

KOMPARASI ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR PADA ANALISIS SENTIMEN TERHADAP ULASAN PENGGUNA APLIKASI TOKOPEDIA

Ryfan Maulana Putra Hertaryawan ^{a,1,*}, Muhammad Raihan ^{b,2}, Imam Santoso ^{c,3}

^{a,b} Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jl Radin Inten II Duren Sawit Seberang Pengadilan Tinggi Agama DKI Jakarta, Jakarta Timur

^c Universitas Teknologi Muhammadiyah Jakarta, Jl. Minangkabau No. 60 – Manggarai Jakarta Selatan

¹ Ryfan@stikomcki.ac.id; ² muhammadraihan@stikomcki.ac.id; ³ imam.santoso@utmj.ac.id

* corresponding author

ARTICLE INFO

ABSTRACT

Keywords

Sentiment Analysis, Naive Bayes, K-Nearest Neighbor, Particle Swarm Optimization, Tokopedia.

Tokopedia is one of the leading e-commerce platforms in Indonesia. The use of e-commerce platforms has increased rapidly in recent years. This is due to technological advances, increased internet access, and consumer behavior that prefers to shop online. The purpose of this research is to conduct sentiment analysis on application performance based on user reviews of the Tokopedia application. Researchers made the decision to use sentiment analysis because it is the most suitable method for processing data sets. From 1019 Tokopedia user reviews on the Play Store that were collected, 176 positive reviews and 843 negative reviews were obtained. Then, the data is classified using the Naive Bayes and K-Nearest Neighbor algorithms, then optimized using Particle Swarm Optimization (PSO). The results of the research conducted obtained an accuracy of 76.30% for the Naive Bayes accuracy value without PSO, 74.09% for Naive Bayes results using PSO. Then the accuracy value obtained for K-Nearest Neighbor without PSO is 83.10%, and with PSO is 83.53%. From the results, the effect of using Particle Swarm Optimization on the two algorithms does not have a big impact, there is an insignificant change in accuracy and AUC values which in the Naive Bayes algorithm actually decreases.

1. Pendahuluan

Di Indonesia, penggunaan platform e-commerce telah meningkat pesat dalam beberapa tahun terakhir. Hal ini disebabkan oleh kemajuan teknologi, peningkatan akses internet, dan perilaku konsumen yang lebih suka berbelanja secara online. Oleh karena itu, penting bagi bisnis e-commerce untuk memahami pengalaman pelanggan mereka saat berbelanja di platform tersebut.

Tokopedia adalah salah satu platform e-commerce terkemuka di Indonesia. Didirikan pada tahun 2009, Tokopedia telah menjadi salah satu tujuan utama bagi konsumen Indonesia untuk berbelanja secara online. Tokopedia telah menjadi topik yang penting dalam konteks industri e-commerce di Indonesia. Sebagai salah satu platform e-commerce terbesar dan terkemuka di negara ini, Tokopedia memiliki jutaan pengguna yang aktif setiap harinya. Platform ini menawarkan berbagai kategori produk, mulai dari fashion, elektronik, kebutuhan sehari-hari, hingga properti dan layanan.

Pada era digital saat ini, ulasan pengguna memiliki peran yang semakin penting dalam membentuk persepsi konsumen terhadap suatu produk atau layanan. Di Indonesia, Tokopedia telah menjadi salah satu platform e-commerce terbesar dan paling populer, dengan jutaan pengguna aktif. Ulasan pengguna Tokopedia yang diterima melalui Play Store memberikan wawasan berharga tentang kepuasan pengguna, masalah yang mungkin dihadapi, dan kualitas pelayanan yang ditawarkan oleh Tokopedia. Oleh karena itu, analisis sentimen terhadap ulasan pengguna Tokopedia pada Play Store menjadi relevan untuk mendapatkan pemahaman yang lebih mendalam tentang kepuasan pengguna dan masalah yang mungkin perlu diperbaiki.

