TESIS

PENGARUH TEXT PREPROCESSING TERHADAP ANALISIS SENTIMEN OPINI PENGGUNA APLIKASI QASIR DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST

Disusun oleh:

Nama : **Dhana Aulia Ayu Kurniawan**
NIM : **19.52.1267**
Konsentrasi : **Business Intelligence**

PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023

TESIS

PENGARUH TEXT PREPROCESSING TERHADAP ANALISIS SENTIMEN OPINI PENGGUNA APLIKASI QASIR DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST

THE EFFECT OF TEXT PREPROCESSING ON SENTIMENT ANALYSIS OF QASIR APPLICATION USERS OPINION USING SUPPORT VECTOR MACHINE AND RANDOM FOREST METHODS

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister

Disusun oleh:

Nama	**: Dhana Aulia Ayu Kurniawan**
NIM	**: 19.52.1267**
Konsentrasi	**: Business Intelligence**

PROGRAM STUDI S2 TEKNIK INFORMATIKA

PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2023

HALAMAN PENGESAHAN

PENGARUH TEXT PREPROCESSING TERHADAP ANALISIS SENTIMEN OPINI PENGGUNA APLIKASI QASIR DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST

THE EFFECT OF TEXT PREPROCESSING ON SENTIMENT ANALYSIS OF QASIR APPLICATION USERS OPINION USING SUPPORT VECTOR MACHINE AND RANDOM FOREST METHODS

Dipersiapkan dan Disusun oleh

Dhana Aulia Ayu Kurniawan

19.52.1267

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 1 Maret 2023

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 1 Maret 2023
Rektor

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

PENGARUH TEXT PREPROCESSING TERHADAP ANALISIS SENTIMEN OPINI PENGGUNA APLIKASI QASIR DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST

THE EFFECT OF TEXT PREPROCESSING ON SENTIMENT ANALYSIS OF QASIR APPLICATION USERS OPINION USING SUPPORT VECTOR MACHINE AND RANDOM FOREST METHODS

Dipersiapkan dan Disusun oleh

Dhana Aulia Ayu Kurniawan
19.52.1267

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 1 Maret 2023

Pembimbing Utama	Anggota Tim Penguji
Prof. Dr. Ema Utami, S.Si., M.Kom.	**Alva Hendi Muhammad, S.T., M.Eng., Ph.D.**
NIK. 190302037	**NIK. 190302493**
	Dhani Ariatmanto, M.Kom., Ph.D.
Pembimbing Pendamping	**NIK. 190302197**
Hanif Al Fatta, M.Kom.	**Prof. Dr. Ema Utami, S.Si., M.Kom.**
NIK. 190302096	**NIK. 190302037**

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 1 Februari 2023
Direktur Program Pascasarjana

Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dhana Aulia Ayu Kurniawan
NIM : 19.52.1267
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:
Pengaruh Text Preprocessing Terhadap Analisis Sentimen Opini Pengguna Aplikasi Qasir Dengan Menggunakan Metode Support Vector Machine Dan Random Forest

Dosen Pembimbing Utama : Prof. Dr. Ema Utami, S.Si., M.Kom.
Dosen Pembimbing Pendamping : Hanif Al Fatta, M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 1 Maret 2023
Yang Menyatakan,

Dhana Aulia Ayu Kurniawan

HALAMAN PERSEMBAHAN

Kedua orang tua tercinta yang selalu mendukung apapun keputusan dan tindakan yang saya ambil, semoga selalu dalam lindungan-Nya dan kelak mendapatkan mahkota terindah di surga.

Suami yang selalu memberikan support, semangat, dan dukungan hingga tesis ini dapat terselesaikan, semoga selalu diberikan kelancaran dalam setiap urusannya.

Adik, Om, Tante dan Sepupu-sepupu di Jogja, Terimakasih atas dukungannya selama ini.

Teman-teman kantor lama (Qasir.id), terimakasih berkat kalian saya bisa mendapatkan ide yang sudah terealisasikan menjadi tesis, menemani dan tidak men-judge saya ketika tesis saya terbengkalai.

Teman-teman kantor (Mitrais), terimakasih atas dukungan dan supportnya, yang selalu mendukung dan memback-up saya ketika izin cuti karena urusan perkuliahan.

Teman-teman kelas Angkatan 2019 dan 2020 yang sempat menjadi rekan kelompok atau diskusi, walaupun tidak pernah bertemu bertatap muka, terimakasih sudah membuat kelas menjadi berwarna.

Mas Ade, Mbak Mala, Fahry, Mbak Fitriani dan Mas Arif. Terimakasih sudah menjadi teman, mentor, teman curhat dan teman yang selalu bisa saya andalkan ketika saya "lost" dan bingung.

HALAMAN MOTTO

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."

(QS Al Baqarah 286)

"Better late than never"

(Geoffrey Chaucer)

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian yang berjudul "Pengaruh Text Preprocessing Terhadap Analisis Sentimen Opini Pengguna Aplikasi Qasir Dengan Menggunakan Metode Support Vector Machine dan Random Forest" dapat diselesaikan. Penulis menyadari bahwa dalam penyusunan Tesis ini masih banyak terdapat kekurangan, oleh karena itu kritik dan saran yang bersifat membangun serta pengembangan kepada penelitian selanjutnya sangat penulis harapkan demi perbaikan isi tesis ini di kemudian hari.

Terimakasih kepada seluruh pihak yang telah memberikan dukungan dalam penyelesaian penelitian ini, semoga apa yang telah diberikan dapat bernilai sebagai amalan baik. Akhir kata, mari jadikan ilmu pengetahuan sebagai kekuatan yang dapat mengembalikan sistem kehidupan menuju arah kebenaran

Yogyakarta, 1 Maret 2023

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Qasir merupakan aplikasi Point-Of-Sale (POS) berbasis android yang bisa diakses secara gratis pada Google Playstore. Dengan banyaknya aplikasi POS yang tersedia, pengguna akan lebih selektif dalam memilih aplikasi yang akan digunakan. Salah satu aspek yang dapat mempengaruhi keputusan memilih aplikasi adalah opini pada aplikasi tersebut. Opini merupakan informasi yang didapatkan setelah menggunakan aplikasi bisa berisi kritik maupun saran. Sehingga berdasarkan hal tersebut pengguna dapat menyimpulkan bagaimana pengguna lain menggunakan aplikasi tersebut. Selain berguna untuk pengguna, opini jika diolah dengan baik akan menghasilkan sebuah informasi yang dapat digunakan untuk evaluasi bagi tim pengembang.

Untuk menganalisa dan menemukan hubungan antar data yang dimiliki dapat menggunakan Data Mining. Penelitian ini akan menggunakan metode Support Vector Machine dan Random Forest dengan menggunakan 4 skenario pengujian text-preprocessing. Karena masing masing metode memiliki kekurangan dan kelebihannya sehingga pada kedua metode tersebut akan dibandingkan nilai berbagai aspeknya.

Hasil akhir dari penelitian ini adalah metode Support Vector Machine lebih baik digunakan dalam dataset Qasir dengan menggunakan skenario pengujian no 1 dan 2 yang akan menghasilkan nilai Akurasi, AUC, Precision dan Recall terbaik diantara seluruh skenario

Kata Kunci: POS, Klasifikasi, Random Forest, Support Vector Machine, Analisis Sentimen.

ABSTRACT

Qasir is an Android-based Point-Of-Sale (POS) application that can be accessed for free on the Google Playstore. With so many POS applications available, users will be more careful in choosing the application to use. One aspect that can influence the decision to choose an application is the opinion on the application. Opinion is information obtained after using the application that can contain criticism or suggestions. So based on this the user can conclude how other users use the application. Apart from being useful for users, opinions if processed properly will produce information that can be used for evaluation for the development team.

To analyze and find relationships between data, we can use Data Mining. This research will use the Support Vector Machine and Random Forest methods using 4 text-preprocessing test scenarios. Because each method has its advantages and disadvantages, the values of various aspects will be compared between the two methods.

The final result of this research is that the Support Vector Machine method is better used in the Qasir dataset by using test scenarios no. 1 and 2 which will produce the best Accuracy, AUC, Precision and Recall values among all scenarios.

Keywords *: POS, Classification, Random Forest, Support Vector Machine, Sentiment Analysis*

BAB I

PENDAHULUAN

1.1. Latar Belakang

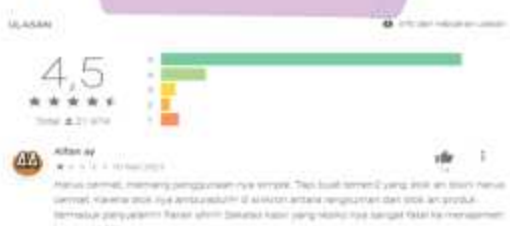
Google Play Store merupakan sebuah aplikasi pada smartphone dimana berbagai jenis kategori aplikasi dapat diunduh secara berbayar maupun gratis. Dalam 1 kategori saja, terdapat ratusan bahkan ribuan aplikasi sejenis dengan masing-masing kelebihan dan kekurangannya. Akan tetapi hal ini menyebabkan pengguna dibuat kebingungan untuk meng-install aplikasi mana yang paling berfungsi dengan baik dikarenakan banyaknya pilihan yang tersedia.

Dalam Google Playstore terdapat kolom rating dan opini pada setiap aplikasi yang memuat berbagai opini dari pengguna yang sudah menggunakan aplikasi, baik saran maupun kritik. Pengguna dapat menambahkan opini dengan maksimal 500 karakter serta rating sesuai dengan pengalaman yang mereka rasakan saat menggunakan aplikasi. Pada rating ini sendiri terdiri dari bintang 1 untuk nilai yang tidak memuaskan sampai dengan bintang 5 untuk nilai yang sangat memuaskan.

Qasir merupakan aplikasi Point-Of-Sale (POS) berbasis android yang bisa diakses secara gratis pada Google Playstore. Aplikasi ini diciptakan untuk membantu usahawan dalam mencatat penjualan, mengelola produk, mengawasi stok dan memantau laporan transaksi. Diawali pada tahun 2015 dan terus berkembang dari tahun ke tahun, saat ini Qasir sudah di download lebih dari 500 ribu lebih pengguna. Yang membedakan aplikasi Qasir dengan aplikasi lainnya adalah aplikasi Qasir mudah digunakan bahkan oleh pemula sekalipun.

Akan tetapi pada saat ini tidak hanya satu atau dua aplikasi POS yang tersedia dan mudah digunakan. Banyaknya aplikasi ini akan membuat pengguna lebih selektif dalam menentukan aplikasi yang akan digunakan untuk mengelola usahanya. Sehingga salah satu aspek yang dapat digunakan untuk mempengaruhi pengguna dalam memilih aplikasi yang sesuai dengan melihat opini dan rating pengguna lain yang terdapat pada halaman aplikasi tersebut. Karena jika rating aplikasi tersebut berada di bawah aplikasi sejenis dan banyak opini negatif maka pengguna akan berpikir 2 kali untuk menggunakan aplikasi tersebut. Berbeda keadaannya jika rating aplikasi tersebut termasuk bagus dan banyak opini positif maka pengguna akan lebih percaya untuk menggunakan aplikasi tersebut.

Selain berpengaruh terhadap pengguna dan calon pengguna aplikasi, opini dan rating ini sangat berpengaruh terhadap pengembang aplikasi. Untuk mempertahankan kualitas aplikasi maka pengembang wajib mengetahui komentar pengguna agar aplikasi jauh lebih baik lagi dari sebelumnya. Kemudian, jika opini-opini ini dikumpulkan lalu diolah maka hasilnya dapat dijadikan kesimpulan bagi perusahaan untuk menentukan sentimen dari aplikasi Qasir dimata pelanggan ini sudah termasuk baik ataupun belum sehingga dapat dijadikan bahan evaluasi kedepannya agar aplikasi Qasir semakin baik.



Gambar 1. 1 Contoh Rating dan Opini aplikasi Qasir pada Google Play

Menganalisis opini untuk mengetahui kinerja atau performa aplikasi bisa dilakukan dengan mudah menggunakan total rating pada aplikasi tersebut, akan tetapi rating tidak bisa mewakili keseluruhan isi opini pengguna sehingga dibutuhkan melihat isi dari opini untuk mendapatkan maksud dari opini pengguna tersebut (Muktafin, Kusrini, & Luthfi, 2020). Aplikasi Qasir memiliki opini dari 21.974 pengguna, jika harus menganalisis opini secara manual dengan melihat satu persatu opini pengguna maka akan membutuhkan waktu yang sangat panjang sehingga dibutuhkan metode yang dapat menganalisis sentimen opini pengguna secara otomatis. Hal ini dapat menanggulangi opini opini yang tidak sesuai dengan rating yang diberikan seperti pada gambar 2. Dan juga menghindari opini yang tidak memberikan penilaian terhadap aplikasi (baik kekurangan maupun kelebihan) bahkan opini yang tidak sesuai dengan topik atau aplikasinya.

Gambar 1. 2 Contoh opini pengguna aplikasi Qasir

Untuk menganalisa dan menemukan hubungan antara data data yang dimiliki dapat menggunakan Data Mining. Dalam prosesnya banyak teknik yang digunakan yaitu statistik, matematika, kecerdasan buatan dan machine learning yang berfungsi untuk mengekstraksi dan identifikasi informasi dari data yang ada. Hasil akhir dari Data Mining berupa model (models) atau pola (pattern). Terdapat dua pendekatan pada Data Mining yaitu supervised learning dan unsupervised learning. Perbedaan antara keduanya adalah adanya data training. (Prasetyowati, 2017)

Diantara banyaknya metode dalam supervised learning yang digunakan untuk menganalisis opini berbentuk teks, yang akan digunakan dalam penelitian kali ini adalah metode Support Vector Machine (SVM) dan Random Forest. Metode Support Vector Machine (SVM) menggunakan konsep mencari hyperline terbaik sebagai pemisah antara dua class data. Metode ini memiliki kelebihan diantaranya implementasi yang mudah, kemampuan untuk mengklasifikasikan pattern yang tidak masuk dalam class dalam training set, kemudian dapat meminimalisir error pada training set, dan kemampuan untuk menghadapi masalah dalam pattern (Sianturi, Hasugian, Simangunsong, & Nadeak, 2020)

Sedangkan Random Forest merupakan pengembangan dari metode CAST. Metode ini akan membangun banyak tree dalam mengklasifikasikan suatu objek berdasarkan atributnya. Keuntungan menggunakan metode ini adalah dapat mengcover dataset besar dengan dimensi tinggi. Mampu mengestimasi missing data dan mempertahankan akurasi secara efektif ketika jumlah missing data yang banyak hal ini dikarenakan Random Forest mempunyai metode tersendiri untuk balancing error. (Tahyudin, 2020)

Pada penelitian sebelumnya yang menganalisis sentimen menggunakan Support Vector Machine (SVM) dengan jumlah dataset 3000 opini aplikasi pada google playstore dan metode evaluasi menggunakan k-fold cross validation yang bernilai 10 mendapatkan nilai akurasi tertinggi sebesar 90,67% dengan menggunakan metode pre-processing data yaitu Tokenization, transform case, Filter tokens (by length), Filter stop words, stemming dan Labeling. (Ahmadi, Apriani, Kurniasari, Handayani, & Gustian, 2020)

Selanjutnya merupakan penelitian terdahulu yang menggunakan metode modifikasi dari Random Forest yaitu Modified Balanced Random Forest. Hal ini dikarenakan data yang dihadapi merupakan imbalanced data yang nantinya akan mempengaruhi performa dari analisis yang akan dibangun. Metode dari pre-processing yang digunakan adalah case folding, cleaning, tokenization, stopword removal, stemming dan remove unknown word. Dari penelitian ini menghasilkan nilai akurasi tertinggi sebesar 79% dengan nilai F1-Scores 74% (Zamzami & Adiwijaya, 2021)

Penelitian sebelumnya membandingkan performa dari metode Naïve Bayes dan Support Vector Machine dengan menggunakan Particle Swarm Optimization (PSO) dan Dataset yang digunakan sebanyak 1364 opini. Hasil penelitian menunjukkan bahwa performa dari metode PSO - Support Vector Machine lebih baik dengan nilai akurasi 93% dan nilai AUC 97% dibandingkan PSO - Naïve Bayes dengan nilai akurasi 69% dan nilai AUC 65%. (Mustopa, et al., 2021)

Kemudian terdapat penelitian sebelumnya yang membandingkan kinerja antara metode metode Random Forest, Naïve Bayes dan Support Vector Machine dengan tweet sebagai dataset dan menggunakan metode pembobotan kata TD-IDF menghasilkan akurasi Support Vector Machine tertinggi diantara metode lainnya dengan nilai 77,58% disusul dengan Random Forest sebesar 75,81% dan Naïve Bayes sebesar 75,22%. Teknik pra - pemrosesan data yang digunakan ada Case-folding, cleaning text, Tokenize, Stemming dan Stopword removal. (Himawan & Eliyani, 2021).

Selanjutnya terdapat penelitian yang membandingkan Naïve Bayes, Support Vector Machine, Decision Trees dan Random Forest menggunakan 400.000 opini amazon sebagai dataset dengan perbandingan 80% untuk data training dan 20% untuk data testing. Metode Pre-processing yang digunakan adalah Tokenization, Removal Stopwords dan Stemming. Menghasilkan metode Support Vector Machine dengan akurasi tertinggi 89% dan sebagai metode yang komplit dikarenakan memiliki nilai yang tinggi untuk semua atribut evaluasi seperti Akurasi, Presisi, Recall dan F1 Score. Disusul dengan metode Random Forest yang menghasilkan akurasi tertinggi 88% kemudian metode Naïve Bayes dengan nilai 85% dan terakhir metode Decision Trees dengan nilai 82%. (Guia, Silva, & Bernardino, 2019)

Yang terakhir adalah penelitian sebelumnya yang membandingkan Naïve Bayes, Random Forest dan Support Vector Machine menggunakan dataset opini aplikasi Ruang Guru pada Google Playstore dengan data sebanyak 1629. Dengan menggunakan metode pre-processing Tokenization, Stemming, Case Folding dan Stopword Removal menghasilkan bahwa Metode Random Forest memiliki akurasi tertinggi dengan nilai 97,16% dan AUC sebesar 0,996 kemudian SVM dengan nilai akurasi 96,01% dengan AUC sebesar 0,543 dan yang terakhir adalah Naïve Bayes dengan akurasi 94,16% dengan AUC sebesar 0,999 (Fitri, Yuliani, Rosyida, & Gata, 2020)

Dalam analisis sentimen menggunakan data text yang dilakukan oleh peneliti sebelumnya dengan berbagai macam metode didapatkan kesimpulan bahwa rata rata untuk metode dengan nilai akurasi tertinggi didapatkan oleh Support

Vector Machine ataupun Random Forest. Sehingga berdasarkan uraian diatas, pada penelitian ini akan membandingkan performa dari kedua metode tersebut (Support Vector Machine dan Random Forest). Selain itu pada setiap penelitian memiliki tahapan pre-prosesing data yang berbeda beda akan menghasilkan hasil akhir yang berbeda juga, hal ini akan menjadi salah satu fokus dari penelitian ini selain membandingkan performa dari kedua metode yang dipilih, penelitian ini akan menganalisis lebih jauh mengenai tahapan apa dalam pre-processing yang dapat meningkatkan kinerja dan juga hasil dari analisis sentimen. Setelah mengetahui performa dan tahapan preprocessing di masing masing metode, yang terakhir adalah menarik kesimpulan dari hasil penelitian tersebut manakah metode yang bekerja paling baik dari keseluruhan tahapan yang telah dilakukan.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan penulis, maka rumusan masalah yang bisa disimpulkan adalah:

a. Apakah algoritma Support Vector Machine dan Random Forest memiliki performa yang baik dalam menganalisis sentimen opini pengguna aplikasi qasir?
b. Apakah tahapan preprocessing yang berpengaruh dalam performa Support Vector Machine dan Random Forest dalam menganalisis sentimen opini pengguna aplikasi Qasir?
c. Metode apa yang bekerja paling baik di keseluruhan skenario tahapan preprocessing?

1.3. Batasan Masalah

Terdapat batasan dalam penelitian yang dilakukan agar penelitian terfokus pada permasalahan yang telah disampaikan sebelumnya. Batasan masalah pada penelitian ini adalah :

a. Objek penelitian adalah PT Solusi Teknologi Niaga (Qasir.id).
b. Penelitian ini menggunakan data opini pengguna Qasir yang ada pada google playstore.
c. Pengambilan opini yang dianalisis merupakan opini dari tahun 2020 - 2022 (Januari 2020 - Desember 2022).
d. Opini yang dianalisa adalah opini berbahasa inggris.
e. Opini yang berisi *Emoticon* atau simbol, hanya terdiri dari satu kata yang berulang tidak digunakan untuk analisis.
f. Algoritma yang digunakan adalah Support Vector Machine (SVM) dan Random Forest.
g. Tahapan pengambilan data akan menggunakan teknik *web scraping* dengan menggunakan phython.
h. Tahapan *Preprocessing* akan dilakukan dalam beberapa skenario yaitu
    a. Case folding, remove frequent dan remove rare word.
    b. Case folding, cleaning, remove frequent dan remove rare word.
    c. Case folding, cleaning, stop word, remove frequent dan remove rare word.
    d. Case folding, cleaning, stop word, remove frequent dan remove rare word, dan stemming.

i. Visualisasi akan ditampilkan dengan word cloud.

j. Pengujian dari penelitian ini akan menggunakan *K-Fold Cross Validation* dan *Confusion Matrix*.

k. Output penelitian ini akan digunakan sebagai bahan evaluasi yang kedepannya bisa digunakan sebagai pertimbangan dalam tahap development aplikasi.

