
PENERAPAN METODE NAÏVE BAYES CLASSIFIER PADA SENTIMEN ANALISIS APLIKASI INVESTASI KEUANGAN DIGITAL (Studi Kasus: Bareksa Dan Bibit)

Jhon Vebrianto Girsang, Indra Kelana Jaya[✉], Harlen Gilbert Simanullang

Fakultas Ilmu Komputer, Universitas Methodist Indonesia, Medan, Indonesia

Email: indraikjs@gmail.com

DOI: <https://doi.org/10.46880/jmika.Vol7No2.pp225-230>

ABSTRACT

Investing online is a very promising opportunity. There are many online investment enthusiasts who do not understand how to invest online correctly and be able to minimize risk. Lack of public understanding of the investment implementation process can lead to fraud by irresponsible parties. So, understanding investing online is very necessary. There are many online investment applications on the Google Play Store, but these investment applications have their own advantages and disadvantages. The objects of research are the applications of Bareksa and Seeds because the news media often report on these applications at the top and selecting an application requires a collection of information obtained from previous user reviews. The method used is Naïve Bayes Classifier. Based on the results, the classification is divided into 3 (three) sentiments, namely positive, negative, and neutral. With a comparison of training and testing data 70%:30% accuracy in the Bareksa application was obtained 54% and 44% in the seed application.

Keyword: *Investing, Google Play Store, Naïve Bayes Classifier.*

ABSTRAK

Investasi secara online merupakan peluang yang sangat menjanjikan. Peminat investasi online belum memahami bagaimana cara berinvestasi online dengan benar dan mampu meminimalkan resiko. Pemahaman Masyarakat yang kurang tentang proses pelaksanaan investasi dapat menimbulkan penipuan oleh pihak yang tidak bertanggung jawab, sehingga diperlukan pemahaman berinvestasi secara online. Ada banyak aplikasi investasi online pada google play store dimana tiap aplikasi investasi memiliki kelebihan dan kekurangan. Objek penelitian adalah aplikasi bareksa dan bibit karena kerap kali media kabar memberitakan aplikasi ini pada urutan teratas dan untuk memilih aplikasi membutuhkan kumpulan informasi yang didapat pada ulasan pengguna sebelumnya. Metode yang digunakan yaitu Naïve Bayes Classifier. Hasil klasifikasi adalah 3 (tiga) sentimen yaitu positif, negatif dan netral. Dengan perbandingan data training dan data testing 70%:30% accuracy pada aplikasi bareksa diperoleh 54% dan pada aplikasi bibit 44%.

Kata Kunci: *Investasi, Google Play Store, Naïve Bayes Classifier.*

PENDAHULUAN

Investasi secara *online* merupakan peluang yang menjanjikan. Peminat investasi *online* belum memahami bagaimana cara berinvestasi online dengan benar yang mampu meminimalkan resiko. Pemahaman masyarakat yang kurang tentang proses investasi dapat menimbulkan penipuan dan kerugian. Jadi pemahaman menyeluruh terkait investasi seperti manfaat, tujuan dan resiko investasi sangat diperlukan. Ada banyak aplikasi investasi di Google Play Store, dengan beragam keunggulan dan kelemahannya. Kelemahan proses seperti pembelian yang membutuhkan waktu akses dan rumit penarikan dana. Masalah ini sangat mempengaruhi layanan kualitas pengguna aplikasi.

Analisis sentiment terhadap aplikasi bibit dan bareksa menggunakan total 998 ulasan. Sentimen positif sebesar 484 dan negatif sebesar 514 terhadap Bareksa. Bibit menggunakan 1063 ulasan yang terdiri dari 541 positif dan 522 negatif. Ulasan melalui tahapan pengolahan data awal dan pemodelan (A. Putra & Dwiki, 2021). Penelitian ini menganalisis 2 (dua) aplikasi investasi online yang tersedia pada Google Play Store, yaitu Aplikasi Bareksa dan Bibit karena kerap kali media kabar memberitakan aplikasi ini pada urutan teratas. Pemilihan aplikasi investasi *online* membutuhkan ulasan (*review*) aplikasi. Pengumpulan data dilakukan dari data ulasan dari kedua aplikasi tersebut. Prediksi analisis sentimen dan akurasi hasil

prediksi menggunakan algoritma *Naive Bayes Classifier*.

