

KLASIFIKASI OPINI PENGGUNA APLIKASI BIBIT PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES

Ferdi Jiranda Sinaga^{1*}, Fariz Prasetya Putra², Edward Pratama³, & Hafiz Irsyad⁴

¹²³⁴ Program Studi, Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang
¹ferdijiranda15@mhs.mdp.ac.id, ²farizp@mhs.mdp.ac.id, ³edopratama0304@mhs.mdp.ac.id, ⁴hafizirsyad@mdp.ac.id

Kata kunci:
data mining; naive bayes; bibit;
klasifikasi opini.

Abstract: To find out the OJK securities of an application when making online investments, a collection of information is needed that is obtained from the experience of users who have used online investment applications. Information is collected based on user experience to find out the OJK securities from the application which in this study is a seed application. The method that can be used to collect information is text mining. Text mining is one of the miners of useful information from data in the form of writing, documents or text in the form of clarification or clustering. The algorithm used in data mining in this study is Naive Bayes. In this study, the data used were 500 comments from seed application users on the Google Play Store. The results obtained using the Naive Bayes algorithm are 373 positive response data and 127 negative response data with an accuracy rate of 75%.

Abstrak: Untuk mengetahui sekuritas OJK sebuah aplikasi saat melakukan investasi online dibutuhkan kumpulan informasi yang didapat dari pengalaman pengguna yang telah menggunakan aplikasi investasi online. Informasi dikumpulkan berdasarkan pengalaman pengguna untuk mengetahui sekuritas OJK dari aplikasi tersebut dimana pada penelitian ini adalah aplikasi bibit. Metode yang dapat digunakan untuk mengumpulkan informasi adalah *Text mining*. *Text mining* merupakan salah satu penambang informasi yang berguna dari data-data yang berupa tulisan, dokumen atau text dalam bentuk klarifikasi maupun *clustering*. Algoritma yang digunakan dalam melakukan data mining pada penelitian ini adalah Naive Bayes. Pada penelitian ini data yang digunakan adalah 500 komentar pengguna aplikasi bibit pada Google Play Store. Hasil yang didapat menggunakan algoritma Naive Bayes adalah 373 data respon positif dan 127 data respon negatif dengan tingkat akurasi sebesar 75%.

Sinaga, dkk. (2023). Klasifikasi Opini Pengguna Aplikasi Bibit Pada Google Play Store Menggunakan Algoritma Naive Bayes. *MDP Student Conference 2023*

PENDAHULUAN

Menurut Cermati, Terdapat beberapa jenis - jenis instrumen yang dapat digunakan untuk berinvestasi sesuai dengan kebutuhan seperti Deposito, Reksadana, Emas, Obligasi, Pendanaan (P2P), Saham, Properti, dan Instrumen Investasi Syariah [1]. Menurut data dari Badan Koordinasi Penanaman Modal (BKPM) capaian realisasi investasi pada periode Januari - Maret pada tahun 2022 yaitu sebesar Rp 282,4 Triliun, 28,5% lebih tinggi dari tahun 2021 menandakan semakin tingginya minat dalam berinvestasi.[2]. Berdasarkan data BKPM tersebut menunjukkan meningkatnya minat masyarakat dalam berinvestasi sehingga muncul banyak aplikasi investasi *online*. Menurut topkarir, terdapat 12 aplikasi investasi yang aman dan telah terdaftar di OJK yaitu Bibit, Stockbit, SymInvest, IPOT, RTI Business, Ajaib, Mirae HOTS, BCAS Best Mobile, MOST Mandiri, Motions Trade, BIONS, dan Poems ID[3]. Namun untuk mengetahui sekuritas OJK diperlukan informasi yang didapat berdasarkan pengalaman pengguna yang telah menggunakan aplikasi investasi berbasis online tersebut karena berdasarkan Kompas.com terdapat banyak kasus - kasus penipuan pada investasi saham online [4]. Pada penelitian ini akan dikumpulkan informasi berdasarkan pengalaman pengguna untuk mengetahui sekuritas OJK dari aplikasi tersebut dimana pada penelitian ini adalah aplikasi bibit.

