

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/354802923>

Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit dan Bareksa dengan Algoritma KNN

Article in JATISI (Jurnal Teknik Informatika dan Sistem Informasi) · June 2021

DOI: 10.35957/jatisi.v8i2.962

CITATIONS

33

READS

708

2 authors, including:



[Safitri Juanita](#)

Universitas Budi Luhur

53 PUBLICATIONS 101 CITATIONS

SEE PROFILE

Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN

Aluisius Dwiki Adhi Putra^{*1}, Safitri Juanita²

Program Studi Sistem Informasi Fakultas Teknologi Informasi Universitas Budi Luhur
e-mail: ¹1712500162@student.budiluhur.ac.id, ²safitri.juanita@budiluhur.ac.id,

Abstrak

Investasi online merupakan kegiatan menanam modal baik langsung maupun tidak dengan harapan pada suatu waktu pemilik modal mendapatkan sejumlah keuntungan yang dilakukan secara online. Terdapat contoh aplikasi investasi online yang sudah banyak diunduh masyarakat menurut google play store yepenaitu bibit dan bareksa. Sehingga Tujuan penelitian ini adalah untuk menganalisa sentimen pada ulasan pengguna aplikasi investasi online yaitu bibit dan bareksa. Jumlah ulasan yang akan digunakan pada penelitian ini sebanyak 998 yang terdiri dari 484 sentimen positif dan 514 sentimen negatif untuk aplikasi bareksa sedangkan untuk aplikasi bibit menggunakan 1063 data yang terdiri dari 541 sentimen positif dan 522 sentimen negatif. Data tersebut juga melewati tahapan preprocessing dan modelling. Pada penelitian ini menggunakan model CRISP-DM (Cross Industry Standard Process for Data Mining) dan algoritma yang digunakan pada penelitian ini adalah K-Nearest Neighbors. Berdasarkan hasil yang diperoleh dari tahapan modelling dengan menggunakan algoritma k-nearest neighbors dan perbandingan 60:40 untuk data training dan data testing, maka nilai akurasi precision dan recall yang dihasilkan dari tiap aplikasi yaitu untuk bibit 85,14% , 91,91%, dan 76,44% sedangkan untuk bareksa yaitu 81,70% , 87,15%, 75,73%.

Kata kunci: analisis sentimen, ulasan, TF-IDF, knn, confusion matrix

Abstract

Online investment is an investment activity, either directly or indirectly, with the hope that at one time the owner of the capital will get a number of benefits made online. There is one example of an online investment application that has been widely used by the public according to the Google Play Store, namely bibit and bareksa. So the purpose of this study was to analyze sentiments on user reviews of online investment applications, namely bibit and bareksa. The number of reviews to be used in this study was 998 consisting of 484 positive sentiments and 514 negative sentiments for the bareksa application, while for the bibit application using 1063 data consisting of 541 positive sentiments and 522 negative sentiments. The data also goes through the preprocessing and modeling stages. This study also uses the CRISP-DM (Cross Industry Standard Process for Data Mining) model and the algorithm used in this study is K-Nearest Neighbors. Based on the results obtained from the modeling stage using the k-nearest neighbors algorithm and a comparison of 60:40 for training data and testing data, the accuracy and recall values generated from each application are 85.14%, 91.91% bibit. and 76.44%, while for bareksa, they were 81.70%, 87.15%, 75.73%.

