

KLASIFIKASI KELUHAN PENGGUNA KAI ACCESS UNTUK PEMESANAN TIKET DENGAN ALGORITMA SVM DAN NAÏVE BAYES

Antonius Yadi Kuntoro¹⁾, Hermanto²⁾, Taufik Asra³⁾

¹ Program Studi Sistem Informasi Fakultas Teknik Informatika, Universitas Nusa Mandiri, Jl. Raya Jatiwaringin No. 2
Jakarta Timur

² Program Studi Teknologi Komputer Fakultas Teknik Informatika, Universitas Bina Sarana Informatika, Jl. Kramat Raya
No. 98 Jakarta Pusat

³ Program Studi Rekayasa Perangkat Lunak Fakultas Teknik Informatika, Universitas Bina Sarana Informatika, Jl. Kramat
Raya No. 98 Jakarta Pusat

Co Responden Email: hermanto.hmt@bsi.ac.id

Article history

Received 22 April 2022

Revised 06 June 2022

Accepted 20 June 2022

Available online 29 June 2022

Keywords

KAI Access, Text Mining, SVM,
Naïve Bayes

Abstract

The development of Information and Communication Technology (ICT) is needed to facilitate and solve various problems faced by humans quickly. Today's society wants everything to be practical and not time-consuming. One example of a problem that is of concern to society today is transportation. Trains are the favorite transportation of Indonesians as evidenced by the increasing number of special train services on various mobile apps. This study focuses on analyzing the satisfaction of users of the KAI Access application on ticket reservations. This study aims to analyze the complaints of users of the KAI Access application in ordering train tickets online. There were 1321 positive and negative comments on kai access application users for ticket booking complaints. By using the SVM Algorithm and Naïve Bayes, a comparison of the positive and negative comments was made. From the testing process, the accuracy of the SVM algorithm is obtained, the accuracy value = 73.36% and the AUC value = 0.794. while for the Naïve Bayes algorithm the accuracy value = 67.10% and the AUC value = 0.573. It can be concluded that the algorithm that has the highest accuracy is the SVM algorithm compared to the Naïve Bayes algorithm.

Riwayat

Diterima 22 April 2022

Revisi 06 Juni 2022

Disetujui 20 Juni 2022

Terbit 29 Juni 2022

Kata Kunci

KAI Access, Text Mining,
SVM, Naïve Bayes

Abstrak

Perkembangan Teknologi Informasi dan Komunikasi (TIK) sangat diperlukan untuk memudahkan dan menyelesaikan berbagai masalah yang dihadapi oleh manusia dengan cepat. Masyarakat zaman sekarang ingin semuanya serba praktis dan tidak menyita banyak waktu. Salah satu contoh permasalahan yang menjadi perhatian masyarakat sekarang adalah transportasi. Kereta api merupakan transportasi favorit orang Indonesia terbukti dengan meningkatnya layanan khusus Kereta Api diberbagai perangkat mobile apps. Penelitian ini fokus terhadap menganalisa kepuasan pengguna aplikasi KAI Access terhadap pemesanan tiket. Penelitian ini bertujuan untuk menganalisis keluhan pengguna aplikasi KAI Access dalam pemesanan tiket kereta api secara online. Terdapat 1321 komentar positif dan negatif pada pengguna aplikasi kai access untuk keluhan pemesanan tiket. Dengan menggunakan Algoritma SVM dan Naïve Bayes dilakukan perbandingan pengujian atas komentar positif dan negatif tersebut. Dari proses pengujian tersebut didapatkan hasil akurasi dari algoritma SVM nilai akurasi = 73.36% dan nilai AUC = 0.794. sedangkan untuk algoritma Naïve Bayes nilai akurasi = 67.10% dan nilai AUC = 0.573. Dapat disimpulkan bahwa algoritma yang lebih unggul adalah memiliki nilai akurasi tertinggi adalah Algoritma SVM dibanding dengan algoritma Naïve Bayes.