Pada penelitian analisis sentimen sebelumnya yaitu :

1. Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbor dimana penelitian tersebut bertujuan untuk melakukan

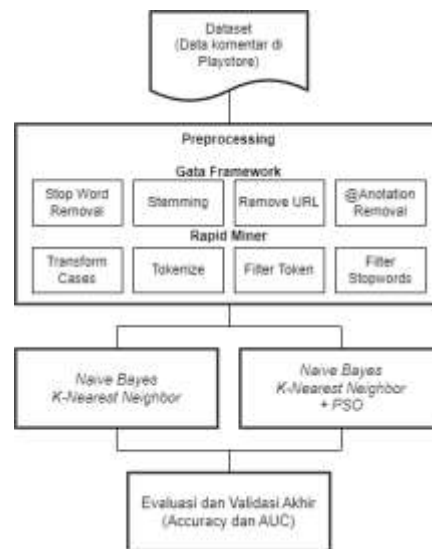
- sentimen analisis atas performa aplikasi yang didasarkan pada komentar pengguna aplikasi dengan menggunakan dataset sebanyak 992 komentar. Hasil yang didapat pada penelitian tersebut dengan algoritma naïve bayes mendapatkan tingkat accuracy sebesar 75,30% dan dengan algoritma k-nearest neighbor mendapatkan tingkat accuracy sebesar 86,09%.
2. Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier dimana penelitian tersebut bertujuan untuk membuktikan analisis dan pengujian komentar pelanggan. Performa yang dihasilkan dari hasil percobaan yang dilakukan oleh rapid evaluator pada 143 data pengujian menghasilkan nilai akurasi 95,10%, dengan class recall: 91,46% (positive class: Positive) sedangkan class precision: 100.00% (positive class: Positive) dan nilai AUC: 0.999 (positive class: Positive).
 3. Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia dimana penelitian tersebut bertujuan untuk mengklasifikasikan komentar pengguna kategori positif dan negatif. Hasil Pengujian berdasarkan nilai class negative, class positive, recall, dan accuracy pada analisis sentimen. dengan nilai accuracy performance yang baik sebesar 97,13%, dengan nilai precision 1 Sementara pada Class Recall dihasilkan nilai 95,49% (positive class: negative). Dan nilai AUC 0,980.
 4. K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia dimana penelitian tersebut bertujuan untuk mengklasifikasikan persepsi pelanggan berdasarkan kelas positif, netral, dan negatif. Hasil yang di dapatkan yaitu 88,11%. Seleksi fitur yang digunakan untuk mendapatkan hasil dengan menggunakan PSO. PSO menggunakan 20 iterasi dan 10 partikel. Hal ini menghasilkan akurasi evaluasi terbaik sebesar 97,9%, presisi 96,17%, recall 96,62%, dan f-measure 96,39%.

Pada penelitian ini yang membedakan penelitian ini dengan penelitian sebelumnya adalah bahwa pada penelitian ini menggunakan 2 algoritma yaitu menggunakan Naive Bayes dan K-Nearest Neighbor, karena pada penelitian sebelumnya hanya menggunakan 1 algoritma. Penggunaan Particle Swarm Optimization (PSO) sebagai seleksi fitur yang diterapkan untuk meningkatkan performa klasifikasi teks pada penelitian sebelumnya hanya pada 1 Algoritma sedangkan penelitian ini pada 2 algoritma. Pengambilan data dalam penelitian ini menggunakan Google Collab sebagai tempat untuk data scraping. Data yang diambil berupa data terbaru ulasan pengguna aplikasi tokopedia tahun 2023.

Penelitian ini diharapkan dapat melakukan analisis sentimen terhadap kinerja aplikasi berdasarkan ulasan pengguna terhadap aplikasi Tokopedia dan Penelitian ini diharapkan akan menghasilkan hasil yang lebih baik dari studi sebelumnya. Selain itu, diharapkan bahwa penelitian ini akan menentukan apakah algoritma Naive Bayes dan K-Nearest Neighbor memiliki hasil yang lebih baik.

2. Metodologi Penelitian

Berikut adalah langkah – langkah metode penelitian yang dijelaskan pada gambar dibawah ini :



Gambar 1. Metode Penelitian

2.1. Pengumpulan dan Pengolahan Data

Penelitian ini akan menggunakan data yang didapat dari penilaian di Play Store. Data ini diambil dari komentar yang ada di aplikasi Tokopedia. Data ini merupakan data terbaru yaitu data pada tahun 2023, berikut source link komentar aplikasi Tokopedia di Play Store <https://play.google.com/store/apps/details?id=com.tokopedia.tkpd&hl=id>. Proses pengambilan data dilakukan dengan menggunakan Data Scraping pada tanggal 2 Juni 2023, data yang digunakan dalam penelitian ini sebanyak 1019 data yang terdiri dari 176 data review positif dan 843 data review negatif berdasarkan nilai pengguna.