Keywords: sentiment analysis, review, TF-IDF, knn, confusion matrix

1. PENDAHULUAN

Menurut HSBC [1] terdapat beberapa instrumen investasi yang cukup populer di Indonesia seperti deposito, emas, properti, saham, reksa dana, dan peer to peer lending. Badan Koordinasi Penanaman Modal (BKPM) melakukan pembukuan realisasi investasi sebesar 214,7 triliun pada triwulan IV (oktober – desember) tahun 2020, naik 3,1% dibandingkan periode yang sama pada tahun 2019 [2]. Berdasarkan data tersebut membuktikan bahwa minat masyarakat terhadap investasi meningkat sehingga mulai bermunculan investasi online. Menurut Namira [3] terdapat 7 aplikasi investasi online untuk pemula yaitu Bibit, Bareksa, HSB Investasi, Tanamduit, Landx, Ajaib, dan KoinWorks. Namun untuk memilih aplikasi investasi online yang terpercaya dibutuhkan kumpulan informasi yang didapat dari pengalaman pengguna aplikasi investasi online tersebut, karena menurut (Arnani, 2021) terdapat beberapa penipuan aplikasi investasi online. Pada penelitian ini hanya akan dibandingkan 2 aplikasi dari 7 aplikasi investasi online pemula menurut (Namira, 2020), yaitu aplikasi investasi online bernama bibit dan bareksa karena 2 aplikasi tersebut memiliki total pengguna bareksa sebesar 11 ribu dan bibit sebesar 85 ribu, total pengguna dua aplikasi tersebut lebih banyak dari 5 aplikasi lainnya berdasarkan data pada google play store serta aplikasi tersebut mempunyai fitur yang tidak jauh beda. Maka tujuan dari penelitian ini yaitu memberikan rekomendasi apakah aplikasi bibit atau bareksa yang lebih unggul dan aman untuk investasi online bagi masyarakat dengan cara analisis sentimen ulasan pengguna pada aplikasi bibit dan bareksa menggunakan algoritma k-nearest neighbors. Berdasarkan penelitian sebelumnya yang dilakukan oleh Siti dan Risa [5] dengan judul penelitian “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel” Dataset yang diperoleh pada penelitian ini dari laman yang berupa kumpulan opini seperti review agen travel, review restoran, review produk, dan lain-lain. Dan didapatkan sebanyak 100 review positif dan 100 review negatif. Hasil dari penelitian ini yang menggunakan algoritma k-nearest neighbors menghasilkan nilai akurasi 87,00% dan titik AUC adalah 0,916 titik AUC dan dinyatakan bahwa K-Nearest Neighbor memiliki hasil yang akurat dalam menganalisis sentiment ulasan aggen perjalanan. Berdasarkan penelitian sebelumnya yang dilakukan oleh Surohman dll [6] dengan judul penelitian “Analisa Sentimen Terhadap Review Fintech Dengan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor” Dataset yang diperoleh dari google play store pada aplikasi Dana. Pada penelitian ini pada aplikasi dana memiliki nilai akurasi terbaik sebesar 84,76% sehingga dapat disimpulkan bahwa review tersebut dapat meyakinkan dari review positif oleh pengguna lain dan untuk review negatif akan menjadi masukan kepada pihak perusahaan sehingga bisa mengembangkan produknya. Berdasarkan penelitian yang dilakukan oleh Indriya dll [7] dengan judul penelitian “Analisis Sentimen pada ulasan Aplikasi BCA Mobile menggunakan BM25 dan Improved K-Nearest Neighbor” Dataset yang diperoleh pada penelitian ini berasal dari google play store. Penelitian ini menggunakan metode BM25 dan dilakukan pengklasifikasian dokumen menggunakan *Improved K-Nearest Neighbor* dan hasil yang diperoleh dengan pengujian 5 fold mendapatkan nilai k values sebesar 10, dengan nilai precision recall dan akurasi sebesar 0,946, 0,934, dan 0,942. Hasil pengujian tersebut menghasilkan nilai yang fluktuatif dikarenakan pengaruh yang besar oleh nilai *k-values*.

2. LANDASAN TEORI

2.1 Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik machine learning untuk menganalisis dan mengekstraksi pengetahuan secara otomatis [8]. Data Mining merupakan tahapan untuk menemukan pola atau informasi dalam sekumpulan data dengan menggunakan teknik dan algoritma tertentu [9].

2.2 Analisis Sentimen

Analisis sentimen merupakan cara mengumpulkan pendapat khalayak umum menggunakan jejaring sosial yang didalamnya terdapat mengandung pelayanan umum serta isu terkini [10]. Analisis Sentimen adalah mengekstrak opini public tentang topik tertentu, produk atau jasa yang di dalam nya terkandung teks-teks yang tidak terstruktur [11]. Tujuan dari analisis sentimen merupakan untuk mengetahui opini seseorang ke arah positif atau negatif yang nantinya dapat digunakan untuk pengambilan keputusan.