PENDAHULUAN

Perkembangan dan kemajuan Teknologi Informasi dan Komunikasi (TIK) sangat diperlukan guna untuk memudahkan dan

menyelesaikan berbagai masalah yang dihadapi oleh manusia dengan cepat dan singkat. Disamping itu, masyarakat zaman sekarang ingin semuanya serba praktis dan tidak menyita banyak waktu. Salah satu contoh

permasalahan sehari-hari yang menjadi perhatian masyarakat sekarang adalah transportasi. Transportasi, adalah sebuah sarana umum yang apapun jenisnya dan dimanapun tempatnya, sangat diperlukan bagi setiap orang yang hendak bepergian, apalagi ke tempat yang tidak mungkin untuk dijangkau hanya dengan berjalan kaki. Begitu banyak transportasi umum yang disediakan baik oleh pemerintah, swasta ataupun perorangan sehingga memudahkan masyarakat untuk mencari alternatif yang terbaik sesuai dengan kebutuhan dan kemampuan mereka. Misalnya kereta api, alat transportasi ini selain memberikan penawaran kenyamanan juga memberikan penawaran berupa tarif yang terjangkau.

PT Kereta Api Indonesia merupakan badan usaha milik Indonesia yang menyediakan jasa di bidang transportasi yaitu kereta api. Seiring dengan perkembangan zaman PT KAI cukup menyadari pentingnya sebuah aplikasi mobile untuk memberikan layanan secara online, dengan adanya aplikasi ini diharapkan dapat memberikan kemudahan dalam mengkases dan melakukan transaksi. Namun pada kenyataannya pengguna aplikasi yang disediakan oleh PT KAI yaitu aplikasi KAI Access cukup rendah berdasarkan rating dan respon dari pengguna pada *Google PlayStore*.

Kereta api nampaknya menjadi salah satu alat transportasi favorit orang Indonesia terbukti dengan meningkatnya layanan khusus Kereta Api Indonesia di berbagai perangkat Android, IOS, dan Windows Phone. Beberapa aplikasi ini seperti Treveloka, Tiket.com, Pegipegi, Tokopedia, dan juga aplikasi favorit yang akan kita bahas yaitu KAI Access. Hal ini yang menjadi salah satu faktor yang ingin dijadikan pedoman oleh peneliti untuk mempelajari lebih lanjut tentang KAI Access terhadap kepuasan para pengguna aplikasi transportasi darat tersebut. Saat ini telah muncul bermacam-macam jenis dari aplikasi mobile yang ada, salah satunya adalah *mobile ticketing*. *Mobile ticketing* merupakan aplikasi *mobile* yang juga dapat digunakan untuk pemesanan serta pembelian tiket.

Aspek penting yang harus diperhatikan dalam pengembangan aplikasi ini adalah persepsi pengguna dalam keluhan pemesanan tiket kereta. Ketika pengguna puas dengan suatu aplikasi maka pengguna akan terus

menggunakan aplikasi tersebut. Evaluasi pengalaman pengguna pada suatu aplikasi dibutuhkan agar dapat melihat kekurangan dari sistemnya, sehingga dapat terus mengembangkan sistem sesuai dengan kebutuhan pengguna.

Keluhan yang diberikan oleh pelanggan dalam pemakaian produk atau jasa merupakan suatu umpan balik dari kualitas produk atau jasa yang digunakan oleh pelanggan. Semakin banyak keluhan yang diberikan oleh pelanggan membutuhkan adanya perhatian ekstra bagi setiap perusahaan yang menerima keluhan untuk memperbaiki dalam pembuatan produk atau jasa. Untuk itu setiap perusahaan harus mampu memberikan kepuasan kepada para pelanggannya dengan cara menyediakan produk yang mutunya lebih baik dan harga yang relatif terjangkau (Hermanto, Mustopa, et al., 2020).

Penelitian ini fokus terhadap menganalisa kepuasan pengguna aplikasi KAI Access terhadap pemesanan tiket, menunjukkan bahwa kualitas layanan penting untuk diperhatikan dianggap sebagai strategi bisnis yang bertujuan untuk memenuhi kepercayaan pengguna terhadap pelayanan yang diterima, keinginan serta kebutuhan sehingga menimbulkan pengaruh positif yaitu loyalitas pelanggan secara terus menerus. Penelitian ini bertujuan untuk menganalisis keluhan pengguna aplikasi KAI Access dalam pemesanan tiket kereta api secara *online*.

Penelitian sebelumnya telah banyak dilakukan dengan menerapkan metode *text mining* oleh peneliti dalam menganalisa sentimen. Seperti penelitian yang dilakukan, Algoritma Klasifikasi *Naïve Bayes* Dan *Support Vector Machine* Dalam Layanan Komplain Mahasiswa (Hermanto, Mustopa, et al., 2020). *Sentiment Analysis On Gojek And Grab User Reviews Using Svm Algorithm Based On Particle Swarm Optimization* (Hermanto, Kuntoro, et al., 2020).

