PERBANDINGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR PADA ANALISIS SENTIMEN ULASAN APLIKASI MYPERTAMINA

Husain Taufiqqurrahman, Fetty Tri Anggraeny, Muhammad Muharrom Al Haromainy
Informatika, Universitas Pembangunan Nasional Veteran Jawa Timur
Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Indonesia
19081010075@student.upnjatim.ac.id

ABSTRAK

MyPertamina merupakan suatu platform layanan keuangan digital yang dikembangkan oleh PT Pertamina. Tujuannya serupa dengan aplikasi seperti OVO dan Dana. Pengguna yang memanfaatkan aplikasi MyPertamina saat bertransaksi di SPBU menghadapi beberapa masalah signifikan. Ini termasuk kesulitan saat mencoba membuka aplikasi MyPertamina, keterbatasan pilihan pembayaran LinkAja di sejumlah SPBU, seringnya kegagalan saat mencoba mendaftar akun, dan kasus di mana transaksi berhasil namun poin hadiah tidak bertambah. Berdasarkan permasalahan tersebut, penelitian ini akan mengimplementasikan dua algoritma klasifikasi yaitu *Naïve Bayes* dan *K-Nearest Neighbor* dalam melakukan analisis sentimen ulasan pada aplikasi MyPertamina. Dataset ulasan yang diujikan terbagi menjadi tiga label sentimen yaitu negatif, netral, dan positif dengan jumlah dataset yang digunakan sebanyak 1500 data ulasan yang discrapping dari Google Play Store. Dari hasil pengujian, didapat hasil berupa algoritma Naïve Bayes memiliki tingkat akurasi 75% dan algoritma K-Nearest Neighbor memiliki tingkat akurasi yang berkisar antara 56% sampai 73%.

Kata kunci: *MyPertamina, Naïve Bayes, K-Nearest Neighbor, Analisis Sentimen*

1. PENDAHULUAN

MyPertamina merupakan suatu platform layanan keuangan digital yang dikembangkan oleh PT Pertamina. Tujuannya serupa dengan aplikasi seperti OVO dan Dana. Melalui MyPertamina, pengguna diberikan kemampuan untuk melakukan pembelian produk dari Pertamina, termasuk bahan bakar (BBM), dan menghubungkannya dengan akun *e-wallet* LinkAja. Saldo yang tersedia di akun LinkAja tersebut kemudian dapat dimanfaatkan untuk melunasi pembelian bahan bakar. Aplikasi MyPertamina dapat diunduh melalui platform Google Play Store dan App Store [1].

Pengguna yang memanfaatkan aplikasi MyPertamina saat bertransaksi di SPBU menghadapi beberapa masalah signifikan. Ini termasuk kesulitan saat mencoba membuka aplikasi MyPertamina, keterbatasan pilihan pembayaran LinkAja di sejumlah SPBU, seringnya kegagalan saat mencoba mendaftar akun, dan kasus di mana transaksi berhasil namun poin hadiah tidak bertambah. Hal ini mengakibatkan penurunan kepercayaan konsumen terhadap MyPertamina dan menyebabkan kurangnya insentif bagi konsumen untuk memanfaatkan MyPertamina dalam transaksi mereka. [2].

Analisis sentimen merupakan suatu teknik yang digunakan untuk secara otomatis mengambil data teks dan mengolahnya dengan tujuan untuk memahami opini atau sudut pandang yang terkait dengan suatu topik atau entitas tertentu, seperti individu, organisasi, produk, atau hal lain, yang terdapat dalam sejumlah data yang dianalisis. [3]. Terdapat berbagai algoritma yang dapat digunakan untuk melakukan klasifikasi dalam konteks analisis sentimen, termasuk tetapi tidak terbatas pada *Naïve Bayes Classifier*, *Random Forest*, *K-Nearest Neighbor*, *Support Vector Machine*, dan sejumlah algoritma lainnya yang dapat berguna dalam proses tersebut.

Algoritma *Naïve Bayes* merupakan algoritma klasifikasi sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan dan menggabungkan nilai dari kumpulan data yang diberikan [4] dengan berdasarkan teorema Bayes [5].

Algoritma *K-Nearest Neighbor* adalah metode klasifikasi objek berdasarkan data pelatihan yang paling dekat dengannya [6]. KNN juga merupakan metode algoritma *supervised learning*, di mana kelas yang paling banyak muncul (mayoritas) yang akan menjadi kelas hasil klasifikasi [7].