1.4. Tujuan Penelitian

Maksud dan tujuan penelitian ini adalah untuk mengetahui perbandingan performa dari analisis sentimen terhadap opini pengguna aplikasi Qasir di google playstore dengan menggunakan metode *Support Vector Machine* dan *Random Forest*.

1.5. Manfaat Penelitian

Manfaat penelitian ini adalah:

a. Dapat digunakan sebagai bahan evaluasi dan *improvement* kedepannya oleh tim pengembang.

b. Dapat digunakan untuk mengukur tingkat kepuasan user melalui opini yang ditulis pada google playstore.

c. Dapat berkotribusi terhadap pengembangan aplikasi dan perusahaan.

d. Dapat dijadikan sarana sebagai pengembangan ilmu pengetahuan dibidang teknologi informasi khususnya dibidang Analisis Sentimen.

e. Dapat digunakan sebagai acuan pagi peneliti lain yang melakukan penelitian sejenis.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Dalam 3 tahun terakhir, terdapat beberapa penelitian yang membahas mengenai analisis sentimen dengan menggunakan metode metode yang akan digunakan pada penelitian ini. Pada penelitian [1] ini melakukan perbandingan opini terhadap beberapa online shop yaitu Tokopedia, JD.ID, Blibli, Shopee dan Lazada untuk mengetahui aplikasi marketplace terbaik yang tersedia pada playstore. Data yang digunakan dalam penelitian ini sejumlah 1500 ulasan dengan perbadingan 50:50 antara ulasan positif dan negative. Sebelum data data diproses menggunakan metode yang sudah ditetapkan, sebelumnya melalui proses data untuk meningkatkan efektivitas dalam proses selanjutnya. Tahapan yang digunakan dalam penelitian ini adalah Tokenization, transform case (Case folding), Filter tokens (min 2- max 25 word), Filter stop words, stemming dan Labeling (positif dan negative). Setelah itu, penelitian ini membagi data kedalam dua bagian yaitu data testing dan data training. Tools yang digunakan adalah Rapidminer versi 9.8. Kemudian masuk kedalam tahap terakhir yaitu evaluasi dengan *k-fold cross validation* dengan k = 10 dan mendapatkan nilai akurasi tertinggi 90,67% untuk *marketplace* Tokopedia dan nilai akurasi terendah didapatkan oleh *marketplace* Lazada dengan nilai 69%. Akan tetapi untuk nilai tertinggi AUC didapatkan oleh JD.ID dengan nilai 85% dan nilai AUC terendah tetap didapatkan oleh *marketplace* Lazada dengan nilai 74%. Untuk tahapan pre-processing mana yang memiliki

pengaruh terhadap penelitian ini tidak dijelaskan dan dijabarkan sehingga hasil tersebut merupakan hasil keseluruhan dari pre-processing dan metode algortima yang digunakan. (Ahmadi, Apriani, Kurniasari, Handayani, & Gustian, 2020)

Selanjutnya merupakan penelitian [2] menggunakan metode modifikasi dari *Random Forest* yaitu *Modified Balanced Random Forest* dikarenakan data yang dihadapi merupakan *imbalanced data* yang nantinya akan mempengaruhi performa dari analisis yang akan dibangun. Model klasifikasi ini sangat bergantung kepada pemilihan parameter, hal ini bergantung kepada jika data yang dimiliki mempunyai dimensi tinggi, tidak terstuktur dan memiliki fitur yang banyak. Karena hal itu, maka penelitian ini menggunakan *Mutual Information* sebagai seleksi fiturnya. Data yang digunakan sebesar 5000 yang berasal dari IMDb, yang terdiri dari 4000 data label negative dan 1000 data label positif. Tahapan preprocessing yang dilakukan adalah case folding, cleaning, tokenization, stopword remola, stemming dan remove unknown word. Kemudian untuk mengidentifikasi fitur pada dokumen menggunakan N-gram, setelah itu masuk kedalam proses pembobotan yang menggunakan adalah TF-IDF. Setelah melakukan ekstraksi data, selanjutnya mengukur ada tidaknya informasi dari sebuah kata kunci atau *term* dengan menggunakan *Mutual Information*. Terdapat 3 skenario pengujian, yang pertama dengan mengukur pengaruh stemming pada dataset yang menghasilkan peningkatan 1% dengan menggunakan stemming. Kemudian scenario kedua untuk mengetahui pengaruh seleksi feature dalam dataset dengan sebelumnya menentukan threshold sebesar 0.01. Setelah itu didapatkan dengan menggunakan seleksi fitur dapat meningkatkan performa akurasi sebesar 1%. Dan yang terakhir

adalah perbandingan performansi Modified Balanced Random Forest dengan Random Forest yang menghasilkan Modified Balanced Random Forest memiliki nilai f1-scores 74%. Setelah melalui 3 skenario pengujian, maka dapat disimpulkan metode Modified Balanced Random Forest dapat menghasilkan nilai akurasi tertinggi sebesar 79% dengan nilai F1-Scores 74%. (Zamzami & Adiwijaya, 2021)

Kemudian penelitian [3] meruapakn penelitian yang membandingkan metode Support Vector Machine dengan Naïve Bayes berdasarkan Particel Swarm Optimization (SPO). *Dataset* yang digunakan sebanyak 1364 opini merupakan jumlah dari 3 bulan (April 2020 – Juni 2020) opini pengguna pada aplikasi Peduli Lindungi. Pada penelitian ini menggunakan 2 step dalam preprocessingnya, yang pertama menggunakan Gata Framework dan yang kedua menggunakan RapidMiner. Gata Framework digunakan untuk mengolah data yang mengandung kata kata berbasa Indonesia, karena pada RapidMiner belum tersedia. Sehingga setelah menggunakan Gata Framework kemudian dataset akan dipre-processing kembali menggunakan RapidMiner. Tahapan pada pre-processing pada RapidMiner adalah Remove duplicates, Nominal to Text, Transform Case (Case Folding), Filter Token (By Length with min 4 chars – 25 chars) dan Filter Stopward (Dictianory). Setelah pre-processing data akan masuk kedalam modelling stage dengan menggunakan RapidMiner 9.1 version. Setelah itu untuk evaluasi metode yang digunakan adalah *k-fold cross validation* dengan nilai k = 10. Dari hasil pengolahan keseluruhan dan evaluasi tersebut menunjukkan bahwa performa dari metode PSO - *Support Vector Machine* lebih baik dengan nilai akurasi 93% dan

nilai AUC 97% dibandingkan PSO - *Naïve Bayes* dengan nilai akurasi 69% dan nilai AUC 65%. (Mustopa, et al., 2021)

Kemudian penelitian [4] ini menggunakan *tweet* sebagai *dataset* dengan jumlah data 14208 baris data yang diambil dari akun @dkijakarta atau akun official Pemerintah Provinsi DKI Jakarta di twitter dalam rentang waktu 9 April 2020 sampai dengan 15 April 2020. Dan seluruh tweet ini akan diolah dengan menggunakan Bahasa Indonesia. Setelah itu data akan dibagi menjadi 3 class, yaitu positif, netral dan negative yang akan diproses dengan 3 metode yaitu SVM, Naïve Bayes dan Random Forest dengan menggunakan metode pembobotan TD-IDF. Metode dengan hasil terbaik akan digunakan untuk memprediksi sentiment kosong. Sebelum itu, data akan masuk kedalam pra-premrosesan data yaitu case folding, cleaning text, tokenization, stemming dan stopword removal. Pada metode SVM akan menggunakan salah satu *class* yang tersedia yaitu LinearSVC. Kemudian pada metode Naïve Bayes akan menggunakan Multinomial Naïve Bayes. Menghasilkan akurasi *Support Vector Machine* tertinggi diantara metode lainnya dengan nilai 77,58% disusul dengan *Random Forest* sebesar 75,81% dan Naïve Bayes sebesar 75,22%. Dan untuk prediksi data kosong menggunakan metode SVM mendapatkan hasil class netral sebesar 83,6%, positif 7,6% dan negative 8,8% (Himawan & Eliyani, 2021)

Selanjutnya terdapat penelitian [5] menggunakan 400.000 *review* amazon sebagai *dataset* dengan perbandingan 80% untuk data *training* dan 20% untuk data *testing*. Pada data terdapat *rating* dan *review*, *rating* akan dibagi menjadi 3 yaitu *rating* 1-2 adalah *Negative*, *rating* 4 -5 adalah Positif sedangkan *rating* 3 akan

dihilangkan. Untuk tahapan preprocessing yang dilakukan adalah *convert to lowercase*, *remove html tags and punctuations*, *tokenization*, *stemming* dan *removing stopwords*. Untuk evaluasi metode *Random Forest* akan menggunakan 50,100,200 dan 400 pohon keputusan/*trees*. Kemudian untuk metode Naïve Bayes akan menggunakan *Multinomial* dan *Bernoulli*. Selanjutnya untuk metode SVM akan menggunakan RBF dan Linier. Hasilnya adalah untuk metode Naïve Bayes hasil yang lebih baik didapatkan oleh Bernoulli, kemudian metode *Random Forest* bisa mendapatkan hasil yang baik walaupun trees yang digunakan sedikit (50), akan tetapi untuk hasil yang lebih baik ditemukan pada 200 trees. Kemudian untuk metode SVM lebih baik menggunakan Linier kernel dibandingkan dengan RBF. Dan hasil performa dari masing masing metode adalah *Support Vector Machine* dengan akurasi tertinggi 89% dan sebagai metode yang komplit dikarenakan memiliki nilai yang tinggi untuk semua atribut evaluasi seperti Akurasi, Presisi, *Recall* dan F1 *Score*. Disusul dengan metode *Random Forest* yang menghasilkan akurasi tertinggi 88% kemudian metode *Naïve Bayes* dengan nilai 85% dan terakhir metode *Decision Trees* dengan nilai 82%. Dan untuk *impact* dari merk dan harga bahwa ZTE merupakan brand dengan positif *review* terbanyak 82,9% dan 1000$ hingga 5000$ merupakan *range* harga dengan positif *review* terbanyak 84,3%. (Guia, Silva, & Bernardino, 2019)

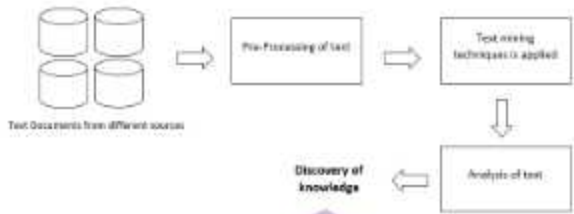
Pada penelitian [6] menggunakan *dataset review* aplikasi Ruangguru pada Google Playstore dengan rentang waktu 01 Maret sampai dengan 30 Maret 2020 dengan data yang diambil sebanyak 1629. Setelah itu data akan masuk kedalam *preprocessing* yaitu *tokenization*, *stemming*, *case folding* dan *stopword removal*.

Kemudian saat pelabelan ditemukan ketidakseimbangan pada class sehingga diterapkannya metode SMOTE untuk mengatasi hal ini. Setelah dilakukan proses ini jumlah dataset sebesar 1515 instance. Pada penelitian ini menggunakan aplikasi RapidMiner untuk memproses pada masing masing algoritma. Setelah itu dilakukan pengujian dengan menggunakan K-Fold 10 dan Cross Validation kemudian hasil dari penelitian tersebut menunjukkan bahwa Metode *Random Forest* memiliki akurasi tertinggi dengan nilai 97,16% dan AUC sebesar 0,996 kemudian SVM dengan nilai akurasi 96,01% dengan AUC sebesar 0,543 dan yang terakhir adalah Naïve Bayes dengan akurasi 94,16% dengan AUC sebesar 0,999. Pada penelitian ini berhasil meningkatkan akurasi dari penelitian sebelumnya sebesar 7,16% akan tetapi tidak dibahas mengenai tahapan apa yang mempengaruhi hasil tersebut. (Fitri, Yuliani, Rosyida, & Gata, 2020).

2.2. Landasan Teori

2.2.1. Text Mining

Berfungsi untuk menghasilkan informasi dari sekumpulan dokumen dengan sumber data yang tidak terstruktur dalam jumlah besar. Sumber data ini dapat diperoleh dari dokumen berbentuk teks, word, pdf, atau format lainnya seperti email, news, media sosial, dan lain sebagainya. (Wanto, Anjar., 2020). Proses pengambilan informasi bisa menghasilkan sebuah analisis sentimen dengan detail identifikasi pernyataan bersifat negatif atau positif (Mesran, et al., 2020). *Text mining* terdiri dari tiga bagian yaitu *text preprocessing, feature selection* dan *text analytic*. (Nugraha, Habibi, & Harani, 2020)



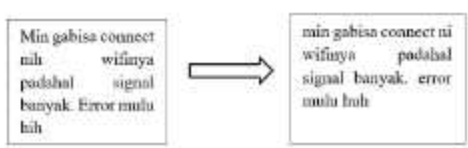
Gambar 2. 1 Bagian utama dari *Text Mining* (Nugraha, Habibi, & Harani, 2020)

2.2.2. Text Preprocessing

Merupakan tahapan untuk mempersiapkan data sebelum digunakan pada proses selanjutnya dengan cara membersihkan data teks mentah, karena data yang diperoleh biasanya bersifat tidak terstruktur dan memiliki banyak *noise* contohnya imbuhan, tanda baca, karakter khusus, angka, dan lain sebagainya. Sehingga pada *text-processing* ini data akan diolah dan menghasilkan bentuk dasar dari masing masing kata untuk keperluan analisis. Berikut merupakan tahapan yang dilakukan pada *text-processing* (Nugraha, Habibi, & Harani, 2020):

A. Case Folding

Berfungsi untuk mengkonversi keseluruhan teks menjadi bentuk *lowercase* (format penulisan huruf kecil). (Nugraha, Habibi, & Harani, 2020) Akan tetapi yang melalui proses ini hanya alfabet saja (huruf 'a' sampai 'z') yang diterima selain itu karakter bisa dihilangkan atau dianggap sebagai pembatas. (Julianto, Bintari, & Indrianti, 2017)



Gambar 2. 2 Contoh Case Folding (Julianto, Bintari, & Indrianti, 2017)

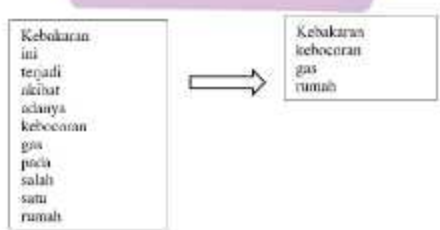
B. Cleaning

Berfungsi untuk menghapus karakter selain ketentuan yang ditentukan seperti imbuhan, karakter selain A – Z, tanda baca, karakter khusus, angka, url (link) dan lainnya. Proses *cleaning* ini dapat memperbaiki atau menangani *missing value.* (Wanto, et al., 2020) Pada tahap ini bisa jadi ketentuan yang diterapkan pada setiap penelitian berbeda antara satu dengan yang lainnya menyesuaikan dengan kebutuhan data yang akan digunakan dalam proses analisis berikutnya.

Gambar 2. 3 Contoh Cleaning (Julianto, Bintari, & Indrianti, 2017)

C. Stopword

Hanya mengambil kata kata penting pada proses sebelumnya. Dalam tahapan ini bisa juga menggunakan metode stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata yang penting). Biasanya untuk metode stoplist peneliti akan menyediakan daftar stoplist yang akan digunakan. Kemudian menambahkan kata kata yang tidak mengandung arti seperti 'ah', 'aaa', 'yey', 'yeah' dan lain sebagainya. (Khairunnisa, Adiwijaya, & Faraby, 2021)



Gambar 2. 4 Contoh Stopword (Julianto, Bintari, & Indrianti, 2017)

D. Stemming

Merupakan proses untuk mengubah term ke bentuk akar katanya. Stem (akar kata) adalah bagian kata yang tersisa setelah dihilangkannya imbuhan (awalan, sisipan, akhiran, kombinasi awalan dan akhiran) (Jumeilah, 2017) Contohnya jika dalam Bahasa inggris, dapat dianalogikan seperti menghilangkan akhiran 's', 'es', 'ing', 'ed'. Akan tetapi akan menjadi sedikit kompleks ketika menghilangkan awalan dan akhiran dalam Bahasa Indonesia seperti buku-buku merupakan hasil jamak dari buku, atau pemerintah merupakan turunan kata dari "perintah" dan terdapat sisipan "em". (Julianto, Bintari, & Indrianti, 2017) Berikut merupakan contoh imbuhan dari Bahasa Inggris yaitu (Suryawinata & Hariyanto, 2016):

- Awalan
  - Berarti tidak atau lawan kata bisa menggunakan 'non-', 'dis-', 'mis-', 'in-', dan lain lain.
  - Berarti lagi/kembali menggunakan 're-'
  - Berarti bekas atau keluaran menggunakan 'ex-'
- Akhiran
  - Untuk membuat kata benday aitu '-er', '-or', '-ness', 'ing'
  - Untuk kata sifat menggunakan '-able,'-less','-ly','ed'
  - Untuk kata kerja menggunakan '-en','-ate'

2.2.3. *Feature Selection*

Merupakan tahapan paling penting dalam keseluruhan proses *text mining* dikarenakan pada tahapan ini seluruh *features* dalam *dataset* yang tidak relevan akan dihapus. Pada tahap ini juga berperan untuk menentukan kata kunci (*term*) yang bisa membedakan suatu dokumen dengan dokumen lainnya dalam satu *corpus*. (Nugraha, Habibi, & Harani, 2020) Terdapat empat pendekatan yang digunakan dalam tahapan ini, yaitu :

A. Document Frequency (DF)

Pendekatan yang digunakan dalam tahap ini adalah membuang *term* umum pada dokumen. Kemudian di dalam dokumen hanya akan tersisa *term* dengan tingkat *overlapping* rendah. (Nugraha, Habibi, & Harani, 2020)

B. Term Frequency (TF)

Pada metode ini hanya menghitung kemunculan dari *term* pada dokumen sehingga *term* dengan frekuensi kemunculan tertinggi akan menjadi tanda dari dokumen tersebut. (Nugraha, Habibi, & Harani, 2020) Secara mudahnya, pada tahap akan menghitung seberapa banyak kemunculan sebuah kata dalam sebuah dokumen. (Khairunnisa, Adiwijaya, & Faraby, 2021) Bobot dari kata t pada dokumen d akan dijelaskan pada Persamaan 1 dimana $f(t_i, d_j)$ adalah frekuensi kemunculan term t_i pada dokumen d_j. (Holle, 2015)

$$TF(t_i, d_j) = f(t_i, d_j) \qquad (1)$$

C. Inverse Document Frequency (IDF)

Konsep yang digunakan sama dengan TF, akan tetapi pada IDF ini akan menghitung kemunculan *term* di dalam keseluruhan *corpus* dokumen. (Nugraha,

Habibi, & Harani, 2020) Atau dalam artian lain ditahap ini akan menghitung bagaimana suatu kata didistribusikan pada koleksi dokumen. (Khairunnisa, Adiwijaya, & Faraby, 2021) Faktor IDF dari term *t* yang akan dijabarkan pada Persamaan 2 dimana *N* merupakan jumlah seluruh dokumen dan *df(t)* adalah jumlah dokumen yang mengandung *term t*. (Holle, 2015)

$$IDF\,(t_i) = 1 + log(N/df(t_i)) \qquad (2)$$

D. Term Frequency/Inverse Document Frequency (TF/IDF)

Pendekatan ini merupakan kombinasi antara TF dan IDF yaitu perhitungannya akan mengambil resiko diantara TF dan IDF. (Nugraha, Habibi, & Harani, 2020) Kombinasi bobot dari *term t* pada dokumen *d* sebagaimana pada persamaan 3 dengan $TF(t_i, d_j)$ adalah frekuensi term ke-1 dan $IDF\,(t_i)$ adalah inverse kemunculan term ke-*i* pada dokumen ke-*j*.

$$TFIDF\,(t_i, d_j) = TF\,(t_i, d_j) \times IDF\,(t_i) \qquad (3)$$

Dengan adanya perumusan tersebut makan bobot akan semakin tinggi saat banyak ditemukan dalam dokumen. *Term* yang sering muncul dalam dokumen, tapi jarang muncul pada keseluruhan dataset akan diberi nilai bobot yang lebih tinggi. (Holle, 2015)

2.2.4. Text Analytic

Merupakan tahapan akhir dari *text mining*. Hasil dari proses tahapan sebelumnya yang sudah di *cleansing* dan diidentifikasi kemudian akan diolah dengan menggunakan algoritma yang sesuai dengan kebutuhan analisis. Dua jenis text analytical yang sering digunakan adalah (Nugraha, Habibi, & Harani, 2020):

A. Topic Modeling

Pendekatan untuk mengelompokkan teks kedalam beberapa kategori berdasarkan tingkat kesamaan term dan kata kunci secara otomatis.