Analysis of user reviews for the peduli lindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization (Mustopa et al., 2020).

Analisis Indeks Kepuasan Pengguna KAI Access Pada Sistem E-Payment Tiket Ka Lokal (Studi Kasus: Daop 8 Surabaya) (Setiawan & Novita, 2021).

Dari beberapa teknik tersebut yang paling sering digunakan untuk klasifikasi data adalah algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* sehingga penelitian ini peneliti menggunakan algoritma *Support Vector Machine* (SVM) dan *naïve bayes* terhadap klasifikasi keluhan pengguna KAI Access pada *google playstore*.

Dari serangkaian latar belakang yang telah diuraikan diatas, peneliti akan melakukan klasifikasi keluhan pengguna KAI Access pada *google playstore* dengan menggunakan metode *Support Vector Machine* (SVM) dan *Naïve Bayes*. Dari Proses pengujian data pada *rapidminer* dimulai dengan pembentukan model dengan data pada bagian pertama pembagian data *training* dan data *testing*. Setelah melakukan pengujian hasil Akurasi dapat diukur dengan menggunakan *confusion matrix*, dan *performance* diukur menggunakan *accuracy* dan AUC serta akan ditampilkan dalam bentuk kurva ROC, tentunya dapat diketahui aplikasi ojek online manakah yang memiliki tingkat akurasi paling tinggi sesuai dengan hasil penelitian menggunakan model algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes*.

METODE PENELITIAN

1. Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait dari berbagai database besar (Ipmawati et al., 2017). Data mining juga diartikan sebagai suatu proses otomatis data-data yang sangat besar dan bertujuan untuk mendapatkan hubungan atau pola yang memberikan manfaat. Data mining juga merupakan suatu proses pendukung pengambil keputusan dimana mencari pola informasi dalam data. Pencarian ini dapat di lakukan oleh pengguna (Sari, 2013).

2. Textmining

Tujuan penting dari *text mining* untuk mendapatkan informasi berkualitas tinggi dari teks. biasanya dilakukan melalui penemuan pola dan *trend* dengan cara seperti pembelajaran pola statistik, pemodelan topik, dan pemodelan bahasa statistik. *Text mining* biasanya memerlukan penataan teks masukan

(misalnya *parsing*, bersamaan dengan penambahan beberapa fitur linguistik turunan dan penghapusan yang lain, dan penyisipan selanjutnya ke dalam *database*). Hal ini diikuti dengan menurunkan pola dalam data terstruktur, dan evaluasi dan interpretasi dari *output*. "Kualitas tinggi" dalam *text mining* biasanya mengacu pada kombinasi relevansi, kebaruan, dan ketertarikan (Witten et al., 2016).

3. Analisa Sentimen

Analisis sentimen adalah sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen positif atau negatif (Hadna et al., 2016). Analisis sentimen mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan text mining yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu (Manalu, 2014). Analisis sentimen juga dapat diartikan sebagai informasi tekstual secara umum dibagi menjadi informasi fakta (objektif) dan opini (subjektif) (Safari, 2012). Analisis sentimen juga merupakan riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual (Ipmawati et al., 2017).

4. E-Ticketing

E-Ticketing merupakan cara untuk mendokumentasikan proses penjualan dari suatu aktifitas perjalanan tanpa harus mengeluarkan dokumen secara fisik. ETicketing mengurangi biaya proses tiket, menghilangkan formulir kertas dan meningkatkan fleksibilitas penumpang dan agen perjalanan dalam membuat perubahan-perubahan dalam jadwal perjalanan (Purba, 2014).

Terdapat banyak keuntungan yang bisa diperoleh penumpang kereta api yang menggunakan fasilitas E-Ticketing, diantaranya adalah :

- a. Pengguna sistem E-Ticketing diuntungkan dengan tidak dibatasi oleh waktu dan tenaga dalam hal pemesanan tiket.

- b. Reservasi tiket kereta api dapat dilakukan dimanapun lokasinya.
- c. Pembayaran dapat dilakukan secara realtime dan online.
- d. Pengguna sistem E-Ticketing KAI dapat memilih tempat duduk sendiri.
- e. Terhindar dari resiko kesalahan input yang dilakukan oleh travel agent atau petugas loket pembelian tiket.
- f. Biaya komunikasi antara penumpang dan petugas reservasi dapat ditekan.