[8] melakukan penelitian untuk membandingkan kinerja algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dalam menganalisis opini publik tentang COVID-19 di platform *Twitter*. Dataset yang digunakan dalam penelitian ini berisi 1.098 opini yang diperoleh melalui API *Twitter*, dengan 610 opini yang bersifat positif dan 488 opini yang bersifat negatif. Hasil penelitian menunjukkan bahwa akurasi yang menggunakan metode klasifikasi *Naïve Bayes* mencapai 63,21%, sementara metode klasifikasi *K-Nearest Neighbor* mencapai 58,10%.

Berdasarkan penjelasan tersebut, penelitian ini akan melakukan pengujian terhadap dataset ulasan yang diujikan terbagi menjadi tiga label sentimen yaitu negatif, netral, dan positif dengan jumlah dataset yang digunakan sebanyak 1500 data ulasan yang discrapping dari Google Play Store.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah suatu teknik yang digunakan untuk secara otomatis mengambil data teks dan mengolahnya dengan tujuan untuk memahami

opini atau sudut pandang yang terkait dengan suatu topik atau entitas tertentu, seperti individu, organisasi, produk, atau hal lain, yang terdapat dalam sejumlah data yang dianalisis [3]. Kegunaan analisis sentimen menurut [9] ialah diantaranya memprediksi hasil pemilu, mendapatkan umpan balik pada obat-obatan, membandingkan produk serupa, menganalisa isu kesehatan mental serta mempelajari ulasan dari produk, film, maskapai penerbangan hingga hotel, dan menganalisis kekurangan suatu produk.

2.2. MyPertamina

MyPertamina merupakan suatu platform layanan keuangan digital yang dikembangkan oleh PT Pertamina. Tujuannya serupa dengan aplikasi seperti OVO dan Dana. Melalui MyPertamina, pengguna diberikan kemampuan untuk melakukan pembelian produk dari Pertamina, termasuk bahan bakar (BBM), dan menghubungkannya dengan akun *e-wallet* LinkAja. Saldo yang tersedia di akun LinkAja tersebut kemudian dapat dimanfaatkan untuk melunasi pembelian bahan bakar. Aplikasi MyPertamina dapat diunduh melalui platform *Google Play Store* dan *App Store* [1].

2.3. *Valance Aware Dictionary and Sentiment Reasoner* (VADER)

VADER (Valance Aware Dictionary and Sentiment Reasoner merupakan sebuah alat analisis sentimen berbasis leksikon atau kosakata yang spesifik digunakan dalam mengekstraksi sentimen yang diekspresikan di media sosial [10]. Penilaian sentimen analisis yang dilakukan oleh VADER terbagi menjadi 4 kategori yaitu *pos, neu, neg* dan *compound.* Penggunaan nilai *compound* ini berguna dalam menetapkan batas standar untuk mengklasifikasikan kalimat sebagai positif, netral, atau negatif. Nilai *compound* yang lebih besar dari atau sama dengan 0,05 dianggap positif, skor *compound* kurang dari atau sama dengan -0,05 dianggap negatif, dan skor *compound* antara -0,05 dan 0,05 dianggap netral [11].

2.4. *Synthetic Minority Over Sampling Technique* (SMOTE)

SMOTE merupakan teknik *oversampling* yang meningkatkan jumlah data dalam kelas minoritas dengan menciptakan data sintetis dari *instance-instance* kelas minoritas yang ada. Dalam SMOTE, *oversampling* dilakukan dengan mengambil contoh dari kelas minoritas dan kemudian mencari k-nearest neighbor untuk setiap contoh tersebut. Hasilnya adalah penciptaan *instance* sintetis daripada sekadar menggandakan *instance* kelas minoritas yang ada [12].