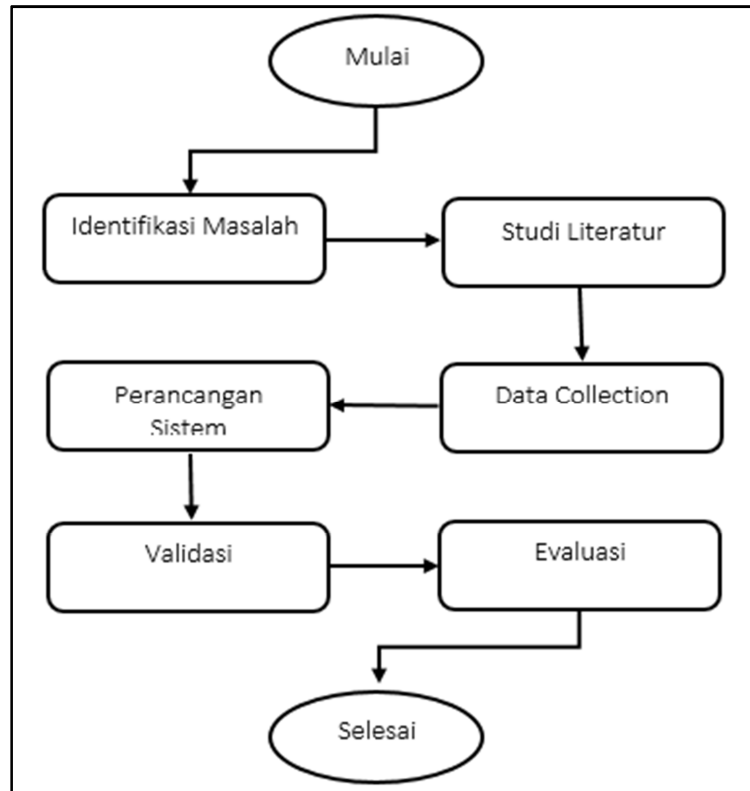
Berdasarkan penelitian terdahulu yang melakukan analisis sentimen terhadap vaksinasi Covid-19 pada twitter yang membandingkan algoritma SVM dan *Naive bayes* dalam analisis terhadap sentimen pengguna twitter menghasilkan algoritma *naive bayes* memiliki performa yang lebih baik dalam analisis sentimen dengan hasil sentimen analisis netral 8,76%, negatif 42,92%, dan positif 48,32% untuk naive bayes dan netral 10,56%, negatif 41,28%, dan positif 48,16% untuk SVM [5].

Berdasarkan penelitian terdahulu yang melakukan analisis sentimen masyarakat terhadap pembelajaran daring pada twitter selama pandemi Covid-19 menunjukkan hasil 30% sentimen positif, 69% sentimen negatif, dan 1% netral dengan algoritma *naive bayes* [6].

Berdasarkan penelitian terdahulu yang melakukan analisis sentimen terhadap pengguna by.u pada twitter menghasilkan pengujian akurasi tertinggi metode *naive bayes* dengan model multinomial sebesar 85% [7].

Berdasarkan penelitian terdahulu yang telah di jelaskan dengan akurasi yang sangat baik didapatkan dari algoritma *naive bayes*, maka dilakukanlah Klasifikasi Opini Pengguna Aplikasi Bibit Pada Google Play Store Menggunakan Algoritma Naive Bayes.

METODE



Gambar 1. Tahapan Penelitian

Penelitian menggunakan menggunakan metode *Cross-Industry Standard Process for Data Mining* yang akan memproses data dengan tahapan sebagai berikut:

Identifikasi Masalah

Identifikasi masalah memiliki tujuan untuk menentukan masalah yang terjadi, yang membuat (judul) dijadikan sebagai judul dengan langkah dilakukan mencari komentar pengguna pada Google Playstore.

Studi Literatur

Studi literatur adalah tahap dilakukannya pencarian terhadap jurnal bacaan terkait mengenai topik yang diklasifikasi.

Koleksi Data (Data Collection)

Koleksi data dilakukannya pengambilan data ulasan dari pengguna aplikasi bibit pada Google Playstore menggunakan modul *snsrape* yang terdapat pada Google Collab menggunakan bahasa pemrograman *Python*.

Perancangan Sistem

Perancangan sistem membuat tahap *preprocessing* menggunakan bahasa pemrograman *Python* pada Google Colabs menggunakan modul *NLTK* dan *Sastrawi Indonesia* dan membuat model menggunakan algoritma *Naive Bayes* menggunakan *software* RapidMiner Studio.

Validasi

Validasi dilakukannya tahapan untuk mengecek hasil yang dikeluarkan sistem apakah benar dan dipastikan telah sesuai dengan yang diinginkan berdasarkan penelitian terdahulu ataupun lebih baik dengan *Cross Validation* $k = 10$. *Cross Validation* merupakan cara mendapatkan nilai akurasi sebuah model yang dibuat menggunakan data yang ditentukan. Model yang digunakan biasanya digunakan untuk memprediksi atau klasifikasi model data tersebut [8].