2.3 Algoritma K-Nearest Neighbors

Algoritma K-Nearest Neighbors adalah penentu klasifikasi berdasarkan contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, tetapi bergantung pada label kategori yang melekat pada dokumen pelatihan mirip dengan dokumen tes [12]. Algoritma K-Nearest Neighbors adalah sebuah metode klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat. K-Nearest Neighbors memiliki prinsip sederhana, bekerja berdasarkan jarak terpendek dari sampel uji ke sampel latih [13]. Menurut Safitri [14] algoritma k-nearest neighbors memiliki tahapan:

1. Menentukan parameter k (jumlah tetangga paling dekat).
2. Tentukan bobot untuk setiap term dengan menggunakan Term Weighting TF-IDF.
3. Hitung kemiripan antar dokumen dengan menggunakan cosine similarity (Persamaan 2.1).

$$\cos(\theta_{ij}) = \frac{\sum_k (d_{ik} d_{jk})}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}}$$

4. Urutkan hasil perhitungan *cosine similarity* dari besar ke kecil
5. Ambil sebanyak K yang paling tinggi kemiripannya dengan dokumen yang diklasifikasikan, tentukan kelasnya.

2.4 Cross Validation Confusion Matrix

Cross Validation adalah salah satu teknik untuk menilai atau memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Pembuatan model biasanya bertujuan untuk melakukan prediksi maupun model biasanya bertujuan untuk melakukan prediksi maupun klasifikasi terhadap suatu data baru yang boleh jadi belum pernah muncul di dalam dataset [15].

Confusion matrix adalah sebuah metode yang biasa digunakan untuk perhitungan akurasi. Dalam pengujian keakuratan hasil pencarian akan dievaluasi nilai recall, precision, accuracy, dan error rate [16].

Tabel 1. Model *Confusion Matrix* [8]

		True Class	
		Positive	Negative
Predicted Class	Positive	True positive(TP)	False negative(FN)
Predicted Class	Negative	False positive (FN)	True negative(TN)

Keterangan :

TP (True Positive) = Jumlah prediksi yang benar dari data yang relevant.
FP (False Positive) = Jumlah prediksi yang salah dari data yang tidak relevant.
FN (False Negative) = Jumlah prediksi yang salah dari data yang tidak relevant.
TN (True Negative) = Jumlah prediksi yang benar dari data yang relevant.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

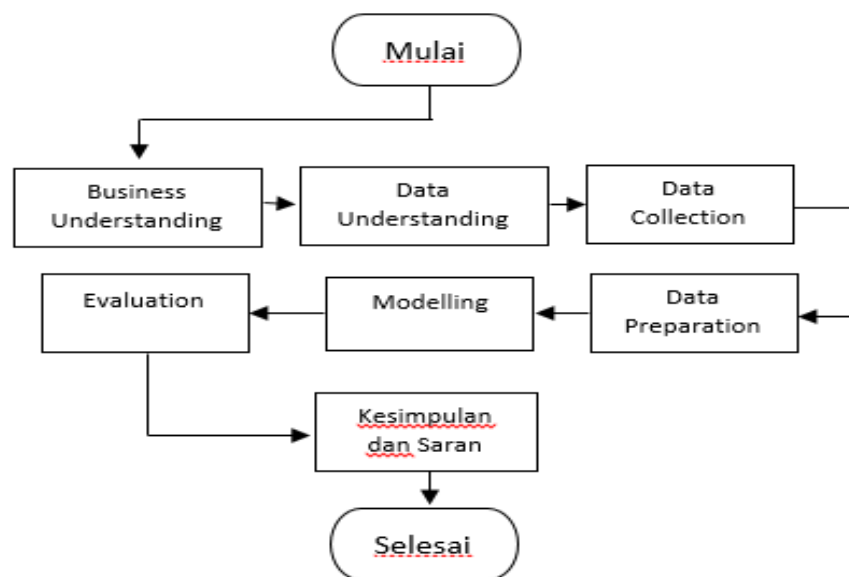
$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

2. 5 Cross-Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang di kembangkan tahun 1996 oleh analisis dari beberapa industry seperti Daimler Chrysler, SPSS dan NCR. CRISP-DM menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian [17]. CRISP-DM adalah metode yang dapat diterapkan ke dalam strategi pemecahan masalah umum serta metodologi yang menyediakan standar baku untuk data mining.

3. METODE PENELITIAN



Gambar 2. Tahapan Penelitian

Pada penelitian ini terdapat beberapa tahapan seperti yang ditunjukkan pada Gambar 3.1, penelitian ini menggunakan metode Cross-Industry Standard Process for Data Mining (CRISP-DM), dimana data akan diproses dengan melakukan beberapa tahapan seperti berikut:

1. Pemahaman Bisnis (Business Understanding)

Pada tahapan ini dilakukan penentuan permasalahan yang akan diangkat yaitu analisis sentimen terhadap ulasan pengguna pada aplikasi bibit dan bareksa, langkah yang dilakukan yaitu mencari data ulasan yang terdapat pada play store.