Setelah pengumpulan data, dilakukan proses pengolahan data, juga dikenal sebagai preprocessing. Proses ini mencakup pekerjaan membangun data dan pekerjaan pembersihan data untuk memungkinkan pengelolaan tahap selanjutnya. Tahapan preprocessing termasuk:

2.1.1 Stop Word Removal

Membuang kata: kata-kata yang biasanya terdiri dari kata sambung dan keterangan yang diabaikan oleh analis.

2.1.2 Stemming

Dengan menghilangkan imbuhan dari awalan dan akhiran, tahapan ini berfungsi untuk mengubah kata-kata menjadi kata dasar.

2.1.3 Remove URL

Proses menghilangkan Uniform Resource Locator yang ada pada dataset hasil pengumpulan data. Ini biasanya digunakan untuk merujuk pada sumber atau sumber informasi atau berita.

2.1.4 Annotation Removal

Menghapus tanda "Mention" dan teks di belakangnya. Biasanya berfungsi untuk menghubungi pengguna lain di media sosial.

2.1.5 Transform Cases

Mengubah semua huruf kapital dalam dataset menjadi huruf kecil. Tujuannya adalah untuk menjaga keseragaman teks pada model klasifikasi dan mencegah kesalahan dalam proses tokenisasi.

2.1.6 Tokenize

Tokenisasi adalah proses memotong sejumlah karakter atau kalimat menjadi potongan-potongan karakter atau kata-kata sesuai kebutuhan.

2.1.7 Transform (Not Negative).

Mengubah kata berarti menggabungkan kata negatif dengan tanda garis bawah ().

2.1.8 Filter Tokens (By Length)

Menghapus kata—kata yang memiliki panjang karakter tertentu, biasanya hanya dua karakter, tidak memiliki arti.

Pengolahan awal dilakukan menggunakan gata framework, pengolahan data teks berbasis web yang dikodekan menggunakan bahasa pemrograman php, yang dapat ditemukan di www.gataframework.com/textmining. Gata Framework memiliki banyak fitur, termasuk penghapusan annotation dan penghapusan URL. Gata Framework dipilih oleh peneliti karena memiliki keunggulan dalam melakukan proses stemming dalam bahasa Indonesia. Data kemudian diproses dengan alat Rapidminer.

2.2. Algoritma

Algoritma yang akan digunakan dan dijadikan perbandingan atau komparasi dalam penelitian ini adalah algoritma Naïve Bayes dan algoritma K-Nearest Neighbor. Peneliti menggunakan algoritma Naïve Bayes dan algoritma K-Nearest Neighbor karena keduanya merupakan algoritma yang baik dalam data mining, dan sangat baik dalam melakukan proses pengklasifikasian data.

2.3. Evaluasi dan Validasi

Pada langkah pengujian metode, peneliti menggunakan RapidMiner untuk menjalankan eksperimen. Data pelatihan yang digunakan adalah komentar aplikasi Tokopedia di Play Store tentang hal-hal baik dan buruk tentang aplikasi Tokopedia. Kemudian data dibagi menjadi dua kelompok: opini positif dan negatif.

Untuk validasi, peneliti menggunakan 10 Fold Cross Validation. Selanjutnya, untuk mengukur akurasi, peneliti menggunakan Confusion Matrix, yang membandingkan akurasi Naive Bayes dengan K-Nearest Neighbor. Selanjutnya, nilai AUC diukur dengan Kurva ROC.

2.3.1 Confusion Matrix

Metrik yang menginformasikan hasil prediksi secara keseluruhan dari nilai akurasi dan untuk melihat kinerja pengklasifikasi, yaitu seberapa sering kasus class X yang benar diklasifikasikan sebagai class X atau kesalahan klasifikasi class lainnya. Dalam kasus di mana dataset hanya memiliki dua kelas—kelas positif dan negatif—maka dapat dibuatkan seperti gambar berikut:

Jumlah record positif dalam dataset yang diklasifikasikan positif disebut sebagai positif positif (TP), negatif positif (TN), negatif negatif (fn), dan false negatif (fn) didefinisikan sebagai jumlah record negatif dalam dataset yang diklasifikasikan negatif

Tabel 1. Tabel Confusion Matrix

Correct Classification	Classified As	
	+	-
+	True Positive	False Negative
-	False Positive	True Negative

