



Analisis Sentimen Jamaah Umrah Di Media Sosial X Menggunakan Algoritma *Naive Bayes*

Surya Syahputra*¹, Sriani², Muhammad Siddik Hasibuan³

*^{1,2,3}Jurusan Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

*Email Penulis Korespondensi: syahputrasurya007@gmail.com

Abstrak

Umroh salah satu hal yang rukun sunnah yang pahalanya setara dengan haji yang membutuhkan biaya besar. Oleh karena itu, pemilihan biro perjalanan umroh harus dilakukan dengan hati-hati, mengingat banyaknya oknum nakal yang melakukan penipuan, seperti penggelapan uang calon jamaah atau penundaan keberangkatan tanpa kepastian. Salah satu penyebab utamanya adalah tawaran biaya murah atau promosi menarik yang tidak disertai dengan izin biro yang jelas. Analisis sentimen terhadap ulasan jamaah dapat menjadi solusi dalam mengidentifikasi biro perjalanan umroh yang terpercaya. Metode data mining, khususnya dengan algoritma *Naive Bayes*, dapat digunakan untuk mengklasifikasikan sentimen ulasan jamaah di media sosial X. Penelitian ini menggunakan 501 data tweet yang diperoleh melalui teknik crawling. Data ulasan dibagi menjadi tiga kategori, yaitu positif, negatif, dan netral. Pembobotan kata dilakukan menggunakan metode *TF-IDF* untuk meningkatkan akurasi hasil analisis. Hasil dari pengujian model pada data testing menunjukkan bahwa terdapat 18 ulasan dengan sentimen positif, 3 ulasan negatif, dan 80 ulasan netral. Evaluasi model *Naive Bayes* ini menghasilkan nilai *precision* 0,80, *recall* 0,80, dan *F1-score* yaitu 0,80.

Kata kunci— Umrah, Biro Perjalanan, *Naive Bayes*, Crawling, Media Sosial X

Abstract

Umrah is one of the important acts of worship for Muslims, which requires a significant amount of money. Therefore, the selection of an Umrah travel agency must be done carefully, considering the number of unscrupulous individuals who commit fraud, such as embezzling prospective pilgrims' money or delaying departures without certainty. One of the main reasons is the offer of low-cost Umrah packages or attractive promotions that are not accompanied by proper agency permits. Sentiment analysis of pilgrim reviews can be a solution in identifying trustworthy Umrah travel agencies. Data mining methods, particularly the *Naive Bayes* algorithm, can be used to classify the sentiment of pilgrims' reviews on social media platform X. This study utilized 501 tweet data obtained through crawling techniques. The review data were classified into three categories: positive, negative, and neutral. Term weighting was performed using the *TF-IDF* method to improve the accuracy of the analysis. The results from testing the model showed that there were 18 positive reviews, 3 negative reviews, and 80 neutral reviews. The evaluation of the *Naive Bayes* model produced a *precision* value of 0.80, a *recall* of 0.80, and an *F1-score* of 0.80.

Keywords— *Umrah, Travel Agency, Naive Bayes, Crawling, Sosial Media X*

1. PENDAHULUAN

Umrah merupakan salah satu ibadah yang dianjurkan dilakukan bagi umat muslim. Berbeda dengan haji yang bersifat wajib dan dilakukan pada waktu tertentu serta memerlukan biaya yang relatif lebih mahal, umrah bisa dilakukan kapan saja sepanjang tahun. Pelaksanaannya yang fleksibel membuatnya menjadi pilihan yang menarik bagi umat Islam di luar pelaksanaan ibadah haji. Namun, pelaksanaan umrah diperlukan persiapan yang matang, terutama dalam pemilihan biro perjalanan umrah yang menyediakan layanan yang sesuai dengan kebutuhan jamaah.