2. Pemahaman Data (Data Understanding)

Pada tahapan ini peneliti bertujuan untuk mengumpulkan, mengidentifikasi, dan memahami data yang dimiliki. Data tersebut juga harus dapat diverifikasi kebenarannya. Data yang digunakan dalam penelitian ini adalah data ulasan yang terdapat pada play store mengenai aplikasi investasi online yaitu bibit dan bareksa.

3. Koleksi Data (Data Collection)

Pada tahapan ini peneliti melakukan pengambilan data ulasan pada play store pada tanggal 21 April 2021. Sebelum melakukan pengambilan data peneliti harus mengkonfigurasi pada aplikasi web scraping agar pengambilan data bisa dilakukan, peneliti harus memilih data mana saja yang akan diambil seperti rating, tanggal, nama, dan yang terpenting ulasan. Proses pengambilan data ulasan yang dilakukan pada penelitian ini diambil dengan menggunakan aplikasi webharvy. Data yang diambil pada penelitian ini hanyalah data ulasan yang terdapat pada aplikasi bibit dan bareksa.

4. Persiapan Data (Data Preparation)

Pada tahapan ini data yang sudah didapatkan sebelumnya, akan diberikan sentimen. Setelah data yang telah didapatkan diberikan sentiment, setelah itu data akan masuk tahap data preprocessing. Berikut adalah tahapan-tahapan yang dilakukan pada tahapan data preprocessing:

- *Remove duplicates* merupakan proses untuk menghapus data yang sama, atau diambil secara berulang pada saat kita scraping.
- *Case Folding* merupakan proses untuk mengubah semua data menjadi huruf kecil.
- *Tokenizing* merupakan proses pemecahan sekumpulan karakter dalam suatu teks kedalam satuan kata, pada proses ini juga dapat menghilangkan karakter pembatas, menghapus angka, dan menghapus tanda baca atau karakter yang bukan merupakan huruf.
- *Stopwords* merupakan proses menghilangkan kata yang tidak memiliki makna atau tidak penting dengan menggunakan stopword list dari Tala [18]. Setelah menggunakan Stopword list tersebut, lalu ditambahkan lagi beberapa kata pada data kita seperti “kok, lah, dll” yang tidak memiliki makna pada penelitian ini.
- *Stemming* merupakan proses pencarian kata dasar dari tiap kata. Pada penelitian ini peneliti menggunakan sebuah file yang berisi imbuhan yang bertujuan untuk menghilangkan imbuhan untuk mendapatkan akar kata.
- *Term Weighting (TF-IDF)* merupakan proses untuk memberikan pembobotan kata dengan mencari nilai Term Frequency (TF), kemudian mencari nilai Document Frequency (DF), lalu mencari nilai Invers Document Frequency (IDF) setelah itu baru menghitung bobot.

5. Pemodelan (Modeling)

Pada tahapan ini akan dilakukan pemodelan terhadap data ulasan yang sudah dilakukan preprocessing. Data ulasan tersebut akan dipecah menjadi data training dan data testing, pada tahapan ini akan dipecah menjadi dua tahapan yaitu split data dan cross fold validation. Pada split data peneliti akan membagi data ulasan tersebut dengan perbandingan data training dan data testing yaitu 60:40, 70:30, dan 80:20. Lalu pada tahap cross fold validation kita akan membagi secara acak kedalam 10 bagian. Kemudian pada tahap pemodelan ini peneliti menggunakan algoritma k-nearest neighbors untuk mendapatkan nilai sentimen positif dan sentimen negatif dari sebuah data ulasan.

6. Evaluasi (Evaluation)

Pada tahapan ini akan dilakukan evaluasi metode klasifikasi dengan mengukur performa menggunakan confusion matrix terhadap algoritma k-nearest neighbors.

7. Kesimpulan dan Saran

Pada tahapan ini akan dilakukan pemberian kesimpulan yang telah didapat selama melakukan penelitian ini serta memberikan saran untuk penelitian lanjutan yang akan dilakukan oleh peneliti lain.

