JUMINTAL: Jurnal Manajemen Informatika dan Bisnis Digital
https://journal.literasisains.id/index.php/JUMINTAL
ISSN 2830-3016 (Media Online)
Vol. 3 No. 2 (November 2024) 72-84
DOI: 10.55123/jumintal.v3i2.3965

Perbandingan Analisis Sentimen Ulasan Aplikasi Ajaib Kripto Menggunakan Metode *Naïve Bayes* dan *K-Nearest Neighbor*

Calvin Wendy[1], Ade Maulana[2]
[1,2]Informatika, Fakultas Ilmu Komputer, Universitas Pelita Harapan, Medan, Indonesia
Email: [1]03082200010@student.uph.edu, [2]ade.maulana@lecturer.uph.edu

Informasi Artikel
Diterima : 06-11-2024
Disetujui : 16-11-2024
Diterbitkan : 25-11-2024

ABSTRACT

Developments that occur in the investment sector make people interested in investing. The platform that is often used is the crypto magic application on the Google Play Store. Potential users see app reviews, which are based on user opinions and used as consideration before deciding to use the app. So data analysis is needed. One way of analyzing data is called sentiment analysis. There are many methods that can be used for sentiment analysis. So the author uses the Naive Bayes and K-Nearest Neighbor methods to find out a more accurate method. The data taken for this study is a review of the crypto magic application collected from 2022 to 2024 with 4784 data. Dataset through the stages of data division using k-fold cross validation with k=5, then the preprocessing stage, TF-IDF transformation, training Naive Bayes and KNN models with training data. The model was then tested on test data. The result is that the Naïve Bayes method provides an accuracy rate of 82%, precision of 83%, recall of 98% and an f1-score of 90%. While the KNearest Neighbor method provides an accuracy rate of 80%, precision of 84%, recall of 94% and f1-score of 89%. The conclusion of this study is that the Naïve Bayes method has a better level of accuracy than the K-Nearest Neighbor method.

Keyword: *K-Nearest Neighbor, Naive Bayes, Sentiment Analysis, Ajaib Kripto*

ABSTRAK

Perkembangan yang terjadi di bidang investasi membuat masyarakat tertarik untuk melakukan investasi. Platform yang sering dipakai adalah aplikasi ajaib kripto pada Google Play Store. Calon pengguna melihat ulasan aplikasi, yang dibuat berdasarkan pendapat pengguna dan digunakan sebagai pertimbangan sebelum memutuskan menggunakan aplikasi tersebut. Sehingga diperlukan

analisis data. Salah satu cara menganalisis data disebut analisis sentimen. Terdapat banyak sekali metode yang dapat digunakan untuk analisis sentimen. Sehingga penulis menggunakan metode Naive Bayes dan K-Nearest Neighbor untuk mengetahui metode yang lebih akurat. Data yang diambil untuk penelitian ini merupakan ulasan dari aplikasi ajaib kripto tersebut yang dikumpulkan dari tahun 2022 sampai 2024 dengan data sebanyak 4784. Dataset melalui tahapan pembagian data dengan menggunakan k-fold cross validation dengan k=5, kemudian tahapan preprocessing, transformasi TF-IDF, melatih model Naive Bayes dan KNN dengan data latih. Model tersebut kemudian dicoba pada data uji. Hasilnya adalah metode Naïve Bayes memberikan tingkat akurasi sebesar 82%, presisi sebesar 83%, recall sebesar 98% dan f1-score sebesar 90%. Sedangkan metode K-Nearest Neighbor memberikan tingkat akurasi sebesar 80%, presisi sebesar 84%, recall sebesar 94% dan f1-score sebesar 89%. Kesimpulan dari penelitian ini adalah metode Naïve Bayes memiliki tingkat akurasi yang lebih baik dibandingkan metode K-Nearest Neighbor.