TINJAUAN PUSTAKA

Text Mining

Text Mining adalah eksplorasi pola pada sekumpulan dokumen untuk mendapatkan informasi dan pengetahuan dari sejumlah besar data (P. Putra & Yudhoatmojo, 2021). Tujuannya untuk membantu dalam menganalisis keterhubungan dokumen berdasarkan kata-kata yang mewakili isi untuk mendapatkan hasil yang berguna. Data yang digunakan dalam proses *text mining* adalah data yang tidak terstruktur, sedangkan *data mining* menggunakan data terstruktur (Lutfianti, 2023).

Text Preprocessing

Text Preprocessing adalah tahapan persiapan teks yang akan diolah selanjutnya. Algoritma klasifikasi tidak dapat mengolah data mentah (Wahyuni, Saepudin, & Falentino, 2022), sehingga dibutuhkan proses awal yaitu pengubahan teks menjadi data *numeric*. Tahapan pengolahan data awal ini dilakukan dengan beberapa tahapan antara lain (Zai, 2022):

- a. *Cleansing*
Cleansing identik dengan proses pembersihan data, yaitu menghilangkan tanda baca dan karakter khusus. *Cleansing* bertujuan agar proses klasifikasi berjalan dengan baik.
- b. *Case Folding*
Case Folding adalah menjaga keseragaman yaitu menyajikan teks dalam bentuk huruf kecil secara keseluruhan. Seragamnya teks akan memudahkan dalam tahapan *tokenizing*.
- c. *Tokenizing*
Tokenizing atau tokenisasi merupakan pemenggalan kalimat ke token atau potongan kata-kata. Proses pemotongan kata akan memudahkan dalam proses *Stopword* dan *Stemming*.
- d. *Stemming*
Stemming merupakan proses untuk menghasilkan kata dasar dari dokumen yang diproses.

Term Weighting

Term Weighting adalah pengubahan teks menjadi nilai numerik. Proses ini menggunakan *tf-idf*. Pembobotan *tf-idf* (*Term Frequency dan Inverse Document Frequency*) merupakan ekstraksi fitur dimana pada setiap kata pada data latih diberikan nilai (Bayhaqy, Sfenrianto, Nainggolan, & Kaburuan, 2018). Rumus (*formula*) yang digunakan untuk menghitung

nilai (*W*) dari tiap dokumen berbanding kata kunci (*pencarian*), dapat dilihat pada persamaan 1.

$$W_{at} = TF_{at} * (IDF_{ft} + 1) \dots (1)$$

Keterangan:

W_{at} = Nilai dokumen ke-a pada kata ke-t

TF_{at} = Jumlah kata yang dicari dalam dokumen.

IDF_{ft} = *Inverse Document Frequency* ($\log(\frac{N}{df})$).

N = Jumlah dokumen.

df = Jumlah dokumen yang mengandung pencarian

Analisis Sentimen

Analisis sentimen (*sentiment analysis*) merupakan penilaian atau sentimen terhadap opini, peristiwa dan atribut (Sholekha, Faqih, & Bahtiar, 2022). Analisis sentimen dilakukan untuk melihat kecenderungan emosi suatu hipotesa awal ke kategori positif, negatif dan netral (Nugroho, 2018).