B. Sentiment Analysis

Sebuah proses untuk menentukan sentimen dan mengelompokkan polaritas teks ke dalam kalimat sehingga dapat menentukan kategori dari teks berupa sentimen positif, negatif maupun netral. Analisis sentimen ini juga bisa dibilang *option mining* dikarenakan berfokus pada pendapat dengan kategori negatif atau positif. (Mesran, et al., 2020)

2.2.5. Machine Learning

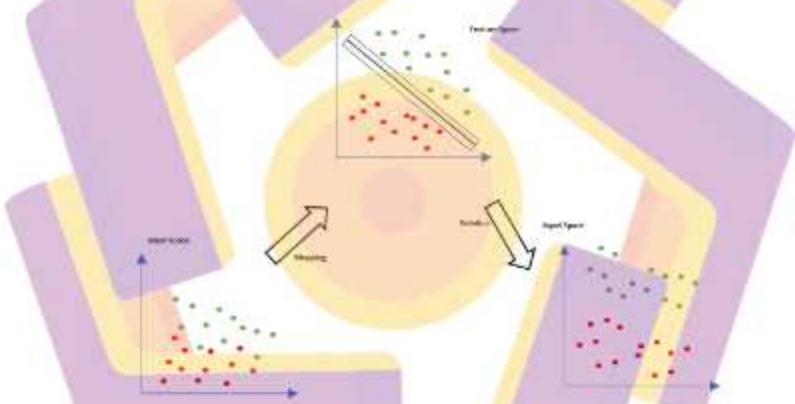
Bagian dari kecerdasan buatan yang berfungsi untuk membuat sistem dengan kemampuan belajar otomatis berdasarkan dari pengalaman (data data terdahulu yang dimiliki) dan dapat meningkatkan kemampuannya sendiri secara eksplisit tanpa harus melalui proses *development* ulang. (Kusuma, 2020)

2.2.6. Klasifikasi

Merupakan proses untuk menyatakan suatu objek atau fungsi yang menjelaskan suatu kelas data, dengan tujuan untuk memperkirakan kelas ke salah satu kelas yang sudah didefinisikan sebelumnya. Klasifikasi bertujuan agar data yang tidak diketahui kelasnya dapat dikelompokkan secara akurat (Suyanto, 2017).

2.2.7. *Support Vector Machine* (SVM)

Merupakan teknik prediksi, klasifikasi maupun regresi yang berkembang sejak tahun 1960-an. Metode ini banyak digunakan pada banyak aplikasi seperti bioinformatika, pengenalan tulisan tangan, dan sebagainya. Implementasi dari SVM ini memerlukan *training* dan *testing* sehingga digolongkan dalam *supervised learning*. Konsep dasar SVM adalah untuk memaksimalkan *margin* yaitu jarak yang memisahkan antarkelas data dengan mencari *hyperplane* terbaik. (Werdiningsih, Nuqoba, & Muhammadun, 2020)

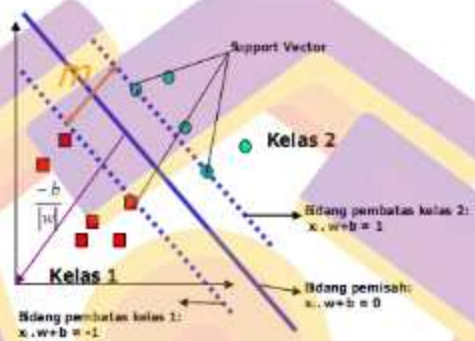


Gambar 2. 5 Konsep dasar algoritma SVM (Nomleni, 2015)

Pada gambar 2.6 terlihat bahwa dasar dari klasifikasi SVM ialah menemukan fungsi pemisah (hyperlane) terbaik diantara dua kelas yang dapat dilakukan dengan cara mencari margin dari hyperlane tersebut dan mencari juga titik maksimalnya. Pada awalnya prinsip kerja dari SVM itu mengklasifikasi secara linier saja, akan tetapi teknologi semakin berkembang begitu pula dengan SVM sehingga saat ini SVM dapat bekerja pada klasifikasi non liniear.

A) Klasifikasi Liniear

Sederhananya SVM diartikan sebagai usaha untuk mencari hyperlane terbaik yang berfungsi sebagai pemisah dua buah kelas pada input space. Dua kelas tersebut adalah +1 dan -1. Digambarkan pada gambar 9 untuk kelas -1 adalah symbol kotak merah sedangkan kelas +1 untuk symbol kotak hijau.



Gambar 2. 6 Klasifikasi Linear SVM (Widyawati, 2012)

Pada gambar diatas dijelaskan bahwa untuk mengklasifikasi dan mendapatkan hasil terbaik dengan menggunakan cara mengukur margin hyperlane dan mencari titik maksimalnya. Margin adalah jarak antara hyperlane terdekat dengan pattern terdekat dari masing masing kelas dan pattern yang paling dekat dengan hyperlane disebut support vector. Untuk mencari margin dapat dirumuskan menjadi

$$m = \frac{1-b-(-1-b)}{|w|} = \frac{2}{|w|} \qquad (4)$$

Dimana m merupakan margin atau jarak antara dua bidang, kemudian w merupakan bidang normal, b merupakan posisi relative terhadap origin. Jarak garis dirumuskan wx+b=c dan ke origin adalah (c-b)/|w|. Kemudian margin m dapat

dimaksimalkan dengan memenuhi konstrain 2 bidang pembatas yang sejajar. Bidang pembatas kelas 1 membatasi kelas 1 dan sebaliknya untuk kelas 2 sehingga

$$x_i.w + b \geq +1 \; for \; y_i = +1$$

$$x_i.w + b \geq +1 \; for \; y_i = -1 \qquad (5)$$

Nilai maksimal dari margin harus memenuhi persamaan (4) dan (5) dan nilai b dan w dikalikan dengan sebuah konstanta yang akan menghasilkan nilai margin yang dikalikan dengan konstanta yang sama. Konstrain adalah scaling constraint yang dipenuhi dengan rescaling b dan w, sehingga pada persamaan (6) memaksimalkan dan menimimalkan w dirumuskan dalam pertidaksamaan.

$$y_i(x_i.w + b) - 1 \geq \quad (6)$$

Untuk mencari nilai margin terbesar dapat dirumuskan menjadi masalah optimasi konstrain yaitu:

$$min \frac{1}{2} |w|^2$$

$$s.t \; y_i \, (x_i . w + b) - 1 \geq 0 \quad (7)$$

Sederhananya, untuk menyelesaikan permasalahan optimasi ini dapat diubah kedalam formulasi lagrangian yang menggunakan lagrange multiplier :

$$min^{Lp}_{w,b}(\mathrm{w,b,a}) = \tfrac{1}{2}|w|^2 - \textstyle\sum_{i=1}^{n} ai \, y_i \, (x_i.w + b) \sum_{i=1}^{n} ai \quad (8)$$

Kemudian menambahkan konstrain $\propto \geq 0$ (nilai dari koefisiensi lagrange). Dengan meminimunkan Lp terhadap w dan b, maka $\frac{\partial}{\partial b} \, Lp \, (w, b, a) = 0$ diperoleh pada persamaan (9) dan $\frac{\partial}{\partial b} \, Lp \, (w, b, a) = 0$ diperoleh pada persamaan (10)

$$\textstyle\sum_{i=1}^{n} a_i y_i = 0 \qquad (9)$$

$$w = \sum_{i=1}^{n} a_i y_i x_i \qquad (10)$$

Akan tetapi terkadang vector w bernilai besar hingga tak terhingga. Sehingga formulasi lagrangian Lp harus diubah kedalam dual problem LD dengan mendistribusikan persamaan (10) ke Lp

$$L_D(\alpha) \equiv \alpha - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j \qquad (11)$$

Sehingga permasalahan pencarian bidang pemisah atau margin terbaik dapat dirumuskan menjadi (Nomleni, 2015)

$$max_{\alpha} L_D(\alpha) \equiv \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j$$

$$s.t \sum_{i=1}^{n} = \alpha_i y_i = 0 \alpha_i \geq 0 \qquad (12)$$

B) Klasifikasi Non-Liniear

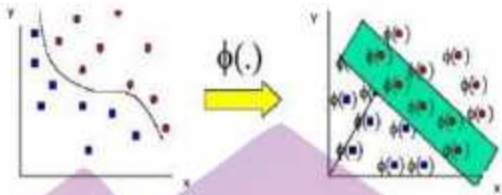
Pada dasarnya, SVM merupakan metode non linier yang mengubah data x yang diterapkan pada fungsi Φ (x) dalam ruang vektor berdimensi tinggi. Sehingga faktor objektifnya merepresentasikan data dalam ruang vektor baru. Proses dalam SVM adalah mencari support vector dengan dot product dari data vector space yang baru. Adanya kernel trick bertujuan untuk menentukan support vector data non linier dalam proses pembelajaran SVM sebagaimana yang didefinisikan pada persamaan 13.

$$K (Xi.Xj) = \Phi(xi).\Phi(xj) \qquad (13)$$

Dua kernel yang digunakan adalah fungsi polinomial dan Gaussian RBF seperti pada persamaan dibawah

$$K (Xi.Xj) = (xi. xj + 1)p \qquad (14)$$

$$K(Xi.Xj) = \exp\left(-\left(\frac{||Xi - Xj||2}{2\sigma 2}\right)\right) \quad (15)$$



Gambar 2. 7 Transformasi dari input space ke feature space (Lidya, 2014)

Kelebihan dari algoritma SVM adalah dapat mencapai optimum dan menangani masalah dimensi tinggi. Namun, pengaturan parameter dan kernel dapat mempengaruhi kinerja proses pembelajaran dan kinerja umum SVM. Pengujian validasi menggunakan cross-fold dan bootstrapping dengan kombinasi parameter tertentu dapat menyebabkan estimasi tingkat kesalahan pada data tertentu. Adapun parameter SVM dengan polynomial kernel yang digunakan adalah gamma (Y) dan cost (C) yang membutuhkan nilai optimum. Parameter cost digunakan untuk mengontrol penalti untuk data training yang salah diklasifikasikan, mengakibatkan kompleksitas fungsi prediksi. Sehingga tingginya nilai C dapat memaksa SVM untuk membuat fungsi prediksi yang cukup kompleks untuk mengklasifikasikan kesalahan pada data training sesedikit mungkin, sedangkan jika nilai C rendah dapat mengarah ke fungsi prediksi yang lebih sederhana. Kemudian untuk parameter gamma dapat dianggap kebalikan dari radius pengaruh sampel yang dipilih oleh model sebagai support vector. Pada penelitian ini, nilai parameter terbaik pada gamma (Y) dan cost (C) berada pada nilai tertentu untuk diimplementasikan. (Muflikah, Widodo, Mahmudi, & Solimun, 2021)

2.2.8. *Random Forest*

Random Forest merupakan teknik pengembangan dari *decision tree*. Pada decision tree, input dimasukkan pada root kemudian turun ke bawah untuk menentukan data dan jenis urutan kelas. Algoritma ini terdiri dari jumpulan pengklasifikasi pohon terstruktur dimana setiap pohon mengeluarkan unit suara untuk kelas paling popular di input x. (Mustaqim, 2020)

Random Forest termasuk dalam metode *Supervised Learning* klasifikasi yang dibangun dari beberapa decision tree dengan memilih sejumlah F fitur secara random sehingga digunakan sebagai *node* untuk membangun *decision tree*. Nilai F akan mempengaruhi hasil akhir dari kinerja algoritma *Random Forest*. Jika nilai F terlalu kecil, maka nilai korelasi dari tree tersebut semakin kecil. Kemudian jika nilai F nya terlalu besar maka nilai korelasi dari tree tersebut akan semakin besar. Adapun untuk mencari nilai F dapat menggunakan M (jumlah seluruh atribut) yang ditentukan dengan persamaan

$$F = \log_2 (M + 1) \qquad (16)$$

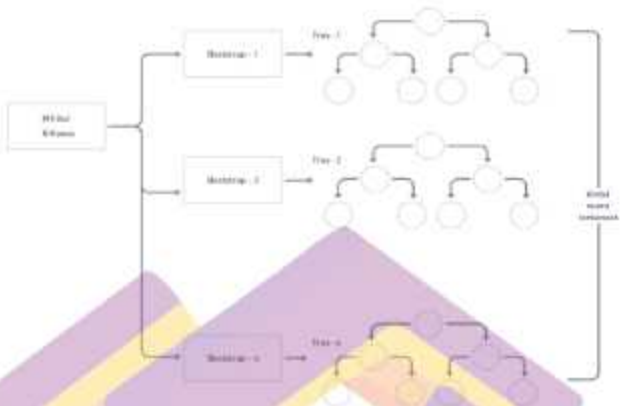
Selain itu, pada metode ini akan memilih sejumlah data training secara acak dengan menggunakan teknik *bootstrapping*. Teknik ini dalam pengambilan sampel digunakan untuk membangun setiap decision tree dengan kandidat yang terpilih sebelumnya. Secara garis besar, cara kerja dari Random Forest adalah : (Muflikah, Widodo, Mahmudi, & Solimun, 2021)

1. Ambil sampel *bootstrap* dari data *training*
2. Konstruksi *decision tree*

3. Setiap node dalam pemilihan fitur split, kemudian pilih hanya diantara fitur m > M
4. Buat *decision tree* dari setiap sampel bootstrap
5. Kemudian ambil suara terbanyak

Contohnya variabel input random forest berupa list 4,3,2,8,3,1,8 maka ketika dilakkukan bagging variabel yang diambil menjadi 2,3,3,8,1,6,4. Pada list yang diambil terdapat duplikasi dari variabel 3. Hal ini diperlukan untuk membuat masing masing algoritma decision tree tidak saling terhubung dan mempengaruhi satu sama lain.

Kemudian proses pengacakan juga berlaku pada algoritma Random Forest. Contohnya terdapat fitur 1, fitur 2 dan fitur 3 maka pada salah satu decision tree hanya akan menggunakan fitur 1 dan 2, kemudian pada decision tree lainnya akan menggunakan fitur 2 dan 3 begitupula seterusnya. Fungsi lain dari pengacakan ini untuk menghindari error dari data yang homogen. Hasil akhir dari Random Forest berupa perhitungan vote dominan dari seluruh decision tree yang digunakan. Secara garis besar seluruh proses random forest dapat dilihat pada gambar 2.9 (Mustaqim, 2020)



Gambar 2. 8 Ilustrasi Algoritma Random Forest
(Muflikah, Widodo, Mahmudi, & Solimun, 2021)

2.2.9. Python

Python merupakan Bahasa pemrograman yang disusun oleh Guido van Rossum pada tahun 1989. Python hadir karena ketidakpuasan Rossum terhadap kinerja Bahasa C dalam pembuatan program komputer. Sehingga Python sendiri ditulis dengan Bahasa C. Beberapa keunggulan Bahasa python dibandingkan dengan Bahasa pemrograman lainnya:

1. Mudah dikuasai oleh pemula sekalipun karena Python adalah Bahasa tingkat tinggi yang lebih dekat dengan Bahasa alami manusia. Tatanan perkalimat yang lebih sederhana dibandingkan Java, Perl, C++, dan lain lain.
2. Python merupakan program sumber terbuka atau opensource. Sehingga tidak memerlukan biaya untuk menggunakannya.

3. Waktu yang diperlukan untuk menulis code dalam Bahasa Python lebih cepat dibandingkan Bahasa pemrograman lain. (Khiatuddin & Muhammad, 2021)

2.2.10. Rapid Miner

Rapid Miner merupakan aplikasi yang dapat digunakan untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. Selain itu aplikasi ini bersifat terbuka atau opensource sehingga tidak diperlukan biaya untuk menggunakannya. RapidMiner memiliki kurang lebih 500 operator data preprocessing dan visualisasi. Selain itu, aplikasi ini juga menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis. GUI ini juga akan menghasilkan file XML yang mendifinisikan proses analitis yang diterapkan ke data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analisis secara otomatis. Beberapa fitur dari RapidMiner antara lain:

- Memuat banyak algoritma data mining.
- Grafis yang canggih.
- Banyak variasi plugin.
- Menyediakan prosedur data mining dan machine learning termasuk ETL, data preprocessing, visualisasi, modelling dan evaluasi/
- Proses data mining yang tersusun atas operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI.
- Mengintegrasikan proyek data mining Weka dan statistika R. (Siregar & Puspabhuana)

2.2.11. *K-Fold Cross Validation*

Cross Validation merupakan teknik validasi yang digunakan untuk mengukur bagaimana hasil analisa akan menggeneralisasi data independen. Pada dasarnya, teknik ini akan memprediksi model kemudian memperkirakan seberapa akurat hasil dari analisis model. Salah satu teknik dari *cross validation* adalah *k-fold cross validation*. Fokus utamanya adalah menghila ngkan bias pada data yang digunakan sehingga pendekatan yang dilakukan adalah memecah data dengan ukuran yang sama berdasarkan k bagian set data. (Tempola, Muhammad, & Khairan, 2018)

2.2.12. *Confusion Matrix*

Untuk mengukur tingkat akurasi pada teknik klasifikasi salah satunya menggunakan Confusion Matrix. Yaitu dengan cara membandingkan nilai aktual dengan nilai prediksi. Metode ini dapat digunakan untuk output yang terdiri dari 2 kelas atau lebih, dengan tabel yang berisi 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. 4 kombinasi tersebut terdiri dari True-Positives, prediksi positif dan benar. False-Positives, prediksi positif dan salah. True-Negatives, prediksi negative dan benar. Dan False-Negatives, prediksi negative dan salah. (Narkhede, 2018)

Tabel 2. 1 *Confusion matrix*

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Negative (FN)	True Negative (TN)

- True-Positives (TP), hasil prediksi maupun aktual menyatakan bahwa pasien hidup.
- False-Positives (FP), hasil prediksi menyatakan meninggal akan tetapi hasil aktual menyatakan bahwa pasien hidup.
- True-Negatives (TN), hasil prediksi maupun aktual menyatakan bahwa pasien meninggal.
- False-Negatives (FN), hasil prediksi menyatakan hidup akan tetapi hasil aktual menyatakan bahwa pasien meninggal.

Berdasarkan hasil confusion matrix dapat menghitung performance matrix untuk mengukur model yang telah dibuat. Adapun beberapa komponen dalam performance matrix adalah:

a) Akurasi

Digunakan untuk menggambarkan seberapa akurat model yang digunakan dalam mengklasifikasikan dengan benar dan tepat. Rumus yang biasa digunakan:

$$Accuracy = (TP+TN)/(TP+FP+FN+TN) * 100\% \quad (17)$$

b) Presisi

Digunakan untuk menghitung akurasi antara data data yang diminta dengan hasil prediksi yang diberi oleh model. Rumus yang bisa digunakan:

$$Precision = (TP) / (TP + FP) \quad (18)$$

c) Recall

Untuk menghitung tingkat keberhasilan model dalam menemukan informasi dari data yang sudah diolah sebelumnya. Rumus yang bisa digunakan:

$$Recall = (TP) / (TP + FN) \quad (19)$$

2.2.13. Area Under Curve (AUC)

Grafik ROC adalah grafik dua dimensi yang memuat hubungan antara True Positive (TPR) atau Sensitivity (sumbu Y) dengan False Positive Rate (FPR) atau Specifity (sumbu X). (Nugroho, 2020) Berikut merupakan formulasi dari sensitivity dan specifity yang dipaparkan dalam persamaan 20 dan 21. (Qadrini, Seppewali, & Aina, 2021)

$$Sensitivity = (TP/(TP+FN) \times 100\% \quad (20)$$

$$Specifity = (TN/(TN+FN) \times 100\% \quad (21)$$

Dari grafik ROC inilah akan menghasilkan sebuah area yang berada dibawah kurva yang merupakan wilayah yang menunjukkan tingkat keakuratan dari model prediksi dan dihitung dengan metode perhitungan yang disebut Area Under Curve (AUC). AUC selalu memiliki nilai yang berada diantara 0 dan 1. Jika nilai AUC yang dihasilkan < 0.5, maka model yang dievaluasi memiliki tingkat keakuratan yang sangat rendah dan mengindikasikan bahwa model tersebut buruk digunakan. (Nugroho, 2020) Apabila nilai AUC semakin mendekati 1, maka model klasifikasi yang terbentuk semakin akurat. Kurva ROC yang baik berada diatas dari garis diagonal (0,0) dan (1,1), sehingga tidak ada nilai AUC yang lebih kecil dari 0,5. (Pratiwi, 2018)

Tabel 2. 2 Klasifikasi Nilai AUC

Nilai AUC	Keterangan
>0,9 – 1	Luar Biasa
>0,8 – 0,9	Sangat Baik
>0,7 – 0,8	Baik
>0,6 - 0,7	Cukup Baik
0,5 – 0,6	Tidak Baik

2.2.14. Populasi, Sampel dan Teknik Pengambilan Sampel

a) Populasi dan Sampel

Populasi adalah totalitas semua nilai yang mungkin, hasil perhitungan atau pengukuran, kuantitatif maupun kualitatif mengenai karakteristik tertentu dari semua anggota kumpulan yang lengkap dan jelas yang ingin dipelajari sifat sifatnya. Sedangkan sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi. (Waskito, 2014)

Untuk menentukan jumlah sampel pada penelitian dapat didasarkan pada perhitungan yang dikemukakan Slovin:

$$n = N / 1+N(e)^2 \qquad (20)$$

Keterangan:

n: Jumlah sampel minimal

N: Jumlah sampel keseluruhan

$(e)^2$: Batas toleransi kesalahan (error tolerance), presentase kelonggaran ketelitian karena kesalahan pengambilan sampel (1% atau 5% atau 10%). (Vitaloka, 2019)

b) Teknik pengambilan sampel

Terdapat beberapa teknik pengambilan sampel, salah satunya adalah simple random sampling. Teknik simple random samping adalah teknik yang sederhana karena pengambilan anggota sampel dari populas dilakukan secara acak tanpa melihat dan memperhatikan kesamaan atau strata yang ada dalam populasi. (Vitaloka, 2019).