Kata Kunci: *K-Nearest Neighbor*, *Naive Bayes*, Analisis Sentimen, Ajaib Kripto

1. PENDAHULUAN

Perkembangan yang terjadi di bidang investasi membuat masyarakat tertarik untuk melakukan investasi. "Dari data yang dikeluarkan oleh KSEI, terdapat 4,515,103 dari penduduk Indonesia, yaitu 272,229,372 atau sekitar 1,65% yang menjadi investor dalam Pasar Modal" (Dian et al., 2022). Alat tukar tidak selalu berbentuk fisik seperti uang logam atau kertas. Sekarang ada uang kripto.

Teknologi telah maju dengan sangat cepat. Pada tahun 2020, terdapat 196,7 juta pengguna internet di Indonesia, menurut data dari Asosiasi Penyelenggara Jasa Internet Indonesia. Aplikasi ini dapat diunduh pada perangkat pengguna melalui *Google Play Store* (Diki Hendriyanto, Ridha and Enri, 2022).

Salah satu *platform* yang sering dipakai untuk melakukan investasi saat ini yaitu aplikasi Ajaib Kripto. Sudah tercatat 50 ribu orang yang telah mengunduh aplikasi Ajaib Kripto, yang merupakan *platform* untuk melakukan transaksi aset kripto dan merupakan salah satu *platform* yang paling sering digunakan untuk melakukan investasi saat ini (Muzamziyah, 2022).

Rating yang didapatkan dari aplikasi di *Google Play Store* diikuti dengan ulasan dari pengguna terhadap aplikasi tersebut (Diki Hendriyanto, Ridha and Enri, 2022). Calon pengguna melihat ulasan aplikasi, yang dibuat berdasarkan pendapat pengguna dan digunakan sebagai pertimbangan sebelum memutuskan untuk menggunakan aplikasi tersebut. Suatu teknik diperlukan untuk mengetahui bagaimana ulasan pengguna terhadap suatu

aplikasi yang sangat banyak dan tidak terstruktur. Karena itu, data ulasan harus dianalisis (Diki Hendriyanto, Ridha and Enri, 2022).

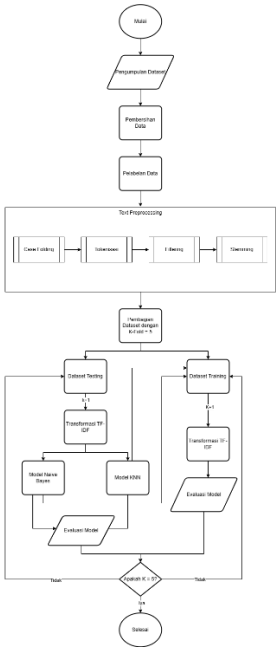
Analisis sentimen adalah fase di mana data teks secara otomatis dianalisis untuk memahami dan mengekstrak sentimen dalam ulasan (Adminlp2m, 2022). Beberapa metode yang dapat digunakan untuk melakukan analisis sentimen termasuk Naive Bayes dan K-Nearest Neighbor.

Pada penelitian yang telah ada, Naive Bayes memiliki tingkat akurasi yang lebih tinggi sebesar 71% dibandingkan K-Nearest Neighbor sebesar 70% saat diterapkan untuk analisis sentimen pengguna aplikasi Shopee (Khatib Sulaiman, Shopee Salman Alfaris and Amikom Yogyakarta, 2023). Sedangkan pada penelitian lainnya, metode K-Nearest Neighbor memiliki tingkat akurasi yang lebih tinggi sebesar 92,8% dibandingkan dengan Naive Bayes sebesar 91,4% saat diterapkan untuk analisis klasifikasi sentimen ulasan pada *e-commerce* Shopee berbasis *world cloud* (Josen Limbong et al., 2022).

Berdasarkan penelitian sebelumnya, terdapat perbedaan pada akurasi antara kedua metode yang digunakan, maka pada penelitian ini akan dilakukan untuk mengetahui perbandingan akurasi antara metode Naive Bayes dan K-Nearest Neighbor, maka dengan ini penulis melakukan penelitian dengan judul "Perbandingan Analisis Sentimen Ulasan Aplikasi Ajaib Kripto Menggunakan Metode Naïve Bayes dan K-Nearest Neighbor".