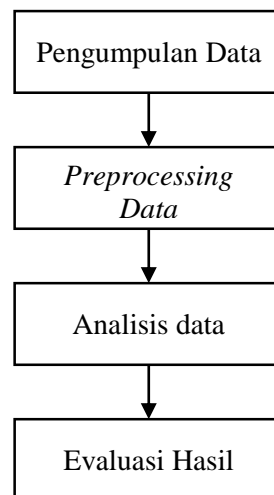
Pemilihan biro perjalanan umrah merupakan salah satu langkah penting dalam memastikan kelancaran dan kenyamanan ketika akan melaksanakan ibadah umrah. Banyaknya biro perjalanan membuat calon jamaah mengalami kesulitan dalam memilih yang terbaik. Kualitas layanan yang bervariasi, dari akomodasi, transportasi, hingga bimbingan ibadah, menjadi faktor-faktor penting yang harus dipertimbangkan. Selain itu, calon jamaah harus mencari tahu izin serta rekam jejak dari biro perjalanan untuk menghindari modus penipuan dari biro perjalanan umrah, diantaranya menggunakan skema ponzi dalam manajemen bisnisnya, tidak rasionalnya paket yang ditawarkan dan fasilitas serta pelayanan yang akan diberikan melebihi harga paket umrah. Beberapa penyimpangan penyelenggara ibadah umrah lainnya yaitu jadwal keberangkatan yang diundur dan tanpa disertai keterangan yang jelas, penambahan biaya dari ketetapan promo yang ditawarkan, keberangkatan yang gagal, pengembalian dana atau *refund* yang mustahil diwujudkan [1]. Hal ini bisa diatasi dengan ulasan dari jamaah yang telah menggunakan layanan tersebut di media sosial, khususnya X. Banyak jamaah umrah yang memanfaatkan X untuk menyampaikan kepuasan atau ketidakpuasan mereka terhadap biro perjalanan umrah yang mereka pilih mampu memberikan informasi kepada calon jamaah lainnya. Namun banyaknya ulasan yang diberikan dapat menyita banyak waktu serta tenaga. Analisis sentimen dapat menjadi solusi untuk mengatasi banyaknya ulasan yang terkandung. Analisis sentimen proses untuk menentukan apakah teks memiliki polaritas positif, negatif, atau netral. Dapat dilakukan pada data teks seperti dokumen, kalimat, atau paragraf. [2]. Penggunaan analisis sentimen dapat diterapkan pada berbagai opini seperti opini kebencian, opini umum, film, politik, merek produk dan lainnya pada media sosial [3]. Algoritma *Naive Bayes* adalah algoritma yang umum digunakan untuk klasifikasi. Pengklasifikasi *Naive Bayes* digunakan sebagai pengklasifikasi probabilistik [4].

Penelitian analisis sentimen yang dilakukan Ernianti Hasibuan dan Elmo Allistair Heriyanto [5] pada ulasan aplikasi amazon *shopping* di *google play store* menggunakan *Naive Bayes classifier* data yang mana menghasilkan 1.156 (53,8%) ulasan bersentimen negatif, dan 991 (46,2%) ulasan bersentimen positif dari keseluruhan ulasan sebanyak 2.147. selain itu, Muhammad Ariel Djamaludin dan rekan-rekannya [6] melakukan klasifikasi analisis sentimen *tweet* kri nanggala 402 di *twitter* menggunakan metode *Naive Bayes classifier*. Hasilnya diperoleh masyarakat cenderung bereaksi netral dan sentimen positif serta negatif sama. Akurasi yang didapatkan dengan metode ini sebesar 73%.

Keefektifan algoritma *Naive Bayes* pada analisis sentimen ini dapat menjadi solusi untuk mengklasifikasi ulasan pada media sosial X. Adapun tujuan penelitian ini adalah melakukan analisis sentimen terhadap *tweet* jamaah umrah dan calon jamaah umrah di media sosial X dengan metode *crawling*. Melalui penelitian ini diharapkan dapat diperoleh informasi yang berguna bagi calon jamaah umrah sehingga mempengaruhi persepsi dalam memilih biro perjalanan yang sesuai dengan harapan mereka.