Naive Bayes Classifier

Naive Bayes Classifier mengadopsi teori kemungkinan / peluang yang digunakan untuk meramal keadaan yang akan datang (Bayhaqy et al., 2018). Algoritma pertama sekali dikemukakan oleh Thomas Bayes. Metode ini melakukan pembelajaran dari data masa lampu untuk menentukan hasil yang diinginkan berdasarkan aturan yang ditentukan. (Nugroho et al., 2018). *Naive Bayes Classifier* (NBC) juga digunakan dalam penentuan kelompok kelas dokumen. Tingkat performa menggunakan algoritma *Naive Bayes Classifier* (NBC) bergantung pada data yang tersedia. Pembagian data dan pemilihan data latih baik akan menghasilkan performa optimal.

Confusion Matrix

Confusion Matrix merupakan matriks yang digunakan untuk perhitungan akurasi dengan membandingkan data actual dan prediksi. *Precision* mengevaluasi kemampuan sistem yang didefenisikan sebagai persentase dokumen yang relevan terhadap *query* (Han, J., Kamber, M., & Pei, 2011).

Tabel 1. Model *Confusion Matrix 3x3*

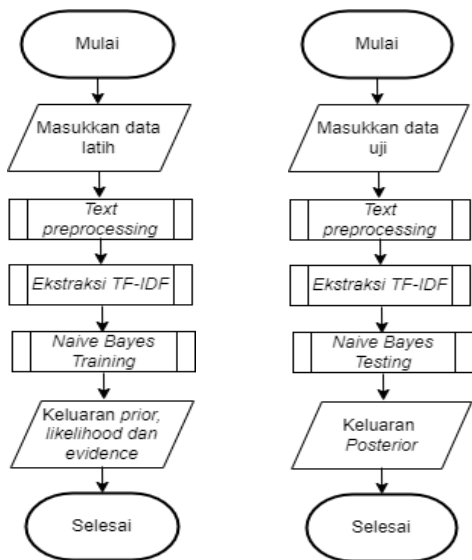
		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	Tfpos	TfPosNeg	FfPosNet
	Negatif	FfNegPos	TfNeg	FfNegNet
	Netral	FfNetPos	FfNetNeg	TfNeg

METODE PENELITIAN

Analisis Sistem

Analisis sistem adalah penguraian menjadi bagian elemen-elemen dengan tujuan mengurai dan

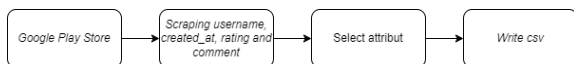
evaluasi masalah lebih terstruktur. Proses dapat dilihat pada Gambar 1.



Gambar 1. Analisis Sistem

Collecting Data

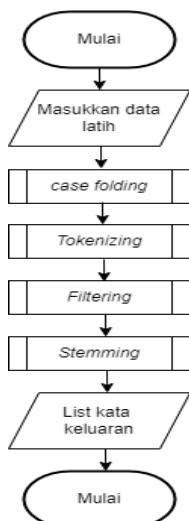
Collecting Data merupakan tahap pengambilan data di Google Play Store. Flow chart collecting data, seperti ditunjukkan pada Gambar 2.



Gambar 2. Collecting Data

Text Preprocessing

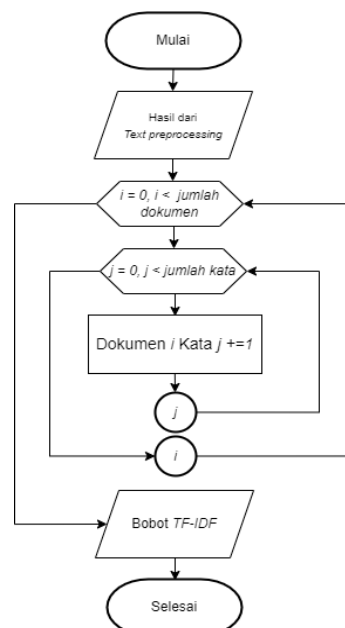
Tahapan awal dalam melakukan analisis (Yulita et al., 2021) yang terdiri dari case folding, tokenizing, filtering/stopword removal dan stemming. Flowchart text preprocessing, pada Gambar 3.