2.5. *Text Preprocessing*

Menurut [13], *Text Preprocessing* merupakan tahap yang dilakukan untuk menyiapkan data awal sebelum menggunakan algoritma untuk melakukan proses klasifikasi. Data dari Internet adalah data yang tidak terstruktur, sehingga diperlukan proses atau preprocessing agar data tersebut dapat digunakan. Karena hal tersebut, diperlukanlah tahap text preprocessing sebelum nantinya dilakukan penentuan fitur [14]. Tahapan text preprocessing yang akan dilakukan pada penelitian ini ialah *case folding*, *cleaning*, normalisasi, tokenisasi, *stopword removal*, dan *stemming.*

2.6. Ekstraksi Fitur (TF-IDF)

Dataset yang telah mengalami proses pelabelan harus diubah menjadi bentuk angka. Dataset tersebut dikonversi menjadi angka menggunakan metode TF-IDF [13]. Metode TF-IDF (*Term Frequency – Inverse Document Frequency*) merupakan suatu metode yang menggabungkan 2 cara untuk memberikan bobot pada kata [15]. Perhitungan TF-IDF dapat dilakukan menggunakan library *sklearn* yaitu *TfidfTransformer* [16] dengan persamaan sebagai berikut.

$$W = tf_{t,d} \times idf_t = tf_{t,d} \times ln\left(\frac{1+N}{1+\ df_t}\right) \quad (1)$$

Keterangan :

W = Bobot TF-IDF
$tf_{t,d}$ = Jumlah frekuensi kata
idf_t = Jumlah inverse frekuensi dokumen tiap kata
df_t = Jumlah frekuensi dokumen tiap kata
N = Jumlah total dokumen

2.7. SEMMA

SEMMA merupakan singkatan dari (*Sample*, *Explore*, *Modify*, *Model*, and *Access*) merupakan metode data mining yang dikembangkan oleh lembaga SAS (*Suite of Analytic*) *Institute*. Metode SEMMA ini dimasukkan ke dalam platform perangkat lunak komersial *SAS Enterprise Miner* [17].

2.8. Naïve Bayes

Algoritma *Naïve Bayes* merupakan algoritma klasifikasi sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan dan menggabungkan nilai dari kumpulan data yang diberikan [4] serta didasarkan pada Teorema Bayes [18]. Walaupun klasifikasi Algoritma *Naïve Bayes* dapat dikatakan sebagai algoritma klasifikasi yang sederhana, tetapi hasil *Naïve Bayes Classifier* merupakan metode *supervised learning*, sehingga setiap data perlu diberikan label sebelum proses *training* dilakukan [15]. Teorema Bayes dapat dijabarkan sebagai berikut.

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)} \quad (2)$$

Keterangan :

P(A|B) = probabilitas terjadinya A jika B terjadi
P(A) = probabilitas A terjadi
P(B|A) = probabilitas terjadinya B jika A terjadi
P(B) = probabilitas B terjadi

2.9. K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* adalah metode klasifikasi objek berdasarkan data pelatihan yang paling dekat dengannya [6]. Ada beberapa cara untuk mengukur jarak kedekatan antar data pada algoritma *K-Nearest Neighbor* diantaranya menggunakan *euclidean distance* dan *cosine similiarity*. Jarak antara tetangga, baik dekat atau jauh, sering diukur menggunakan *euclidean distance*. Untuk pengklasifikasian teks, pengukuran jarak menggunakan Cosine Similarity lebih umum digunakan [19]. Langkah-langkah dalam menghitung metode K-Nearest Neighbor menggunakan cosine similarity ialah sebagai berikut.

1) Membagi dua data menjadi data uji dan data latih
2) Menghitung jarak antara data uji di setiap label data menggunakan cosine similarity dengan jarak semua data latih. Rumus *cosine similarity* adalah sebagai berikut.

$$cosSim(A,B) = \frac{A \,.\, B}{||A|| \times ||B||} \qquad (3)$$

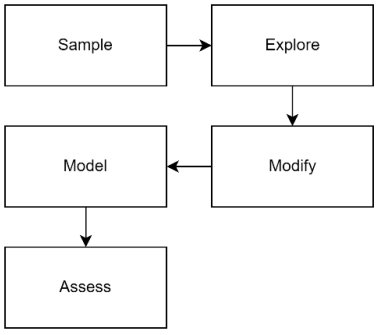
Keterangan :

A . B = perkalian (dot) dari vektor 'A' dengan 'B'
|| A || = panjang (besaran) dari vektor 'A'
|| B || = panjang (besaran) dari vektor 'B'

3) Urutkan hasil *cosine similarity* berdasarkan parameter k yang ditentukan. Jika k=3 artinya akan dipilih 3 nilai terbesar dari hasil perhitungan *cosine similarity*.
4) Selanjutnya, gunakan mayoritas atribut kelas berdasarkan k tetangga terdekat yang sudah dipilih untuk menentukan prediksi kelas.