Evaluasi

Evaluasi dilakukan penilaian terhadap hasil yang didapatkan berdasarkan sistem yang dirancang. Pada tahap ini digunakan metode Confusion Matrix yang akan menghitung *precision*, *recall* dan *accuracy*. *Confusion Matrix* merupakan metode untuk menghitung nilai akurasi pengujian yang melakukan perhitungan untuk mendapatkan nilai *Precision*, *Recall*, dan *Accuracy* [9].

HASIL DAN PEMBAHASAN

Koleksi Data (Data Collection)

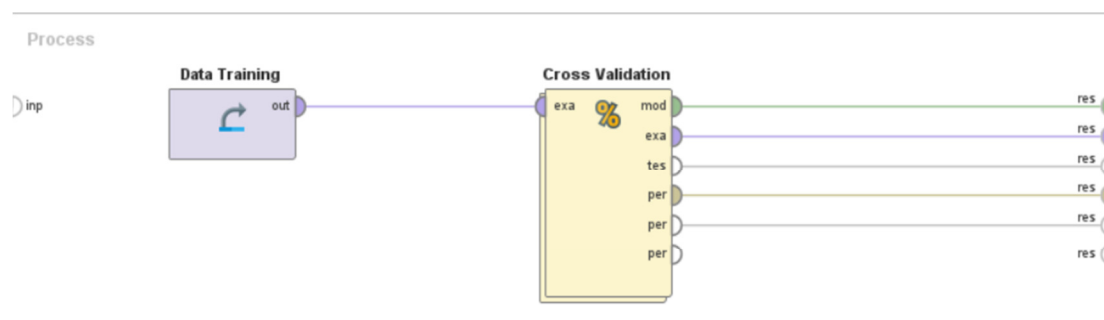
Pada tahap ini data yang didapat merupakan hasil *crawling* dari Google Playstore yang mana berasal dari kolom ulasan aplikasi Bibit yang dimana data yang diambil berjumlah 500 dataset yang kemudian diberi label positif dan negatif.

Tabel 1. Proses Labeling Data

Text	Sentimen
Aman rekomendasi bintang 5	Positif
Hasilnya jadi negative bintang 1	Negatif

Perancangan Sistem

Perancangan sistem dilakukan perancangan *preprocessing* data untuk untuk mengubah teks dari bahasa manusia menjadi format yang dapat dibaca mesin, kemudian membuat model menggunakan algoritma *Naive Bayes* untuk mengklasifikasi *accuracy* data [10].



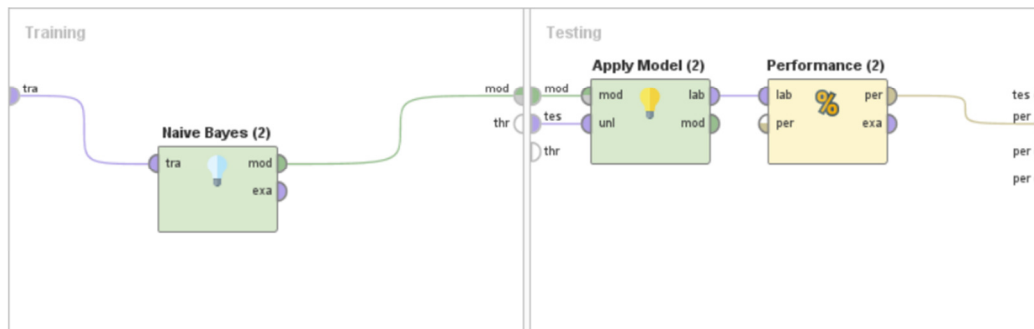
Gambar 2. Cross Validation

Pada gambar 2 diatas, data training yang telah dibagi menjadi akan dilanjutkan ketahapan *cross validation*, dimana *cross validation* bertujuan agar mendapatkan hasil yang lebih akurasi yang maksimal dalam penelitian ini.

Cross Validation juga merupakan teknik validasi dari pengembangan model *Split Validation* dimana validasinya mengukur training *error* dengan menguji dengan test data atau data uji. *Cross Validation* sendiri dikembangkan karena pada model sebelumnya terdapat kelemahan yaitu pengambilan sampelnya diambil

secara acak, kemudian pengambilan sampel test error nya juga tidak bisa mendistribusikan kelas secara terstruktur. Meskipun hasil yang di peroleh bisa maksimal tapi tidak bisa menjangkau pengujian yang lebih efisien.