2. METODE

Pada bab sebelumnya, terdapat latar belakang dan rumusan masalah yang sudah dijelaskan, sehingga dibentuk kerangka berpikir yang akan dilakukan pada penelitian ini dalam gambar 1 di bawah ini.



Gambar 1. *Flowchart* Pengerjaan

2.1. Pengumpulan Data

Pengumpulan data merupakan langkah penting yang berfungsi sebagai bahan untuk melakukan analisis. Dengan teknik yang tepat, maka data yang akurat dapat diperoleh untuk menyelesaikan penelitian yang ada(Rangkuti, 2024). Dataset yang digunakan merupakan data yang diambil dari salah satu aplikasi di Google Play Store yaitu Ajaib Kripto. Data yang diambil merupakan ulasan dari aplikasi tersebut. Data yang dikumpulkan memiliki rentang waktu dari tahun 2022 sampai 2024 sebanyak 4784 data.

2.2. Pembersihan Data

Pembersihan data merupakan proses untuk mengidentifikasi, mendeteksi, dan mengoreksi kesalahan atau ketidakakuratan yang terdapat dalam kumpulan data. Pembersihan data dilakukan agar dapat meningkatkan kualitas dan keakuratan data saat menganalisis data. Data yang tidak bersih terdiri dari beberapa hal, yaitu data duplikat, data hilang, data yang tidak valid, dan kolom yang tidak diperlukan.

2.3. Pelabelan Data

Pelabelan data merupakan proses untuk mengkategorikan kumpulan data ke dalam kelompok tertentu (Telnoni, 2021). Label tersebut berupa label positif yang ditandai dengan angka "1" dengan jumlah sebanyak 3568 data, label negatif yang ditandai dengan angka "-1" dengan jumlah sebanyak 929 data, dan label netral yang ditandai dengan angka "0" dengan jumlah sebanyak 287 data.

2.4. *Text Preprocessing*

Tahap *text preprocessing* merupakan sebuah proses untuk menyeleksi data teks agar menjadi lebih terstruktur, mengubah data teks menjadi angka, dan mengumpulkan informasi yang penting sehingga dapat dianalisis. Tugas utamanya yaitu untuk menghilangkan dan menangani *noise* data agar perhitungan menjadi optimal (Fathiah Annisa et al., n.d.). *Text preprocessing* terdiri atas beberapa tahap, antara lain:

2.4.1. *Case Folding*

Case folding berfungsi untuk menyamaratakan penggunaan huruf kapital dikarenakan data yang kita miliki tidak selalu terstruktur dan konsisten. Dalam penelitian ini mengkonversi teks ke dalam format kecil(*lowercase*). Hal tersebut bertujuan untuk memberikan bentuk standar pada teks.

2.4.2. Tokenisasi

Tokenisasi merupakan proses memecah kalimat-kalimat menjadi kata atau disebut dengan token sehingga dapat mempermudah proses analisis data. Tokenisasi dapat membedakan mana antara pemisah kata atau bukan.

2.4.3. *Filtering*

Filtering merupakan proses menghapus bagian teks yang tidak penting sehingga mengganggu dalam proses analisis. Penghapus tersebut mencakup penghapusan tautan web, kata-kata umum, tanda baca, atau angka. Dalam proses *filtering*, terdapat empat metode yang digunakan, yaitu:

a. *URL Removal*

URL removal merupakan proses menghapus tautan web dari teks karena tautan tersebut biasanya tidak relevan dalam analisis teks.

b. *Stopword Removal*

Stopword removal merupakan proses penghapusan kata-kata umum yang tidak memberikan banyak informasi penting dalam analisis teks. Kata-kata tersebut sering muncul secara umum tetapi tidak memberikan kontribusi signifikan terhadap makna teks.

c. *Punctuation Removal*

Punctuation removal merupakan proses penghapusan tanda baca dari teks. Tanda baca seperti koma, titik, atau tanda seru sering kali tidak diperlukan dalam analisis teks.

d. *Number Removal*

Number removal merupakan proses pembersihan teks dari angka. Angka juga tidak memberikan informasi yang berguna dalam pemrosesan teks. Data angka hanya akan menambah kompleksitas pada proses pemrosesan dan memengaruhi akurasi model secara negatif.