4. HASIL DAN PEMBAHASAN

4.1 Business Understanding

Pada tahapan business understanding ini maka peneliti mencoba untuk memahami objek atau masalah yang diangkat selama penelitian. Masalah yang diangkat oleh peneliti adalah seberapa besar akurasi yang dihasilkan dari penerapan algoritma k-nearest neighbors dalam mengklasifikasikan sentimen dari sebuah ulasan pengguna yang berkaitan dengan aplikasi investasi online yaitu bibit dan bareksa. Penelitian ini akan menggunakan data yang didapat melalui google play store.

4.2 Data Understanding

Pada tahapan data understanding ini maka peneliti mencoba untuk memahami data yang akan digunakan. Pada tahapan ini peneliti juga melakukan pengambilan data ulasan pengguna pada tanggal 3 April 2021 dan 23 April 2021 untuk aplikasi bibit dan bareksa, data yang diambil yaitu berupa nama, tanggal, rating, serta ulasan. Hasil crawling memperoleh data ulasan sebanyak 998 terdiri 484 sentimen positif dan 514 sentimen negatif untuk bareksa dan 1063 terdiri 541 sentimen positif dan 522 sentimen negatif untuk bibit. Dalam memberikan label pada ulasan peneliti menggunakan pelabelan berdasarkan bintang.

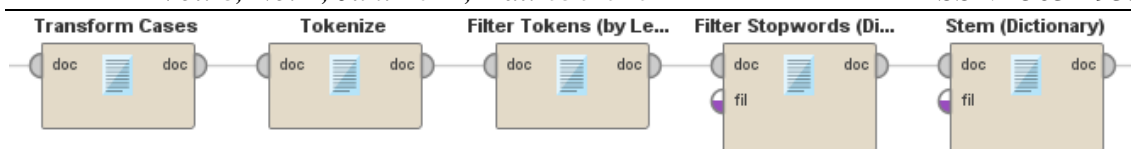
4.3 Data Collection

Pada tahapan data collection ini maka peneliti mengumpulkan data yang akan digunakan selama penelitian dan mempelajari apa yang dapat dilakukan dengan data tersebut. Pada tahapan ini peneliti mengumpulkan data ulasan pengguna dari google play store dengan menggunakan webharvy. Data yang digunakan pada penelitian ini menggunakan data pada 13 Februari – 20 April 2021 dan berbahasa Indonesia

4.3 Data Preparation

Pada tahapan data preparation ini maka data yang sudah terkumpul akan dilakukan data preprocessing oleh peneliti. Berikut ini adalah langkah-langkah yang dilakukan selama data preprocessing:

1. *Remove duplicates*, disini peneliti melakukan proses remove duplicates untuk menghapus data yang muncul secara berulang.
2. *Case folding*, disini peneliti melakukan proses case folding dengan menggunakan operator transform cases pada rapid miner untuk mengubah semua data ulasan menjadi huruf kecil.
3. *Tokenizing*, disini peneliti melakukan proses tokenizing untuk memecah ulasan dari kalimat menjadi perkata, menghilangkan karakter yang bersifat pembatas, dan menghapus angka atau tanda baca.
4. *Stopwords*, disini peneliti melakukan stopwords untuk menghilangkan kata yang tidak penting atau tidak deksriptif dengan menggunakan stoplist dari Tala.
5. *Stemming*, disini peneliti melakukan proses stemming untuk mencari akar kata dari sebuah token dengan membuat file untuk menghilangkan imbuhan atau mencari akar kata.



Gambar 2. PreProcessing Data

4.3 Modelling

Pada tahapan modelling ini, maka peneliti akan melakukan pengukuran performa klasifikasi dalam dua tahap:

1. Modelling menggunakan split data

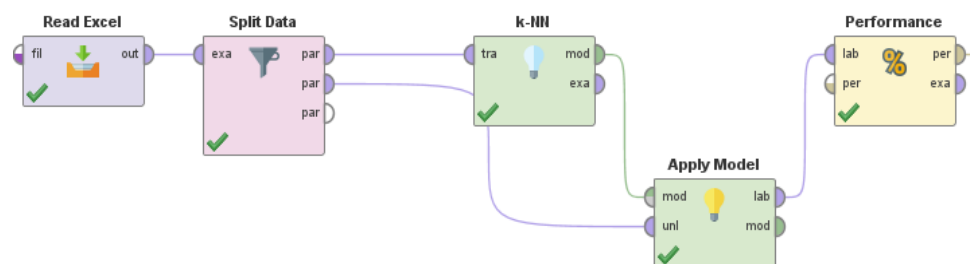
Pada tahapan ini maka data akan di split menjadi 3 model yaitu 60:40, dan 70:30