2.3.2 Kurva ROC

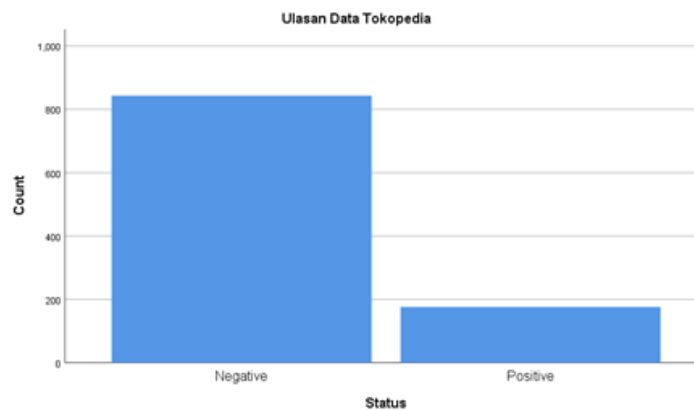
Kurva ROC adalah grafik dua dimensi yang menunjukkan confusion matrix dan menunjukkan akurasi dan perbandingan klasifikasi secara visual. Kurva ROC digambarkan sebagai garis horizontal false positif dan garis vertikal true positif. Metode umum untuk mengklasifikasikan keakuratan pemeriksaan dengan AUC [4].

- 0.90 - 1.00 = Excellent Classification;
- 0.80 - 0.90 = Good Classification;
- 0.70 - 0.80 = Fair Classification;
- 0.60 - 0.70 = Poor Classification;
- 0.50 - 0.60 = Failure

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Hasil pengumpulan data yang didapat dari komentar aplikasi Tokopedia di Play Store mengenai komentar aplikasi Tokopedia, peneliti mengumpulkan 1019 data opini pengguna Tokopedia. Kemudian data tersebut dikumpulkan dalam format file .xls, setelah itu data dikelompokkan menjadi dua bagian yaitu data positif dan data negatif. Pengelompokan data ini dilakukan oleh beberapa koresponden, sehingga dapat dihasilkan 176 opini positif dan 843 opini negatif.



Gambar 2. Statistik Hasil Pengumpulan Data

3.2. Pengolahan Data

Setelah data dikumpulkan dan dibagi menjadi komentar positif dan negatif, proses pengolahan data, juga dikenal sebagai preprocessing, dimulai. Berikut adalah tahapan preprocessing :

3.2.1 Stop Word Removal

Membuang kata: kata-kata yang biasanya terdiri dari kata sambung dan keterangan yang diabaikan oleh analisis.

Tabel 2. Tabel Hasil *Stop Word Removal*

Sebelum	Sesudah
Cukup baik, tampilan aplikasi bagus dan ringan. Hanya saja saat ini sudah tidak bisa menggunakan double promo seperti sebelumnya	tampilan aplikasi bagus ringan double promo

3.2.2 Stemming

Dengan menghilangkan imbuhan dari awalan dan akhiran, tahapan ini berfungsi untuk mengubah kata-kata menjadi kata dasar.

Tabel 3. Tabel Hasil *Stemming*

Sebelum	Sesudah
Dengan menghilangkan imbuhan dari awalan dan akhiran, tahapan ini berfungsi untuk mengubah kata-kata menjadi kata dasar.	gratis ongkir sekarang satu pesan estafet beberapa expedisi makin lama sampe dulu cukup satu expedisi bisa pilih yg paling cepat dan dekat dari rumah kecewaaa

3.2.3 *Remove URL*

Proses menghilangkan Uniform Resource Locator yang ada pada dataset hasil pengumpulan data. Ini biasanya digunakan untuk merujuk pada sumber atau sumber informasi atau berita.

Tabel 4. Tabel Hasil *Remove URL*

Sebelum	Sesudah
daripada tokopedia juga mending https://youtu.be/iYPELEaHi3U kalo kaya gini gimana tuh	Nih kompas juga bilang mirip kalo kaya gini gimana tuh

3.2.4 *Annotation Removal*

Menghapus tanda "Mention" dan teks di belakangnya. Biasanya berfungsi untuk menghubungi pengguna lain di media sosial.

Tabel 5. Tabel Hasil *Annotation Removal*

Sebelum	Sesudah
daripada tokopedia juga mending https://youtu.be/iYPELEaHi3U kalo kaya gini gimana tuh	Nih kompas juga bilang mirip kalo kaya gini gimana tuh

3.2.5 *Transforms Cases*

Mengubah semua huruf kapital dalam dataset menjadi huruf kecil. Tujuannya adalah untuk menjaga keseragaman teks pada model klasifikasi dan mencegah kesalahan dalam proses tokenisasi.