3. METODE PENELITIAN

Metodologi yang digunakan pada penelitian ini ialah *Sample, Explore, Modify, Model, and Access* (SEMMA).



Gambar 1. Proses SEMMA

3.1. *Sample*

Sample merupakan tahap pertama pada SEMMA. Tahap ini berfokus pada pengambilan sampel data. Pada tahap ini dilakukan proses *scrapping* data ulasan MyPertamina di *Google Play Store*.

3.2. *Explore*

Tahap kedua yaitu *Explore*, pada tahap ini menjelaskan mengenai hasil scraping yang ada pada tahap Sample. Tahapan ini menggunakan software Microsoft Excel dalam melihat hasil scraping.

	A	B	C	D	E	F	G	H	I	J	K
1	at,"content"										
2	2023-06-25 10:36:41,"Saya coba isi bbm mngunakan mypertamina sdh isi saldo tapi tidak bisa di transaksi . Sudah 3										
3	2023-06-24 04:08:36,"Sdh 3thn pake mypertamina amanÃ² aja saya lihat yg komen rataÃ² kasih bintang 1 ada apa d										
4	2023-06-24 00:48:43,"Mau rubah data aja hampir 2 bln baru berhasil.."										
5	2023-06-23 23:48:05,"Updatenya kira2 donk...masa iya tiap 3hr sekali..."										
6	2023-06-23 21:22:29,"Tidak bisa di sambungkan ke aplikasi linkaja"										

Gambar 2. Data ulasan mypertamina

Hasil dari eksplorasi ini menggunakan data ulasan aplikasi MyPertamina tanggal 30 September 2018 sampai dengan 25 Juni 2023 sebanyak 1.500 ulasan. Data-data yang diambil terdiri dari kolom *at*, dan, *content* seperti yang terdapat pada Gambar 1.

3.3. *Modify*

Tahap ketiga yaitu *Modify*, merupakan proses yang digunakan untuk memodifikasi data. Tahapan yang dilakukan ialah *text preprocessing* yang berupa *case folding*, *cleaning*, normalisasi, tokenisasi, *stopword removal* dan *stemming*. Setelah tahapan text preprocessing selesai dilakukan tahapan translasi, pelabelan sentimen, ekstraksi fitur, *handling imbalance data*, dan pembagian data pada skema pengujian.

3.4. *Model*

Tahapan selanjutnya merupakan tahap pemodelan yaitu dengan melakukan komputasi pada data ulasan yang sudah melalui tahap *modify* dengan menggunakan model algoritma Naïve Bayes dan *K-Nearest Neighbor*

3.5. *Assess*

Tahap terakhir merupakan tahap *assess*, dimana tahapan ini merupakan tahap untuk mengukur kinerja serta mengevaluasi algoritma klasifikasi pemodelan dari komputasi yang sudah dilakukan. Pengukuran yang dilakukan berupa *accuracy*, *precision*, *recall* dan *F1-score*.

4. HASIL DAN PEMBAHASAN

Penelitian ini pada menggunakan bahasa pemrograman *Python* yang melakukan perbandingan penggunaan algoritma *Naïve Bayes* dan *K-Nearest Neighbor*.

4.1. *Sample*

Pada penelitian ini, data yang digunakan yaitu data ulasan aplikasi MyPertamina versi android yang berasal dari Google Play Store. Data yang digunakan merupakan data teks berbahasa Indonesia sebanyak 1500 ulasan yang paling relevan. Data tersebut diambil menggunakan *library* google-play-scrapper dengan bahasa pemrograman Python. Data yang diambil terdiri dari dua kolom yaitu kolom *at* yang berisikan waktu ulasan tersebut ditulis dan kolom *content* yang berisikan ulasan dari pengguna aplikasi MyPertamina.

4.2. *Explore*

Pada tahap Explore, data yang digunakan dalam penelitian ini berjumlah 1500 data. Berdasarkan kolom *at* data yang digunakan yaitu data pada rentang waktu 30 September 2018 sampai dengan 25 Juni 2023. Pada tahap ini, peneliti juga melihat kata-kata yang pengejaannya salah maupun kata dalam bahasa gaul untuk kemudian dibuat kamus normalisasinya.

4.3. *Modify*

Pada tahap *Modify*, peneliti hanya menggunakan data kolom content yang berisikan seluruh ulasan pengguna aplikasi MyPertamina. Proses yang akan dilakukan pada tahapan Modify ini terdiri dari preprocessing data berupa *case folding*, *cleaning*, normalisasi, tokenisasi, *stopword removal*, dan *stemming*.