2. METODE PENELITIAN

Penelitian ini dibagi menjadi ke dalam 3 tahap, yaitu tahap pengumpulan data, *preprocessing data* dan pengujian metode. Seperti terlihat pada Gambar 1 adalah langkah-langkah pada penelitian ini.



Gambar 1 Langkah penelitian

Gambar 1 adalah tahap-tahap penelitian yang dilakukan. Tahap pertama adalah pengumpulan data dilanjutkan dengan tahap *preprocessing data* dan kemudian tahap analisis data.

2.1 Pengumpulan Data

Tahap pengumpulan data dilakukan dengan metode *crawling*. Dalam konteks teknologi dan internet, merujuk pada proses di mana mesin pencari (seperti *Google*) mengunjungi dan mengindeks halaman web. Mesin pencari menggunakan program yang disebut "*crawler*" atau "*spider*" untuk menjelajahi internet, menemukan konten baru atau yang diperbarui, dan mengumpulkan informasi tentangnya. [7]. Pengumpulan data menggunakan metode *crawling* dengan *application interface* (API) sebagai perantara mendapatkan data *tweet* tentang umrah dengan kata kunci yang digunakan #Umroh pada media sosial X. *API key* ini bersifat rahasia karena setiap pengguna memiliki *key* yang berbeda-beda. *API key* yang sudah didapat digunakan untuk *crawling* data menggunakan *Google Colab* dengan bahasa pemrograman *Python*. Data yang diambil dari tahun 2022-03-01 sampai tahun 2023-06-28 dengan bentuk format *Comma Separated Value* (CSV).

2.2 Preprocessing Data

Setelah didapat data dalam bentuk teks, selanjutnya dilakukan *preprocessing data*. *Preprocessing data* adalah tahapan awal yang digunakan untuk menghilangkan kesalahan pada data mentah yang dapat mengganggu saat pemrosesan data. Tujuan dari *preprocessing* adalah untuk membersihkan, transformasi data, dan mempersiapkan data sehingga dapat diolah lebih efektif oleh algoritma analisis atau model *machine learning* [8]. Pada tahap ini dilakukan *text processing* data dimana terdiri dari *cleaning*, *case folding*, normalisasi kata, *stemming*, *filtering*, *tokenizing*.

1. *Cleaning* adalah tahapan membersihkan atau menghilangkan karakter yang tidak diperlukan pada data dokumen. Tahapan ini mengubah, menghapus kesalahan atau ketidakakuratan yang biasanya pada data mentah mengandung informasi yang tidak akurat atau tidak lengkap. Pada tahap ini dilakukan proses mengisi nilai yang hilang, menghaluskan *noisy data*, dan menyelesaikan inkonsistensi yang ditemukan. Pada tahap ini juga data dibagi ke dalam beberapa segmen yang memiliki ukuran.
2. *Case folding* adalah proses mengubah atau mengkonversi huruf kapital ke dalam huruf kecil (*lowercase*) pada semua data yang terdapat di dalam dokumen. Tujuan ini untuk menjaga konsistensi saat analisis teks.
3. Normalisasi kata adalah proses mengubah kata yang disingkat ke dalam arti yang sama dan konsisten berdasarkan Kamus Besar Bahasa Indonesia (KBBI) agar dapat menjadi sebuah

informasi yang sama. Contoh penerapan normalisasi kata yaitu kata “utk” menjadi “untuk” dan “yg” menjadi “yang”.