Tabel 6. Tabel Hasil *Transforms Cases*

Sebelum	Sesudah
Mau belanja di tokopedia gak jd krn ada biaya layanan 1000 dan biaya jasa aplikasi 2000-3000. Besok2 mau ditambah biaya apalagi lg ya sm tokpedia	mau belanja di tokopedia gak jd krn ada biaya layanan 1000 dan biaya jasa aplikasi 2000-3000. besok2 mau ditambah biaya apalagi lg ya sm tokpedia

3.2.6 *Tokenize*

Tokenisasi adalah proses memotong sejumlah karakter atau kalimat menjadi potongan-potongan karakter atau kata-kata sesuai kebutuhan.

3.2.7 *Transform (Not Negative)*

Mengubah kata berarti menggabungkan kata negatif dengan tanda garis bawah (_).

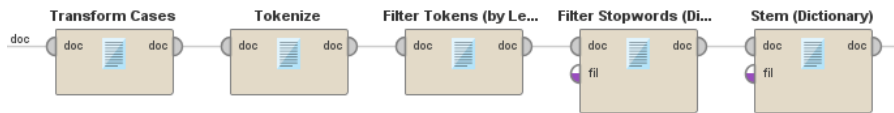
Tabel 7. Tabel Hasil *Transform Negative*

Sebelum	Sesudah
Aplikasi tidak optimal di tablet saya.. banyak fitur yg tidak muncul di tablet termasuk untuk tarik saldo ke rekening, jadi agak menyulitkan	aplikasi tidak_optimal di tablet saya.. banyak fitur yg tidak_muncul di tablet termasuk untuk tarik saldo ke rekening, jadi agak menyulitkan

3.2.8 *Filter Tokens (By Length)*

Menghapus kata—kata yang memiliki panjang karakter tertentu, biasanya hanya dua karakter, tidak memiliki arti.

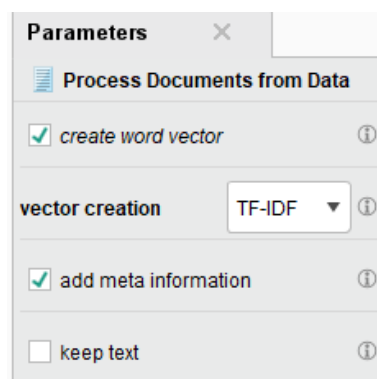
Berikut gambar proses pengolahan data pada tools rapidminer



Gambar 3. Proses Pengolahan Data Rapidminer

Proses pembobotan kata akan digunakan untuk mengubah data preprocessing yang berupa kata ke dalam bentuk angka. Proses ini menentukan seberapa berat setiap kata yang akan digunakan sebagai fitur, dan jumlah fitur yang dihasilkan terkait dengan jumlah dokumen yang akan diproses. Ada dua komponen proses pada tahap ini: TF (Frekuensi Terma) dan IDF (Frekuensi Inverse Dokumen).

Nilai TF adalah jumlah tiap kata yang muncul pada sebuah dokumen, dan nilai Df adalah jumlah nilai dokumen pada tiap kata, yang berbanding terbalik, yaitu nilai IDF lebih besar apabila suatu kata jarang muncul pada sebuah dokumen daripada apabila suatu kata sering muncul.

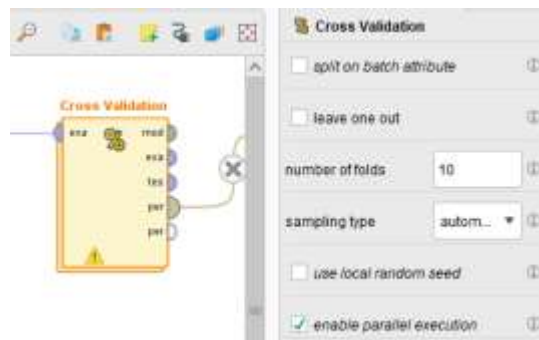


Gambar 4. Proses Tf-Idf

3.3. Proses Validasi Algoritma

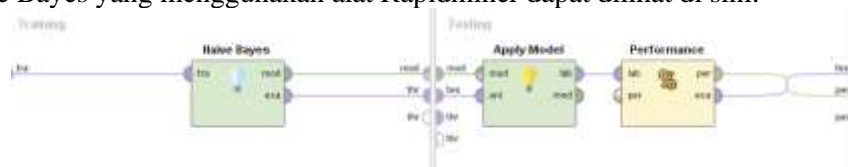
Selanjutnya adalah penggunaan algoritma Naïve Bayes dan K-Nearest Neighbor dilakukan secara otomatis dengan menggunakan aplikasi Rapidminer. Tujuan dari tahapan ini adalah untuk menghasilkan cross-validation dan confusion matrix sebagai proses validasi dimana evaluasi dilakukan dengan membagi data ke dalam dua kelas. Dalam proses ini, bobot setiap kata pada data *training* dibandingkan dengan kata pada data *testing*. Hasil dari setiap dokumen *training* adalah jumlah dari bobot kata probabilitas positif dan probabilitas negatif, kemudian bobot dokumen dibandingkan, jika bobot dokumen probabilitas positif lebih besar maka hasil sentimennya positif. Sebaliknya, jika probabilitas negatif lebih besar maka hasil sentimennya negatif.