2.3. Keaslian Penelitian

Tabel 2. 3 Matriks literatur review dan posisi penelitian

Pengaruh Text Preprocessing Terhadap Analisis Sentimen Opini Pengguna Aplikasi Qasir Dengan Menggunakan Metode Support Vector Machine dan Random Forest

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	*Sentiment Analysis Online Shop on The Play Store Using Method Support Vector Machine* (SVM)	Muhammad Iqbal Ahmadi, Fuji Apriani, Mia Kurniasari, Siti Handayani, Dudih Gustian., SEMNASIF, 2020	Membangun sistem sentimen analisis untuk mengetahui aplikasi online marketplace terbaik pada playstore.	Pada penelitian ini melakukan perbandingan opini terhadap bebcrapa online shop yaitu Tokopedia, JD.ID, Blibli, Shopee dan Lazada. Kemudian model evaluasi menggunakan *k-fold cross validation* dengan k = 10 dan mendapatkan nilai akurasi tertinggi 90,67% untuk *marketplace* Tokopedia dan nilai akurasi terendah didapatkan oleh *marketplace* Lazada dengan nilai 69%. Akan tetapi untuk nilai tertinggi AUC didapatkan oleh JD.ID dengan nilai 85% dan nilai AUC terendah tetap didapatkan oleh *marketplace* Lazada dengan nilai 74%	Dapat menambahkan dataset yang lebih banyak dikarenakan data yang digunakan untuk analisis masing masing aplikasi hanya berjumlah 300 data. Kemudian bisa membandingkan dengan metode *supervised learning* lainnya seperti C45, *Random Forest*, dll.	Pada penelitian ini akan membandingkan kinerja dari metode *Random Forest* dengan *Support Vector Machine*. Dan *dataset* yang digunakan hanya berasal dari 1 review aplikasi GooglePlaystore saja yaitu aplikasi Qasir. Kemudian pada penelitian ini akan membandingkan kinerja text-preprocessing dengan 4 skenario.

Tabel 2. 3 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2	Analisis Sentimen Terhadap Review Film Menggunakan Metode *Modified Balanced Random Forest* dan *Mutual Information*	Firdausi Nuzula Zamzami, Adiwijaya, Mahendra Dwifebri P., Jurnal Media Informatika Budidarma, 2021.	Membantu konsumen untuk memetakan informasi yang penting dan tidak penting dalam review terhadap sebuah film.	Penggunaan stemming pada preprocessing membantu untuk meningkatkan performance akurasi pada penelitian. Kemudian metode *Mutual Information* juga berhasil untuk mengurangi fitur kurang relevan yang digunakan pada proses klasifikasi. Nilai akurasi tertinggi sebesar 79% dengan nilai F1-Scores tertinggi 74%. Kemudian metode ini bekerja baik pada review imbalance data berbahasa inggris, dengan meningkatkan F1-Scores sebesar 27% dan proses stemming mampu meningkatkan nilai F1-Scores sebesar 1%.	Dengan menerapkan metode under-sampling yang dilakukan secara acak terhadap majority class untuk modified balanced random forest. Kemudian menambahkan dataset yang pada dataset movie review berbahasa inggris agar mendapatkan hasil analisis yang lebih baik.	Pada penelitian ini akan membandingkan kinerja dari metode *Random Forest* dengan *Support Vector Machine*. Kemudian *dataset* yang akan digunakan berasal dari *review* aplikasi Google Play store. Pada penelitian ini hanya melakukan pengujian pada stemming dengan hasil akhir sedangkan pada penelitian ini akan membandingkan kinerja text-preprocessing dengan 4 skenario.

Tabel 2. 3 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm Based on Particle Swarm Optimization	Ali Mustopa, Hermanto, Anna, Eri Bayu Pratama, Ade Hendini, Deni Risdiansyah., IEEE, 2021.	Mengklasifikasikan komentar pengguna pada aplikasi Peduli Lindungi sehingga bisa meningkatkan performa aplikasi lebih baik dimasa mendatang.	*Dataset* yang digunakan sebanyak 1364 opini merupakan komentar pengguna selama 3 bulan terakhir dengan metode evaluasi yang digunakan 10-fold cross validation sehingga menghasilkan performa dari metode PSO - *Support Vector Machine* lebih baik dengan nilai akurasi 93% dan nilai AUC 97% dibandingkan PSO - *Naïve Bayes* dengan nilai akurasi 69% dan nilai AUC 65%.	Bisa dengan menambahkan dataset yang lebih banyak dikarenakan data yang digunakan hanya tiga bulan terakhir. Kemudian bisa membandingkan dengan metode *supervised learning* lainnya seperti C45, *Random Forest*, dll.	Pada penelitian ini tidak menggunakan RapidMiner sebagai aplikasi penunjang untuk klasifikasi, *dataset* yang digunakan dari tahun 2020 – 2021 kemudian metode yang akan dibandingkan adalah metode *Random Forest* dengan *Support Vector Machine*. Kemudian pada penelitian ini akan membandingkan kinerja text-preprocessing dengan 4 skenario.

Tabel 2. 3 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
4	Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi	Ragil Dimas Himawan, Eliyani., JEPIN, 2021.	Untuk mengetahui sentimen data tweet masyarakat terhadap akun Pemprov DKI Jakarta di masa pandemi, dan juga untuk mengetahui metode terbaik yang digunakan dalam memprediksi sentimen pada data yang kosong.	Menggunakan *tweet* sebagai *dataset* dengan jumlah data 14208 data dan menggunakan metode pembobotan kata TD-IDF, sehingga menghasilkan akurasi *Support Vector Machine* tertinggi diantara metode lainnya dengan nilai 77,58% disusul dengan *Random Forest* sebesar 75,81% dan Naïve Bayes sebesar 75,22%. Kemudian hasil prediksi nilai kelas sentiment pada data kosong menggunakan metode SVM menghasilkan prediksi netral 83,6%, negative 8,8% dan positif 7,6%.	Menggunakan lebih banyak lagi dataset sehingga dapat menghasilkan nilai akurasi yang lebih baik, Kemudian penelitian bisa dilakukan percobaan pada tahap *text processing* mana yang paling berpengaruh terhadap akurasi serta bisa menggunakan metode lain atau metode yang sudah dimodifikasi.	Data yang digunakan adalah review google play store, kemudian pada penelitian ini akan membandingkan kinerja dari *Support Vector Machine* dan *Random Forest* saja. Kemudian pada penelitian ini akan membandingkan kinerja text-preprocessing dengan 4 skenario.

Tabel 2. 3 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
5	Comparison of Naïve Bayes, Support Vector Machine, Decision Trees and Random Forest on Sentiment Analysis	Márcio Guia, Rodrigo Rocha Silva, Jorge Bernardino., SCITEPRESS, 2019.	Untuk mengevaluasi metode klasifikasi dengan membandingkan kinerja dari masing masing metode serta mengevaluasi dampak *Brand* and *Price* berdasarkan Hasil Review Polaritas.	*Dataset* sebesar 400.000 opini amazon sebagai dengan 80% untuk data *training* dan 20% untuk data *testing*. Kemudian hasilnya adalah metode *Support Vector Machine* dengan akurasi tertinggi 89% dan sebagai metode yang komplit dikarenakan memiliki nilai yang tinggi untuk semua atribut evaluasi seperti Akurasi, Presisi, *Recall* dan F1 *Score*. Disusul dengan metode *Random Forest* yang menghasilkan akurasi tertinggi 88% kemudian metode *Naïve Bayes* dengan nilai 85% dan terakhir metode *Decision Trees* dengan nilai 82%.	Untuk metode evaluasi ditambahi dengan k-fold cross validation untuk menghindari bias dari data yang digunakan saat proses analisis berjalan.	*Dataset* yang akan digunakan merupakan review aplikasi Qasir, kemudian metode yang dibandingkan pada penelitian ini hanya SVM dan *Random Forest*. Kemudian pada penelitian ini akan membandingkan kinerja text-preprocessing dengan 4 skenario.
6	Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naïve Bayes, Random Forest dan Support Vector Machine	Evita Fitri, Yuri Yuliani, Susy Rosyida, Windu Gata., Transformatika, 2020.	Untuk mengetahui analisis sentimen dari review pengguna aplikasi belajar online (Ruangguru)	*Dataset* review aplikasi Ruangguru pada Google Playstore dengan dengan 1629 data. Kemudian hasil dari penelitian, *Random Forest* memiliki akurasi tertinggi dengan nilai 97,16% dan AUC sebesar 0,996 kemudian SVM dengan nilai akurasi 96,01% dengan AUC sebesar 0,543 dan yang terakhir adalah Naïve Bayes dengan akurasi 94,16% dengan AUC sebesar 0,999.	Menggunakan lebih banyak lagi *Dataset*. Kemudian, untuk analisis nya bisa menggunakan aplikasi selain RapidMiner agar lebih beragam lagi hasil dari analisis yang dilakukan.	*Dataset* yang digunakan adalah aplikasi Qasir, dengan SVM dan *Random Forest*. Tidak menggunakan aplikasi RapidMiner untuk pengolahan data. Akan membandingkan kinerja text-preprocessing dengan 4 skenario.

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Adapun jenis, sifat dan pendekatan penelitian yang akan dilakukan pada penelitian ini sebagai berikut:

a. Jenis Penelitian Eksperimen

Penelitian ini dilakukan untuk mengetahui dan memperoleh performa terbaik dalam menghasilkan nilai akurasi tertinggi dalam menganalisa opini pelanggan aplikasi Qasir.

b. Sifat Penelitian Deskriptif

Penelitian ini menjelaskan tahapan-tahapan yang akan digunakan dalam analisis opini pengguna aplikasi Qasir.

c. Pendekatan Penelitian Kuantitatif

Penelitian ini menggunakan pendekatan kuantitatif yaitu hasil dari penerapan metode yang diimplementasikan berupa angka yang menunjukkan nilai akurasi sebagai tolak ukur performa dari model yang digunakan pada analisis.

3.2. Metode Pengumpulan Data

Dalam penelitian ini akan menggunakan data yang berada di kolom komentar google playstore pada aplikasi Qasir. Metode yang akan digunakan dalam pengambilan data adalah *web scraping* yang merupakan proses untuk pengambilan atau ekstraksi data dari sebuah website yang disimpan dalam format tertentu. Kemudian data tersebut akan dievaluasi sudah sesuai atau belum. Setelahnya data

yang akan disimpan adalah dalam format csv. Pada penelitian ini, akan menggunakan Bahasa pemrograman python untuk mengambil data yang sudah tersedia pada aplikasi Qasir.

3.3. Metode Analisis Data

Metode analisis data yang dilakukan untuk mengetahui metode yang menghasilkan nilai akurasi dan waktu proses terbaik yaitu:

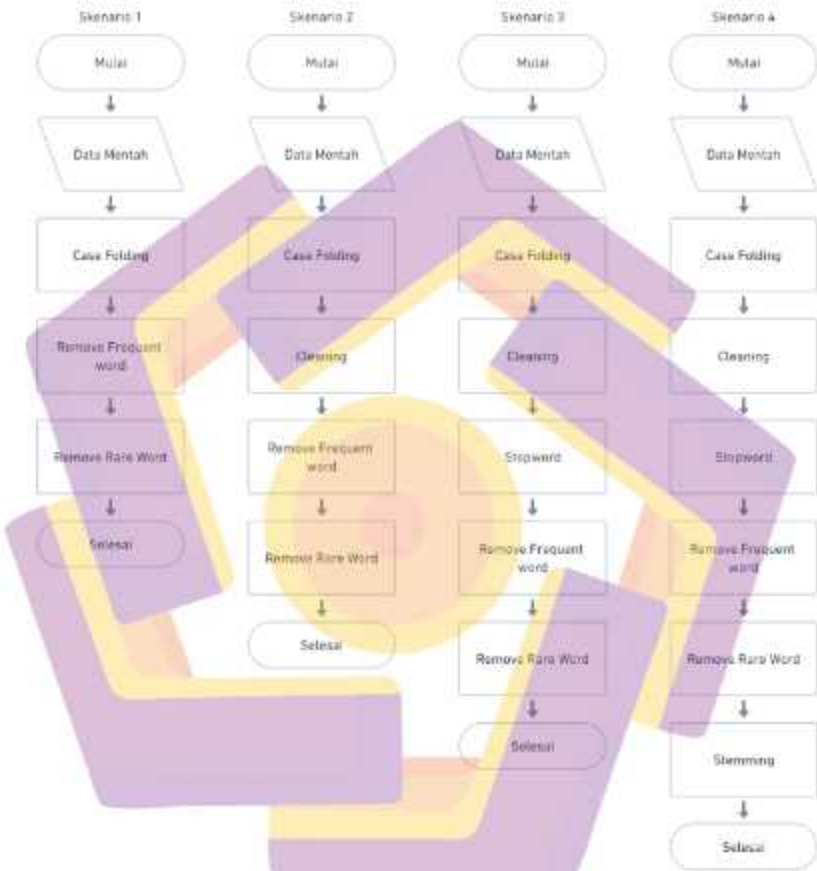
A. Text Preprocessing

Pada proses ini bertujuan untuk mempersiapkan data-data yang nantinya digunakan sebagai input diproses selanjutnya. Pada tahapan ini akan menjalankan 4 scenario text preprocessing yaitu :

- Case folding, remove frequent word, dan remove rare word.
- Case folding, cleaning remove frequent word, remove rare word.
- Case folding, cleaning, stopword, remove frequent word dan remove rare word.
- Case folding, cleaning, stopword, remove frequent word, remove rare word dan stemming.

Seluruh data (dataset mentah) akan dimasukkan kedalam system kemudian dilakukan proses case folding atau menyeragamkan bentuk huruf mulai A-Z, kemudian huruf kapital akan diseragamkan menjadi huruf kecil. Setelah itu masuk kedalam proses cleaning yaitu menghilangkan seluruh tanda baca, angka, symbol dan link URL. Kemudian data akan dihilangkan dari kata yang tidak dianggap penting dalam teks pada proses stopword. Setelah itu kata yang sering muncul akan dihilangkan pada remove frequent word. Selain itu, kata yang jarang muncul juga

akan dihilangkan pada remove rare word. Dan yang terakhir, data akan diubah menjadi kata dasarnya pada proses stemming.



Gambar 3. 1 Diagram alur proses text-processing tiap skenario

B. Labeling Sentiwordnet

Proses untuk memberikan label terhadap setiap data yang nantinya akan dikategorikan menjadi kelas positif, negative atau netral. Proses labeling akan dilakukan dengan menggunakan *lexicon based* dengan SentiWordNet yang

merupakan hasil anotasi otomatis dari WordNet atau disebut sebagai database leksikal untuk Bahasa Inggris. Karena menggunakan lexicon based maka pelabelan dilakukan pada setiap kata dalam kalimat, jika satu kata memiliki lebih dari satu arti maka akan dipilih berdasarkan FirstSense dari SentiWordNet yang muncul paling atas atau popular. Kemudian pada kata akan dilakukan pencarian nilai sentimen yang akan menghasilkan bobot. Bobot dari tiap kalimat inilah yang akan digunakan sebagai acuan untuk melakukan proses perbandingan sehingga kalimat bisa dikategorikan sebagai positif atau negative.

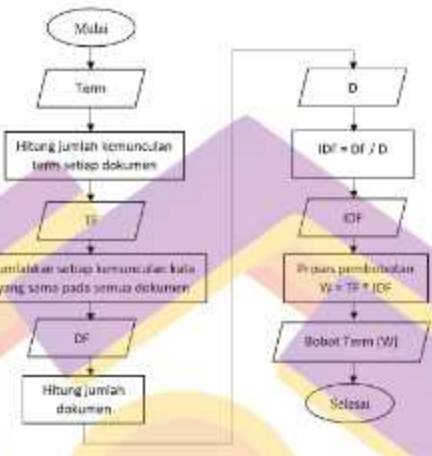
C. Labeling Manual

Selanjutnya untuk memastikan bahwa labeling yang dilakukan oleh Sentiwordnet sudah benar, maka akan dilakukan labeling manual dengan metode simple random sampling. Untuk menentukan jumlah sample bisa dengan memasukkan total data kedalam rumus yang telah diketahui pada bab 2. Selanjutnya, untuk memilih data dengan komposisi data dengan label positif dan data dengan label negative berjumlah sama. Kemudian akan diambil data pertama dan terakhir dari masing masing label untuk dicek manual.

D. Pembobotan

Menggunakan metode TF-IDF, proses yang paling awal dilakukan adalah menghitung terlebih dahulu nilai TF perkata dengan bobot masing masing kata = 1. Kemudian nilai dari IDF diformulasikan dalam rumus IDF (word) = $\log \frac{td}{df}$ dimana IDF (word) adalah nilai IDF dari setiap kata yang akan dicari. TF adalah jumlah keseluruhan dokumen dan DF adalah jumlah kemunculan kata pada semua dokumen. Setelah mendapatkan nilai dari masing masing TF dan IDF, kemudian

untuk mendapatkan bobot akhir maka masuk kedalam rumus akhir yaitu W = TF * IDF. Berikut merupakan diagram alur proses perhitungan TF-IDF.

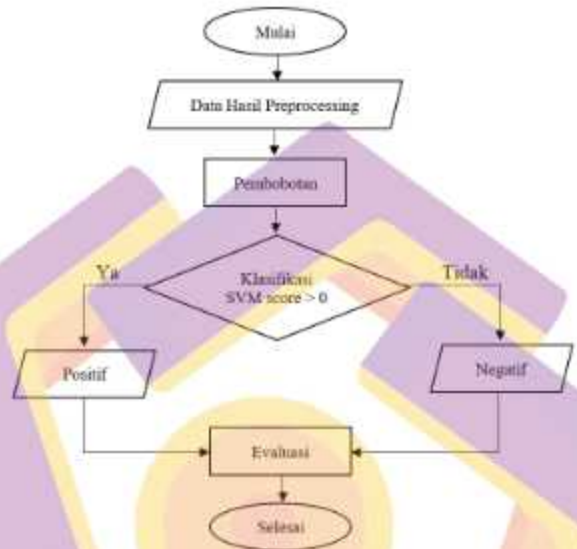


Gambar 3. 2 Alur proses pembobotan TF-IDF

E. Support Vector Machine (SVM)

Bagian ini menggambarkan proses klasifikasi menggunakan algoritma SVM dengan inputnya adalah hasil dari *preprocessing* yang sudah memiliki bobot. Setelah itu kata tersebut akan diklasifikasikan berdasarkan nilai term yang dimiliki, jika kata > 0 maka masuk kedalam komentar positif, selain itu akan dianggap negatif. Klasifikasi menggunakan SVM dimulai dengan mengubah teks menjadi data vector yang memiliki 2 dimensi yaitu word id dan bobot dengan tujuan akhir dapat menemukan margin terbesar yang memisahkan kelas. Pada prosesnya, hasil pembobotan kata kata akan dihitung similaritasnya kemudian diurukan berdasarkan hasil perhitungan dari similiritas tersebut. Setelah itu menghitung nilai pada masing masing kategori dan menghitung probabilitas dokumen yang diuji di masing masing kategori dan mencari probabilitas paling besar, kemudian menentukan sentiment

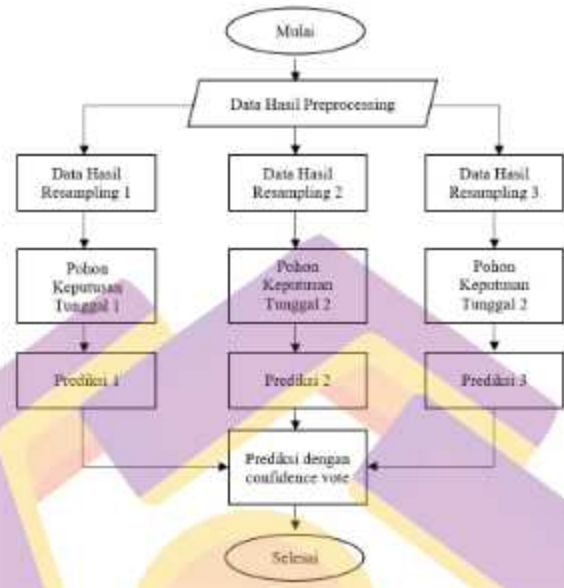
dokumen yang diuji. Pada dibawah ini merupakan cara kerja SVM yang dilakukan pada aplikasi Rapid Miner.



Gambar 3. 3 Diagram Alur Sederhana SVM

F. Random Forest

Bagian ini menggambarkan proses klasifikasi menggunakan algoritma *Random Forest* dengan inputnya adalah hasil dari *preprocessing* dan sudah memiliki bobot yaitu kelas positif dan negatif. Kemudian, beberapa data tersebut akan masuk kedalam proses cross validation yang akan dibagi menjadi beberapa bagian (sesuai dengan nilai k nya) yang akan dilatih dan ditest. Data akan diuji terhadap beberapa pohon keputusan yang nantinya akan menghasilkan sebuah prediksi. Kemudian setiap prediksi ini akan dikumpulkan dan yang diambil sebagai hasil akhir adalah prediksi dengan confidence vote. Pada dibawah ini merupakan cara kerja Random Forest yang dilakukan pada aplikasi Rapid Miner.