Setelah selesai melakukan tahap normalisasi, data akan terbagi menjadi dua yaitu untuk tahap translasi untuk dilakukan pelabelan sentimen dengan *VADER* dan tahap tokenisasi yang akan digunakan untuk proses ekstraksi fitur.

Tabel 1. Hasil pelabelan sentimen

Label	Jumlah Sentimen
Negatif	892
Netral	167
Positif	441

Pada tahap pelabelan sentimen, label sentimen minoritas dalam hal ini label Netral memiliki jumlah perbandingan yang sangat jauh dengan label Negatif. Untuk itu diperlukan penyeimbangan data minoritas dengan cara melakukan teknik *oversampling* menggunakan SMOTE yang hasilnya dapat dilihat pada Tabel 2.

Tabel 2. Perbandingan hasil *oversampling*

Label	Sebelum SMOTE	Setelah SMOTE
Negatif	892	892
Netral	167	892
Positif	441	892

Setelah selesai melakukan tahapan *preprocesing*, terdapat tahapan pembobotan kata ekstrasi fitur dengan TF-IDF. Proses terakhir pada tahap *Modify* ini ialah membagi data untuk skema pengujian.

Pembagian data pada algoritma *Naïve Bayes* akan terbagi menjadi dua yaitu data latih sebanyak 70% dan data uji sebanyak 30%. Pada algoritma K-Nearest Neighbor, pembagian data akan terbagi menjadi dua yaitu data latih sebanyak 70% dan data uji sebanyak 30% serta menggunakan k atau jumlah tetangga terdekat sebanyak 3,5,7,9, dan 11 seperti yang tertera pada Tabel 3.

Tabel 3. Skema pengujian *K-Nearest Neighbor*

Algoritma	Jumlah k	Data latih	Data uji
K-Nearest Neighbor	3	70%	30%
	5		
	7		
	9		
	11		

4.4. *Model*

Setelah selesai melakukan tahap *Modify*, tahapan selanjutnya ialah Model. Tahapan ini akan melakuan komputasi dengan model algoritma Naïve Bayes dan K-Nearest Neighbor.

Algoritma Naïve Bayes yang akan digunakan pada penelitian ini ialah algoritma Multinomial Naïve Bayes. Hal ini didasarkan pada penelitian yang dilakukan oleh [20]. Penelitian tersebut menghasilkan kesimpulan yaitu tingkat akurasi yang paling tinggi ialah algoritma Multinomial Naïve Bayes dengan tingkat akurasi 63,74% .

Algoritma kedua sebagai model pembanding dari Naïve Bayes ialah *K-Nearest Neigbors*. Penggunaan model KNN pada penelitian ini akan menggunakan metode *Cosine Similarity* dalam mengukur tingkat kemiripan antar kalimat.

4.5. *Assess*

Tahap terakhir pada SEMMA ialah *Assess*. Tahapan ini akan mengulas hasil pengujian dari kedua model yang ada. Tahapan Assess yang akan dilakukan ialah *accuracy*, *precision*, *recall* dan *F1-score*. yang terdapat pada *library classification report*.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.71      0.64      0.67       259
           1       0.79      0.89      0.84       271
           2       0.75      0.71      0.73       273

    accuracy                           0.75       803
   macro avg       0.75      0.75      0.75       803
weighted avg       0.75      0.75      0.75       803
```

Gambar 3. *Classification Report Multinomial Naïve Bayes*

Hasil pengujian pertama menggunakan model algoritma *Multinomial Naïve Bayes* dengan data latih 70%, dan data uji 30%. Pada Gambar 3. didapatkan prediksi *accuracy* sebesar 75%, *precision* 75%, *recall* 75% dan *f1-score* 75%.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.79      0.39      0.52       259
           1       0.68      1.00      0.81       271
           2       0.77      0.78      0.78       273

    accuracy                           0.73       803
   macro avg       0.75      0.72      0.70       803
weighted avg       0.75      0.73      0.70       803
```

Gambar 4. *Classification Report* KNN k=3

Hasil pengujian kedua menggunakan model algoritma *K-Nearest Neighbor* dengan jumlah k sebanyak 3, data latih 70%, dan data uji 30%. Pada Gambar 4. didapatkan prediksi *accuracy* sebesar 73%, *precision* 75%, *recall* 72% dan *f1-score* 70%.