Penggunaan Algoritma Naïve Bayes dan K-Nearest Neighbor untuk klasifikasi data dihubungkan dengan pengujian cross-validation sepuluh kali lipat dimana pengujian ini mengevaluasi proses kerja algoritma dengan membagi data secara acak ke dalam *10-fold* untuk mendapatkan sepuluh data yang sama, kemudian digunakan *9-fold* untuk data *training* dan *1-fold* untuk data *testing*



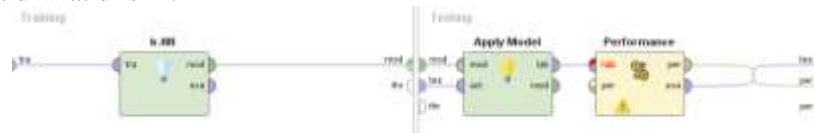
Gambar 5. Proses Cross Validation dengan Rapidminer

Pada tahap selanjutnya, proses evaluasi dilakukan untuk mengevaluasi kinerja dari permodelan dan perhitungan yang telah digunakan. Ini dilakukan dengan menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor. Cross Validation digunakan untuk mengevaluasi kinerja model algoritma yang dipilih sehingga dapat menghasilkan nilai akurasi. Proses cross validation dapat dilihat pada Model pengujian Naïve Bayes yang menggunakan alat Rapidminer dapat dilihat di sini.



Gambar 6. Model Pengujian Naïve bayes dengan Rapidminer

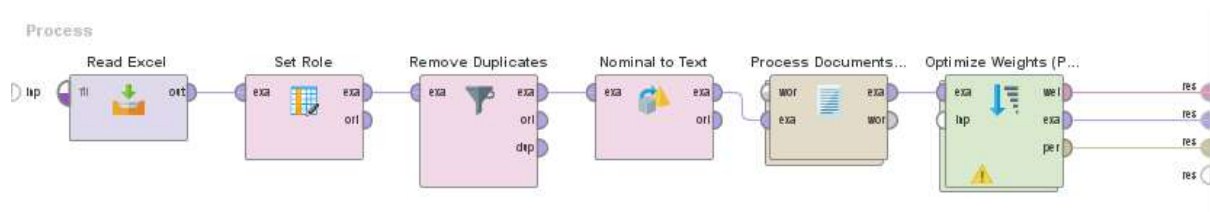
. Lalu proses cross validation pada Model pengujian K-Nearest Neighbor yang menggunakan alat Rapidminer dapat dilihat di sini.



Gambar 7. Model Pengujian K-Nearest Neighbor dengan Rapidminer

3.4. Fitur seleksi Particle Swarm Optimization

Peneliti mencoba meningkatkan akurasi pengklasifikasi Algoritma Naïve Bayes dan K-Nearest Neighbor dengan menggunakan Particle Swarm Optimization. Untuk hasil evaluasi, penelitian ini nantinya didapat akurasi dan nilai AUC dengan menggunakan aplikasi RapidMiner untuk hasil evaluasi. Algoritma dan Cross Validation Terjadi didalam PSO



Gambar 8. Model Pengujian PSO pada Rapidminer

3.5. Evaluasi

Penelitian ini menggunakan algoritma Naive Bayes dan K-Nearest Neighbor untuk melakukan analisis sentimen. Kemudian dalam menentukan evaluasinya peneliti menggunakan Accuracy dan AUC (Area Under Curve). Selanjutnya peneliti akan membandingkan hasil akurasi dari Naive Bayes dan K-

Nearest Neighbor lalu dibandingkan pula Naïve Bayes dan K-Nearest Neighbor dengan menggunakan Particle Swarm Optimization.

Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 1019 komentar opini pengguna dari aplikasi Tokopedia di Play Store mengenai penilaian aplikasi Tokopedia, maka hasil Akurasi Algoritma Naive Bayes tanpa menggunakan fitur Seleksi Particle Swarm Optimization yaitu sebesar 76.30 %, sedangkan untuk nilai AUC sebesar 0.500.