2.4.4. *Stemming*

Stemming merupakan proses mengkonversi kata-kata ke bentuk dasarnya. Hal tersebut membantu mengurangi dimensi fitur dan meningkatkan konsistensi dalam data teks.

2.4.5. Pembagian *Dataset* dengan *K-Fold Cross Validation*

Pembagian *dataset* merupakan langkah penting dalam persiapan data. *Dataset* dibagi menjadi dua bagian utama yaitu data pelatihan dan data pengujian. Pembagian *dataset* ini diperlukan untuk membantu mengidentifikasi apakah model memiliki tingkat akurasi yang baik (Hafid, 2023). Dalam kasus ini, pembagian *dataset* dilakukan dengan menggunakan *k-fold cross validation*. *K-fold cross validation* adalah proses dimana dataset dipecah menjadi sejumlah *K fold* dan digunakan untuk mengevaluasi kemampuan model saat diberikan data baru (Nugroho and Amrullah, 2023). Nilai optimal untuk parameter K pada *K-fold cross validation* berkisar antara 5 sampai 10. Jika nilai k adalah 5, maka validasi silang dilakukan 5 kali lipat. Setiap lipatan digunakan sebagai set pengujian pada satu titik dalam proses (hebohseo, 2024).

2.4.6. *Transformation* (TF-IDF)

TF-IDF merupakan teknik untuk melakukan pembobotan data yang berbasis pada statistik kemunculan kata dan tingkat kepentingan dokumen yang mengandungnya (Adhiatma and Qoiriah, n.d.).

2.4.7. Metode Naïve Bayes

Dalam penelitian ini, metode Naive Bayes digunakan untuk mencari hasil akhir dari data yang ada. Tipe yang digunakan adalah multinomial Naive Bayes. Multinomial Naïve Bayes digunakan untuk memecahkan masalah klasifikasi dokumen(UMA, 2021). Data hasil normalisasi TF-IDF pada data *training* digunakan untuk melatih model Naïve Bayes.

2.4.8. Metode K-Nearest Neighbor

Dalam penelitian ini, metode K-Nearest Neighbor digunakan untuk mencari hasil akhir dari data yang ada. K-Nearest Neighbour adalah model *machine learning* yang digunakan untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut(Baharuddin, Azis and Hasanuddin, 2019). Data hasil normalisasi TF-IDF pada data training digunakan untuk melatih model K-Nearest Neighbor.

2.4.9. Evaluasi Model

Setelah proses pelatihan model dilakukan, maka kita dapat menggunakan *confusion matrix* untuk mengukur atau melakukan perhitungan akurasi pada konsep data mining.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dilakukan dengan teknik *scraping*. Teknik *scraping* tersebut menggunakan bahasa *Python* dan menggunakan *library google_play_scraper* yang digunakan untuk mengambil data dari Google Play Store. Atribut yang akan diambil yaitu username, score, at, dan content. Untuk mengambil data sesuai atribut yang diinginkan.

Setelah mendapatkan data dengan atribut yang kita inginkan, data tersebut dapat dilihat pada tabel 6 di bawah ini.

Tabel 6. *Dataframe* sesuai atribut yang diinginkan

userName	*score*	*at*	*content*
Wildan 5898	4	2024-03-18 12:17:24	pakai kode ini ”lanaxdxt” untuk dapetin tambahan
Yardi Bringaz	1	2024-03-18 05:25:28	Aset ny kagak lengkap

3.2 Pembersihan Data

Dalam penelitian ini, terdapat 4 atribut, yaitu "*username*", "*score*", "*at*", dan "*content*". Kemudian atribut "*username*" dan "*at*" akan dihapus dikarenakan tidak relevan dengan penelitian yang ada. Hal ini dapat dilihat pada tabel 7 di bawah ini.