Tabel 2. Perbandingan *Accuracy* Berdasarkan Nilai K

Perbandingan 60:40	AKURASI KNN BIBIT	Perbandingan 60:40	AKURASI KNN BAREKSA
$K=1$	61.32%	$K=1$	59.90%
$K=3$	58.49%	$K=3$	57.39%
$K=5$	73.82%	$K=5$	77.44%
$K=7$	82.78%	$K=7$	81.70%
$K=9$	85.14%	$K=9$	81.70%

Tabel 3. Perbandingan *Accuracy* dengan Split Data

Dataset Bibit yang sudah dilakukan <i>preprocessing</i>				Dataset Bareksa yang sudah dilakukan <i>preprocessing</i>			
Perbandingan	Decision Tree	Naïve Bayes	KNN	Perbandingan	Decision Tree	Naïve Bayes	KNN
80 : 20	73.11%	78.30%	83.49%	80 : 20	65.50%	75.50%	70.00%
70 : 30	73.90%	77.36%	82.70%	70 : 30	68.56%	72.91%	79.26%
60 : 40	76.89%	75.71%	85.14%	60 : 40	69.76%	73.68%	81.70%
Dataset Bibit yang belum dilakukan <i>preprocessing</i>				Dataset Bareksa yang belum dilakukan <i>preprocessing</i>			
Perbandingan	Decision Tree	Naïve Bayes	KNN	Perbandingan	Decision Tree	Naïve Bayes	KNN
80 : 20	50.94%	50.94%	50.94%	80 : 20	51.50%	51.50%	51.50%
70 : 30	50.78%	50.78%	50.78%	70 : 30	51.51%	51.51%	51.51%
60 : 40	50.82%	50.82%	50.82%	60 : 40	51.50%	51.50%	51.50%



Gambar 3. Proses Modelling Split Data

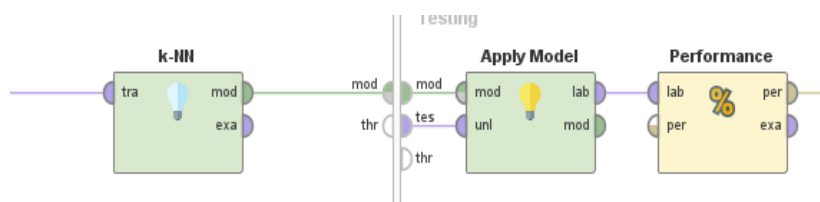
Berdasarkan Tabel 3 maka disimpulkan bahwa akurasi tertinggi sebesar 85,14% pada aplikasi bibit dan 81,70% pada aplikasi bareksa, yang dihasilkan dengan pemodelan terhadap *dataset* yang sudah dilakukan *preprocessing*, menggunakan algoritma *k-nearest neighbors*, dan menggunakan perbandingan 60:40 untuk data *training* dan data *testing*.

2. Modelling menggunakan cross validation

Pada tahapan ini maka data training dan data testing di split secara acak menjadi 10 bagian (number of folds = 10)

Tabel 4. Perbandingan Accuracy Dengan Cross Validation

Dataset Bibit yang sudah dilakukan <i>preprocessing</i>		Dataset Bareksa yang sudah dilakukan <i>preprocessing</i>	
Decision Tree	74.48%	Decision Tree	68.30%
Naïve Bayes	77.02%	Naïve Bayes	75.23%
KNN	79.94%	KNN	76.33%
Dataset Bibit yang belum dilakukan <i>preprocessing</i>		Dataset Bareksa yang belum dilakukan <i>preprocessing</i>	
Decision Tree	50.89%	Decision Tree	51.50%
Naïve Bayes	50.89%	Naïve Bayes	51.50%
KNN	50.89%	KNN	51.50%



Gambar 4. Proses Modelling Cross Validation

Berdasarkan Tabel 3 dan Tabel 4 yang berisi hasil pengukuran performa model klasifikasi *k-nearest neighbors* dengan 2 cara yaitu pembagian *dataset* dengan *split data* dan *cross validation*, maka dapat disimpulkan bahwa pemodelan dengan *split data* 60:40 memiliki akurasi sebesar 81,70% pada aplikasi bareksa dan 85,14% pada aplikasi bibit sedangkan pemodelan dengan *cross validation* memiliki akurasi sebesar 76,33% pada aplikasi bareksa dan 79,94% pada aplikasi bibit sehingga akurasi tertinggi yang dihasilkan pada penelitian ini adalah dengan pemodelan menggunakan *split data* 60:40.