Classification Report:

	precision	recall	f1-score	support
0	0.74	0.36	0.48	259
1	0.63	0.99	0.77	271
2	0.76	0.69	0.72	273
accuracy			0.68	803
macro avg	0.71	0.68	0.66	803
weighted avg	0.71	0.68	0.66	803

Gambar 5. *Classification Report* KNN k=5

Hasil pengujian ketiga menggunakan model algoritma *K-Nearest Neighbor* dengan jumlah k sebanyak 5, data latih 70%, dan data uji 30%. Pada Gambar 5. didapatkan prediksi *accuracy* sebesar 68%, *precision* 71%, *recall* 68% dan *f1-score* 66%.

Classification Report:

	precision	recall	f1-score	support
0	0.73	0.35	0.47	259
1	0.58	0.99	0.73	271
2	0.73	0.60	0.66	273
accuracy			0.65	803
macro avg	0.68	0.64	0.62	803
weighted avg	0.68	0.65	0.62	803

Gambar 6. *Classification Report* KNN k=7

Hasil pengujian keempat menggunakan model algoritma *K-Nearest Neighbor* dengan jumlah k sebanyak 7, data latih 70%, dan data uji 30%. Pada Gambar 6. didapatkan prediksi *accuracy* sebesar 65%, *precision* 68%, *recall* 64% dan *f1-score* 62%.

Classification Report:

	precision	recall	f1-score	support
0	0.66	0.32	0.43	259
1	0.57	0.97	0.72	271
2	0.71	0.57	0.63	273
accuracy			0.62	803
macro avg	0.65	0.62	0.59	803
weighted avg	0.65	0.62	0.60	803

Gambar 7. *Classification Report* KNN k=9

Hasil pengujian keempat menggunakan model algoritma *K-Nearest Neighbor* dengan jumlah k sebanyak 9, data latih 70%, dan data uji 30%. Pada Gambar 7. didapatkan prediksi *accuracy* sebesar 62%, *precision* 65%, *recall* 62% dan *f1-score* 59%.

Classification Report:

	precision	recall	f1-score	support
0	0.66	0.31	0.42	259
1	0.54	0.94	0.69	271
2	0.63	0.49	0.56	273
accuracy			0.59	803
macro avg	0.61	0.58	0.55	803
weighted avg	0.61	0.59	0.56	803

Gambar 8. *Classification Report* KNN k=11

Hasil pengujian keempat menggunakan model algoritma *K-Nearest Neighbor* dengan jumlah k sebanyak 11, data latih 70%, dan data uji 30%. Pada Gambar 8. didapatkan prediksi *accuracy* sebesar 59%, *precision* 61%, *recall* 58% dan *f1-score* 55%.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis sentimen ulasan terhadap aplikasi MyPertamina di *Google Play Store*, dapat disimpulkan bahwa algoritma Naïve Bayes mendapatkan hasil *accuracy* sebesar 75%, *precision* sebesar 75%, *recall* sebesar 75%, dan *f1-score* sebesar 75% jika dibandingkan dengan semua model yang diuji menggunakan algoritma K-Nearest Neighbor yang memiliki nilai *accuracy* berkisar antara 59% sampai dengan 73%, nilai *precision* berkisar antara 61% sampai dengan 75%, nilai *recall* berkisar antara 58% sampai dengan 72%, dan nilai *f1-score* berkisar antara 55% sampai dengan 70%. Adapun saran untuk penelitian berikutnya ialah dapat melakukan komparasi dengan berbagai algoritma klasifikasi yang lain, agar memiliki hasil komparasi yang beragam.

DAFTAR PUSTAKA

[1] A. Lutfi, A. Annisa Fitriani, I. Ramadani, N. Azahra Putri, and Y. Shizuka Nelsi, "Efekvitas Penggunaan Aplikasi My Pertamina Di Era Kenaikan Bbm Bersubsidi," vol. 1, no. 2, 2022.

[2] R. Muhammad Ibrahim and N. Novandriani Karina Moeliono, "Persepsi Konsumen Pada My Pertamina (Studi Pada Penggunaan My Pertamina Kota Bandung)," *J. Ilm. Mhs. Ekon. Manaj. Accredit. SINTA*, vol. 4, no. 2, pp. 396–413, 2020, [Online]. Available: http:jim.unsyiah.ac.id/ekm

[3] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, [Online]. Available: http://jim.teknokrat.ac.id/index.php/informatika

[4] R. Y. Hayuningtyas, "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita," *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019, doi: 10.31311/ji.v6i1.4685.