Tabel 7. *Dataframe* setelah penghapusan atribut

Score	*Content*
1	Aplikasi opo yo, jgn pakai ini jelek
1	lama amat proses verifikasi pembukaan data

Setelah melakukan penghapusan kolom yang tidak diperlukan, pembersihan data lainnya seperti memeriksa apakah terdapat data kosong pada setiap elemen yang ada, menghapus baris duplikat, dan lain-lain.

3.3 Pelabelan Data

Pelabelan data pada penelitian ini menggunakan *Excel*. Dataset dengan rating 1 sampai 2 diberi label -1 (negatif), rating 3 diberi label 0 (netral), dan rating 4 sampai 5 diberi label 1 (positif). Data setelah pemberian label dapat dilihat pada tabel 8 di bawah ini.

Tabel 8. *Dataframe* setelah pemberian label

	score	*content*	label
0	5	tidak tau cara masuk ganti nama n password	1
1	5	ok	1
…	…	…	…
4779	1	saran saya mending pakai aplikasi saja	-1
4780	5	moga aplikasi ini dapat manfaat untuk kita semua	1

3.3.1 *Text Preprocessing*

Pada tahap *text preprocessing*, data yang sudah terkumpul dan telah diberi label akan dilakukan preprocessing seperti pada tabel 9 dibawah ini :

Tabel . Contoh data *pre-processing*

Data *pre-processing*	Hasil
Data	Aplikasi apa ya, jangan pakai ini jelek
Case Folding	aplikasi apa ya, jangan pakai ini jelek
Tokenisasi	[”aplikasi”, ”apa”, ”ya”, ”jangan”, ”pakai”, ”ini”, ”jelek”]
Filtering	[”aplikasi”, ”jangan”, ”pakai”, ”jelek”]
Stemming	[”aplikasi”, ”jangan”, ”pakai”, ”jelek”]

1. Pembagian *Dataset* dengan *K-Fold Cross Validation*

   Pembagian *dataset* dengan *K-Fold Cross Validation* dapat dilakukan dengan menggunakan kelas *Kfold* dari modul *model_selection* dari *library sklearn.*

2. Transformasi (TF-IDF)

   Proses perhitungan TF-IDF dapat dilakukan dengan menggunakan *class TfidfVectorizer()* yang terdapat dalam modul *feature_extraction.text* dari *library sklearn*.

3. Model Naive Bayes

   Membuat model Naïve Bayes dapat dilakukan dengan menggunakan kelas *MultinomialNB* dari modul *naive_bayes* dari *library sklearn.*

4. Model K-Nearest Neighbor

   Membuat model K-Nearest Neighbor dapat dilakukan dengan menggunakan kelas *KNeighborsClassifier* dari modul *neighbors* dari *library sklearn*.