5. Keluhan (Complaint)

Keluhan secara definisi diartikan sebagai satu pernyataan atau ungkapan rasa kurang puas terhadap satu produk atau layanan jasa, baik secara lisan maupun tertulis, dari penyampai keluhan baik internal maupun eksternal. Atau sebuah ungkapan ketidakpuasan antara harapan dengan fakta terhadap apa yang diterima dalam bentuk produk maupun layanan jasa (Hermanto, Mustopa, et al., 2020).

Keluhan yang diberikan oleh pelanggan dalam pemakaian produk atau jasa merupakan suatu umpan balik dari kualitas produk atau jasa yang digunakan oleh pelanggan. Semakin banyak keluhan yang diberikan oleh pelanggan membutuhkan adanya perhatian ekstra bagi setiap perusahaan yang menerima keluhan untuk memperbaiki dalam pembuatan produk atau jasa. Untuk itu setiap perusahaan harus mampu memberikan kepuasan kepada para pelanggannya dengan cara menyediakan produk yang mutunya lebih baik dan harga yang relatif terjangkau (Muljono et al., 2018).

6. Support Vector Machine

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition*. Sebagai salah satu metode *pattern recognition*, usia SVM terbilang masih relatif muda (Rahmayuni, 2014). Walaupun demikian, evaluasi kemampuannya dalam berbagai aplikasinya menempatkannya sebagai *state of the art* dalam *pattern recognition*, dan dewasa ini merupakan salah satu tema yang berkembang dengan pesat. SVM adalah metode *learning machine* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang

memisahkan dua buah *class* pada input *space* (Nugroho et al., 2003).

7. Naïve Bayes

Teori *Naïve Bayes* memiliki kemampuan klasifikasi yang serupa dengan *Decision Tree* dan *Neural Network* bahkan algoritma *Naïve Bayes* memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Fahlapi & Rianto, 2020).

HASIL DAN PEMBAHASAN

Business Understanding

Pada tahap *business understanding* bisa disebut juga sebagai tahap pemahaman penelitian, menentukan tujuan proyek penelitian dalam perumusan mendefinisikan masalah data mining. Semakin banyak masyarakat yang menggunakan media sosial maka semakin banyak data berupa opini yang di hasilkan, dalam opini-opini tersebut tidak hanya membicarakan satu topik saja yang di sampaikan.

Pada penelitian ini memiliki tujuan untuk membuat sebuah Analisis sentimen untuk mengetahui komentar masyarakat terhadap belajar online pada masa covid-19 melalui opini yang disampaikan dalam media sosial seperti twitter dan mengimplementasikan metode klasifikasi dalam *text mining*.

Data Understanding

Tahap ini adalah proses memahami data yang Tahapan ini adalah proses memahami data yang akan digunakan sebagai bahan yang akan diteliti untuk bisa dilakukan ke tahap setelahnya yaitu *processing*. Data yang dikumpulkan dalam proses crawling data tweet ini kemudian disimpan dalam format *file excel*. Adapun sumber data utama yang digunakan dalam penelitian ini menggunakan dataset komentar pengguna aplikasi KAI Access pada *google play*. Dataset yang diambil sejumlah 1321 data. Dari komentar pengguna aplikasi KAI Access pada *google play* tersebut kemudian membagi statusnya menjadi dua bagian yaitu 673 dengan label negatif dan 683 label positif agar status menjadi kondisi normal dalam penelitian *text mining*. Semua data komentar tersebut dikelompokkan menjadi satu baik itu positif atau negatif dan disimpan dalam bentuk ekstensi *.xls*.

Data Preparation

Persiapan data sebelum data akan dilakukan modelling yaitu mempersiapkan data untuk melakukan langkah-langkah text preprocessing.