4. *Stemming* adalah mengubah kata pada data *tweet* yang memiliki imbuhan prefiks, sufiks dan konfiks menjadi kata dasar atau kata akar. Tujuan dari *stemming* adalah mengubah variasi kata sehingga variasi tersebut dapat digabungkan ketika analisis. Proses ini menggunakan bantuan kamus pada daftar *package* kata dasar.
5. *Filtering* adalah proses pengambilan kata-kata penting dan menghapus kata-kata umum yang tidak memberikan informasi signifikan saat analisis. Pada penelitian menggunakan metode *stopword*. Metode *stopword* berisi kata-kata yang biasa dipakai tetapi tidak jelas dan dapat dibuang seperti "dari", "yang", "untuk", "dan", "di" dan sebagainya.
6. *Tokenizing* adalah proses membagi atau memecah teks besar yang berupa kalimat, paragraf menjadi token-token atau unit yang lebih kecil. Umumnya acuan pemisah antar kata atau kalimat adalah spasi atau tanda baca. Tujuan dari tokenisasi yaitu untuk mempermudah analisis teks lebih lanjut dalam memahami struktur bahasa. Pada penelitian ini data yang diambil menjadi per kata.

2.3 Analisis Data

Analisis data adalah proses evaluasi data yang sesuai untuk digunakan. Tahap analisis adalah langkah sistematis untuk mengidentifikasi data yang sesuai dan metode optimal untuk memanfaatkannya. Tahap ini melibatkan berbagai metode atau teknik dalam pengolahan data. [9]. Pada tahap ini dibagi menjadi 3 proses yaitu pelabelan kelas sentimen, pembobotan TF-IDF, dan penerapan metode.

2.3.1 Pelabelan Kelas Sentimen

Pada tahap ini dilakukan pemberian kelas sentimen pada data yaitu positif, netral dan negatif. Pelabelan data diproses dengan melakukan perhitungan nilai skor dengan kamus *lexicon* yang mana berisi kumpulan kata positif dan negatif yang berbahasa Indonesia. Pelabelan kelas sentimen penting dalam analisis sentimen, yang digunakan untuk memahami opini publik, mengevaluasi produk, atau memantau reputasi merek. Proses ini bisa dilakukan secara manual oleh manusia atau secara otomatis menggunakan algoritma *machine learning* dan teknik pemrosesan bahasa alami. Formula (1) memperlihatkan rumus yang digunakan untuk memperoleh skor.

$$\text{Skor} = (\sum \text{kata positif} - \sum \text{kata negatif}) \quad (1)$$

2.3.2 Pembobotan TF-IDF

Metode TF-IDF adalah teknik untuk mengukur pentingnya suatu kata dalam dokumen relatif terhadap kumpulan dokumen. *Term Frequency* menghitung seberapa sering kata muncul dalam dokumen. *Inverse Document Frequency* mengukur seberapa umum atau langka kata dalam seluruh dokumen. TF-IDF dihitung dengan mengalikan TF dan IDF dan sering digunakan untuk mengekstrak kata kunci serta meningkatkan relevansi dalam pencarian informasi dan analisis teks. [10]. Tahap pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) menggunakan teknik menghitung pembobotan kata dalam data dokumen. Bobot setiap kata dihitung dan dikalikan dengan IDF. Pada dasarnya untuk perhitungan TF-IDF terbagi menjadi dua yaitu TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*). Nilai *Term Frequency* (TF) merupakan teknik dalam mencari bobot dari dokumen dengan mencari banyaknya kemunculan *term* pada dokumen. Semakin sering sebuah *term* muncul maka akan mempengaruhi nilai pembobotan. Sedangkan *Inverse Document Frequency* (IDF) adalah metode untuk menghitung penyebaran *term* pada keseluruhan dokumen dan diakhir perhitungan TF dan IDF akan digabungkan. Adapun rumus TF-IDF seperti yang disampaikan oleh [11] dapat dilihat pada formula (2), formula (3) dan formula (4).

$$tf_{t,d} = \frac{\text{Jumlah kemunculan term } t \text{ dalam dokumen}}{\text{Total jumlah term dalam dokumen}} \quad (2)$$

Keterangan:

Tf = frekuensi kemunculan kata pada sebuah dokumen

$$idf_t = \log\left(\frac{N}{df_t}\right) \quad (3)$$

Keterangan:

Idf = mengukur penting/tidak sebuah kata dalam dokumen

N = total jumlah dokumen dalam dokumen tersebut

$$tfidf_{t,d} = tf_{t,d} * idf_t \quad (4)$$

Keterangan:

TF-IDF = hasil penggabungan antara TF dan IDF

2.3.3 Klasifikasi Algoritma Naive Bayes

Algoritma *Naive Bayes* adalah metode klasifikasi yang menggunakan teorema *Bayes* dengan asumsi bahwa fitur-fitur bersifat independen. Meskipun asumsi ini sederhana, *Naive Bayes* sering memberikan hasil yang baik, terutama dalam klasifikasi teks seperti analisis sentimen dan pengklasifikasian email. Keunggulannya meliputi kecepatan dan efektivitas dalam menangani data besar [12]. Algoritma *Naive Bayes* terdiri dari dua tahapan dalam klasifikasi teks: tahap pelatihan data dan tahap klasifikasi. Pada tahap pelatihan, data sampel diolah untuk menciptakan representasi yang tepat. Probabilitas awal untuk setiap kategori ditentukan berdasarkan data sampel. Pada tahap klasifikasi, nilai kategori data ditentukan berdasarkan kemunculan term dalam data yang diklasifikasikan [13]. Adapun bentuk umum persamaan dari algoritma *Naive Bayes* dapat dilihat dari persamaan formula (5) [14].

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (5)$$

Keterangan:

X = data dengan kelas yang belum

H = hipotesis data X merupakan suatu kelas spesifik

P(H|X) = probabilitas hipotesis H berdasar kondisi X (*posteriorprobability*)

P(H) = probabilitas hipotesis H (*priorprobability*)

P(X|H) = probabilitas X berdasarkan kondisi pada hipotesis H

2.4 Evaluasi Hasil

Tahap ini adalah evaluasi hasil klasifikasi yang menggunakan data yang dikumpulkan. Evaluasi hasil adalah proses menilai seberapa baik kinerja penerapan metode mengklasifikasikan data atau memprediksi kategori yang benar berdasarkan data pelatihan yang diberikan. Pada penelitian ini evaluasi menggunakan *confusion matrix*. *Confusion matrix* adalah alat yang memberikan gambaran detail tentang prediksi model dengan menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas. *Confusion matrix* berguna untuk melihat gambaran akurasi yang lebih kompleks [15]. Dalam *confusion matrix*, akan dihitung akurasi, presisi, *recall*, dan *f-measure* yang dinyatakan dalam formula (6), formula (7), formula (8), dan formula (9).

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (6)$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (7)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (8)$$

$$F\text{-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% \quad (9)$$

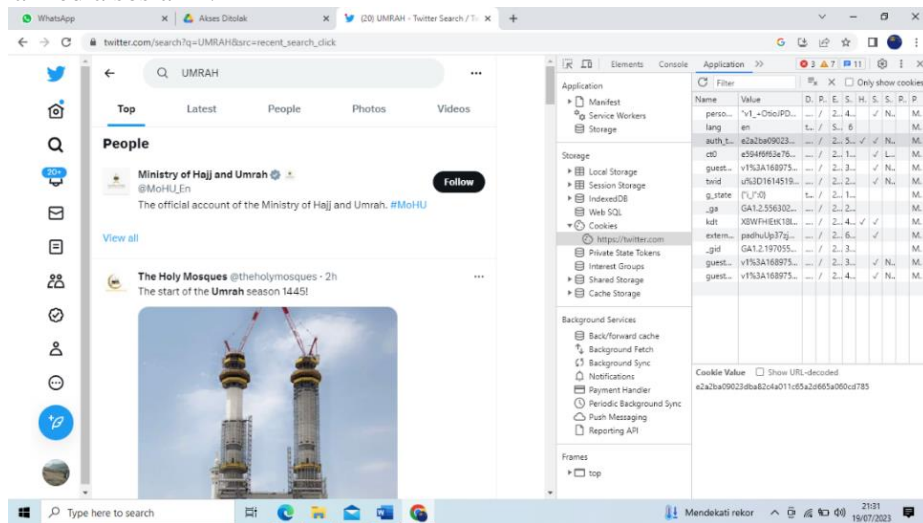
Nilai dari f -measure dengan rentang nilai dari 0 sampai 1. Nilai 1 menunjukkan performa model yang sempurna, di mana presisi dan *recall* keduanya adalah 1. Sedangkan nilai 0 memperlihatkan bahwasannya model tidak bisa melakukan klasifikasi dengan baik.