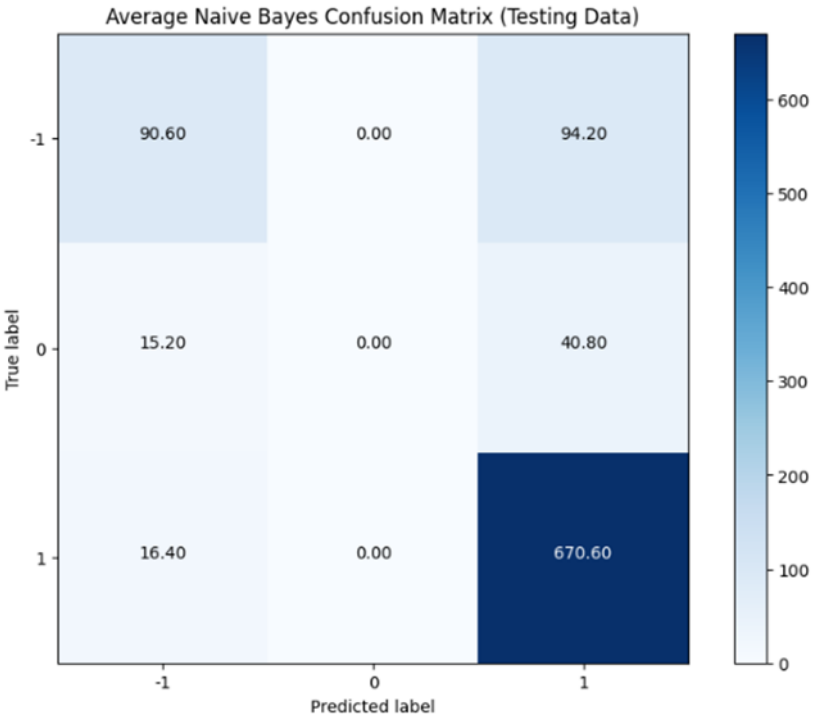
5. Hasil Evaluasi Model

   Hasil prediksi pada data uji akan ditampilkan dalam bentuk *confusion matrix* serta dalam matriks evaluasi. Setelah melalui tahapan-tahapan di atas, hasil yang diperoleh dapat dilihat di bawah ini.

1. Data Uji

a. Naive Bayes

   Rata-rata hasil prediksi pada data uji untuk label positif dengan model Naive Bayes dalam bentuk *confusion matrix* dapat dilihat pada gambar 4 di bawah ini.



Gambar 4. Rata-rata confusion matrix pada data uji dengan model Naïve Bayes

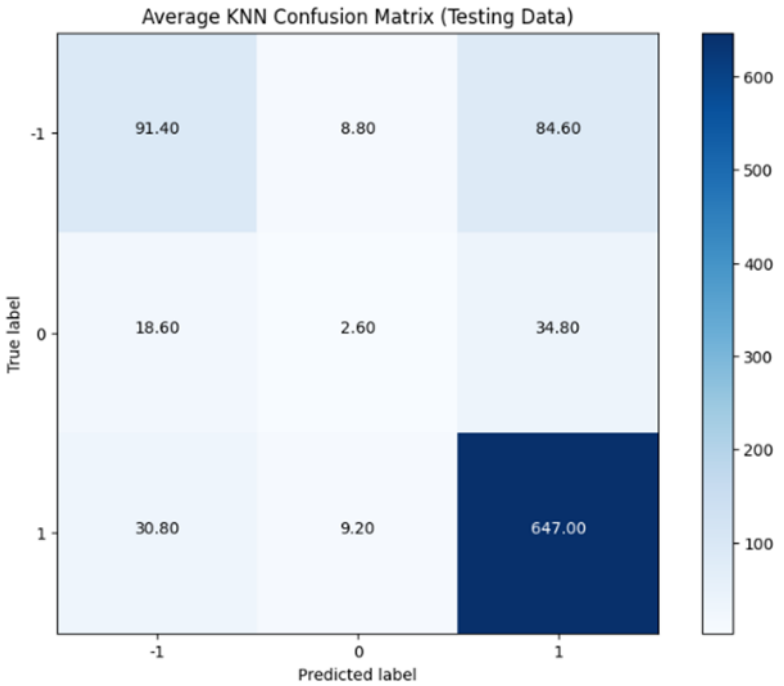
Hasil prediksi pada data uji untuk label positif dengan model Naive Bayes dalam bentuk matriks evaluasi untuk setiap K dapat dilihat pada tabel 11 di bawah ini.

Tabel 11. Matriks evaluasi pada data uji label positif dengan model Naive Bayes

K	*Accuracy*	*Precision*	*Recall*	*F1-Score*
1	0,81	0,82	0,97	0,89
2	0,84	0,85	0,98	0,91
3	0,82	0,82	0,98	0,89
4	0,82	0,83	0,98	0,90
5	0,82	0,84	0,97	0,90

b. K-Nearest Neighbor

Rata-rata hasil prediksi pada data uji untuk label positif dengan model KNN dalam bentuk *confusion matrix* dapat dilihat pada gambar 5 di bawah ini.



