yang keliru diklasifikasikan, dilakukan perhitungan untuk mendapatkan akurasi, presisi, recall, dan F1 score. Formulasi dari rumus akurasi, presisi, recall, dan F1-score adalah sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \tag{4}$$

$$Presisi = \frac{TP}{TP+FP} * 100\% \tag{5}$$

$$Recal = \frac{TP}{TP+FN} * 100\% \tag{6}$$

$$skor F1 = 2 * \frac{Presisi*Recal}{Presisi+Recal} \tag{7}$$

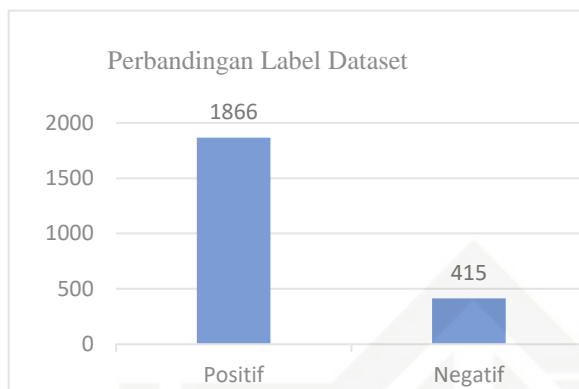
3. HASIL DAN PEMBAHASAN

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



3.1 Pembahasan

Dataset yang dikumpulkan berjumlah 3000 tweets. Kemudian dilakukan pembersihan seperti menghapus data yang ganda sehingga tersisa sebanyak 2281. Kemudian melakukan pelabelan data yang divalidasi oleh pakar guru bahasa Indonesia yang hasilnya pada gambar berikut.



Gambar 2. Diagram Perbandingan Dataset

Pada gambar 2 merupakan hasil dari dari proses pelabelan tweet yang termasuk kedalam kelas positif sebanyak 1866 dan kelas negatif sebanyak 415 tweet.

Setelah proses pelabelan, kemudian dilakukan pengolahan data dengan google colab menggunakan bahasa pemrograman python. Adapun salah satu contoh hasil dari tahap text preprocessing pada tabel berikut:

Tabel 1. Preprocessing

Komentar awal	b'Biaya haji tahun 2023 kenaikan tinggi dibandingkan sebelumnya, yaitu total Rp 84 juta – Rp 96 juta. Dari jumlah itu, Rp 44,3 juta – Rp 55,9 juta ditanggung langsung jemaah. Biaya mahal ini mesti diikuti peningkatan layanan. \nhttps://t.co/0DKmJjhmQk\n\n#Opini #AdadiKompas'
Cleaning	b Biaya haji tahun kenaikan tinggi dibandingkan sebelumnya yaitu total Rp juta Rp juta Dari jumlah itu Rp juta Rp juta ditanggung langsung jemaah Biaya mahal ini mesti diikuti peningkatan layanan n
Casefolding	b biaya haji tahun kenaikan tinggi dibandingkan sebelumnya yaitu total rp juta rp juta dari jumlah itu rp juta rp juta ditanggung langsung jemaah biaya mahal ini mesti diikuti peningkatan layanan n
Tokenisasi	'b, biaya, haji, tahun, kenaikan, tinggi, dibandingkan, sebelumnya, yaitu, total, rp, juta, rp, juta, dari, jumlah, itu, rp, juta, rp, juta, ditanggung, langsung, jemaah, biaya, mahal, ini, pasti, diikuti, peningkatan, layanan, dan,
Stopword removal	biaya, haji, tahun, kenaikan, tinggi, dibandingkan , total, rp, juta, rp, juta rp, juta, rp, juta, ditanggung jemaah, biaya, mahal , diikuti, peningkatan, layanan
Stemming	biaya, haji, tahun, naik, tinggi, banding , total, rp, juta, rp, juta rp, juta, rp, juta, tanggung jemaah, biaya, mahal , ikut, tingkat, layan

Pada tabel 1 merupakan hasil dari *preprocessing* yang diolah dengan google colab menggunakan bahasa pemrograman python yang dimulai dari data awal kemudian melalui *proses cleaning, Case Folding, Tokenisasi, Stopword removal* dan *Steming*.