Gambar 3. Preprocessing

Ekstraksi Tf-Idf

Flowchart pembobotan / ekstraksi tf-idf, pada Gambar 4.

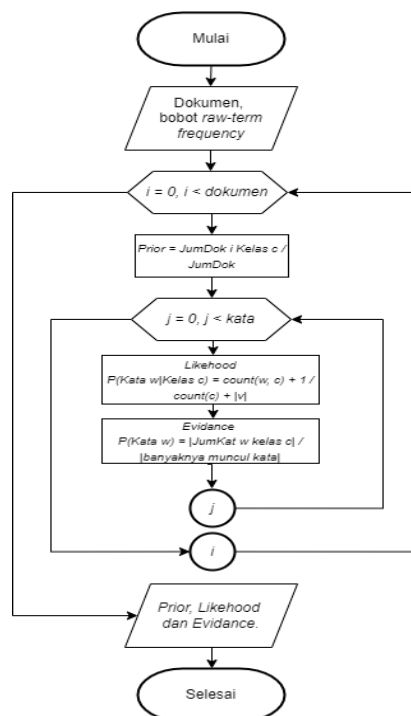


Gambar 4. Tf-IDF

Naive Bayes Classifier

Naive Bayes Training

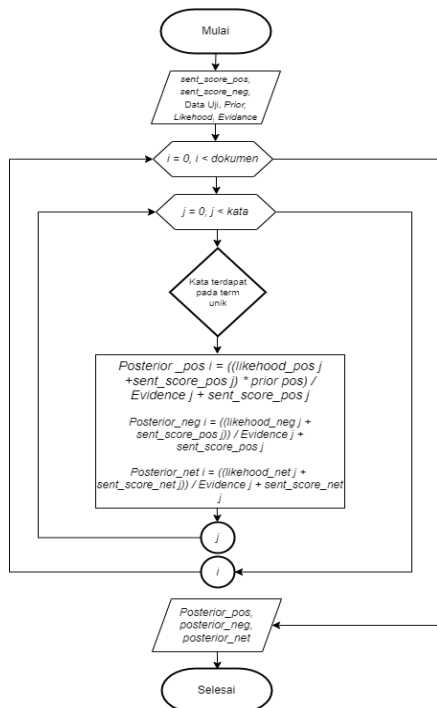
Pada bagian ini merupakan perhitungan peluang kemungkinan data sebelumnya dan bukti pada dokumen, lihat Gambar 5.



Gambar 5. Naive Bayes Training

Naïve Bayes Testing

Pada *Naïve Bayes Testing*, dilakukan perhitungan pembobotan *tf-idf*. Proses dapat dilihat pada Gambar 6.



Gambar 6. Flowchart *Naïve Bayes Testing*

**HASIL DAN PEMBAHASAN
 Penerapan *Naïve Bayes Classifier***

Tahapan dalam penerapan metode *Naïve Bayes Classifier* dimulai dari pembagian data training dan data testing, text preprocessing, term weighting data training yang digunakan sebagai berikut:

1. Aplikasi yang super lengkap, Tersedia berbagai Jenis Reksadana dari berbagai Manager Investasi. – Positif
2. Sejauh ini Paling Nyaman pakai bareksa. – Positif
3. Kasih 3 dulu baru coba kalo bagus nanti Di kasih full,, - Netral
4. Sementara bintang 3 dulu :) – Netral
5. Proses pembelian katanya maksimal 2 hari, saya sudah 6 hari kok gak selesai transaksinya.... – Negatif
6. Kenapa mau login Tidak bisa,, Udah reset Password bolak balik masih gak bisa. – Negatif

Selanjutnya adalah melakukan *text preprocessing* terhadap data training dan data testing kemudian dilakukan pembobotan *TF-IDF*. Hasilnya akan diolah menggunakan metode *Naïve Bayes Classifier* dengan menghitung *prior*, *likelihood* dan *evidence*. Proses akhir adalah menghitung *posterior*. Proses perhitungan *posterior* dilakukan menggunakan

data uji ke seluruh kelas sentiment. Hasil ditampilkan paada Tabel 2.