3. HASIL DAN PEMBAHASAN

Pada bagian ini menjelaskan proses pengumpulan data pada media sosial X, *preprocessing data*, analisis data, dan evaluasi hasil.

3.1 Pengumpulan Data

Tahap ini merupakan tahap awal dimana melakukan pengumpulan *dataset* yang akan digunakan pada proses selanjutnya. *Dataset* yang digunakan berupa *tweet* atau unggahan pengguna media sosial X.



Gambar 2 Mengambil Data Menggunakan API

Pada Gambar 2 merupakan pengambilan *API key* pada media sosial X dengan mengambil data *tweet* nya yang akan digunakan pada metode *crawling*. *Keyword* yang digunakan pada data *tweet* media sosial X yaitu #Umroh.

```

--Taking a break, waiting for 10 seconds...
Got some tweets, saving to file...
Your tweets saved to: /content/tweets-data/umrah.csv
Total tweets saved: 449

Got some tweets, saving to file...
Your tweets saved to: /content/tweets-data/umrah.csv
Total tweets saved: 466

Got some tweets, saving to file...
Scrolling more...

Got some tweets, saving to file...
Scrolling more...
Scrolling more...

Got some tweets, saving to file...
Your tweets saved to: /content/tweets-data/umrah.csv
Total tweets saved: 484

Got some tweets, saving to file...
Your tweets saved to: /content/tweets-data/umrah.csv
Total tweets saved: 501
Already got 501 tweets, done scrolling...

```

Gambar 3 Proses *Crawling* Menggunakan *Google Colab*

Pada Gambar 3 merupakan pengumpulan data menggunakan metode *crawling* dengan *application interface* (API) *key* pada media sosial X dengan mengambil *tweet* menggunakan

bahasa pemrograman Python. Data yang diperoleh berupa data teks dengan jumlah data 501. Hasilnya dapat dilihat pada Gambar 4.

Gambar 4 Data Hasil Crawling

3.2 Preprocessing Data

Pada bagian *preprocessing data* menjelaskan proses tahapan mengolah dan mempersiapkan data yang akan digunakan dari data mentah. Pada tahap ini dilakukan *cleaning*, *case folding*, normalisasi kata, *stemming*, *filtering*, dan *tokenizing*

Tabel 1 Hasil Preprocessing Data

Data Tweet		Hasil Filtering
@masgubrak @AlissaWahid Nah itu...banyak di depan atau sekitaran masjidil haram tapi umroh aja antrinya panjang apalagi hajipesan albaik harus pesan dulu...	➔	banyak di depan sekitar masjidil haram tapi umroh aja antri panjang apalagi haji pesan albaik harus pesan dulu
Tibaâ€• pengen umroh sama istri, padahal jodohnya aja belum ada ðŸ¥†	➔	tiba pengen umroh sama istri padahal jodoh aja belum ada
@lilaccount Bener bgt. Umroh dan haji panggilan/undangan Allah.	➔	bener banget umroh dan haji panggilan undangan allah

Pada Tabel 1 merupakan hasil dari proses *preprocessing data cleaning*, *case folding*, normalisasi kata, *stemming* dan *filtering*. Data yang awalnya penuh dengan *noise*, karakter tidak relevan, dan variasi kata informal menjadi lebih terstruktur, rapi dan bersih. Hasil akhirnya diperoleh data yang lebih akurat, terfokus pada pesan dan mempermudah untuk proses selanjutnya. Tahap selanjutnya yaitu *tokenizing*.