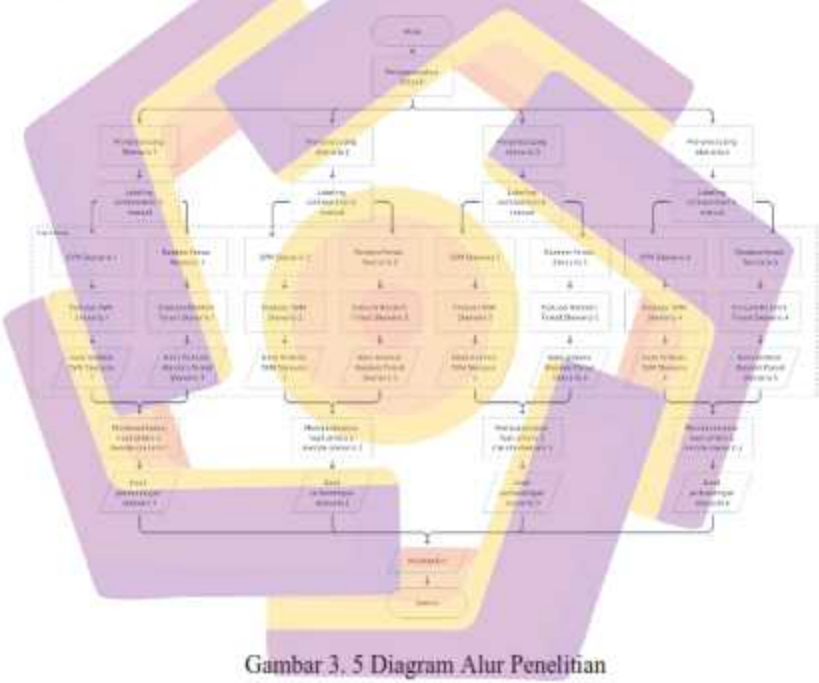
Gambar 3. 4 Alur Kerja Sederhana *Random Forest*

3.4. Metode Evaluasi

Pada tahap ini akan mengukur performa dari model yang telah dibangun, terdapat 2 model yaitu *Training* dan *Testing*. Tahap *training* untuk melatih sistem yang telah dibangun dengan dataset untuk menghasilkan model yang digunakan untuk melakukan klasifikasi pada tahap *testing* dengan memberi data pada model tanpa adanya label. Pada *testing* ini akan mengukur seberapa baik model yang telah dihasilkan. Metode pengujian yang akan digunakan pada penelitian ini yaitu *k-fold cross validation* dengan k sama dengan 5 dan *confusion matrix*.

3.5. Alur Penelitian

Adapun untuk alur penelitian yang akan diterapkan seperti gambar 3.5. Untuk tahapan pre-processing di masing masing scenario sudah dilampirkan pada gambar 3.1 sub bab text pre-processing, dengan catatan pada bagian yang didalam kotak garis putus putus merupakan proses yang dilakukan pada aplikasi RapidMiner.



Gambar 3. 5 Diagram Alur Penelitian

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Studi Literatur

Untuk menyelesaikan penelitian ini harus dilakukan sesuai dengan metode penelitian yang telah dipaparkan sebelumnya, studi literatur merupakan langkah awal yang dilakukan dengan mempelajari arsip/buku, jurnal, makalah dan laporan penelitian sebelumnya sebagai pedoman/landasan dalam menyusun penelitian ini. Studi literatur berisi teori yang digunakan sehingga membantu dalam menyelesaikan penelitian ini. Studi Literatur terdapat pada Bab II Landasan Teori.

4.2. Pengumpulan Data

Pada tahap pengumpulan data dilakukan kegiatan untuk mengumpulkan data yang digunakan pada penelitian ini. Data diambil dengan teknik web scrapping dengan menggunakan google colab sebagai platform/aplikasi dan phyton sebagai bahasa pemrogramannya. Karena data yang dibutuhkan adalah komentar pada aplikasi google play store sehingga untuk mengambil datanya cukup dengan memanggil library google-play-scraper dan instance atau ID dari aplikasi tersebut yang bisa diketahui saat mengakses google play store di web version.

Gambar 4. 1 Instal library, memanggil library dan mengambil data

Kemudian data yang sudah diambil tersebut ditampilkan untuk memastikan data yang sudah diambil benar atau tidak. Pada gambar 4.1 merupakan syntax untuk menampilkan data sedangkan pada gambar 4.2 merupakan hasil dari data yang berhasil didapatkan.

```
df_baru = pd.DataFrame(np.array(result),columns=['review'])
df_baru = df_baru.join(pd.DataFrame(df_baru.pop('review').tolist()))
df_baru.head()
```

Gambar 4. 2 Menampilkan data

Gambar 4. 3 Hasil *get-all-data*

Berikut merupakan isi dari masing masing kolom bisa dilihat pada tabel 4.1

Tabel 4. 1 Kolom yang tersedia

No	Nama Kolom	Penjelasan
1	reviewId	Id dari review yang diberikan
2	userName	Nama user
3	userImage	Gambar profil dari user
4	content	Komentar user
5	Score	Bintang
6	thumbUpCount	Jempol pada komentar
7	reviewCreatedVersion	Komentar diberikan pada versi aplikasi keberapa
8	at	Tanggal dan waktu komentar dibuat
9	replyContent	Komentar balasan dari pihak CS kepada user
10	repliedAt	Tanggal dan waktu komentar balasan dibuat

Pada penelitian ini hanya menggunakan 4 kolom yaitu username, at, content dan score sehingga kolom lainnya akan dihilangkan karena tidak digunakan.

Gambar 4. 4 Kolom yang akan digunakan

Total data yang terkumpul adalah 2775. Setelah itu data di export ke bentuk csv untuk digunakan dalam proses selanjutnya.

Gambar 4. 5 Total data yang terkumpul

Berikut merupakan hasil akhir dari pengambilan data dari google play store dengan menggunakan teknik web scrapping dengan sorting most relevan.

Tabel 4. 2 Contoh dataset yang digunakan

No	Username	At	Content	Score
1	Alvin Sanjaya Putra	1/22/2021 4:32:00 PM	great app to think that this apps are mostly free and only cost you a little to upgrade. Would be great if they can add a royalty system for customers UPDATE been using this apps for over 5 months now the inventory system need a fix please When u have 2 outlets stocks if often mixed when u transfer between outlets it doesnt automatically transferred well sometimes it does partially i have put a feedback in regard of this but havent seen any improvement	3
2	Niko Sutiono	6/27/2020 1:45:00 AM	For now I give it a 4 stars but potential to be 5 stars Some inputs 1 Configuration should be able in both apps web and mobile for now mobile apps has better option to configure 2 Might consider the web apps to be a back office apps with better configuration and customisation 3 Before finalizing transaction it is better to have confirmation and review screen not immediate posting also able to go back for editing 4 Need to consider discount with buy 2 get 3 options Great Apps	4
3	SB Wong	3/21/2022 6:04:00 PM	Had been trying here and there finally manage to found an application whereby can add variants onto my items Thank you very much	5

Tabel 4. 2 Contoh dataset yang digunakan (Lanjutan)

No	Username	At	Content	Score
4	Lysander Cisco	3/19/2022 5:39:00 AM	Hope the search scan and pending order button on the bottom not top	3
5	Dwi Permana	1/14/2022 11:56:00 PM	It was a really good app and simple to used I subscribed for pro version and get the grab merchant integration but as well as you guys need to know the grab option group feature was not available in this version So be aware if your store really need those option	3

4.3. Pre-processing Data

Tahap preparasi data atau *preprocessing* data ini data akan diproses terlebih dahulu sebelum digunakan untuk analisis. Sebagaimana yang sudah dipaparkan pada batasan masalah, pada penelitian ini akan menggunakan 4 skenario dalam tahapan *preprocessing*, yaitu :

a) Case folding, remove frequent word, dan remove rare word.
b) Case folding, cleaning, remove frequent word dan remove rare word.
c) Case folding, cleaning dan stopword removal, remove frequent word, remove rare word.
d) Case folding, cleaning, stopword removal, remove frequent word, remove rare word dan stemming.

Untuk keseluruhan tahapan *preprocessing* ini akan dilakukan di google colab dengan menggunakan bahasa pemrograman python. Pada laporan penelitian ini akan dipaparkan tahapan *preprocessing* pada skenario 4, karena pada skenario tersebut seluruh tahapan *preprocessing* dilakukan. Tahap pertama dari pre-processing ini adalah case folding atau mengubah seluruh huruf pada data menjadi kecil. Pada gambar 4.5 terdapat 2 kolom yaitu English yang merupakan kolom asli

dari dataset, sedangkan content merupakan kolom hasil dari case folding ini. Contoh hasilnya adalah : *Very useful and help me in selling!*, menjadi *very useful and help me in selling*.

Gambar 4. 6 Tahapan *Case lower*

Setelah itu data akan masuk dalam tahap *cleaning* atau menghapus tanda baca yang tidak digunakan. Untuk melakukan *cleaning* pada python menggunakan method *maketrans* yang berfungsi untuk me*replace* karakter tertentu yang sudah di *define* pada variabel string.punctuation (!"#$%&'()*+, -./:;<=>?@[\]^_`{|}~). Pada gambar 4.6, hasil dari tahapan cleaning terdapat pada kolom text_wo_punct. Contoh hasilnya adalah : *…..in selling!* menjadi *….in selling.*

Gambar 4. 7 Tahapan *cleaning*

Persiapan yang dapat dilakukan untuk tahapan *stopword* adalah *import library* NLTK atau *Natural Language Toolkit*. Karena dataset yang digunakan berbahasa Inggris, sehingga library yang digunakan adalah corpus.

Gambar 4. 8 *Import library* NLTK corpus

Selanjutnya masuk kedalam tahap *stopword* yaitu menghilangkan kata yang dianggap tidak memiliki makna sehingga hanya tersisa kata kata penting sebagai gantinya. Pada gambar dibawah ini, hasilnya tersimpan pada kolom text_wo_stop. Contoh : *very useful and help me in selling* menjadi *useful help selling.*

Gambar 4. 9 Tahapan *Stopword*

Tahapan selanjutnya adalah menghilangkan kata kata yang sering muncul. Sebelum itu, kata kata yang ada pada dataset dihitung terlebih dahulu. Setelah mendapatkan jumlah kemunculan per kata pada dataset, 10 kata teratas akan dihapus. Setelah itu, hasil dari proses ini dapat dilihat pada gambar 4.10 di kolom text_wo_stopfreq. Contohnya: karena kata helpful masuk kedalam salah satu list kata yang sering muncul, maka kata tersebut akan dihilangkan. Dari *quite helpful financial process* menjadi *quite financial prosess.*

Gambar 4. 10 Menghilangkan kata yang sering muncul

Karena pada variabel cnt.most_common yang digunakan pada tahapan sebelumnya sudah ada jumlah perkata sehingga pada tahapan menghilangkan kata yang jarang muncul, variabel tersebut dapat digunakan kembali. Sama dengan tahapan sebelumnya, tahapan ini pun akan menghilangkan 10 kata yang jarang muncul pada dataset. Setelah itu, hasil dari proses ini dapat dilihat pada gambar 4.11 di kolom text_wo_stopfreqrare. Contohnya: karena kata startups masuk kedalam salah satu list kata yang jarang muncul, maka kata tersebut akan dihilangkan. Dari *helpful for business startups* menjadi *helpful for business.*

Gambar 4. 11 Menghilangkan kata yang jarang muncul

Tahapan terakhir pada pre-processing ini adalah *Stemming*. Tahapan ini bertujuan untuk menghilangkan kata imbuhan, sehingga kata akan ditampilkan

dalam bentuk kata dasarnya. Pada penelitian ini, algoritma yang digunakan adalah Porter. Dalam python, terdapat library stemming porter sehingga untuk menggunakannya perlu di import terlebih dahulu. Hasilnya terdapat pada kolom text_stemmed di gambar 30. Contoh hasilnya: Dari *useful help selling* menjadi *use help sell*.

Gambar 4. 12 Stemming

Berikut merupakan hasil akhir pada masing masing skenario.

Tabel 4. 3 Hasil akhir dataset yang digunakan

No	Scenario 1	Scenario 2	Scenario 3	Scenario 4
1	useful help me in selling! thank god there are finally apps like this!	useful help me in selling thank god there are finally apps like this	useful help selling god finally apps like	use help sell god final app like
2	these apps really make easier me manage my distributor business forward etalastic!	these apps really make easier me manage my distributor business forward etalastic	apps really make easier manage distributor business forward etalastic	app realli make easier manag distributor busi forward etalast
3	quite in financial process, i say, job (y)	quite in financial process i say job y	quite financial process say job	quit financi process say job
4	why doesn't help function work? how add menu how? i'm using asus z4c. thx :)	why doesnt help function work how add menu how im using asus z4c thx	doesnt help function work add menu im using asus z4c thx	doesnt help function work add menu im use asu z4c thx

Tabel 4. 3 Hasil akhir dataset yang digunakan (Lanjutan)

No	Scenario 1	Scenario 2	Scenario 3	Scenario 4
5	new version better than one that hasn't been updated, so discount feature can be synchronized without manual input make easier	new version better than one that hasnt been updated so discount feature can be synchronized without manual input make easier	new version better one hasnt updated discount feature synchronized without manual input make easier	new version better one hasnt updat discount featur synchron without manual input make easier

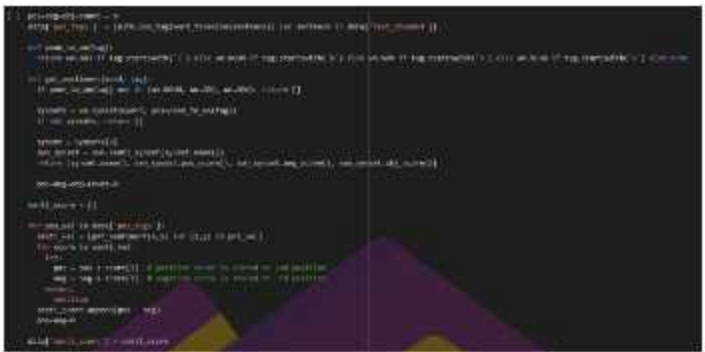
4.4. Labeling

Setelah dataset selesai diproses, langkah selanjutnya adalah menentukan data tersebut memiliki sentimen positif atau negatif. Untuk itu labeling dilakukan dengan menggunakan sentiwordnet dan *crosscheck* manual pada penelitian ini.

4.4.1. Labeling Sentiwordnet

Cara kerja dari *Sentiwordnet* adalah memecah setiap kata pada kalimat, kemudian diberikan bobot dimasing masingnya. Setelah itu bobot per kata akan dijumlahkan. Pada gambar 4.13 terdapat proses labeling *Sentiwordnet* dengan menggunakan python.

Proses awal labeling adalah menentukan tiap kata pada kalimat termasuk dalam *Noun, Adj, Adv* dan *Verb* dan diberikan *tag* sesuai katanya. Kemudian, jika kata kata tersebut tidak ada didalam list *Sentiwordnet* maka akan di-*return* kosong (*empty*). Setelah itu menggunakan *synset* untuk mengelompokkan kata kata sinonim yang mengekspresikan konsep yang sama. Sesudah dikelompokkan, kemudian kata akan dinilai dan dihitung.



Gambar 4. 13 Proses Labeling *Sentiwordnet*

Setelah menghitung bobot seluruh kalimat, selanjutnya adalah pemberian label dengan aturan jika memiliki nilai >= -0.05 dianggap negatif, <= 0.05 dianggap positif, dan selain itu dianggap netral.

Gambar 4. 14 Pemberian label

Akan tetapi pada penelitian ini tidak akan menggunakan label netral, sehingga setelah proses labeling selesai dilakukan data netral akan dihapus.

Gambar 4. 15 Menghapus neutral data

Hasil labeling dari skenario pertama yang menghasilkan 1250 data Positif dan 480 data Negatif yang bisa dilihat pada chart dibawah ini.

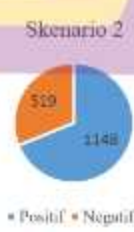
Gambar 4. 16 Perbandingan data pada skenario 1

Kemudian pada Tabel 4.4 terdapat hasil labeling santiwornet yang telah di export dengan beberapa contoh kalimat, tags, senti score dan label akhir.

Tabel 4. 4 Labeling pada skenario 1

No	Text	Tags	Senti Score	Label
1	these apps really make easier me manage my distributor business. forward etalastic!	[('these', 'DT'), ('apps', 'VBP'), ('really', 'RB'), ('make', 'VB'), ('easier', 'JJR'), ('me', 'PRP'), ('manage', 'VB'), ('my', 'PRP$'), ('distributor', 'NN'), ('business', 'NN'), ('.', '.'), ('forward', 'RB'), ('etalastic', 'JJ'), ('!', '.')]	1.25	Positive
2	quite in financial process, i say, job (y)	[('quite', 'RB'), ('in', 'IN'), ('financial', 'JJ'), ('process', 'NN'), (',', ','), ('i', 'NNS'), ('say', 'VBP'), (',', ','), ('job', 'NN'), ('(', '('), ('y', 'NN'), (')', ')')]	0.375	Negative

Hasil labeling dari skenario kedua yang menghasilkan 1148 data Positif dan 519 data Negatif yang bisa dilihat pada chart dibawah ini.



Gambar 4. 17 Perbandingan data pada skenario 2

Kemudian pada Tabel 4.5 terdapat hasil labeling santiwornet yang telah di export dengan beberapa contoh kalimat, tags, senti score dan label akhir.

Tabel 4. 5 Labeling pada skenario 2

No	Text	Tags	Senti Score	Label
1	these apps really make easier me manage my distributor business forward etalastic	[('these', 'DT'), ('apps', 'VBP'), ('really', 'RB'), ('make', 'VB'), ('easier', 'JJR'), ('me', 'PRP'), ('manage', 'VB'), ('my', 'PRPS'), ('distributor', 'NN'), ('business', 'NN'), ('forward', 'RB'), ('etalastic', 'JJ')]	1.25	Positive
2	quite in financial process i say job y	[('quite', 'RB'), ('in', 'IN'), ('financial', 'JJ'), ('process', 'NN'), ('i', 'NNS'), ('say', 'VBP'), ('job', 'NN'), ('y', 'NN')]	0.375	Negative

Hasil labeling dari skenario ketiga yang menghasilkan 997 data Positif dan 451 data Negatif yang bisa dilihat pada chart dibawah ini.

Gambar 4. 18 Perbandingan data pada skenario 3

Kemudian pada Tabel 4.6 terdapat hasil labeling santiwornet yang telah di export dengan beberapa contoh kalimat, tags, senti score dan label akhir.

Tabel 4. 6 Labeling pada skenario 3

No	Text	Tags	Senti Score	Label
1	apps really make easier manage distributor business forward etalastic	[('apps', 'NNS'), ('really', 'RB'), ('make', 'VBP'), ('easier', 'JJR'), ('manage', 'NN'), ('distributor', 'NN'), ('business', 'NN'), ('forward', 'RB'), ('etalastic', 'JJ')]	1.25	Positive
2	quite financial process say job	[('quite', 'RB'), ('financial', 'JJ'), ('process', 'NN'), ('say', 'VBP'), ('job', 'NN')]	0.375	Negative

Hasil labeling dari skenario keempat yang menghasilkan 740 data Positif dan 473 data Negatif yang bisa dilihat pada chart dibawah ini.

Gambar 4. 19 Perbandingan data pada skenario 4

Kemudian pada Tabel 4.7 terdapat hasil labeling santiwornet yang telah di export dengan beberapa contoh kalimat, tags, senti score dan label akhir.

Tabel 4. 7 Labeling pada skenario 4

No	Text	Tags	Senti Score	Label
1	app realli make easier manag distributor busi forward etalast	[('app', 'NN'), ('realli', 'NNS'), ('make', 'VBP'), ('easier', 'JJR'), ('manag', 'NN'), ('distributor', 'NN'), ('busi', 'NN'), ('forward', 'RB'), ('etalast', 'VBD')]	0.625	Positive
2	doesnt help function work add menu im use asu z4c thx	[('doesnt', 'NN'), ('help', 'NN'), ('function', 'NN'), ('work', 'NN'), ('add', 'VBP'), ('menu', 'NNS'), ('im', 'VBP'), ('use', 'NN'), ('asu', 'NN'), ('z4c', 'NN'), ('thx', 'NN')]	0.5	Positive

4.4.2. Labeling Manual (*Crosscheck*)

Sebelum melakukan labeling manual, langkah pertama adalah menghitung jumlah minimal sampel yang akan digunakan. Karena jumlah data pada tiap tiap scenario berbeda, sehingga perhitungan jumlah sampel juga mengikuti banyaknya data yang tersedia.

- Skenario 1 $n = 1730 / 1 + 1730 (10\%)^2 = 94.53$
- Skenario 2 $n = 1667 / 1 + 1667 (10\%)^2 = 94.34$

- Skenario 3 $n = 1448 / 1 + 1448 (10\%)^2 = 93.54$
- Skenario 4 $n = 1213 / 1 + 1213 (10\%)^2 = 92.38$

Sehingga berdasarkan perhitungan diatas maka scenario 1 memiliki minimal sampel sejumlah 94.53 data, kemudian scenario 2 sejumlah 93.34, scenario 3 sejumlah 93.54 dan yang terakhir scenario 4 sejumlah 92.38. Dan jika dibulatkan untuk seluruh scenario bisa menggunakan minimal 100 data. Akan tetapi pada penelitian ini jumlah minimal akan dikalikan 2 sehingga mendapatkan 200 data.