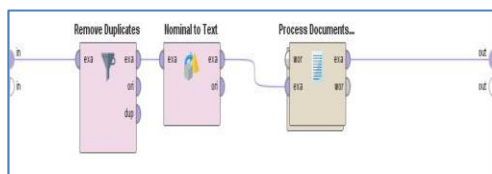
Dalam *pre-processing* teks berbahasa Indonesia yang dikombinasikan dengan aplikasi RapidMiner untuk memproses kata-kata dalam bahasa Indonesia (Hermanto, Kuntoro, et al., 2020). Menghilangkan *noise* dalam komentar dengan memilih menu *annotation removal (delete @username, #hashtag, Transformation remove URL (http://), Tokenizing (regex) simbol-simbol (emoticon dan nomor), Indonesian Stemming (menghapus bahasa asing), Transformation not negative* atau kata yang di depannya terdapat kata “tidak” akan di normalisasi dengan menggunakan tanda garis bawah “_” agar memiliki arti yang jelas, misalnya “tidak baik” menjadi “tidak_baik” karena kata “tidak” bermakna negatif sedangkan kata “baik” bermakna positif dan terakhir.

Remove Duplicates

Ini merupakan tahapan *data preparation* selanjutnya yang digunakan pada *software rapidminer*. *Remove duplicates* digunakan untuk menghilangkan *text* yang sama atau duplikat. Hal ini dilakukan agar data tidak dipenuhi oleh *text* yang sama sehingga memperlambat proses *running software* untuk menganalisa model.

Nominal to Text

Ini merupakan operator yang ada dalam rapidminer yang berfungsi untuk mengubah semua angka yang ada dalam *text* menjadi sebuah *text*. Sehingga angka yang ada akan dianggap jenis data text bukan *numeric* atau *nominal*. Gambar 1. memperlihatkan bagaimana penggunaan operator ini digunakan pada proses yang ada pada *rapidminer*.



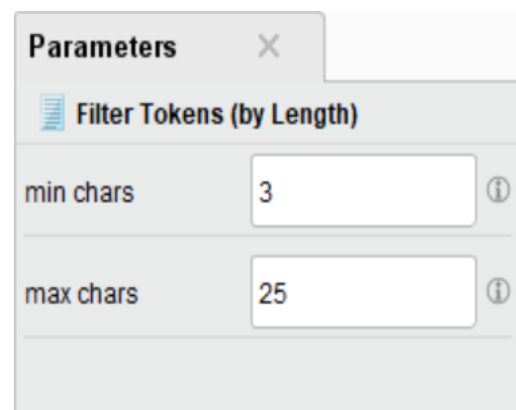
Gambar 1. Desain Model Preprocessing Data Local menggunakan operator Remove Duplicates dan Nominal to Text Transform Case.

Operator yang digunakan pada tahapan ini adalah untuk mengubah huruf kapital yang

masih ada pada *text* akan diubah menjadi huruf kecil semua. Hal ini dilakukan agar ketikan dilakukan proses ke dalam model klasifikasi terdapat keseragaman huruf dan tidak terjadi kesalahan dalam proses *tokenize*.

Filter Token (by Length)

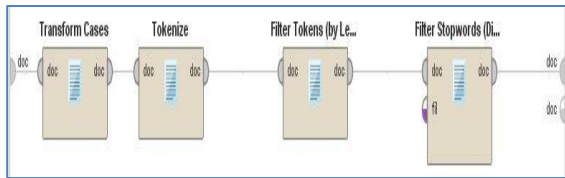
Ini adalah proses yang ada pada data preparation untuk menghilangkan sejumlah kata (setelah proses *tokenize*) dengan panjang karakter tertentu. Pada penelitian ini panjang minimum karakter yang digunakan adalah 3 karakter dan panjang maksimum 25 karakter. Artinya kata yang panjangnya kurang dari 3 karakter dan lebih dari 25 karakter akan dihilangkan. Untuk mendapatkan hasil seperti ini maka dilakukan setting pada Parameters dari operator ini (Gambar 2).



Gambar 2. Parameters dari Filter Tokens (by Length)

Filter Stopword (Dictionary)

Selanjutnya adalah penggunaan operator *Stopword Removal (by Directory)* yang berfungsi untuk menghilangkan kata-kata yang tidak hubungan dengan isi *text*. Pada tahapan sebelumnya dengan menggunakan *service text mining Gataframework* telah dilakukan namun ada beberapa kata yang belum dapat bisa dihilangkan oleh *service* sebelumnya karena belum dimasukkan sebagai kata yang harus dihapus. Maka dengan operator *Stopword Removal (by Directory)* peneliti dapat mendaftarkan kata yang harusnya dihapus dari *text*. Gambar 3. merupakan penjelasan dari penggunaan operator pada proses *rapidminer*.



Gambar 3. Desain dari Penggunaan operator untuk Data Preparation.