Tabel 2. Tabel Posterior

No	$P(\text{Sentimen} \mid \text{Data_uji-ke-}i)$	Nilai Posterior
1	$P(\text{Positif} \mid \text{Data_uji1})$	3,3093508
2	$P(\text{Negatif} \mid \text{Data_uji1})$	0,3
3	$P(\text{Netral} \mid \text{Data_uji1})$	1,3377450

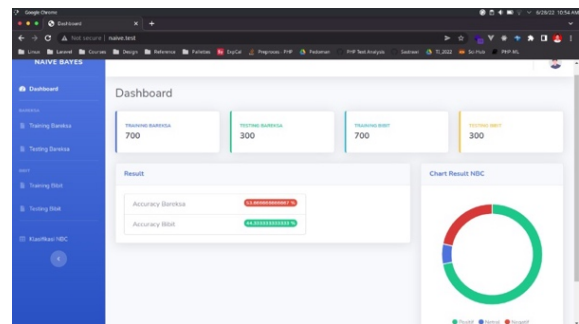
Penentuan sentimen data testing dapat dilihat dari hasil perhitungan dari *posterior* tertinggi yaitu sentiment positif.

Implementasi Sistem

Implementasi klasifikasi dengan menggunakan metode *Naïve Bayes Classifier* ini akan menampilkan Dashboard, Training Bareksa, Testing Bareksa, Training Bibit, Testing Bibit dan Classification NBC.

Halaman Dashboard

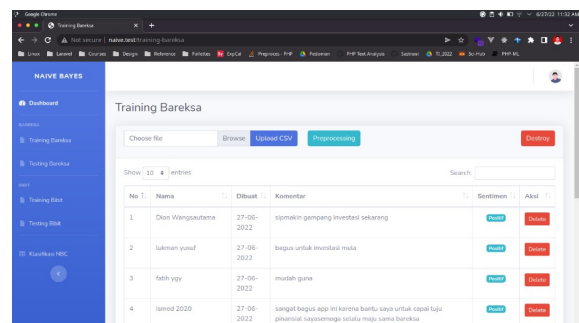
Pada halaman ini akan ditampilkan informasi *dataset training* dan *testing*. Adapun tampilan pada Gambar 6.



Gambar 6. Dashboard

Halaman Training Bareksa

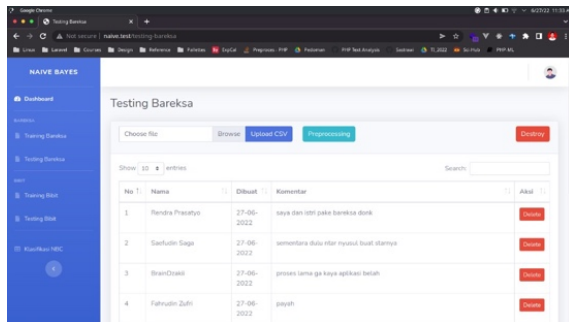
Halaman yang menampilkan *data* sebagai *data training*. User dapat *upload file* berbentuk *csv*, melakukan *text preprocessing* dan menghapus seluruh data, ditunjukkan pada Gambar 7.



Gambar 7. Training Bareksa

Halaman Testing Bareksa

Tampilan ini berisi *dataset* sebagai *data testing*. User dapat *upload file csv*, melakukan *text preprocessing* dan menghapus seluruh data.