Tabel 2 Hasil Tokenizing Data

Data Tweet		Hasil Tokenizing		
banyak di depan sekitaran masjidil haram tapi umroh aja antri panjang apalagi haji pesan albaik harus pesan dulu	➔	banyak di depan atau sekitar	masjidil haram tapi umroh aja antri panjang	Apalagi haji pesan albaik harus pesan dulu
tiba pengen umroh sama istri padahal jodoh aja belum ada	➔	tiba pengen	sama istri	jodoh aja

		umroh	padahal	belum ada
bener banget umroh dan haji panggilan undangan allah	→	bener banget umroh	dan haji panggilan	undangan allah

Pada Tabel 2 merupakan proses dari *tokenizing data* menggunakan data hasil *filtering*. Pada tahap ini kalimat dipecah menjadi beberapa unit-unit yang lebih kecil berdasarkan spasi antar kata.

3.3 Analisis Data

Pada bagian ini menjelaskan penerapan metode algoritma *Naïve Bayes* pada *dataset*. Dimulai dengan melakukan pelabelan data pada *dataset* dengan jumlah 501 data dengan data *training* sebanyak 400 (79,84%) dan data *testing* sebanyak 101 (20,16%).

Tabel 3 Sentimen Data

Prediksi	Nilai		
	Positif	Negatif	Netral
Positif	41	0	0
Negatif	0	8	0
Netral	0	0	351

Terdapat kelas sentimen positif, negatif dan netral. Selanjutnya adalah perhitungan *confusion matrix* pada data *training*. Dengan mengacu pada Tabel 4, maka diperoleh nilai Accuracy seperti diperlihatkan pada Gambar 5.

Accuracy: 0.8348868175765646

Gambar 5 Accuracy data training

Selanjutnya adalah pelabelan kelas pada data *testing* dengan jumlah data sebanyak 101 data. Hasil pelabelan kelas dapat dilihat pada Tabel 5.

Tabel 5 Hasil Pelabelan Kelas

Kelas Sentimen Sementara	Jumlah
Positif	18
Negatif	3
Netral	80
Total	101

Pada Tabel 5 Dapat diamati bahwa jumlah klasifikasi data dalam 101 opini masyarakat di media sosial X terbagi menjadi tiga bagian: sentimen positif, sentimen negatif, dan sentimen netral. Jumlah sentiment positif 18, negatif 3, dan netral 80. Namun, yang perlu diperhatikan adalah data sentimen positif dan negatif karena berkontribusi secara signifikan.

	precision	recall	f1-score
0	0.80	0.80	0.80
1	0.80	0.80	0.80
accuracy			0.80
macro avg	0.80	0.80	0.80
weighted avg	0.80	0.80	0.80

Gambar 6 *Confusion Matrix* data *Testing*

Hasil akurasi atau performa metric yang didapat menggunakan algoritma *Naive Bayes* pada *dataset* opini masyarakat di media sosial X sebesar 0,80 untuk *precision*. Nilai 0,80 untuk *recall*, dan 0,80 untuk *f-1 score* seperti yang terlihat pada Gambar 6. Metode pengujian dalam penelitian ini menggunakan metrik akurasi untuk mengevaluasi tingkat keakuratan prediksi yang diperoleh. Akurasi penerapan algoritma *Naive Bayes* dalam analisis sentimen jamaah umrah di media sosial X menunjukkan tingkat keakuratan yang tinggi. Penggunaan algoritma *Naive Bayes* pada opini masyarakat di media sosial X efektif.

4. KESIMPULAN

Pengujian dilakukan pada data dengan sebanyak 501 buah data dari opini masyarakat pada media sosial X dengan data *training* sebanyak 400 dan data *testing* sebanyak 101. Pada data *testing* jumlah data sentimen positif berjumlah 18, jumlah data negatif 3 dan jumlah data netral 80. Hasil evaluasi penerapan algoritma *Naive Bayes* menghasilkan nilai 0,80 untuk *precision*, 0,80 untuk *recall* dan 0,80 untuk *f-1 score*.