Gambar 5. Rata-rata confusion matrix pada data uji dengan model KNN

Hasil prediksi pada data uji untuk label positif dengan model KNN dalam bentuk matriks evaluasi untuk setiap K dapat dilihat pada tabel 12 di bawah ini.

Tabel 12. Matriks evaluasi pada data uji label positif dengan model KNN

K	*Accuracy*	*Precision*	*Recall*	*F1-Score*
1	0,80	0,84	0,95	0,89
2	0,80	0,85	0,94	0,89
3	0,80	0,86	0,92	0,89
4	0,80	0,83	0,95	0,89
5	0,80	0,84	0,95	0,89

Setelah melakukan pelatihan dan pengujian dengan menggunakan model Naive Bayes dan KNN, kita dapat mengambil rata-rata matriks evaluasi pada data uji dan data latih untuk label positif dengan model Naive Bayes dan KNN. Hal tersebut dapat dilihat pada tabel 13 di bawah ini.

Tabel 13. Rata-rata Matriks evaluasi pada data uji dan data latih label positif pada model Naive Bayes dan KNN

	Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Data Testing	Naïve Bayes	0,82	0,83	0,98	0,90
	KNN	0,80	0,84	0,94	0,89

Hasil pengujian menunjukkan bahwa model Naive Bayes memiliki tingkat akurasi yang lebih besar dibandingkan dengan model KNN. Dalam model Naive Bayes, semakin banyak data pelatihan yang dilatih, maka model tersebut akan lebih sulit untuk mengidentifikasi data. Sedangkan dalam model KNN, semakin banyak data pelatihan yang dilatih, maka model tersebut akan lebih efektik untuk mengidentifikasi data.

4. PENUTUP

4.1. Kesimpulan

Berdasarkan hasil penelitian yang ada, dapat disimpulkan bahwa metode Naïve Bayes memiliki tingkat akurasi yang lebih baik dibandingkan metode K-Nearest Neighbor. Metode Naïve Bayes memberikan tingkat akurasi sebesar 82%, presisi sebesar 83%, recall sebesar 98% dan f1-score sebesar 90%. Sedangkan metode KNearest Neighbor memberikan tingkat akurasi sebesar 80%, presisi sebesar 84%, recall sebesar 94% dan f1-score sebesar 89%.

Berdasarkan hasil penelitian yang ada, perbandingan keakuratan antara kedua metode dilihat berdasarkan hasil dari f1-score. Hal tersebut dikarenakan data yang digunakan terdiri dari distribusi kelas yang tidak seimbang antara sentimen positif, negatif, dan netral. Jika sebagian besar data adalah sentimen positif, maka model yang memprediksi sentimen positif dapat memiliki akurasi yang lebih tinggi sehingga akurasi tidak dapat diandalkan. Sehingga berdasarkan hasil dari f1-score, metode Naive Bayes memiliki performa yang lebih baik dibandingkan dengan metode KNN dengan nilai f1-score sebesar 90%:89%.

4.2. Saran

Berdasarkan hasil penelitian yang ada, diberikan beberapa saran sebagai berikut :

1. Peneliti selanjutnya dapat mengembangkan aplikasi berdasarkan model pada penelitian ini agar dapat digunakan oleh masyarakat.
2. Peneliti selanjutnya dapat menggunakan teknik-teknik pemrosesan teks yang lebih canggih seperti penggunaan *deep learning* atau *word embeddings* kemudian membandingkan hasil dengan hasil penelitian ini.

DAFTAR PUSTAKA

Adhiatma, F.D. and Qoiriah, A., n.d. Penerapan Metode TF-IDF dan Deep Neural Network untuk Analisa Sentimen pada Data Ulasan Hotel. *Journal of Informatics and Computer Science*.