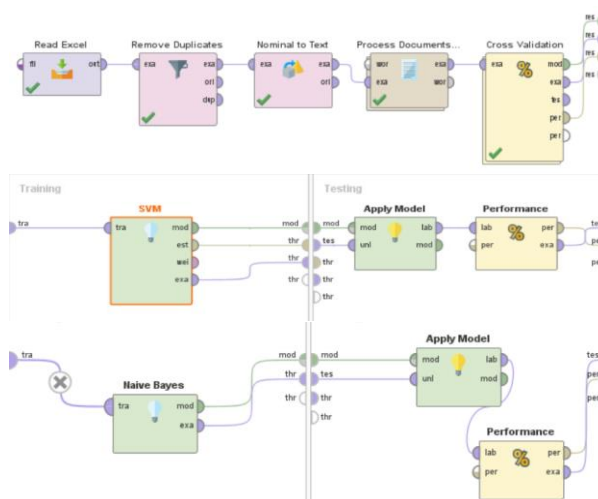
Tahapan Pemodelan

Merupakan tahap pemilihan teknik mining dengan menentukan algoritma yang akan digunakan. Tool yang digunakan adalah RapidMiner versi 9.1. Hasil pengujian model yang dilakukan adalah menganalisa komentar pengguna aplikasi KAI Access dengan label data komentar positif dan negatif menggunakan algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* untuk mendapatkan nilai akurasi terbaik.

Pengujian Model dengan Algoritma SVM Dan NB

Pengaturan dan penggunaan operator serta parameter dalam *frameworks RapidMiner* sangat berpengaruh terhadap akurasi dan model yang terbentuk, secara lebih jelas pengujian terhadap algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* Berbasis *Particle Swarm Optimization* tersebut adalah sebagai berikut.

Proses model ini algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* dilakukan proses *10 fold validation* dapat dilihat pada gambar 4.



Gambar 4. Desain Model Algoritma (SVM) dan NB

Gambar 4 diatas adalah model pengujian algoritma *Support Vector Machine (SVM)* dan

Naive Bayes menggunakan *rapidminer*, diawali dari memasukan data kemudian mengatur *set role* yang nantinya menentukan label disana dan nominal text lalu keproses dokumen. Pada pengujian ini, data digunakan adalah data bersih yang telah melalui *preprocessing*. Data tersebut diambil dari operator Read Excel, hal ini dilakukan karena dataset disimpan dalam bentuk Excel (.xlsx). *Process documents from files* untuk mengkonversi files menjadi dokumen. *Process validasi* terdiri dari data training dan data testing. Kemudian masuk ke model desain algoritmanya diproses ke dalam operator *cross validation Support Vector Machine (SVM)* dan *Naive Bayes* didalamnya ada perhitungan algoritmanya kemudian modelnya diapply setelah itu masuk ke penilaian performancenya barulah muncul hasil nilai *accuracy* dan *auc* nya.

Model Evaluasi

Tahapan evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil dibuat pada langkah sebelumnya. Untuk evaluasi digunakan *10-fold cross validation*. Dari hasil pengujian model dari dua algoritma yang dipakai adalah untuk menghasilkan sebuah nilai *Accuracy (Confusion Matrix)* dan *AUC (Area Under Curve)*. Maka mendapatkan hasil grafik ROC dengan nilai *AUC (Area Under Curve)*.

Nilai Accuracy algoritma Support Vector Machine (SVM) Dan Naïve Bayes

Berdasarkan hasil percobaan yang dilakukan dengan menggunakan data komentar pengguna aplikasi KAI Access pada google play dimana dalam percobaan ini menggunakan Algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* dengan menggunakan 1321 data komentar pengguna aplikasi KAI Access pada *google play* maka dihasilkan nilai *Accuracy (Confusion Matrix)* sebagai berikut:

Tabel 1. Nilai Accuracy Algoritma SVM

accuracy: 73.26% +/- 3.03% (micro average: 73.26%)			
	true Negatif	true Positif	class precision
pred. Negatif	458	154	74.84%
pred. Positif	176	446	71.70%
class recall	72.24%	74.33%	

Acc (Accuracy)