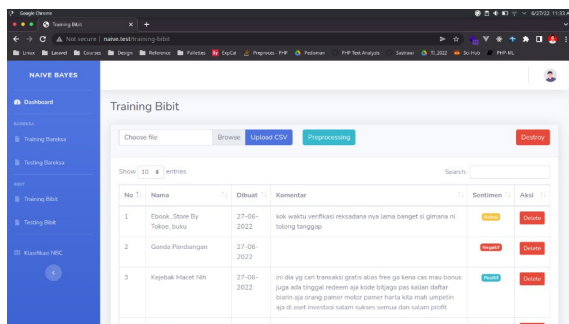


Gambar 8. Testing Bareksa

a

Halaman Training Bibit

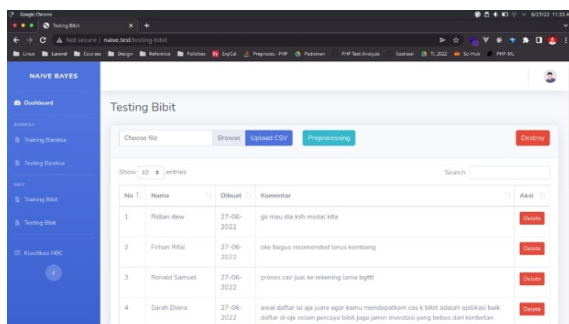
Tampilan yang berisi *data training*. User dapat *upload file* dengan format *csv*, melakukan *text preprocessing* dan menghapus seluruh data. Tampilan *Training Bibit* dapat dilihat pada Gambar 9.



Gambar 9. Training Bibit

Halaman Testing Bibit

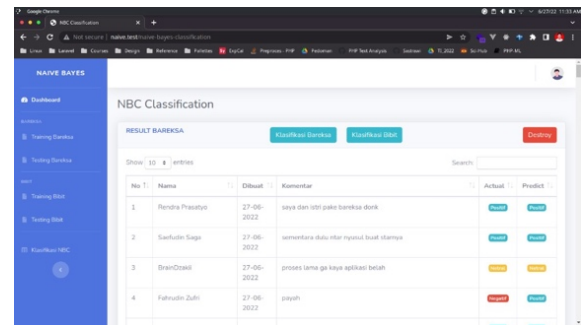
Halaman berisi *dataset* sebagai *data testing*. User dapat *upload file* berbentuk *csv*, melakukan *text preprocessing* dan menghapus seluruh data. Tampilan dari *testing bibit* dapat dilihat pada Gambar 10.



Gambar 10. Testing Bibit

Halaman Klasifikasi

Tampilan halaman ini berisi hasil klasifikasi antara kedua aplikasi, dapat dilihat pada Gambar 11.



Gambar 11. Halaman Klasifikasi

Evaluasi Model Algoritma

Hasil yang diperoleh dari metode *Naive Bayes Classifier*, selanjutnya hasil tersebut akan di evaluasi dengan confusion matrix, dari 300 data bareksa yang digunakan untuk mengevaluasi model terdapat 220 data yang berlabel positif, 23 data yang berlabel netral dan 57 data yang berlabel negatif. Hasil ini ditunjukkan pada Tabel 3.

Tabel 3. Hasil Prediksi Bareksa

		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	140	23	57
	Negatif	38	17	2
	Netral	11	8	4

sesuai dengan persamaan berikut:

$$\text{Accuracy} : (140+17+4)/300 = 0,536 = 53,6\%$$

Precision dari masing-masing sentimen adalah sebagai berikut:

- Precision sentimen positif $140/189 = 0,74 = 74\%$
- Precision sentimen negatif $17/48 = 0,354 = 35,4\%$
- Precision sentimen netral $4/63 = 0,063 = 0,63\%$

Recall dari masing-masing sentimen adalah sebagai berikut:

- Recall sentimen positif $140/220 = 0,636 = 63,6\%$
- Recall sentimen negatif $17/57 = 0,298 = 29,8\%$
- Recall sentimen netral $4/23 = 0,173 = 17,3\%$

F1 Score dari masing-masing sentimen adalah sebagai berikut:

- F1 Score sentimen positif $\frac{1}{2}(1/0,74 + 1/0,636) = 1,461 = 14,61\%$
- F1 Score sentimen negatif $\frac{1}{2}(1/0,354 + 1/0,298) = 3,090 = 30,90\%$

- F1 Score sentimen netral $\frac{1}{2}(1/0,063 + 1/0,173) = 10 = 100\%$

Pada aplikasi bibit data yang digunakan sebanyak 300 dengan hasil adalah 185 data berlabel positif, 26 data berlabel netral dan 89 data berlabel negatif. Sehingga didapatkan hasil confusion matrix yang dapat dilihat pada Tabel 4.