Setelah seluruh dataset dari keempat skenario terlabeli, untuk mengecek apakah data sudah memiliki label yang benar dilakukan *crosscheck* dengan dataset asli pada 200 data acak dengan komposisi 100 data dengan label positif dan 100 data dengan label negatif. Kemudian akan diambil 50 data pertama dan terakhir dari masing masing label untuk di-*crosscheck* manual.

Pada Tabel 4.8 terdapat 4 komentar yang akan menjadi contoh dalam proses *crosscheck* manual dengan komposisi 2 label positif dan negatif.

Tabel 4. 8 Hasil Labeling Sentiwordnet

No	Komentar	Label
1	Quite helpful in the financial process, I say, good job (y)	Negative
2	easy & simple.	Positive
3	Why doesn't the help function work? how to add menu how? I'm using asus z4c. thx :)	Negative
4	Very helpful, good luck to the founders.	Negative

Kemudian dibandingkan dengan dataset asli yang terdapat bintang (star).

Tabel 4. 9 Hasil Labeling Sentiwordnet

No	Komentar	Star
1	Quite helpful in the financial process, I say, good job (y)	4
2	easy & simple.	3
3	Why doesn't the help function work? how to add menu how? I'm using asus z4c. thx :)	2
4	Very helpful, good luck to the founders.	5

Jika label sama maka tidak akan diubah. Jika label berbeda, maka akan diubah sesuai dengan bintang pada komentar dataset asli. Jika label berbeda akan tetapi bintang pada komentar tersebut adalah 3, maka akan dilakukan penilaian menurut pendapat pribadi lebih condong kemanakah komentar tersebut. Pada tabel 4.10 terdapat 2 kolom label dengan perbedaan Label S untuk Label yang dihasilkan oleh sentiwordnet, sedangkan Label M untuk Label dataset asli.

Tabel 4. 10 Hasil Crosscheck Labeling

No	Komentar	Label S	Label M
1	Quite helpful in the financial process, I say, good job (y)	Negative	Positive
2	easy & simple.	Positive	Positive
3	Why doesn't the help function work? how to add menu how? I'm using asus z4c. thx :)	Negative	Negative
4	Very helpful, good luck to the founders.	Negative	Positive

Pada tabel diatas terdapat 2 perbedaan label yang dihasilkan oleh Sentiwordnet dan Crosscheck labeling, yaitu data ke 1 dan data ke 4. Pada data 1 label yang dihasilkan oleh sentiwordnet adalah Negatif sedangkan label aslinya adalah Positif, sehingga label pada data ke 1 akan diubah ke Positif mengikuti label aslinya. Pun juga pada data ke 4, label aslinya adalah Positif sehingga label akan disesuaikan dengan label aslinya. Sedangkan pada data ke 2, hasil pelabelannya

adalah tetap Positive walaupun bintang yang dihasilkan pada komentar aslinya 3. Pada data ke 2 lebih merujuk pada komentar positif karena terdapat 2 kata yang berarti baik, *easy* untuk kemudahan dan *simple* untuk sederhana atau mudah. Setelah *crosscheck* manual, berikut merupakan hasil akhir dari komposisi di masing masing skenario:

- Skenario 1 : 1414 data Positif dan 316 data Negatif.
- Skenario 2 : 1354 data Positif dan 313 data Negatif.
- Skenario 3 : 1201 data Positif dan 247 data Negatif.

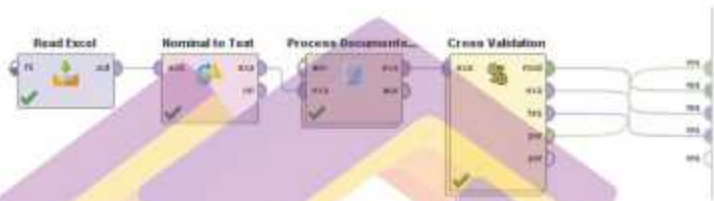
Skenario 4 : 973 data Positif dan 240 data Negatif.

4.5. Proses Analisis dengan Rapid Miner

Data yang sudah melalui tahapan pre-processing dan labeling, kemudian diimport pada aplikasi Rapid Miner. Pada aplikasi ini akan menggunakan 4 operator yaitu Read Excel untuk membaca file excel, Nominal to Text untuk mengkonversi seluruh atribut nominal kedalam atribut sting. Kemudian Process Documents from Data untuk memproses dataset yang sudah masuk dengan TF-IDF, dan yang terakhir adalah Cross Validation yang memiliki dua subproses yaitu subproses Pelatihan dan subproses Pengujian. Subproses Pelatihan digunakan untuk melatih model. Model yang dilatih kemudian diterapkan dalam subproses Pengujian. Kinerja model diukur selama fase Pengujian.

Seluruh operator ini dihubungkan dengan konektor exa atau example set yang merupakan konektor untuk data table/dataset. Kemudian pada operator cross validation terdapat beberapa konektor output yaitu mod atau model, exa atau example set, tes atau test result data dan per yang merupakan performance. Masing

masing konektor output akan dihubungkan dengan res atau port akhir untuk mengeluarkan hasil sesuai dengan konektor yang terkoneksi. Untuk alur yang digunakan kedua algoritma ini sama. Hanya berbeda pada metode yang digunakan saja, yang satu menggunakan SVM, yang satunya menggunakan Rapid Miner.



Gambar 4. 20 Proses analisis pada Rapid Miner

4.6. Pembobotan dengan menggunakan TF-IDF

Selanjutnya akan masuk kedalam tahap pembobotan dengan menggunakan TF-IDF pada Process Documents from Data yang nantinya hasil ini akan digunakan pada algoritma SVM maupun Random Forest. Karena proses pembobotannya tidak terlihat, sehingga ditambahkan perhitungan manual. Pada tabel 4.11 merupakan contoh komentar yang akan dihitung bobotnya.

Tabel 4. 11 Contoh Komentar

No	Komentar
1	nice simple display easy understand works well
2	reliable professional cashier job
3	easy use beginners

Langkah pertama pada perhitungan TF.IDF adalah dengan mengihitung TF terlebih dahulu. Pertama, kata perlu dipecah dan di hitung berapa kali muncul pada seluruh komentar yang dicontohkan pada tabel 4.12.

Tabel 4. 12 Perhitungan TF

Term	TF		
	D1	D2	D3
beginners	0	0	1
cashier	0	1	0
display	1	0	0
easy	1	0	1
job	0	1	0
nice	1	0	0
professional	0	1	0
reliable	0	1	0
simple	1	0	0
understand	1	0	0
use	0	0	1
well	1	0	0
works	1	0	0

Kemudian dihitung Panjang dokumen pada masing masing contoh yang menghasilkan data 1 adalah 7, data 2 adalah 4 dan data 3 adalah 3. Setelah itu, setiap angka pada tabel 4.12 akan dibagi dengan panjang masing masing dokumen.

Tabel 4. 13 Perhitungan Normalisasi TF

Term	TF		
	D1	D2	D3
beginners	0	0	0.33
cashier	0	0.25	0
display	0.14	0	0
easy	0.14	0	0.33
job	0	0.25	0
nice	0.14	0	0
professional	0	0.25	0
reliable	0	0.25	0
simple	0.14	0	0
understand	0.14	0	0
use	0	0	0.33
well	0.14	0	0
works	0.14	0	0

Setelah mengetahui nilai dari masing masing kata, kemudian selanjutnya adalah menghitung DF. Untuk menghitung DF cukup dengan menghitung kemunculan tiap kata pada seluruh dokumen yang akan ditampilkan pada tabel 14.

Tabel 4. 14 Perhitungan DF

Term	DF
beginners	1
cashier	1
display	1
easy	2
job	1
nice	1
professional	1
reliable	1
simple	1
understand	1
use	1
well	1
works	1

Untuk menghitung nilai IDF dengan memasukkan rumus *log(jumlah dokumen/df)*. Sehingga akan diketemukan hasilnya pada tabel 4.15.

Tabel 4. 15 Perhitungan IDF

Term	DF
beginners	0.477
cashier	0.477
display	0.477
easy	0.176
job	0.477
nice	0.477
professional	0.477
reliable	0.477
simple	0.477
understand	0.477
use	0.477
well	0.477
works	0.477

Setelah diketahui seluruh nilai TF dan IDF maka bisa dihitung untuk bobotnya menggunakan TF.IDF dengan mengkalikan nilai TF dan IDF.

Tabel 4. 16 Perhitungan TF/IDF

Term	TF.IDF		
	D1	D2	D3
beginners	0	0	0.159
cashier	0	0.119	0
display	0.068	0	0
easy	0.025	0	0.059
job	0	0.119	0
nice	0.068	0	0
professional	0	0.119	0
reliable	0	0.119	0
simple	0.068	0	0
understand	0.068	0	0
use	0	0	0.159
well	0.068	0	0
works	0.068	0	0

Sehingga jika melihat pada tabel diatas, tiap kata pada masing masing kalimat bisa jadi memiliki nilai yang berbeda. Pada dokumen 1 kata easy bernilai 0.025 sedangkan pada dokumen 3 bernilai 0.059. Hal ini bisa saja terjadi, tergantung dengan jumlah dokumen, jumlah kemunculan kata dan juga berapa panjang dokumen. Dan juga 1 kata bisa memiliki nilai yang berbeda seperti kata easy.

4.7. Analisis menggunakan algoritma Support Vector Machine

Setelah itu data akan proses dengan menggunakan algoritma Support Vector Machine menggunakan aplikasi RapidMiner. Pada proses ini dilakukan dalam operator Cross Validation. Operator SVM digunakan pada sub-proses Training, sedangkan operator Apply Model dan Performance digunakan pada sub-proses Testing. Input yang digunakan pada operator SVM adalah tra atau training set yang

didapatkan dari proses sebelumnya, kemudian akan menghasilkan output mod atau model, est atau estimated performance, wei atau weight dan exa atau example. Untuk parameter yang digunakan dalam operator SVM adalah parameter standart tanpa ada tambahan lain seperti kernel type dot, C = 0, convergence epsilon = 0,1 dan seluruh checkbox (scale, balance cost, dll) tidak dicentang.

Selesai pada proses training, masuk kedalam proses testing yang diperlukan adalah input mod (model) sehingga yang dengan menggunakan input tersebut proses training dan testing bisa terhubung. Uni merupakan unlabelled data juga dihubungkan dari konektor tes yang memuat data tidak berlabel. Dari proses Apply Model ini akan menghasilkan output data yang sudah memiliki label (lab), yang kemudian di hitung performancenya pada operator Performance yang akan memberikan output per atau performance. Untuk nilai K yang digunakan dalam performance ini adalah 5.

Gambar 4. 21 Alur Support Vector Machine dalam Rapid Miner

Setelah di-*run* dengan alur yang sama tetapi dengan dokumen yang berbeda beda (4 skenario) mendapatkan hasil yang bisa dilihat pada tabel 4.17.

Tabel 4. 17 Hasil Akurasi dari algoritma SVM

No	Skenario	Jumlah Data	Akurasi
1	Skenario 1	1730	87.86%
2	Skenario 2	1667	87.46%
3	Skenario 3	1448	85.91%
4	Skenario 4	1213	85.24%

Skenario 1 yang diolah menggunakan metode Support Vector Machine dengan jumlah data 1730 mendapatkan akurasi sebesar 87.86%. Kemudian skenario 2 dengan jumlah data sebanyak 1667 mendapatkan akurasi sebesar 87.46%. Setelah itu skenario 3 yang diproses dengan jumlah data 1448 mendapatkan akurasi sebesar 85.91% dan yang terakhir pada skenario 4 dengan jumlah data 1213 mendapatkan akurasi sebesar 85.24%.

Dari tabel 4.17 dapat dilihat bahwa akurasi terbaik untuk metode Support Vector Machine terdapat pada skenario ke 1 dengan akurasi sebesar 87.86%. Sedangkan untuk akurasi terburuk terdapat pada skenario ke 4 dengan akurasi sebesar 85.24%. Selain itu, dengan nilai akurasi yang diatas 80% untuk seluruh skenario maka dapat dikatakan bahwa kinerja analisis adalah baik. Walaupun belum merepresentasikan keseluruhan analisis, akan tetapi bisa dibilang bahwa data yang memiliki proses pre-processing lebih sederhana didalamnya mendapatkan nilai yang lebih tinggi dibandingkan dengan tahapan pre-processing yang kompleks.

Setelah diketahui nilai akurasi dari masing masing skenario, berikutnya adalah nilai dari confussion matrix yang dapat dilihat pada tabel dibawah ini.

Tabel 4. 18 Hasil Confusion Matrix Algoritma SVM Skenario 1

	True Positive	True Negative
Pred Positive	1373	169
Pred Negative	41	147

Tabel 4. 19 Hasil Confusion Matrix Algoritma SVM Skenario 2

	True Positive	True Negative
Pred Positive	1315	170
Pred Negative	39	143

Tabel 4. 20 Hasil Confusion Matrix Algoritma SVM Skenario 3

	True Positive	True Negative
Pred Positive	1165	168
Pred Negative	36	79

Tabel 4. 21 Hasil Confusion Matrix Algoritma SVM Skenario 4

	True Positive	True Negative
Pred Positive	945	151
Pred Negative	28	89

Berdasarkan tabel diatas pada Algoritma Support Vector Machine bisa dikatakan bahwa:

a. Hasil prediksi dari dataset aplikasi Qasir yang memiliki nilai True Positive dan Pred Positive sebanyak 1373 pada skenario 1. Kemudian 1315 pada skenario 2. Selanjutnya 1165 pada skenario 3 dan yang terakhir 945 pada skenario 4.
b. Hasil prediksi dari dataset aplikasi Qasir yang memiliki nilai True Positive dan Pred Negative adalah sebanyak 41 pada skenario 1. Kemudian 39 pada skenario 2. Selanjutnya 36 pada skenario 3 dan 28 pada skenario 4.
c. Hasil prediksi dari dataset aplikasi Qasir yang memiliki nilai True Negative dan Pred Positive sebanyak 169 pada skenario 1. Kemudian 170 pada skenario 2. Selanjutnya 168 pada skenario 3 dan 151 pada skenario 4.
d. Hasil prediksi dari dataset aplikasi Qasir yang memiliki nilai True Negative dan Pred Negative sebanyak 53 pada skenario 1. Kemudian 60 pada skenario 2. Selanjutnya 14 pada skenario 3 dan 21 pada skenario 4.

Setelah diketahui nilai confusion matrix berikutnya adalah nilai AUC, Precision dan Recall. Nilai yang dihasilkan oleh AUC dapat digunakan sebagai patokan untuk melihat kinerja algoritma. Kemudian Recall memiliki fungsi untuk mengevaluasi prediksi yang dilakukan. Dan precision digunakan untuk mengevaluasi seberapa baik ketepatan algoritma untuk memprediksi.

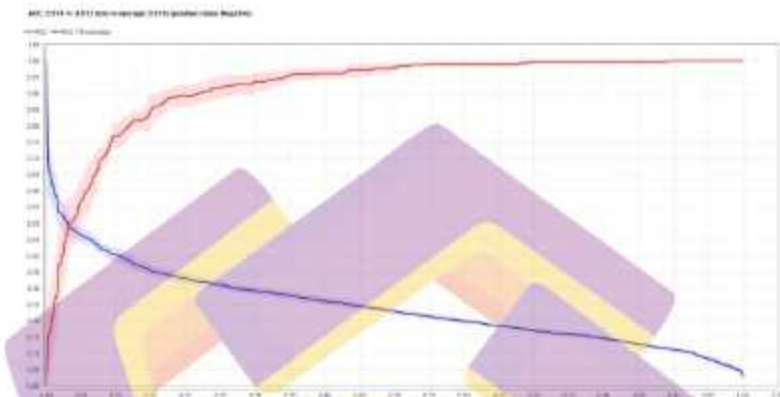
Nilai AUC didapatkan dari sebuah area dibawah kurva ROC yang dibentuk dari nilai True Positif yang merupakan jumlah data dengan prediksi nilai positif dan nilai aslinya positif dan False Positif yang merupakan jumlah data dengan prediksi nilai negatf dan nilai aslinya positif yang nanti setiap scenario akan ditampilkan grafiknya pada halaman selanjutnya. Nilai ini dapat digunakan untuk menggambarkan tingkat keakuratan klasifikasi yang dilakukan. Semakin mendekati 100% semakin baik klasifikasi yang dilakukan. Jika akurasi menilai seberapa akurat hasil klasifikasi yang dilakukan, maka AUC merupakan nilai dari seberapa akurat model atau klasifikasi untuk menghasilkan nilai akhir yang dilakukan.

Tabel 4. 22 Hasil AUC, Precision, dan Recall dari algoritma SVM

No	Skenario	AUC	Precision	Recall
1	Skenario 1	91.80%	89.04%	97.10%
2	Skenario 2	91.90%	88.55%	97.12%
3	Skenario 3	89.30%	87.40%	97.00%
4	Skenario 4	88.10%	86.22%	97.12%

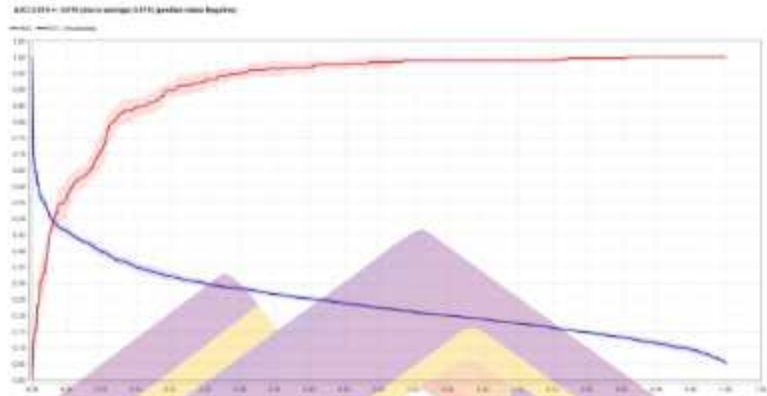
Berdasarkan penjelasan diatas, maka skenario 1 memiliki ketepatan dalam memprediksi sebesar 89,04% (precision), kemudian nilai performa untuk prediksi yang sudah dilakukan sebesar 97.10% (recall) dan yang terakhir kinerja dari keseluruhan algoritma SVM dinilai sebesar 91.80% (nilai AUC). Sehingga dapat

disimpulkan bahwa skenario 1 bekerja dengan sangat baik pada masing masing aspeknya.



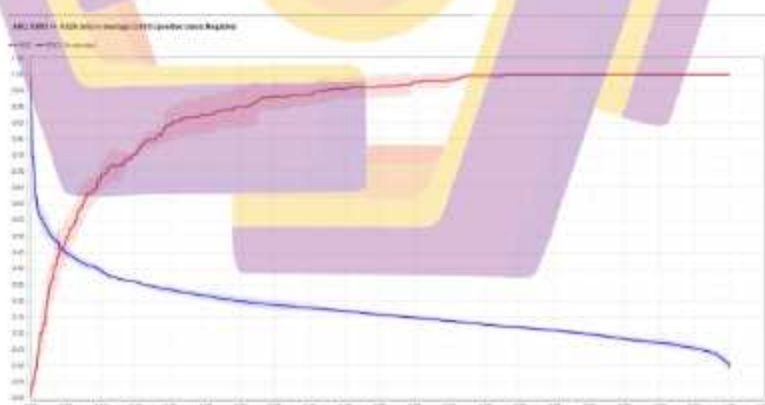
Gambar 4. 22 AUC pada skenario 1

Selanjutnya skenario 2 memiliki ketepatan dalam memprediksi sebesar 88.55% (precision), kemudian nilai performa untuk prediksi yang sudah dilakukan sebesar 97.12% (recall) dan yang terakhir kinerja dari keseluruhan algoritma SVM dinilai sebesar 91.90% (nilai AUC). Sehingga dapat disimpulkan bahwa skenario 2 bekerja dengan sangat baik pada masing masing aspeknya.



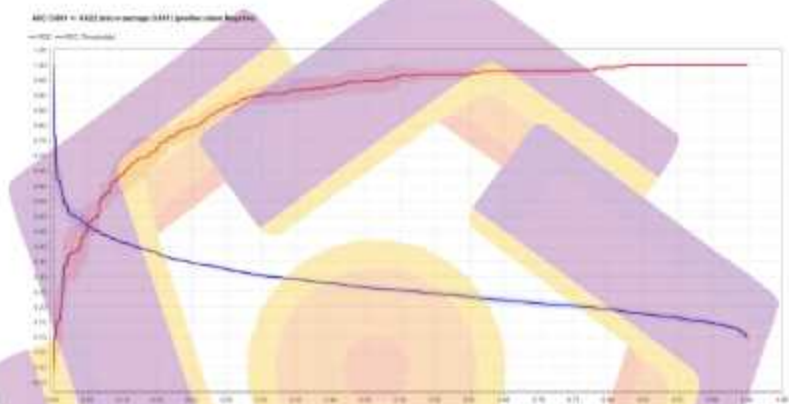
Gambar 4. 23 AUC pada skenario 2

Adapun skenario 3 memiliki ketepatan dalam memprediksi sebesar 87.40% (precision), kemudian nilai performa untuk prediksi yang sudah dilakukan sebesar 97.00% (recall) dan yang terakhir kinerja dari keseluruhan algoritma SVM dinilai sebesar 89.30% (nilai AUC). Sehingga dapat disimpulkan bahwa skenario 3 bekerja dengan sangat baik pada masing masing aspeknya.