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

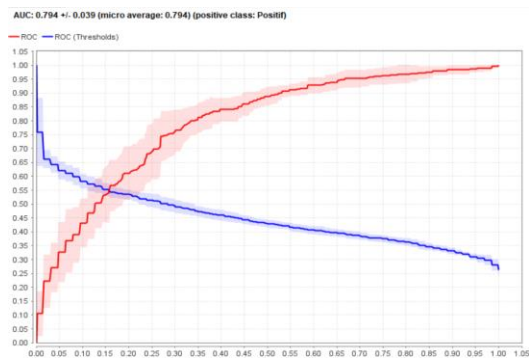
$$= \frac{446 + 458}{446 + 458 + 154 + 176}$$

$$= \frac{904}{1.234} = 0,732$$

Jumlah *True Negatif* (TN) adalah 458 record diklasifikasikan sebagai *Negatif* dan *False Positif* (FP) adalah 176 record diklasifikasikan sebagai *Negatif*. Berikutnya 446 *True Positif* (TP) diklasifikasikan sebagai *True Positif* dan 154 record *False Positif* diklasifikasikan sebagai *Positif*. Berdasarkan tabel 1 diatas menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma SVM adalah sebesar 73.26%.

Nilai AUC dari Algoritma SVM

Berikut ini akan dijelaskan Kurva ROC dan Confusion Matrix dari algoritma *Support Vector Machine*:



Gambar 5. Nilai AUC dalam Algoritma SVM

Dari hasil pengujian model yang telah dilakukan adalah untuk mendapatkan hasil akurasi dan *Area Under Curve* (AUC). Maka mendapatkan hasil grafik ROC dengan nilai *Area Under Curve* (AUC) sebesar 0.794 dengan *performance* akurasi yaitu *excellent*.

Tabel 2. Nilai Accuracy Algoritma NB

	true Negatif	true Positif	class precision
pred. Negatif	440	212	67.48%
pred. Positif	194	388	66.67%
class recall	69.40%	64.67%	

Acc (Accuracy)

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

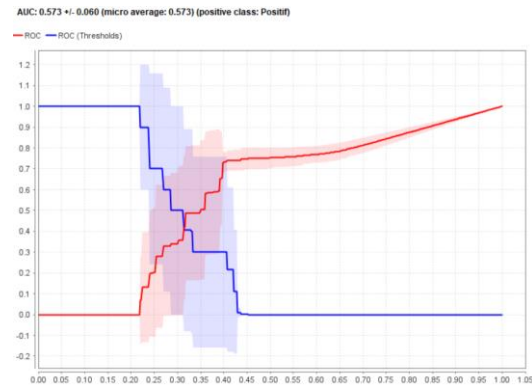
$$= \frac{338 + 440}{338 + 440 + 194 + 212}$$

$$= \frac{778}{1.184} = 0,657$$

Jumlah *True Negatif* (TN) adalah 440 record diklasifikasikan sebagai *Negatif* dan *False Positif* (FP) adalah 194 record diklasifikasikan sebagai *Negatif*. Berikutnya 338 *True Positif* (TP) diklasifikasikan sebagai *True Positif* dan 212 record *False Positif* diklasifikasikan sebagai *Positif*. Berdasarkan tabel 1 diatas menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma SVM adalah sebesar 67.10%.

Nilai AUC dari Algoritma NB

Berikut ini akan dijelaskan Kurva ROC dan Confusion Matrix dari algoritma *Naïve Bayes*:



Gambar 6. Nilai AUC dalam Algoritma Naïve Bayes

Dari hasil pengujian model yang telah dilakukan adalah untuk mendapatkan hasil akurasi dan *Area Under Curve* (AUC). Maka mendapatkan hasil grafik ROC dengan nilai *Area Under Curve* (AUC) sebesar 0.573 dengan *performance* akurasi yaitu *excellent*.

KESIMPULAN

Dalam penelitian ini setelah dilakukan *preprocessing* dan dilakukan pengujian model dengan metode data mining yaitu *support vector machine* (SVM) dan *Naïve Bayes* hasil evaluasi dan validasi, diketahui bahwa nilai akurasi untuk menentukan bahwa klasifikasi komentar pengguna KAI Acces keluhan pemesanan tiket dilakukan pengujian dengan

algoritma *support vector machine* (SVM) dan Naïve Bayes dengan data komentar dibagi menjadi dua kategori label positif dan negatif, dapat dibuktikan dengan nilai akurasi dan nilai *AUC* dari algoritma yaitu untuk SVM nilai akurasi = 73.36% dan nilai *AUC* = 0.794. sedangkan untuk algoritma Naïve Bayes nilai akurasi dan nilai *AUC* dari algoritma yaitu untuk SVM nilai akurasi = 67.10% dan nilai *AUC* = 0.573.