Berikut ini adalah Tabel Confusion Matrix Algoritma Naive Bayes tanpa fitur seleksi Particle Swarm Optimization.

Tabel 8. Tabel *Confusion Matrix* algoritma Naïve baiyes

	True Positive	True Negative	
Pred Positive	48	113	29.81%
Pred Negative	128	728	85.05%
	22.27%	86.56%	

Pada tabel 8 Confusion Matrix dapat dilihat, sebanyak 48 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class positif, sebanyak 128 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif.

Dan sebanyak 113 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 728 data diprediksi class negatif sesuai yaitu termasuk kedalam prediksi class negatif.



Gambar 9. Grafik *Area Under Curve* (AUC) NB.

Berikutnya adalah pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 1019 komentar opini pengguna dari aplikasi Tokopedia di Play Store mengenai penilaian aplikasi Tokopedia, maka hasil Akurasi Algoritma K-Nearest Neighbor yaitu sebesar 74.09%, sedangkan untuk nilai AUC sebesar 0.511.

Berikut ini adalah Tabel Confusion Matrix Algoritma Naive Bayes dengan fitur seleksi Particle Swarm Optimization.

Tabel 9. Tabel *Confusion Matrix* algoritma Naïve baiyes + PSO

	True Positive	True Negative	
Pred Positive	684	108	86.36%
Pred Negative	155	68	30.49%
	81.53%	38.64%	

Pada tabel 9 *Confusion Matrix* dapat dilihat, sebanyak 684 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class positif, sebanyak 155 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif.

Dan sebanyak 108 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 68 data diprediksi class negative.



Gambar 10. Grafik *Area Under Curve* (AUC) NB dan PSO.

Selanjutnya, pengujian telah dilakukan dengan menggunakan dataset sebanyak 1019 komentar opini pengguna dari aplikasi Tokopedia di Play Store mengenai penilaian aplikasi Tokopedia. Hasilnya menunjukkan akurasi Algoritma K Nearest Neighbor sebesar 83,10 persen tanpa menggunakan Fitur Optimization Particle Swarm Selection, dengan nilai AUC sebesar 0.677.

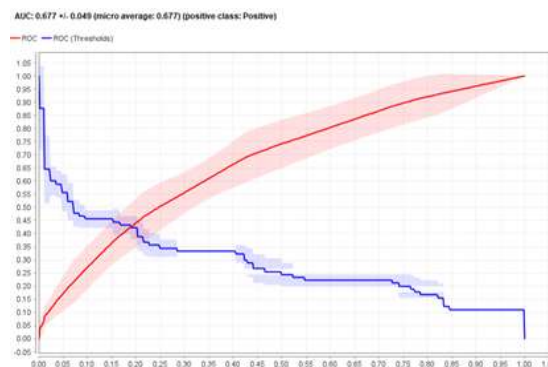
Berikut ini adalah Tabel *Confusion Matrix* Algoritma K Nearest Neighbor tanpa fitur seleksi Particle Swarm Optimization.

Tabel 10. Tabel *Confusion Matrix* algoritma KNN

	True Positive	True Negative	
Pred Positive	826	156	84.11%
Pred Negative	16	20	55,56%
	98.10%	11.36%	

Pada tabel 10 *Confusion Matrix* dapat dilihat, sebanyak 826 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class positif, sebanyak 16 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif.

Dan sebanyak 156 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 20 data diprediksi class negatif sesuai yaitu termasuk kedalam prediksi class negatif.



Gambar 11. Grafik Area Under Curve (AUC) KNN

Yang terakhir pengujian yang sudah dilakukan dengan menggunakan yang menggunakan dataset yang terdiri dari 1019 komentar pengguna tentang aplikasi Tokopedia di Play Store, maka hasil Akurasi Algoritma K Nearest Neighbor dengan Fitur Seleksi Particle Swarm Optimization sebesar 83.53 persen, dengan nilai AUC sebesar 0.601.

Berikut ini adalah Tabel Confusion Matrix Algoritma K Nearest Neighbor dengan fitur seleksi Particle Swarm Optimization.

Tabel 11. Tabel Confusion Matrix algoritma KNN dan PSO

	True Positive	True Negative	
Pred Positive	834	162	83.73%
Pred Negative	5	13	72,22%
	99.40%	7.43%	

Pada tabel 11 Confusion Matrix dapat dilihat, sebanyak 834 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class positif, sebanyak 5 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif.

Dan sebanyak 156 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 20 data diprediksi class negatif sesuai yaitu termasuk kedalam prediksi class negatif.