Selanjutnya dilakukan pembobotan kata menggunakan TF-IDF yang mengubah data teks menjadi data numerik agar dapat memberikan bobot pada setiap kata atau fitur.

```

from sklearn.feature_extraction.text import TfidfVectorizer
data_x=[]

for isi in fitur:
    data_x.append(isi)
vectorizer=TfidfVectorizer()
X=vectorizer.fit_transform(data_x)
#data_x
    
```

Gambar 3. Pembobotan Kata TF-IDF



Pada Gambar 3, dapat dilihat program ini mengonversi kumpulan teks (*fitur*) menjadi representasi numerik menggunakan metode TF-IDF dengan bantuan *TfidfVectorizer* dari *scikit-learn*. Digunakan untuk melatih model machine learning dalam pengklasifikasian teks sentimen.

3.2 Hasil

Dari 2281 data, dilakukan pengujian dengan membagi data menjadi dua model yaitu data training (data latih) dan data testing (data uji). Pada penelitian ini menggunakan tiga kali percobaan pembagian data yang berbeda, yaitu yang pertama 30% data uji dan 70% data latih, yang kedua 20% data uji dan 80% data latih, yang ketiga 10% data uji dan 90% data latih. Kemudian dilakukan tahapan pengujian kinerja menggunakan algoritma naïve bayes classifier.

```
>>> #data = ubahdata(fitur)
>>> from sklearn.model_selection import train_test_split
>>> x_train, x_test, y_train, y_test = train_test_split(X, kelas, test_size=0.3, random_state=0)
>>> x_test
<685x3782 sparse matrix of type '<class 'numpy.float64''>
with 9028 stored elements in Compressed Sparse Row format>

>>> [43] from sklearn.naive_bayes import MultinomialNB
>>> model_naive = MultinomialNB().fit(x_train, y_train)
>>> predicted_naive = model_naive.predict(x_test)
>>> print(model_naive)
MultinomialNB()

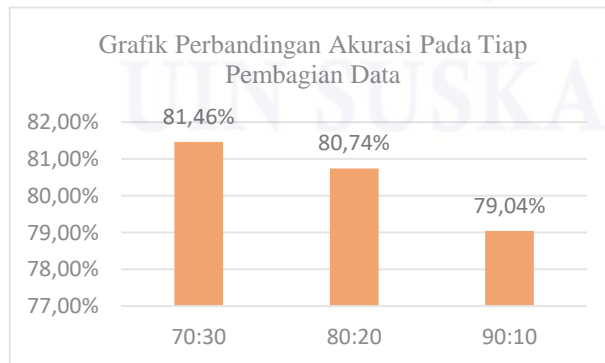
>>> [44] from sklearn.metrics import accuracy_score
>>> score_naive = accuracy_score(predicted_naive, y_test)
>>> print("Accuracy with Naive-bayes: ",score_naive)
>>> print(score_naive*100)
Accuracy with Naive-bayes:  0.8145985401459854
81.45985401459855
```

Gambar 4. Program Naïve Bayes

Pada gambar, program ini dapat dipecah menjadi beberapa langkah:

- Pertama, data dibagi menjadi data latih dan data uji. Data X (*fitur*) dan kelas (*label*) digunakan sebagai input. Pembagian data uji ditentukan dengan *test_size=0.3*, yang berarti 70% data digunakan untuk pelatihan dan 30% untuk pengujian.
- Kemudian, model klasifikasi Naive Bayes (*MultinomialNB*) diinisialisasi dan dilatih menggunakan data pelatihan dengan *MultinomialNB().fit(x_train, y_train)*.
- Model yang telah dilatih digunakan untuk melakukan prediksi pada data uji dengan *model_naive.predict(x_test)*.
- Hasil prediksi tersebut diukur akurasi dengan menggunakan metrik akurasi (*accuracy*) dari *sklearn* melalui *accuracy_score(predicted_naive, y_test)*.
- Hasil akurasi adalah sekitar 81,46%, yang berarti model Naive Bayes mampu memprediksi dengan benar sekitar 81,46% dari data uji.