Adminlp2m, 2022. Analisis Sentimen, Definisi Tipe Dan Cara Kerja.

Baharuddin, M.M., Azis, H. and Hasanuddin, T., 2019. ANALISIS PERFORMA METODE K-NEAREST NEIGHBOR UNTUK IDENTIFIKASI JENIS KACA. *ILKOM Jurnal Ilmiah*, 11(3), pp.269–274. https://doi.org/10.33096/ilkom.v11i3.489.269-274.

Dian, S., Permana, H., Trilogi, U., Kampus, J., No, T., Selatan, J., Bayu, K. and Bintoro, Y., 2022. Memulai Investasi Melalui Platform Investasi Digital. *JIKIS)*, 02, p.1.

Diki Hendriyanto, M., Ridha, A.A. and Enri, U., 2022. ANALISIS SENTIMEN ULASAN APLIKASI MOLA PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE SENTIMENT ANALYSIS OF MOLA APPLICATION REVIEWS ON GOOGLE PLAY STORE USING SUPPORT VECTOR MACHINE ALGORITHM. *Journal of Information Technology and Computer Science (INTECOMS)*, 5(1).

Fathiah Annisa, N.D., Alam, S., Jamil, M. and Negeri Makassar, U., n.d. ANALISIS SENTIMEN APLIKASI INVESTASI A Sentiment Analysis of Investment Applications. [online] https://doi.org/10.24252/best.v4i2.49416.

Hafid, H., 2023. *Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor pada Data Kasus Covid-19 di Indonesia*. [online] *Journal of Mathematics*, Available at: <http://www.ojs.unm.ac.id/jmathcos>.

hebohseo, 2024. *K-Fold Cross Validation dalam Machine Learning*. STATSIDEA.

Josen Limbong, J.A., Sembiring, I., Dwi Hartomo, K., Kristen Satya Wacana, U. and Korespondensi, P., 2022. ANALISIS KLASIFIKASI SENTIMEN ULASAN PADA E-COMMERCE SHOPEE BERBASIS WORD CLOUD DENGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR ANALYSIS OF REVIEW SENTIMENT CLASSIFICATION ON E-COMMERCE SHOPEE WORD CLOUD BASED WITH NAÏVE BAYES AND K-NEAREST NEIGHBOR METHODS. *r (JTIIK)*, 9, pp.347–356. https://doi.org/10.25126/jtiik.202294960.

Khatib Sulaiman, J., Shopee Salman Alfaris, A. and Amikom Yogyakarta, U., 2023. Komparasi Metode KNN dan Naive Bayes Terhadap Analisis Sentimen Pengguna. *Indonesian Journal of Computer Science Attribution*, 12(5), pp.2023–2766.

Muzamziyah, I.L., 2022. *TINJAUAN HUKUM ISLAM TERHADAP TRANSAKSI MATA UANG KRIPTO PADA APLIKASI AJAIB KRIPTO SKRIPSI*.

Nugroho, A. and Amrullah, A., 2023. *EVALUASI KINERJA ALGORITMA K-NN MENGGUNAKAN K-FOLD CROSS VALIDATION PADA DATA DEBITUR KSP GALIH MANUNGGAL*. *JINTEKS*, .

Rangkuti, M., 2024. *Teknik-Teknik Pengumpulan Data dalam Penelitian: Panduan Lengkap untuk Peneliti*. umsu.

Telnoni, P.A., 2021. Pelabelan Data Dengan Latent Dirichlet Allocation dan K-Means Clustering pada Data Twitter Menggunakan Bahasa Indonesia. *Jurnal Elektro dan Telekomunikasi Terapan*, [online] 7(2), p.885. https://doi.org/10.25124/jett.v7i2.3442.

UMA, B., 2021. *Mengenal Algoritma Naive Bayes dan Kegunaannya*. BIRO ADMINISTRASI MUTU AKADEMIK DAN INFORMASI.