Gambar 12. Grafik Area Under Curve (AUC) NB dan PSO

Dari hasil pengujian diatas, didapat perbedaan hasil akurasi dan AUC dari Naive bayes dengan K Nearest Neighbor. Nilai akurasi K Nearest Neighbor memiliki hasil yang lebih baik dari Naive Bayes, dengan atau tanpa menggunakan fitur seleksi Particle Swarm Optimization.

Nilai akurasi Naive Bayes tanpa menggunakan fitur seleksi Particle Swarm Optimization adalah 76.30% dan nilai AUC 0.500, sedangkan Naive Bayes dengan menggunakan fitur seleksi Particle Swarm Optimization memiliki akurasi sebesar 74.09% dan nilai AUC 0.511, selisih peningkatan akurasi tanpa dan dengan menggunakan fitur seleksi adalah 4.29% dan selisih peningkatan AUC 0.011.

Kemudian pada pengujian menggunakan K Nearest Neighbor tanpa fitur seleksi Particle Swarm Optimization, nilai akurasi 83.10%, dan nilai AUC 0.677. sedangkan K Nearest Neighbor dengan menggunakan fitur seleksi Particle Swarm Optimization memiliki hasil akurasi sebesar 83.53% dan nilai AUC 0.601. Selisih peningkatan akurasi tanpa dan dengan menggunakan fitur seleksi adalah 0.43% dan selisih AUC 0.066. Berikut adalah tabel perbandingannya.

Tabel 12. Tabel Perbandingan Hasil Pengujian

	NB	NB + PSO	KNN	KNN + PSO
Accuracy	76.30%	74.09%	83.10%	83.53%
AUC	0.500	0.511	0.677	0.601

4. Kesimpulan

Berdasarkan dari penelitian yang sudah dilakukan, dapat disimpulkan bahwa hasil akurasi dan AUC dengan algoritma K Nearest Neighbor lebih baik dari algoritma Naive Bayes, Begitu juga pengaruh penggunaan fitur seleksi Particle Swarm Optimization terhadap kedua algoritma tersebut juga tidak berdampak besar. dapat dilihat dari hasil pengujian diatas, terjadi perubahan nilai akurasi dan AUC yang tidak signifikan yang mana pada algoritma Naive Bayes justru mengalami penurunan. Dan pada penelitian ini juga, hasil akurasi dan AUC dengan menggunakan algoritma Naive Bayes dan K Nearest Neighbor memiliki hasil yang kurang baik dari penelitian sebelumnya yang menjadi referensi.

Daftar Pustaka

- [1] Setiawan, Kiki, Beni Rahmatullah, Burhanuddin Burhanuddin, Atik Budi Paryanti, and Fariz Fauzi. "Komparasi Metode Naive Bayes Dan Support Vector Machine Menggunakan Particle Swarm Optimization Untuk Analisis Sentimen Mobil Esemka." *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)* 4, no. 3 (2020): 102-111.
- [2] BASARI, Abd Samad Hasan, et al. Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. *Procedia Engineering*, 2013, 53: 453-462.
- [3] Grosan, Crina, Ajith Abraham, and Monica Chis. *Swarm intelligence in data mining*. Springer Berlin Heidelberg, 2006.
- [4] ZHANG, Lei; LIU, Bing. Sentiment Analysis and Opinion Mining. *Encyclopedia of machine learning and data mining*, 2017, 1: 1152-1161.
- [5] Romadloni, Nova Tri, Imam Santoso, and Sularso Budilaksono. Perbandingan Metode Naive Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *ikraith-informatika*, 2019, 3.2: 1-9.
- [6] El Firdaus, Muhammad Farid, Nurfaizah Nurfaizah, and Sarmini Sarmini. "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbor." *JURIKOM (Jurnal Riset Komputer)* 9, no. 5 (2022): 1329-1336.

- [7] Salsabila, Sifa Melina, Aang Alim Murtopo, and Nurul Fadhilah. "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier." *Jurnal Minfo Polgan* 11, no. 2 (2022): 30-35.
- [8] Pajri, Dicki, Yuyun Umaidah, and Tesa Nur Padilah. "K-nearest neighbor berbasis particle swarm optimization untuk analisis sentimen terhadap Tokopedia." *Jurnal Teknik Informatika dan Sistem Informasi* 6, no. 2 (2020).
- [9] Apriani, Rita, and Dudih Gustian. "Analisis Sentimen dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia." *Jurnal Rekayasa Teknologi Nusa Putra* 6, no. 1 (2019): 54-62.