4.4 Evaluation

Penelitian ini akan melakukan pengujian dengan menggunakan confusion matrix yang diperoleh dari tahapan modelling dengan menggunakan algoritma *k-nearest neighbors* dimana pada Sub bab 4.3 akurasi dengan split data 60:40 lebih tinggi dibandingkan cross validation. Sehingga confusion matrix yang disajikan pada Tabel 5 adalah hasil dari evaluasi pengukuran klasifikasi *k-nearest neighbors* dengan dataset yang berjumlah 4362, terdapat 3489 data training, dan 873 data testing.

Tabel 5. Confusion Matrix Aplikasi Bibit

	True positive	True negative	Class precision
Pred. positive	159	14	91.91%
Pred. negative	49	202	80.48%
Class recall	76.44%	93.52%	

Perhitungan dari Tabel 5 adalah sebagai berikut:

- $Precision = \frac{TP}{TP+FP} = \frac{159}{159+14} = 0.9191 = 91.91\%$
- $Recall = \frac{TP}{TP+FN} = \frac{159}{159+49} = 0.7644 = 76.44\%$
- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{159+202}{159+202+14+49} = 0.8514 = 85.14\%$

Tabel 6. Confusion Matrix Aplikasi Bareksa

	True positive	True negative	Class precision
Pred. positive	156	23	87.15%
Pred. negative	50	170	77.27%
Class recall	75.73%	88.08%	

Perhitungan dari Tabel 6 adalah sebagai berikut:

- $Precision = \frac{TP}{TP+FP} = \frac{156}{156+23} = 0.8715 = 87.15\%$
- $Recall = \frac{TP}{TP+FN} = \frac{156}{156+50} = 0.7573 = 75.73\%$
- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{156+170}{156+170+23+50} = 0.8170 = 81.70\%$

5. KESIMPULAN

Berdasarkan hasil yang didapat setelah melakukan penelitian ini, maka dapat diambil kesimpulan bahwa berdasarkan proses klasifikasi model dengan algoritma *k-nearest neighbors*, menggunakan model split data dengan perbandingan 60:40 terhadap dataset ulasan pengguna pada aplikasi bibit dan bareksa, maka nilai *accuracy*, *precision*, dan *recall* yang dihasilkan sebesar 85,14% , 91,91%, dan 76,44% untuk bibit sedangkan untuk bareksa yaitu 81,70% , 87,15%, 75,73%. Pada aplikasi bareksa opini yang didapatkan untuk setiap kelas menghasilkan diantaranya pada kelas negatif yaitu kurangnya pilihan pembayaran yang tersedia, sulitnya dalam melakukan registrasi terutama dalam verifikasi akun yang memakan cukup banyak waktu dan ketidak bermanfaat aplikasi setelah diupdate ke versi yang baru. Pada kelas positif menghasilkan tanggapan pelanggan yang merasa puas karena kemudahan saat menggunakan aplikasi serta daftar investasi yang cukup lengkap. Pada aplikasi bibit opini yang didapatkan untuk setiap kelas negatif yaitu dalam hal pencairan dana dan registrasi yang memakan cukup banyak waktu sehingga masyarakat harus menunggu dalam melakukan pencairan serta registrasi sedangkan untuk kelas positif yaitu masyarakat cukup dibantu dengan adanya robo advisor yang dipunyai oleh bibit, masyarakat juga dibuat nyaman dengan kemudahan serta fitur yang cukup lengkap pada aplikasi tersebut.

6. SARAN

Dari kesimpulan diatas dapat diberikan saran-saran sebagai berikut:

1. Penulis sangat menyarankan agar dilakukan penelitian atau pengembangan lebih lanjut terhadap penelitian ini dengan metode lainnya dan juga menggunakan dataset yang lebih baru dan lebih banyak dari penelitian sebelumnya.
2. Hasil dari proses analisis sentiment ini dapat digunakan sebagai pertimbangan masyarakat dalam mengambil keputusan lebih lanjut dalam memilih aplikasi investasi online yang tersedia dalam google play store.