Maka muncul lah *cross validation* yang mampu bekerja dengan cepat dengan pengambilan sampel yang lebih struktur, jadi dalam jumlah pengujian beberapa pun set data latih dan set data uji akan diambil dengan data yang berbeda dengan percobaan atau literasi sebelumnya.



Gambar 3. Model Sistem

Validasi

Pada tahap validasi dilakukan pembagian data dan dengan metode *Cross Validation* dengan k fold = 10 dilakukan klasifikasi menggunakan metode *Naïve Bayes* sebagai algoritma pembuatan model. Kemudian menggunakan *Confusion Matrix* untuk mendapatkan *Recall*, *Precision*, dan *Accuracy* terhadap model yang dibuat [11].



Gambar 4. Validasi *Cross Validation*

Evaluasi

Pada tahap evaluasi dilakukan pengujian menggunakan *confusion matrix* yang diperoleh dari tahapan modelling menggunakan algoritma *Naive Bayes* dimana pada evaluasi ini didapatkan nilai true positif dan true negatif [12].

Tabel 2. *Confusion Matrix*

	True Positive	True Negative	Class Precision
Pred. Positive	373	125	74.90%
Pred. Negative	0	2	100.00%
Class recall	100.00%	1.57%	

SIMPULAN

Berdasarkan Klasifikasi Opini Pengguna Aplikasi Bibit Pada Google Play Store Menggunakan Algoritma Naive Bayes didapat kesimpulan. (1) Berdasarkan komentar pengguna Bibit pada Google Play Store respon yang dari 500 data didapat data positif 373 dan 127 data negatif. Dikarenakan hasil tersebut didapat bahwa data positif lebih banyak dari pada data negatif, hal ini membuktikan bahwa aplikasi bibit terbukti aman secara OJK. (2) Berdasarkan model yang telah dibuat dengan data uji dan data latih yang dibuat dari 500 dataset, klasifikasi menggunakan metode algoritma *Naive Bayes* mampu memprediksi dengan cukup baik, ditunjukkan dengan nilai akurasi yang didapat sebesar 75%

DAFTAR PUSTAKA

- [1] A. Putra, D. Haeirudin, H. Khairunnisa, and R. Latifah, “Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma Svm,” *Semin. Nas. Sains dan Teknol.* 2021, No. November, pp. 1–6, 2021.
- [2] E. Fitri, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest dan Support Vector Machine,” *J. Transform.*, Vol. 18, No. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [3] H. Nuraliza, O. N. Pratiwi, and F. Hamami, “Analisis Sentimen IMBd Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance,” *J. Mirai Manaj.*, Vol. 7, No. 1, pp. 1–17, 2022.
- [4] N. Yolanda, I. H. Santi, D. Fanny, and H. Permadi, “Analisis Sentimen Popularitas Aplikasi Moodle dan Edmodo Menggunakan Algoritma Support Vector Machine,” Vol. 3, No. 1, pp. 48–59, 2022.
- [5] D. Alita and A. R. Isnain, “Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier,” *J. Komputasi*, Vol. 8, No. 2, pp. 50–58, 2020, doi: 10.23960/komputasi.v8i2.2615.
- [6] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, “Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB,” *J. Inform. Upgris*, Vol. 7, No. 1, pp. 6–11, 2021, doi: 10.26877/jiu.v7i1.7099.
- [7] N. Bahrawi, “Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based,” *J. Inf. Technol. Its Util.*, Vol. 2, No. 2, p. 29, 2019, doi: 10.30818/jitu.2.2.2695.
- [8] A. I. Tangraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naive Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, Vol. 9, No. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [9] A. K. Dewi, “Analisis Sentimen Ekspedisi Sicepat dari Ulasan Google Play Menggunakan Algoritma Naive Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, Vol. 9, No. 2, pp. 796–805, 2022, doi: 10.35957/jatisi.v9i2.1802.

- [10] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, Vol. 8, No. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [11] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment Analysis Of Social Media Twitter With Case of Anti-LGBT Campaign in Indonesia Using Naïve Bayes, Decision Tree, and Random Forest Algorithm,” *Procedia Comput. Sci.*, Vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.
- [12] D. D. Tran, T. T. S. Nguyen, and T. H. C. Dao, “Sentiment Analysis of Movie Reviews Using Machine Learning Techniques,” *Lect. Notes Networks Syst.*, Vol. 235, No. August, pp. 361–369, 2022, doi: 10.1007/978-981-16-2377-6_34.
- [13] Abdul Muiz Khalimi, “Pengujian Data Dengan Cross Validation,” 2023