5. SARAN

Penelitian ini hanya membatasi diri pada opini dalam bahasa Indonesia. Oleh karena itu, perlu melakukan pengembangan kembali dengan menerjemahkan bahasa asing dan menambahkan metode sebagai pembandingan efektivitas terbaik pada dataset ini.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim Redaksi Jurnal Teknika Politeknik Negeri Sriwijaya atas kesempatan untuk menerbitkan artikel ilmiah ini.

DAFTAR PUSTAKA

- [1] F. M. Arif, "Penyelenggaraan Ibadah Umroh Berbasis Maslahat," *Al-Amwal J. Islam. Econ. Law*, vol. 4, no. 1, hal. 22–39, 2019, doi: 10.24256/alw.v4i1.1180.
- [2] M. A. Fauzi, "Random forest approach fo sentiment analysis in Indonesian language," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, hal. 46–50, 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
- [3] M. Cindo dan D. P. Rini, "Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Literatur Review: Metode Klasifikasi Pada Sentimen Analisis," *Januari*, hal. 66–70, 2019.
- [4] Merinda Lestandy, Abdurrahim Abdurrahim, dan Lailis Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, hal. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.
- [5] Ernianti Hasibuan dan Elmo Allistair Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *J. Tek. dan Sci.*, vol. 1, no. 3, hal. 13–24, 2022, doi: 10.56127/jts.v1i3.434.

-
- [6] M. A. Djamaludin, A. Triayudi, dan E. Mardiani, “Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan Metode Naïve Bayes Classifier,” *J. JTİK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 2, hal. 161–166, 2022, doi: 10.35870/jtik.v6i2.398.
- [7] E. Putri, “An Impact of the Use Instagram Application Towards Students Vocabulary,” *J. Ilm. Pustaka Ilmu*, vol. 2, no. 2, hal. 1–10, 2022.
- [8] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, dan L. Efrizoni, “Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, hal. 273–281, 2024, doi: 10.57152/malcom.v4i1.1085.
- [9] Y. Septiawan dan C. Chairani, “Perbandingan Akurasi Metode Deteksi Ujaran Kebencian dalam Postingan Twitter Menggunakan Metode SVM dan Decision Trees yang Dioptimalkan dengan Adaboost,” *Teknika*, vol. 17, no. 2, hal. 287 – 299–287 – 299, 2023.
- [10] A. Afda, “Analisa Dokumen Menggunakan Metode TF-IDF,” vol. 5, no. 5, hal. 461–465, 2024.
- [11] K. Tri Putra, M. Amin Hariyadi, dan C. Crysdiyan, “Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis Svm,” *J. Cahaya MAndalika*, hal. 1449, 2023.
- [12] A. F. Watratan, A. Puspita B, dan D. Moeis, “IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID,” *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, hal. 7–14, 2020, doi: 10.55606/jurritek.v1i1.127.
- [13] Rina Noviana dan Isram Rasal, “Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter,” *J. Tek. dan Sci.*, vol. 2, no. 2, hal. 51–60, 2023, doi: 10.56127/jts.v2i2.791.
- [14] A. C. Khotimah dan E. Utami, “Comparison Naive Bayes Classifier, K-Nearest Neighbor Aand Support Vector Machine In The Classification Of Individual On Twitter Account,” *J. Tek. Inform.*, vol. 3, no. 3, hal. 673–680, 2022.
- [15] J. Khatib Sulaiman, M. Dzakwan Ar Rosyid, dan I. Artikel Abstrak, “Klasifikasi Tingkat Risiko Kesehatan Ibu Hamil Menggunakan Algoritma Support Vectore Machine Universitas AMIKOM Yogyakarta,” *Indones. J. Comput. Sci. Attrib.*, vol. 12, no. 5, hal. 2023–2798, 2023.