Gambar 4. 24 AUC pada skenario 3

Terakhir, skenario 4 memiliki ketepatan dalam memprediksi sebesar 86.22% (precision), kemudian nilai performa untuk prediksi yang sudah dilakukan sebesar 97.12% (recall) dan yang terakhir kinerja dari keseluruhan algoritma SVM dinilai sebesar 88.10% (nilai AUC). Sehingga dapat disimpulkan bahwa skenario 4 bekerja dengan sangat baik pada masing masing aspeknya.



Gambar 4. 25 AUC pada skenario 4

Dari tabel 17 dapat disimpulkan bahwa AUC terbaik dari seluruh skenario yang diproses dengan algoritma Support Vector Machine dimiliki oleh skenario 2 (91.90%). Kemudian nilai precision terbaik dimiliki oleh skenario 1 (89.04%), dan nilai Recall terbaik dimiliki oleh skenario 2 (97.12%).

Berbeda dengan penilaian akurasi yang semakin sedikit tahapan pre-processing dilakukan, maka semakin tinggi nilai akurasi yang didapatkan, nilai AUC/Precision dan Recall ini memiliki kesimpulan yang berbeda beda. Skenario 2 memiliki nilai AUC dan Recall yang lebih baik dibandingkan skenario 1. Sebagai perbandingan skenario 1 dan 2 hanya pada tahapan cleaning data (pre-processing). Jika dapat disimpulkan bahwa tahapan cleaning data dapat meningkatkan nilai

AUC dan Recall pada analisis dengan metode SVM. Akan tetapi hal ini perlu digali dan dikaji lebih dalam pada penelitian selanjutnya sejenis.

Kemudian dari nilai yang didapatkan dari Confusion Matrix dapat dihitung nilai akurasi, recall, dan precision untuk memastikan perhitungan dari Rapid Miner adalah valid. Untuk mengujinya akan dihitung salah satu skenario saja yaitu skenario 2.

a. Nilai Akurasi

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

$$Akurasi = \frac{1315 + 143}{143 + 39 + 170 + 1315}$$

$$Akurasi = \frac{1458}{1667} = 0.8746$$

b. Nilai Recall

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

$$Recall = \frac{1315}{1315 + 39}$$

$$Recall = \frac{1315}{1354} = 0.97119$$

c. Nilai Precision

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

$$Precision = \frac{1315}{1315 + 170}$$

$$Precision = \frac{1315}{1485} = 0.8855$$

Dari ketiga perhitungan diatas disimpulkan bahwa nilai yang didapatkan pada aplikasi rapid miner adalah valid. Ketiga nilai diatas merupakan representative kedekatan antara hasil penelitian dengan nilai yang sebenarnya. Akurasi merupakan nilai yang didapatkan perbandingan nilai hasil klasifikasi dengan nilai yang sebenarnya. Kemudian Precision merupakan kecocokan antara bagian klasifikasi dengan informasi yang dibutuhkan, dan yang terakhir Recall merupakan tingkatan keberhasilan system dalam memanggil informasi. Sedangkan untuk melihat seberapa baik kinerja klasifikasi dapat diukur dengan menggunakan nilai Precision, Recall dan AUC.

Dari keseluruhan nilai akurasi, recall, precision dan AUC yang dihasilkan oleh masing masing skenario dengan menggunakan algoritma Support Vector Machine dapat disimpulkan bahwa skenario 1 dan 2 memiliki performa yang stabil dan lebih baik dibandingkan skenario yang lainnya. Mengapa 2 skenario yang di highlight, karena keduanya sama sama unggul pada 2 kategori yaitu skenario 1 unggul dalam nilai Akurasi dan Precision, sedangkan skenario 2 unggul dalam nilai AUC dan Recall. Walaupun nilai perbedaan antara skenario dan kategori atau aspek tidaklah jauh. Akan tetapi dari point ini bisa ditarik kesimpulan bahwa semakin minim text pre-processing yang digunakan semakin baik pula hasilnya, dengan skenario 1 yang menggunakan case folding, remove frequent word, remove rare word dan skenario 2 menggunakan case folding, remove frequent word, remove rare word dan cleaning.

4.8. Analisis menggunakan algoritma Random Forest

Data akan diproses dengan menggunakan algoritma Random Forest di aplikasi Rapid Miner. Pada proses ini dilakukan dalam operator Cross Validation. Operator Random Forest digunakan pada sub-proses Training, sedangkan operator Apply Model dan Performance digunakan pada sub-proses Testing. Input yang digunakan pada operator Random Forest adalah tra atau training set yang didapatkan dari proses sebelumnya, kemudian akan menghasilkan output mod atau model, wei atau weight dan exa atau example. Untuk parameter yang digunakan dalam operator Random Forest adalah parameter standart tanpa ada tambahan lain seperti number of trees 100, criterion = gain_ratio, voting strategy = confidence vote dan seluruh checkbox (apply pruning, prepruning dan guess subset ratio) tidak dicentang.

Selesai pada proses training, masuk kedalam proses testing yang diperlukan adalah input mod (model) sehingga yang dengan menggunakan input tersebut proses training dan testing bisa terhubung. Uni merupakan unlabelled data juga dihubungkan dari konektor tes yang memuat data tidak berlabel. Dari proses Apply Model ini akan menghasilkan output data yang sudah memiliki label (lab), yang kemudian di hitung performancenya pada operator Performance yang akan memberikan output per atau performance. Untuk nilai K yang digunakan dalam performance ini adalah 5.

Gambar 4. 26 Alur Random Forest pada Rapid Miner

Setelah di-*run* dengan alur yang sama tetapi jumlah data yang berbeda-beda mendapatkan hasil yang bisa dilihat pada tabel 4.23.

Tabel 4. 23 Hasil Akurasi dari algoritma Random Forest

No	Skenario	Jumlah Data	Akurasi
1	Skenario 1	1730	83.01%
2	Skenario 2	1667	82.96%
3	Skenario 3	1448	82.67%
4	Skenario 4	1213	81.86%

Skenario 1 yang diolah menggunakan algoritma Random Forest dengan jumlah data 1730 mendapatkan akurasi sebesar 83.01%. Kemudian skenario 2 dengan jumlah data sebanyak 1667 mendapatkan akurasi sebesar 82.96 %. Setelah itu skenario 3 yang diproses dengan jumlah data 1448 mendapatkan akurasi sebesar 82.67% dan yang terakhir pada skenario 4 dengan jumlah data 1213 mendapatkan akurasi sebesar 81.86%.

Dari tabel 4.23 dapat dilihat bahwa akurasi terbaik untuk algoritma Random Forest terdapat pada skenario 1 dengan akurasi sebesar 83.01%. Untuk akurasi terburuk terdapat pada skenario 4 dengan akurasi sebesar 81.86%. Sama dengan algoritma SVM, akurasi pada algoritma Random Forest pun memiliki kesimpulan bahwa data yang tidak memiliki proses pre-processing kompleks didalamnya mendapatkan nilai yang lebih tinggi dibandingkan dengan tahapan pre-processing yang kompleks. Walaupun perbedaan akurasi antar skenarionya sangatlah kecil, akan tetapi tetap saja semakin kompleks tahapan pre-processingnya semakin kecil pula nilai akurasi yang didapat. Selain itu, dengan nilai akurasi yang diatas 80% untuk seluruh skenario maka dapat dikatakan bahwa kinerja analisis adalah baik.

Setelah diketahui masing masing nilai akurasi, selanjutnya adalah nilai dari confussion matrix yang dimiliki oleh masing masing skenario yang dapat dilihat pada tabel dibawah ini.

Tabel 4. 24 Hasil Confusion Matrix Algoritma Random Forest Skenario 1

	True Positive	True Negative
Pred Positive	1406	286
Pred Negative	8	30

Tabel 4. 25 Hasil Confusion Matrix Algoritma Random Forest Skenario 2

	True Positive	True Negative
Pred Positive	1349	279
Pred Negative	5	34

Tabel 4. 26 Hasil Confusion Matrix Algoritma Random Forest Skenario 3

	True Positive	True Negative
Pred Positive	1192	242
Pred Negative	9	5

Tabel 4. 27 Hasil Confusion Matrix Algoritma Random Forest Skenario 4

	True Positive	True Negative
Pred Positive	969	216
Pred Negative	4	24

Berdasarkan tabel diatas pada Algoritma Random Forest bisa ditarik kesimpulan:

a. Hasil prediksi dari dataset aplikasi Qasir yang memiliki nilai True Positive dan Prediksi Positive adalah 1406 untuk skenario 1. Kemudian 1349 untuk skenario 2. Selanjutnya 1192 untuk skenario 3 dan 969 untuk skenario 4.

b. Hasil prediksi dari dataset aplikasi Qasir yang memiliki nilai True Positive dan Prediksi Negative adalah 8 untuk skenario 1. Kemudian 5 untuk skenario 2. Selanjutnya 9 untuk skenario 3 dan 4 skenario 4.

c. Hasil prediksi dari dataset aplikasi Qasir yang memiliki nilai True Negative dan Prediksi Positive adalah 286 untuk skenario 1. Kemudian 279 untuk skenario 2. Selanjutnya 242 untuk skenario 3 dan 216 skenario 4.

d. Hasil prediksi dari dataset aplikasi Qasir yang memiliki nilai True Negative dan Prediksi Negative adalah 30 untuk skenario 1. Kemudian 34 untuk skenario 2. Selanjutnya 5 untuk skenario 3 dan 24 skenario 4.

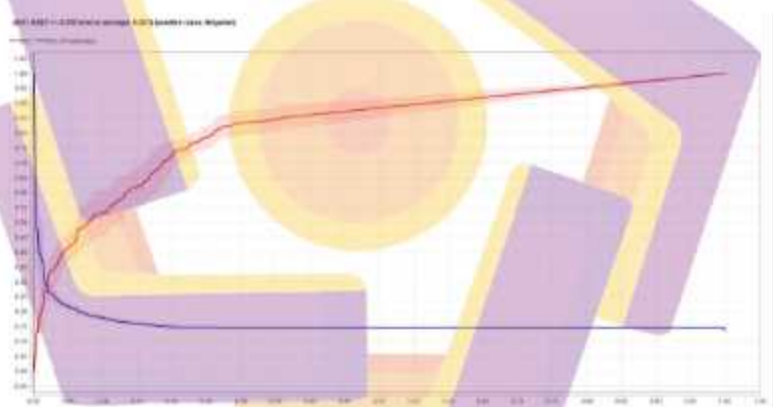
Setelah diketahui nilai confusion matrix berikutnya adalah nilai AUC, Precision dan Recall. Nilai yang dihasilkan oleh AUC dapat digunakan sebagai patokan untuk melihat kinerja algoritma. Kemudian Recall memiliki fungsi untuk mengevaluasi prediksi yang dilakukan. Dan precision digunakan untuk mengevaluasi seberapa baik ketepatan algoritma untuk memprediksi.

Kemudian nilai AUC didapatkan dari sebuah area dibawah kurva ROC yang dibentuk dari nilai True Positif yang merupakan jumlah data dengan prediksi nilai positif dan nilai aslinya positif dan False Positif yang merupakan jumlah data dengan prediksi nilai negatf dan nilai aslinya positif yang nanti setiap scenario akan ditampilkan grafiknya pada halaman selanjutnya. Nilai ini dapat digunakan untuk menggambarkan tingkat keakuratan klasifikasi yang dilakukan. Semakin mendekati 100% semakin baik klasifikasi yang dilakukan. Jika akurasi menilai seberapa akurat hasil klasifikasi yang dilakukan, maka AUC merupakan nilai dari seberapa akurat model atau klasifikasi untuk menghasilkan nilai akhir yang dilakukan.

Tabel 4. 28 Hasil AUC, Precision, dan Recall dari algoritma Random Forest

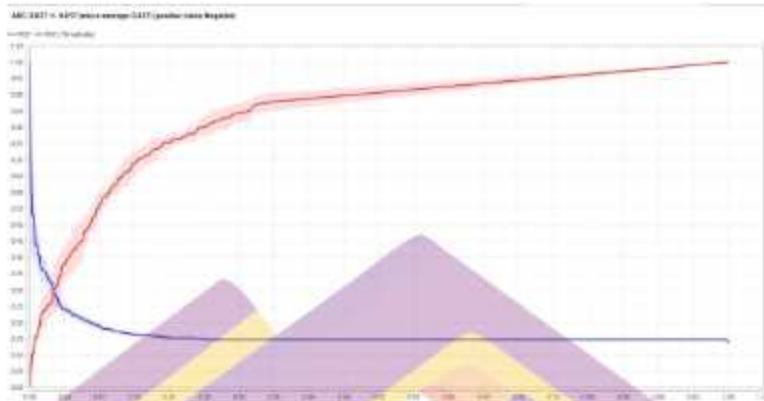
No	Skenario	AUC	Precision	Recall
1	Skenario 1	82.30%	83.10%	99.43%
2	Skenario 2	83.70%	82.86%	99.63%
3	Skenario 3	81.40%	83.12%	99.25%
4	Skenario 4	83.00%	81.77%	99.59%

Skenario 1 dapat menghasilkan nilai AUC sebesar 82.30%, berdasarkan tabel klasifikasi performa nilai yang dihasilkan adalah baik karena berada diatas 80%. Selanjutnya nilai precision sebesar 83.10% dan nilai recall sebesar 99,43%. Dapat disimpulkan bahwa skenario 1 diproses dengan baik pada masing masing aspeknya.



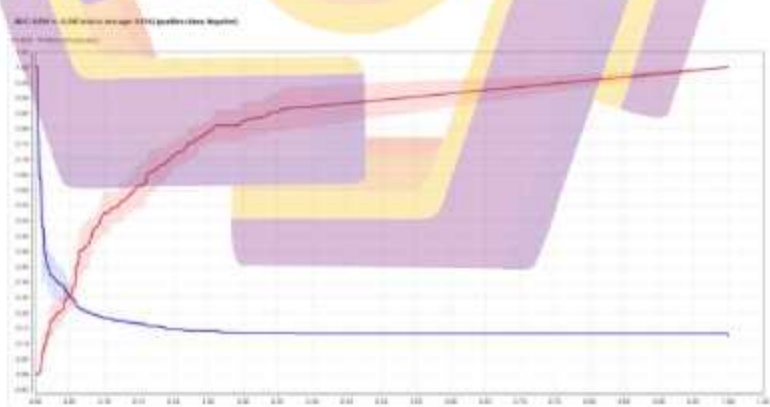
Gambar 4. 27 AUC pada skenario 1 dengan Random Forest

Selanjutnya skenario 2 menghasilkan nilai AUC sebesar 83.70%, berdasarkan tabel klasifikasi performa nilai yang dihasilkan adalah baik karena berada diatas 80%. Selanjutnya skenario 2 memiliki nilai precision sebesar 82.86% dan nilai recall sebesar 99.63%. Sehingga dapat disimpulkan bahwa skenario 2 diproses dengan baik pada masing masing aspeknya.



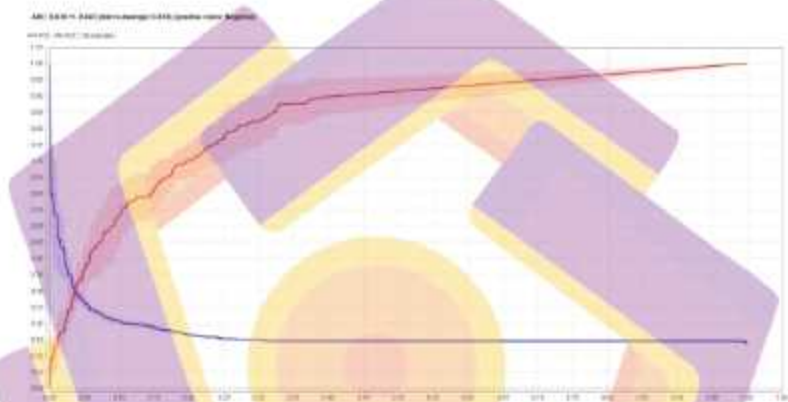
Gambar 4. 28 AUC pada skenario 2 dengan Random Forest

Adapun skenario 3 menghasilkan nilai AUC sebesar 81.40%, sehingga menurut tabel klasifikasi performa nilai yang dihasilkan adalah cukup karena berada diatas 80%. Selanjutnya skenario 3 memiliki nilai precision sebesar 83.12% dan nilai recall sebesar 99.25%. Sehingga dapat disimpulkan bahwa skenario 3 diproses dengan baik pada masing masing aspeknya.



Gambar 4. 29 AUC pada skenario 3 dengan Random Forest

Terakhir, skenario 4 menghasilkan nilai AUC sebesar 83.00%, sehingga menurut tabel klasifikasi performa nilai yang dihasilkan adalah baik karena berada diatas 80%. Selanjutnya skenario 4 memiliki nilai precision sebesar 81.77% dan nilai recall sebesar 99.59%. Sehingga dapat disimpulkan bahwa skenario 4 diproses dengan baik pada masing masing aspeknya.



Gambar 4. 30 AUC pada skenario 4 dengan Random Forest

Pada algoritma ini, nilai AUC tertinggi dimiliki oleh skenario 2 dengan nilai 83.70%, dimana nilai akurasi dari skenario 3 adalah yg terburuk yaitu 81.40%. Sedangkan nilai precision tertinggi dimiliki oleh skenario 3 dengan nilai 83.12%. Dan untuk nilai recall tertinggi dimiliki oleh skenario 2 dengan nilai 99.63%.

Sejalan dengan kesimpulan yang diambil pada algoritma SVM, pada algoritma Random Forest ini nilai AUC/Precision dan Recall ini memiliki kesimpulan yang berbeda-beda. Skenario 2 memiliki nilai AUC dan Recall paling baik. Jika dapat disimpulkan bahwa tahapan cleaning data (dibandingkan dengan tanpa cleaning data pada skenario 1) dapat meningkatkan nilai AUC dan Recall

pada analisis dengan metode Random Forest. Akan tetapi hal ini perlu digali dan dikaji lebih dalam pada penelitian selanjutnya sejenis.

Kemudian dari nilai-nilai yang didapatkan dari Confusion Matrix dapat dihitung nilai akurasi, recall, dan precision untuk memastikan perhitungan dari Rapid Miner adalah valid. Untuk mengujinya akan dihitung salah satu skenario saja yaitu skenario 1.

a. Nilai Akurasi

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

$$Akurasi = \frac{1406 + 30}{30 + 8 + 286 + 1406}$$

$$Akurasi = \frac{1436}{1730} = 0.83005$$

b. Nilai Recall

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

$$Recall = \frac{1406}{1406 + 8}$$

$$Recall = \frac{1406}{141} = 0.9943$$

c. Nilai Precision

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

$$Precision = \frac{1406}{1406 + 286}$$

$$Precision = \frac{1406}{1692} = 0.8309$$

Dari ketiga perhitungan diatas disimpulkan bahwa nilai yang didapatkan pada aplikasi rapid miner adalah valid. Ketiga nilai diatas merupakan representative kedekatan antara hasil penelitian dengan nilai yang sebenarnya. Akurasi merupakan nilai yang didapatkan perbandingan nilai hasil klasifikasi dengan nilai yang sebenarnya. Kemudian Precision merupakan kecocokan antara bagian klasifikasi dengan informasi yang dibutuhkan, dan yang terakhir Recall merupakan tingkatan keberhasilan system dalam memanggil informasi. Sedangkan untuk melihat seberapa baik kinerja klasifikasi dapat diukur dengan menggunakan nilai Precision, Recall dan AUC.

Untuk menentukan mana skenario terbaik dari keempat skenario harus dilihat dari segala aspek, yaitu akurasi, AUC, precision dan recall. Untuk skenario ke 1 memiliki nilai akurasi tertinggi dengan nilai 83.01% tetapi nilai dari aspek lainnya lebih baik ditemukan pada skenario lainnya. Kemudian untuk skenario ke 2 yang memiliki nilai akurasi tertinggi ke 2 dengan nilai 82.96%, juga memiliki nilai AUC dan Recall tertinggi dengan nilai 83.70% dan 99.63%. Sedangkan untuk skenario ke 3, memiliki nilai akurasi tertinggi ke 3 dengan nilai 82.67% dan nilai Precision tertinggi dengan nilai 83.12%. Jika dilihat dari keseluruhan akurasi maka skenario 2 adalah yang terbaik, karena unggul dalam 2 aspek yaitu AUC dan juga Recall. Walaupun nilai akurasi pada skenario 1 dan nilai precision pada skenario 3 lebih baik, akan tetapi jika dilihat dari keseluruhan aspek maka skenario 2 (case folding, remove frequent word, remove rare word dan cleaning) paling baik digunakan untuk dataset aplikasi Qasir dengan menggunakan algoritma Random Forest.

4.9. Wordcloud

Berikut merupakan kumpulan kata yang sering muncul saat pre-processing data, kemudian data tersebut diolah dan divisualisasikan dengan menggunakan wordcloud.