Tabel 4. Hasil Prediksi Bibit

		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	75	26	7
	Negatif	40	51	12
	Netral	70	12	7

accuracy dengan perhitungan sesuai dengan persamaan berikut:

Accuracy : $(75+51+7)/300 = 0,443 = 44,3\%$

Precision dari masing-masing sentimen adalah sebagai berikut:

- Precision sentimen positif $75/185 = 0,405 = 40,5\%$
- Precision sentimen negatif $51/89 = 0,573 = 57,3\%$
- Precision sentimen netral $7/26 = 0,269 = 26,9\%$

Recall dari masing-masing sentimen adalah sebagai berikut:

- Recall sentimen positif $75/108 = 0,694 = 69,4\%$
- Recall sentimen negatif $51/103 = 0,495 = 49,5\%$
- Recall sentimen netral $7/89 = 0,078 = 7,8\%$

F1 Score dari masing-masing sentimen adalah sebagai berikut:

- F1 Score sentimen positif $\frac{1}{2}(1/0,405 + 1/0,694) = 3,910 = 39,10\%$
- F1 Score sentimen negatif $\frac{1}{2}(1/0,573 + 1/0,495) = 1,882 = 18,82\%$
- F1 Score sentimen netral $\frac{1}{2}(1/0,269 + 1/0,078) = 8,268 = 82,68\%$

KESIMPULAN

Berdasarkan hasil pengujian dari penerapan metode *Naïve Bayes Classifier* pada sentimen analisis aplikasi investasi keuangan digital yang dilakukan beberapa tahapan yaitu dengan melakukan *text preprocessing*, pembobotan dokumen dan klasifikasi dengan jumlah komentar 1000 dengan perbandingan data latih dan uji sebesar 70% : 30% menghasilkan akurasi pada BAREKSA 54% dan pada aplikasi Bibit 44%.

DAFTAR PUSTAKA

Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. *2018 International Conference on Orange Technologies, ICOT 2018*, (November 2019).
<https://doi.org/10.1109/ICOT.2018.8705796>

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition*. The Morgan Kaufmann Series in Data Management Systems.

Lutfianti, N. (2023). Penerapan Sentimen Analisis Dengan Algoritma SVM Dalam Tanggapan Netizen Terhadap Berita Resesi 2023. *Sisfotenika*, 13(1), 53–64.

Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 2(2), 200.
<https://doi.org/10.30645/j-sakti.v2i2.83>

Putra, A., & Dwiki, A. (2021). Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(2), 636–646. <https://doi.org/10.35957/jatisi.v8i2.962>

Putra, P., & Yudhoatmojo, S. B. (2021). User Perception Analysis of Online Learning Platform “Zenius” During the Coronavirus Pandemic Using Text Mining Techniques. *Jurnal Sistem Informasi*, 17(2), 33–47.

Sholekha, I., Faqih, A., & Bahtiar, A. (2022). Sentiment Analysis of Public Opinion Covid-19 Vaccine Using Naïve Bayes and Random Forest Methods. *Jurnal Teknik Informatika*, 15(1), 34–43. <https://doi.org/10.15408/jti.v15i1.24847>

Wahyuni, W. A., Saepudin, S., & Falentino, Se. (2022). Sentiment Analysis of Online Investment Applications on Google Play Store using Random Forest Algorithm Method. *Jurnal Mantik*, 5(4), 2203–2209.

Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Portal Data*, 2(3).