Setelah melakukan pengujian dengan google colab menggunakan bahasa pemrograman *python* dari tiga kali percobaan dengan membagi data uji dan data latih, diperoleh hasil akurasi tertinggi pada percobaan pertama yaitu 70:30 dengan akurasi 81,46%.

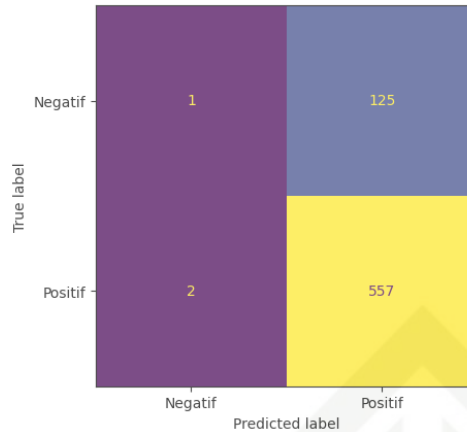


Gambar 5. Perbandingan Akurasi

Pada gambar 5 merupakan hasil akurasi yang didapatkan pada tiap percobaan yaitu 81,46% pada pembagian data 70:30, pada pembagian data 80:20 yaitu 80,74% dan pada pembagian data 90:10 yaitu 79,04%.



Evaluasi dilakukan dengan melihat confusion matrix dengan menghitung nilai accuracy, precision, recall dan F1-Score. Dapat dilihat pada gambar 6 dan 7.



Gambar 6. Confusion Matrix

Pada gambar 6 dengan percobaan pertama dengan pembagian data 70:30 diperoleh hasil confusion matrix dengan nilai TP (True Positif) 557, Nilai TN (True Negatif) 1, Nilai FP (false Positif) 125 dan FN (False Negatif) 2.

Tabel 2. Confusion Matrix

True Positif	557
True Negatif	1
False Positif	125
False Negatif	2
Total	685

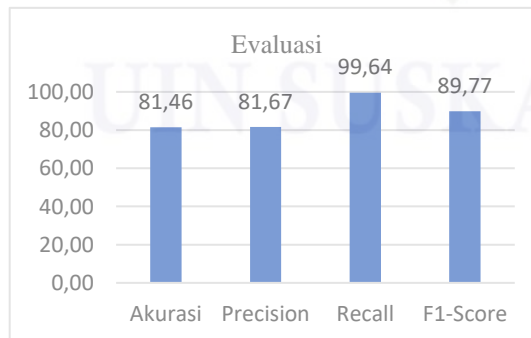
- a. True Positive (TP): Merupakan jumlah sampel yang seharusnya diklasifikasikan sebagai positif dan juga benar-benar diklasifikasikan sebagai positif oleh model.
- b. True Negative (TN): Merupakan jumlah sampel yang seharusnya diklasifikasikan sebagai negatif dan memang benar-benar diklasifikasikan sebagai negatif oleh model
- c. False Positive (FP): Merupakan jumlah sampel yang seharusnya diklasifikasikan sebagai negatif tetapi salah diklasifikasikan sebagai positif oleh model.
- d. False Negative (FN): Merupakan jumlah sampel yang seharusnya diklasifikasikan sebagai positif tetapi salah diklasifikasikan sebagai negatif oleh model.

Berdasarkan tabel 2, kemudian dilakukan perhitungan menggunakan rumus 4 untuk melihat akurasi pada percobaan pertama.

$$\text{Akurasi} = \frac{557+1}{557+1+125+2} * 100\%$$

$$\text{Akursi} = \frac{558}{685} * 100 = 81,46\%$$

Hail akurasi menggunakan rumus 4 pada percobaan pertama perbandingan data 70:30 yaitu 81,46%.



Gambar 7. Evaluasi

Penilaian dilakukan dengan melihat confusion matrix berdasarkan hasil yang diperoleh dari persamaan 4, 5, 6, dan 7 dan diperoleh nilai Akurasi 81,46, nilai Precision 81,67 nilai Recall 99,64 dan nilai F1-Score 89,77 seperti digambar 5.