Dalam penelitian dapat disimpulkan bahwa algoritma yang lebih unggul adalah Algoritma Support Vector Machine dibanding dengan algoritma Naïve Bayes. Untuk itu, penerapan algoritma *Support Vector Machine* pada penelitian ini memiliki akurasi yang lebih tinggi sehingga dapat digunakan untuk memberikan solusi terhadap permasalahan klasifikasi pada komentar pengguna KAI Access untuk keluhan pemesanan tiket.

Untuk Penelitian lanjutan dengan mengambil jumlah data yang lebih besar lagi dalam kurun waktu lebih dari 3 bulan pada google play, sehingga akan mendapatkan nilai akurasi yang lebih baik lagi. Selain itu Menambahkan metode klasifikasi lain untuk membandingkan dengan metode dari Supervised Learning lainnya dari model yang diusulkan, seperti penggunaan Algoritma C-45, KNN dan lainnya.

REFERENSI

- Fahlapi, R., & Rianto, Y. (2020). Twitter Comment Predictions on Dues Changes BPJS Health In 2020. *Sinkron*, 5(1), 170. <https://doi.org/10.33395/sinkron.v5i1.10588>
- Hadna, M. S., Santosa, P. I., & Winarno, W. W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*(Sentika), 57–64. <https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/95.pdf>
- Hermanto, H., Mustopa, A., & Kuntoro, A. Y. (2020). Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa. *JITK (Jurnal Ilmu*

Pengetahuan Dan Teknologi Komputer), 5(2), 211–220. <https://doi.org/10.33480/jitk.v5i2.1181>

- Hermanto, Kuntoro, A. Y., Asra, T., Pratama, E. B., Effendi, L., & Ocanitra, R. (2020). Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique. *Journal of Physics: Conference Series, 1641*(1). <https://doi.org/10.1088/1742-6596/1641/1/012102>
- Ipmawati, J., Kusriani, & Taufiq Luthfi, E. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Indonesian Journal on Networking and Security*, 6(1), 28–36.
- Manalu, B. U. (2014). ANALISIS SENTIMEN PADA TWITTER MENGGUNAKAN TEXT MINING SKRIPSI Boy Utomo Manalu. *Teknologi Informasi Fak. ILKOM UNSUT*.
- Muljono, M., Artanti, D. P., Syukur, A., Prihandono, A., & others. (2018). Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes. *Konferensi Nasional Sistem Informasi (KNSI) 2018*.
- Mustopa, A., Hermanto, Anna, Pratama, E. B., Hendini, A., & Risdiansyah, D. (2020). Analysis of user reviews for the peduli lindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization. *2020 5th International Conference on Informatics and Computing, ICIC 2020*, 2. <https://doi.org/10.1109/ICIC50835.2020.9288655>
- Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support vector machine teori dan aplikasinya dalam bioinformatika. *Kuliah Umum IlmuKomputer. Com*.

- Purba, M. M. (2014). PEMESANAN TIKET KERETA API ONLINE (E-Ticketing) MENGGUNAKAN APLIKASI KAI Access. *Jurnal Sistem Informasi Universitas Suryadarma*.
<https://doi.org/10.35968/jsi.v6i2.324>
- Rahmayuni, I. (2014). Perbandingan Performansi Algoritma C4.5 dan Cart Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang. *Teknoif*, 2(1), 40–46.
- Safari, C. B. (2012). *Skripsi Analisis Sentimen Menggunakan Metode Klasifikasi SVM*. Universitas Pendidikan Indonesia.
- Sari, E. N. (2013). Analisa Algoritma Apriori Untuk Menentukan Merek Pakaian Yang Paling Diminati Pada Mode Fashion Group Medan. *Pelita Informatika Budi Darma*, 4(3), 35–39.
- Setiawan, H., & Novita, D. (2021). Analisis Kepuasan Pengguna Aplikasi KAI Access Sebagai Media Pemesanan Tiket Kereta Api Menggunakan Metode EUCS. *Jurnal Teknologi Sistem Informasi*, 2(2), 162–175.
<https://doi.org/10.35957/jtsi.v2i2.1375>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. In *Data Mining: Practical Machine Learning Tools and Techniques*.
<https://doi.org/10.1016/c2009-0-19715-5>