UCAPAN TERIMA KASIH

Peneliti mengucapkan kepada semua pihak yang telah membantu demi terlaksananya penelitian ini yang tidak bisa peneliti sebutkan namanya satu persatu.

DAFTAR PUSTAKA

- [1] HSBC, "Article: HSBC," 2019. [Online]. Available: https://www.hsbc.co.id/1/PA_esf-ca-app-content/content/indonesia/personal/offers/news-and-lifestyle/files/articles/html/201906/jenis-jenis-investasi-yang-populer-di-indonesia.html.
- [2] BKPM, "Publikasi : BKPM," January 2021. [Online]. Available: <https://www.bkpm.go.id/id/publikasi/siaran-pers/readmore/2413001/68601>.
- [3] Namira, "tech-trend," 11 Agustus 2020. [Online]. Available: <https://www.idntimes.com/tech/trend/izza-namira-1/daftar-aplikasi-investasi-terbaik-untuk-pemula/7>.
- [4] M. Arnani, "tren: Kompas," 7 Maret 2021. [Online]. Available: <https://www.kompas.com/tren/read/2021/03/07/200000665/waspada-ini-daftar-28-investasi-bodong-yang-disetop-ojk?page=all>.
- [5] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, Vol. VI, No. 1, pp. 64–69, 2018, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/2626>.
- [6] S. Surohman, S. Aji, R. Rousyati, and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier dan K- Nearest Neighbor," *EVOLUSI J. Sains dan Manaj.*, Vol. 8, No. 1, pp. 93–105, 2020, doi: 10.31294/evolusi.v8i1.7535.
- [7] I. D. Onantya, Indriati, and P. P. Adikara, "Analisis Sentimen pada Ulasan Aplikasi BCA Mobile Menggunakan BM25 dan Improved K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 3, No. 3, pp. 2575–2580, 2019.
- [8] A. Syakuro, "Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG) Halaman Judul Skripsi Oleh: Abdan Syakuro," *Anal. Sentimen Masy. Terhadap E-Commerce pada Media SOS. Menggunakan Metod. Naive Bayes Classif. Dengan Sel. fitur Inf. gain*, pp. 1–89, 2017.

-
- [9] Ramadina, "Penerapan Fungsi Data Mining Klasifikasi Untuk Prediksi Masa Studi Mahasiswa Tepat Waktu Pada Sistem Informasi Akademik Perguruan Tinggi," *JUPITER (Jurnal Penelit. Ilmu dan Teknol. Komputer)*, Vol. 7, No. 1, pp. 39–50, 2015.
- [10] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 pada Twitter Menggunakan Metode Naïve Bayes dan Knn," *Inti Nusa Mandiri*, Vol. 15, No. 1, pp. 23–28, 2020.
- [11] S. Saidah and J. Mayary, "Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based dan K – Nearest Neighbor," *J. Ilm. Inform. Komput.*, Vol. 25, No. 1, pp. 1–17, 2020, doi: 10.35760/ik.2020.v25i1.2411.
- [12] R. P. Fitrianti, A. Kurniawati, and D. Agusten, "Analisis Sentimen Terhadap Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma K-Nearest Neighbor," *Semin. Nas. Apl. Teknol. Inf.*, pp. 27–32, 2019.
- [13] R. Sari, "Analisis Sentimen pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn)," *EVOLUSI J. Sains dan Manaj.*, Vol. 8, No. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.
- [14] R. N. Safitri, "Analisis Sentimen Review Pelanggan Hotel Menggunakan Metode K-Nearest Neighbor (K-NN) (Studi Kasus: Hotels.com, Booking.com, Agoda.com)," *Tugas Akhir Univ. Din.*, Vol. 21, No. 1, pp. 1–9, 2020, [Online]. Available: <http://repositorio.unan.edu.ni/2986/1/5624.pdf>.
- [15] A. Imron, "Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier," 2019, [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/14268>.
- [16] R. Melita, V. Amrizal, H. B. Suseno, and T. Dirjam, "Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) dan Cosine Similarity pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Syarah Umdatil Ahkam)," *J. Tek. Inform.*, Vol. 11, No. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.
- [17] M. F. Rifai, H. Jatnika, and B. Valentino, "Penerapan Algoritma Naïve Bayes pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *Petir*, Vol. 12, No. 2, pp. 131–144, 2019, doi: 10.33322/petir.v12i2.471.
- [18] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *M.Sc. Thesis, Append. D*, Vol. pp, pp. 39–46, 2003.