4. KESIMPULAN

Berdasarkan hasil penelitian, analisis sentimen masyarakat terhadap kenaikan biaya haji tahun 2023 menggunakan metode naïve bayes classifier dengan dataset 3000 tweets dan data bersih sebanyak 2281. Hasil dari pelabelan data Positif 1866 dan data Negatif 415, dan menghasilkan nilai akurasi 81,46% pada pembagian data 70:30, pada pembagian data 80:20 yaitu 80,74%, dan pada pembagian data 90:10 yaitu 79,04. Pada penelitian ini terdapat lebih banyak respon positif dari masyarakat hal ini membuktikan bahwa kenaikan biaya haji tahun 2023 dapat diterima oleh masyarakat. Sementara nilai akurasi tertinggi pada penelitian ini yaitu 81,46% pada pembagian data 70:30.

REFERENCES

- [1] A. J. Showail, "Solving Hajj and Umrah Challenges Using Information and Communication Technology: A Survey," *IEEE Access*, vol. 10, no. May, pp. 75404–75427, 2022, doi: 10.1109/ACCESS.2022.3190853.
- [2] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier," *J. Linguist. Komputasional*, vol. 5, no. 1, p. 34, 2022, doi: 10.26418/jlk.v5i1.79.
- [3] M. F. Muzakki, J. Jondri, and U. R. Febrian, "Analisis Sentimen Mahasiswa Terhadap Fasilitas Universitas Telkom Menggunakan Metode Jaringan Saraf Tiruan Dan Tf-Idf," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 8608–8616, 2019.
- [4] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.
- [5] S. J. S. Tyas, M. Febianah, F. Solikhah, A. L. Kamil, and W. A. Arifin, "Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," *J. Teknol. Inf. Dan Komun.*, vol. 8, no. 1, pp. 86–99, 2021, [Online]. Available: <https://jurnal.plb.ac.id/index.php/tematik/article/view/576/365>
- [6] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [7] Aldinata, A. M. Soesanto, V. C. Chandra, and D. Suhartono, "Sentiments comparison on Twitter about LGBT," *Procedia Comput. Sci.*, vol. 216, pp. 765–773, 2023, doi: 10.1016/j.procs.2022.12.194.
- [8] M. Wongkar and A. Angdresy, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, pp. 1–5, 2019, doi: 10.1109/ICIC47613.2019.8985884.
- [9] F. E. Purwiantono and A. Aditya, "Klasifikasi Sentimen Sara, Hoaks Dan Radikal Pada Postingan Media Sosial Menggunakan Algoritma Naive Bayes Multinomial Text," *J. Tekno Kompak*, vol. 14, no. 2, p. 68, 2020, doi: 10.33365/jtk.v14i2.709.
- [10] I. Verawati and B. S. Audit, "Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1411, 2022, doi: 10.30865/mib.v6i3.4132.
- [11] M. Hudha, E. Supriyati, and T. Listyorini, "Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 1–6, 2022, doi: 10.33387/jiko.v5i1.3376.
- [12] R. T. Aldisa and P. Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID- 19 Dengan Perbandingan Metode Naive Bayes , Decision Tree dan SVM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.
- [13] F. Wulandari, E. Haerani, M. Fikry, and E. Budianita, "Analisis sentimen larangan penggunaan obat sirup menggunakan algoritma naïve bayes classifier," vol. 4, no. 1, pp. 88–96, 2023.
- [14] A. Aziz, "Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, p. 115, 2022.
- [15] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [16] T. Sanubari, C. Prianto, and N. Riza, *Odol (One Desa One Product Unggulan Online) Penerapan Metode Naïve Bayes Pada pengembangan Aplikasi E-Commerce Menggunakan Codeigniter*. Bandung: Kreatif Industri Nusantara, 2020.
- [17] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.
- [18] D. A. Pratiwi, R. M. Awangga, and M. Y. H. Setyawan, *Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informaika Menggunakan Metode Naive Bayes*, Pertama. Bandung: Kreatif Industri Nusantara, 2020.
- [19] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021, [Online]. Available: <https://ejournal.tunasbangsa.ac.id/index.php/jsakti/article/view/369/348>