[5] I. W. A. Argodi, E. Y. Puspaningrum, and M. M. Al Haromainy, "Implementasi Metode Tf-Idf Dan Algoritma Naive Bayes Dalam Aplikasi Diabetic Berbasis Android," *J. Tek. Mesin, Elektro dan Ilmu Komputer(TEKNIK)*, vol. 3, no. 2, pp. 23–33, 2023, [Online]. Available: http://journal.amikveteran.ac.id/index.php/teknik/article/view/2009%0Ahttps://journal.amikveteran.ac.id/index.php/teknik/article/download/2009/1610

[6] F. Tangguh Admojo, "Indonesian Journal of Data and Science Klasifikasi Aroma Alkohol Menggunakan Metode KNN," *Indones. J. Data an Sci.*, vol. 1, no. 2, pp. 34–38, 2020.

[7] S. Nurjanah, A. M. Siregar, and D. S. Kusumaningrum, "Penerapan Algoritma K – Nearest Neighbor (KNN) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta," *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 2, pp. 71–76, 2020, [Online]. Available:

http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/14

[8] M. Syarifuddin, "Analysis of Public Opinion Sentiment Regarding Covid-19 on Twitter Using the Naïve Bayes and Knn Method," *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.

[9] P. K. Tiwari, M. Sharma, P. Garg, T. Jain, V. K. Verma, and A. Hussain, "A Study on Sentiment Analysis of Mental Illness Using Machine Learning Techniques," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1099, no. 1, p. 012043, 2021, doi: 10.1088/1757-899x/1099/1/012043.

[10] R. Astrada Fathurrahman, "Analisis Sentimen Pasar Saham dari Berita Keuangan menggunakan Algoritma Pencocokan String," no. 13519113, 2021, [Online]. Available: https://ekonomi.kompas.com/read/2017/11/03/221703226/kenaikan-

[11] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," *Proc. 8th Int. Conf. Weblogs Soc. Media, ICWSM 2014*, pp. 216–225, 2014, doi: 10.1609/icwsm.v8i1.14550.

[12] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 379, 2020, doi: 10.26418/jp.v6i3.42896.

[13] S. Mulyani and R. Novita, "Implementation of the Naive Bayes Classifier Algorithm for Classification of Community Sentiment About Depression on Youtube," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1355–1361, 2022, doi: 10.20884/1.jutif.2022.3.5.374.

[14] F. Tri Anggraeny, I. Yuniar Purbasari, and E. Fitria Wulandari, "Undergraduate Thesis Supervisor Recommendation Based On Text Similarity," no. 10, pp. 86–94, 2019, [Online]. Available: https://jurnal.darmajaya.ac.id/index.php/icitb/article/view/2077

[15] A. Sabrani, I. W. Gede Putu Wirarama Wedashwara, and F. Bimantoro, "METODE MULTINOMIAL NAÏVE BAYES UNTUK KLASIFIKASI ARTIKEL ONLINE TENTANG GEMPA DI INDONESIA (Multinomial Naïve Bayes Method for Classification of Online Article About Earthquake in Indonesia)," *Jtika*, vol. 2, no. 1, pp. 91–92, 2020, [Online]. Available: http://jtika.if.unram.ac.id/index.php/JTIKA/

[16] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *JMLR*, pp. 2825–2830, 2011, Accessed: Apr. 09, 2023. [Online]. Available: http://scikit-learn.sourceforge.net.

[17] L. A. Kurgan and P. Musilek, "A survey of Knowledge Discovery and Data Mining process models," *Knowl. Eng. Rev.*, vol. 21, no. 1, pp. 1–24, 2006, doi: 10.1017/S0269888906000737.

[18] A. M. Rahat, A. Kahir, and A. K. M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," *Proc. 2019 8th Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2019*, pp. 266–270, 2020, doi: 10.1109/SMART46866.2019.9117512.

[19] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, "Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor," *Kilat*, vol. 8, no. 1, pp. 81–92, 2019, doi: 10.33322/kilat.v8i1.421.

[20] N. Umar and M. Adnan Nur, "Application of Naïve Bayes Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 4, pp. 585–590, 2022, doi: 10.29207/resti.v6i4.4179.