Gambar 4. 31 Worldcould Positive

Gambar 4. 32 Worldcould Negative

Dari gambar tersebut terdapat beberapa kata yang terhighlight yaitu Cashier, Simple, Menu, Great untuk positif. Error, Problem, Fix, Print, Transaction untuk negatif. Dengan ini, dapat disimpulkan bahwa kata kata yang terhighlight merupakan kata yang banyak muncul dalam dataset yang diolah. Atau mungkin kata kata yang muncul ini juga merupakan kata kunci untuk mengetahui hal apa saja yang positif dan negatif dari aplikasi ini.

Salah satu contoh komentar positif yang memuat salah satu kata yang ada pada wordcloud : *A simple application but very useful. There are still many improvements but applaud the developers who are open to the obstacles we encounter. Hopefully the features will increase*. Komentar ini menganggap aplikasi ini simpel dan sangat berguna walaupun memiliki banyak ruang untuk dikembangkan lagi.

Dan salah satu contoh komentar negatif: *Login as operator or supervisor the receipt is printed only half without subtotal, total, pay and change. Login as owner of the receipt, everything is print-ed. The store cashier must log in as the owner so that the receipt can be printed completely. Fix the application again*. Komentar ini merupakan salah satu contoh bug pada fitur printer dengan spesifik user. Sehingga tim developer dapat memperbaiki sesuai komentar yang dipaparkan. Secara keseluruhan, dalam komentar positif rata rata pengguna menuliskan tentang penggunaan aplikasi Qasir yang simple, useful, dan bagus. Dan pada komentar negatif, pengguna mendapatkan error pada saat menggunakan aplikasi di fitur print, transaksi dan pada saat menggunakan aplikasi Pro (Qasir-Pro/ layanan berbayar). Dari kata kata yang muncul pada gambar 4.29 dan gambar 4.30, mempermudah

pengembang dari aplikasi untuk dapat mengetahui apa saja yang harus diperbaiki dan ditingkatkan berdasarkan banyaknya kata yang muncul sehingga aplikasi Qasir menjadi lebih baik lagi.

4.10. Skenario Pengujian

Skenario Pengujian yang dimaksud adalah untuk membandingkan nilai akurasi, AUC, precision dan recall dari kedua algoritma dengan keempat skenario sehingga bisa ditarik kesimpulan mana algoritma yang paling baik.

Tabel 4. 29 Hasil Perbandingan Akurasi

	Skenario 1	Skenario 2	Skenario 3	Skenario 4
SVM	87.86%	87.46%	85.91%	85.24%
RF	83.01%	82.96%	82.67%	81.86%
Perbandingan	SVM	SVM	SVM	SVM

Pada keseluruhan skenario hasil lebih baik didapatkan oleh algoritma Support Vector Machine dibandingkan dengan Random Forest dengan perbedaan 3% - 5% pada masing masing skenario. Hal ini membuktikan bahwa dari akurasi algoritma Support Vector Machine lebih unggul. Selanjutnya adalah perbandingan nilai AUC dari masing masing skenario untuk kedua algoritma.

Tabel 4. 30 Hasil Perbandingan AUC

	Skenario 1	Skenario 2	Skenario 3	Skenario 4
SVM	91.80%	91.90%	89.30%	88.10%
RF	82.30%	83.70%	81.40%	83.00%
Perbandingan	SVM	SVM	SVM	SVM

Sejalan dengan nilai akurasi, pada nilai AUC ini algoritma Support Vector Machine lebih baik di seluruh skenario dibandingkan dengan algoritma Random Forest. Dengan perbedaan nilai sebesar 5% - 9% pada masing masing skenario, membuktikan bahwa algoritma Support Vector Machine lebih unggul. Ketiga

adalah perbandingan nilai Precision dari masing masing skenario untuk kedua algoritma.

Tabel 4. 31 Hasil Perbandingan Precision

	Skenario 1	Skenario 2	Skenario 3	Skenario 4
SVM	89.04%	88.55%	87.40%	86.22%
RF	83.10%	82.86%	83.12%	81.77%
Perbandingan	SVM	SVM	SVM	SVM

Dengan perbedaan nilai sebesar 4% - 6% pada masing masing skenario, membuktikan bahwa algoritma Support Vector Machine lebih unggul pada seluruh skenario dibandingkan dengan algoritma Random Forest. Dan yang terakhir adalah perbandingan nilai Recall dari masing masing skenario untuk kedua algoritma.

Tabel 4. 32 Hasil Perbandingan Recall

	Skenario 1	Skenario 2	Skenario 3	Skenario 4
SVM	97.10%	97.12%	97.00%	97.12%
RF	99.43%	99.63%	99.25%	99.59%
Perbandingan	RF	RF	RF	RF

Bebeda dengan nilai aspek lainnya seperti akurasi, precision dan AUC, pada nilai recall ini algoritma Randpm Forest lebih baik di seluruh skenario dibandingkan dengan algoritma Support Vector Machine. Dengan memiliki perbedaan nilai sebesar 2% pada masing masing skenario

4.11. Kesimpulan Pengujian

Setelah diketahui hasil perbandingan seluruh aspek (akurasi, AUC, precision, recall) dari keempat skenario dapat disimpulkan bahwa :

- **Akurasi** tertinggi dimiliki oleh algoritma Support Vector Machine dengan nilai 87.86% pada skenario 1.

- Pada perbandingan akurasi, algoritma Support Vector Machine mendominasi dengan seluruh skenarionya memiliki nilai tertinggi,
- **AUC** tertinggi dimiliki oleh algoritma Support Vector Machine dengan nilai 91.90 % pada skenario 2.
- Pada perbandingan AUC, algoritma Support Vector Machine mendominasi dengan seluruh skenarionya memiliki nilai tertinggi,
- **Precision** tertinggi dimiliki oleh algoritma Support Vector Machine dengan nilai 89.04% pada skenario 1.
- Pada perbandingan Precision, algoritma Support Vector Machine memiliki nilai lebih baik pada seluruh skenarionya.
- **Recall** tertinggi dimiliki oleh algoritma Random Forest dengan nilai 99.63% pada skenario 2.
- Pada perbandingan Recall, algoritma Random Forest memiliki nilai lebih baik pada seluruh skenarionya.

Dengan adanya kesimpulan diatas maka seluruh pertanyaan pada rumusan masalah sudah terjawab. Support Vector Machine dan Random Forest memiliki performa yang baik dalam menganalisis sentimen opini pengguna aplikasi Qasir karena keduanya memiliki nilai di masing masing aspeknya diatas 80%.

Kemudian berdasarkan hasil pemaparan diatas bahwa skenario ke 1 (case folding, remove frequent word dan remove rare word) dan skenario ke 2 (case folding, remove frequent word, remove rare word dan cleaning) adalah tahapan skenario yang paling berpengaruh dalam proses penelitian ini. Karena pada kedua skenario tersebut memiliki nilai tertinggi di 2 kategori pada masing masing aspek,

yaitu skenario 1 memiliki nilai tertinggi pada aspek Akurasi dan Precision, sedangkan skenario 2 memiliki nilai tertinggi pada AUC dan Recall. Dan yang terakhir, metode Support Vector Machine merupakan metode yang bekerja paling baik di keseluruhan skenario tahapan preprocessing. Bisa dilihat pada penjabaran kesimpulan diatas bahwa hasil dari algoritma Support Vector Machine lebih baik hampir diseluruh skenario yang ada

4.12. Perbandingan dengan Penelitian sebelumnya

Setelah mendapatkan hasil akhir dari penelitian ini, maka terdapat beberapa perbedaan baik dari jumlah data, algoritma yang digunakan serta hasil analisis dengan penelitian terdahulu. Terdapat 6 penelitian terdahulu yang telah dibahas pada Bab II – Landasan Teori. Nilai akurasi dan AUC pada penelitian ini berkisar di 80% - 90%, sehingga posisi penelitian ini berada ditengah-tengah dengan penelitian lain jika dilihat dari hasil akhirnya (akurasi dan AUC).

Pada penelitian sebelumnya yang dilakukan oleh Fitri,Yuliani dkk dengan jumlah dataset yang tidak jauh berbeda (1629 data) dan algoritma yang sama menghasilkan nilai akurasi pada Random Forest sebesar 97,16% dengan nilai AUC 0.96, kemudian Support Vector Machine menghasilkan nilai akurasi sebesar 96.01% walaupun memiliki nilai AUC yang rendah yaitu 0.54. (Fitri, Yuliani, Rosyida, & Gata, 2020) Jika dibandingkan walaupun memiliki nilai akurasi yang lebih rendah, tetapi penelitian ini memiliki keunggulan nilai AUC yang lebih stabil yang mengindikasikan bahwa penelitian berjalan dengan baik.

Selain itu terdapat penelitian yang dilakukan oleh Eliyani yang menggunakan 3 algoritma yaitu Random Forest, Support Vector Machine dan

Naïve Bayes. Akan tetapi pada perbandingan kali ini algoritma Naïve Bayes tidak akan digunakan karena tidak ada nilai pembanding. Hasil akhir pada penelitian Eliyani menghasilkan nilai akurasi untuk Random Forest sebesar 75,81% dan Support Vector Machine sebesar 77,58% (Himawan & Eliyani, 2021). Berdasarkan hasil tersebut bisa dikatakan bahwa penelitian yang dilakukan saat ini memiliki analisis yang lebih baik, sehingga bisa meningkatkan akurasi pada algoritma Random Forest sebesar 7.2% dan algoritma Support Vector Machine sebesar 10.82%.

Kemudian, penelitian yang dilakukan oleh Firdausi Nuzula yang menuliskan penggunaan stemming pada tahapan preprocessing membantu untuk meningkatkan performance akurasi (Zamzami & Adiwijaya, 2021). Berbeda dengan itu, pada penelitian ini mendapatkan bahwa hasil terbaik pada skenario ke 1 dan 2 yaitu case folding, remove frequent word, remove rare word dan cleaning. (untuk skenario 1 tanpa cleaning) Stemming yang dilakukan pada skenario 4 mendapatkan hasil yang kurang baik dengan menggunakan algoritma Support Vector Machine ataupun Random Forest. Hal ini bisa saja terjadi karena adanya perbedaan struktur dataset, algoritma yang digunakan serta dari proses preprocessing itu sendiri.

BAB V

PENUTUP

5.1. Kesimpulan

Setelah melalui tahap pengujian pada algoritma Support Vector Machine dan Random Forest di keempat skenario pre-processing, maka dapat diambil beberapa kesimpulan antara lain:

1. Kedua algoritma dapat bekerja dengan baik dan menghasilkan nilai akurasi, AUC, Precision dan Recall untuk masing masing skenario diatas 80%.
2. Dan untuk tahapan yang berpengaruh dalam penelitian ini adalah skenario ke 1 (case folding, remove frequent word dan remove rare word) dan skenario ke 2 (case folding, remove frequent word, remove rare word dan cleaning. Karena pada kedua skenario tersebut memiliki nilai tertinggi di 2 kategori pada masing masing aspek, yaitu skenario 1 memiliki nilai tertinggi pada aspek Akurasi dan Precision, sedangkan skenario 2 memiliki nilai tertinggi pada AUC dan Recall.
3. Berdasarkan hasil analisa pada masing masing metode dan scenario bahwa metode Support Vector Machine memiliki performa yang paling baik dibandingkan metode Random Forest.

5.2. Saran

Saran yang dapat digunakan oleh peneliti-peneliti selanjutnya untuk meningkatkan hasil kesimpulan pada penelitian ini, adalah sebagai berikut :

1. Penggunaan data yang lebih banyak.
2. Tahapan pre-processing yang berbeda sehingga memungkinkan data yang akan diolah berbeda.

3. Menggunakan data berbahasa Indonesia melihat pengguna aplikasi Qasir rata rata WNI.
4. Menggunakan software lain (rapid miner) atau bahasa pemrograman lain.
5. Menggunakan pengaturan yang berbeda pada saat bagian klasifikasi pada Rapid Miner untuk mendapatkan hasil yang lebih baik.
6. Menggali lebih lanjut tentang Rapid Miner (tentang masing masing operator).

DAFTAR PUSTAKA

Ahmadi, M. I., Apriani, F., Kurniasari, M., Handayani, S., & Gustian, D. (2020). SENTIMENT ANALYSIS ONLINE SHOP ON THE PLAY STORE USING METHOD SUPPORT VECTOR MACHINE (SVM). *Seminar Nasional Informatika 2020*.

Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *TRANSFORMTIKA*.

Guia, M., Silva, R., & Bernardino, J. (2019). Comparison of Naïve Bayes, Support Vector Machine, Decision Tree and Random Forest on Sentiment Analysis. *SCITEPRESS – Science and Technology Publications*.

Himawan, R. D., & Eliyani. (2021). Perbandingan Akurasi Analisis Sentimen Tweet. *JEPIN ((Jurnal Edukasi dan Penelitian Informatika)*.

Holle, K. F. (2015). PEMBOBOTAN KATA BERBASIS PREFERENCE UNTUK PERANGKINGAN DOKUMEN FIQIH BERBAHASA ARAB. *Institut Teknologi Sepuluh November Repository*.

Julianto, R., Bintari, E. D., & Indrianti. (2017). Analisis Sentimen Layanan Provider Telepon Seluler pada Twitter menggunakan Metode Naïve Bayesian Classification. *Journal of Big Data Analytic and Artificial Intelligence*.

Jumeilah, F. S. (2017). Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian. *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*.

Khairunnisa, S., Adiwijaya, & Faraby, S. A. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *Jurnal Media Informatika Budidarma*.

Khiatuddin, M., & Muhammad. (2021). *Memberantas Buta Program Dengan Bahasa Python*. Bandar Lampung: UPPM Universitas Malahayati.

Kusuma, D. P. (2020). *Machine Learning Teori, Program dan Studi Kasus*. Yogyakarta: Deepublish Publlisher.

Lidya, S. K. (2014). SENTIMENT ANALYSIS PADA TEKS BAHASA INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBOR (K-NN). *Repository Universitas Sumatera Utara*.

Mesran, Sulaiman, O. K., Wijoyo, H., Putra, S. H., Watrianthos, R., Sinaga, R., . . . Indarto, S. L. (2020). *Merdeka Kreatif di Era Pandemi Covid-19: Suatu Pengantar*. Medan: Green Press.

Muflikah, L., Widodo, Mahmudi, W. F., & Solimun. (2021). *Machine Learning Dalam Bioinformatika*. Malang: UB Press.

Muktafin, E., Kusrini, & Luthfi, E. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *EKSPLORA INFORMATIKA*.

Mustaqim, T. (2020). SENTIMENT ANALYSIS OPINI PELANTIKAN KABINET PEMERINTAH INDONESIA TAHUN 2019

MENGGUNAKAN VADER DAN RANDOM FOREST. *Repository UNNES*.

Mustopa, A., Hermanto, Anna, Pratama, E. B., Hendini, A., & Risdiansyah, D. (2020). Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm Based on Particle Swarm Optimization. *IEEE*.

Narkhede, S. (2018, 05 09). *Understanding Confusion Matrix*. Retrieved from Towardsdatascience: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Nomleni, P. (2015). SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE (SVM). *Repository Institut Teknologi Sepuluh November*.

Nugraha, F. A., Habibi, R., & Harani, N. H. (2020). *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Bandung: Kreatif industri Nusantara.

Nugroho, D. (2020). Analisis Kerentanan Tanah Longsor Menggunakan Metode Frequency Ratio di Kabupaten Bandung Barat, Jawa Barat. *itenas*.

Pratiwi, D. L. (2018). Penerapan Metode Combine Sampling pada Klasifikasi Imbalanced Data Biner Status Ketertinggalan Desa di Jawa Timur. *ITS Repository*.

Qadrini, L., Seppewali, A., & Aina, A. (2021). Decision Tree dan Adaboost pada Klasifikasi Penerima Program Bantuan Sosial. *Jurnal Inovasi Pendidikan*.

Sianturi, F., Hasugian, P., Simangunsong, A., & Nadeak, B. (2020). *Data Mining Teori dan Aplikasi Weka*. Indonesia: IOCS Publisher.

Siregar, A. M., & Puspabhuana, A. (n.d.). *Data Mining Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group.

Suryawinata, Z., & Hariyanto, S. (2016). *Translation Bahasa Teori & Penuntun Praktis Menerjemahkan*. Malang: Media Nusa Creative.

Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.

Tahyudin, I. (2020). *Pengenalan Machine Learning Menggunakan Jupyter Notebook*. Purwokerto: Zahira Media Publisher.

Tempola, F., Muhammad, M., & Khairan, A. (2018). Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation. *Jurnal Teknologi Informasi dan Ilmu Komputer*.

Vitaloka, L. (2019). Pengaruh Kualitas Pelayanan dan Literasi Keuangan Syariah Terhadap Minat Nasabah Menggunakan Produk BMT. *UMY Repository*.

Wanto, A., Siregar, M. N., Windarto, A. P., Hartama, D., Ginantra, N. L., Napitupulu, D., . . . Prianto, C. (2020). *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis.

Waskito, A. B. (2014). Pengaruh Penguasaan Teknologi Informasi dan Komunikasi dan Motivasi Terhadap Kemampuan Penelitian Tindakan Kelas. *UPI Repository*.

Werdiningsih, I., Nuqoba, B., & Muhammadun. (2020). *Data Mining Menggunakan Android, Weka dan SPSS*. Surabaya: Pusat Penerbitan dan Percetakan UNAIR.

Widyawati, D. K. (2012). ANALISIS KINERJA SUPPORT VECTOR MACHINE. *Repository Institut Pertanian Bogor*. Retrieved from Docplayer.

Zamzami, F. N., & Adiwijaya, M. D. (2021). Analisis Sentimen Terhadap Review Film Menggunakan Metode Modified Balanced Random Forest dan Mutual Information. *JURNAL MEDIA INFORMATIKA BUDIDARMA*.

LAMPIRAN

Contoh Labeling Manual pada Skenario 1.

English	Sentiment	Star	Manual
Quite helpful in the financial process, I say, good job (y)	Negative	4	Positive
Why doesn't the help function work? how to add menu how? I'm using asus z4c. thx :)	Negative	2	Negative
Very helpful, good luck to the founders.	Negative	5	Positive
Very cool!!!	Negative	5	Positive
In summary, there are numbers other than zero. Even though there is no tax. But the overall number is correct. Over all good	Negative	4	Positive
Cool!!!	Negative	5	Positive
Cool. Very helpful. It still needs to be developed again. Especially the added balance reduction menu and also services. Because of the difficulty and time efficient if you have to go to the website. Make it all in the app to make it easy!	Negative	4	Positive
Very helpful to record sales..but will it be paid in the future? And Batu alone says may see sales history, why is it lamentably close? Please guide.. Thank you very much	Negative	5	Positive
Update review... After finding some bugs as well and consulting the CS section and immediately responding quickly to bug fixes directly with the owner... Continued success for Qasir to develop applications that are always useful.	Negative	5	Positive
It's good, it's just not complete, because you can't directly share it with other applications, for example, you want to share it to WA, FB, etc.	Negative	4	Positive
Admin..please make a cashier category for laundry..because in my opinion the current application is not yet compatible for the laundry business menu	Negative	5	Positive
If you use 2 different cellphones with the same staff, sometimes it's history transactions are not immediately updated. But overall OK	Negative	4	Positive
After using it for 5 days, I just found out the minus of this application, the amount of stock goods that have been sold are sometimes debited directly from the total stock, and sometimes not, so the real stock is 5 but at the qasir it's still 7, that means not terminuskan, then how come the selling price can be zero, even though it's already done input the price, and after saving it goes back to zero, I'm confused with this application, please fix it, I was expecting more same application, good user interface, good menu, but data collection the number of real stock data and sold is different, please fix it immediately, so that maximally help many MSMEs	Negative	4	Positive
The point is that you are satisfied with using this app, simple, easy to use, not complicated. Still waiting some updates â€¢ credit payment feature for customers who can't pay directly â€¢ copy-paste the details of the receipt so that it can be sent via social media	Negative	4	Positive
full of features, cool	Negative	5	Positive
continued success for the qasir application	Negative	5	Positive

Very good POS application, simple but complete. Complete features. Can monitor shop transactions from outside, can be multiple cashiers in one shop, that's what I looked for it and couldn't find it in other apps. Thank you, hopefully more perfected again	Negative	5	Positive
I'm rating it again, I like the features, intuitive ux, for people who are quite old can use it without problems, hopefully the payment method features reproduced and can be modified/styling the receipt. thx.	Negative	5	Positive
Very helpful for all types of businesses. Has quite a lot of features.	Negative	5	Positive
Hello Qasir Team, previously it's been almost 3 months using this app, it's been safe and the system is quite helpful but when updating it turns out a dialog box appears "an error occurred in the cashier system" try again "the outlet you entered is wrong", please help :((	Negative	3	Negative
cool.	Negative	5	Positive
Cool app	Negative	5	Positive
Recommended for more advanced Indonesian SMEs	Negative	5	Positive
why is this application qasir? the statement "the outlet you entered is wrong"...then enter what is correct? If the cellphone/email/subdomain number is correct, everything is correct??	Negative	1	Negative
Keep the server down	Negative	1	Negative
Very helpful in recording my sales transaction data.. Success is always for Qasir and the team... Just enter it, maybe in the future sending the receipt can output text so that it can be sent by WA or other media.. Thank you..	Negative	5	Positive
Very helpful for MSMEs, lightweight with a simple UI, the features are quite complete. After looking for POS apps, Qasir is the most suitable and free. Good job! Thank you Qasir continues to grow.	Negative	5	Positive
It's too risky because pending transactions can be printed and not included in the transaction history list. The risk of fraud. After the pending transaction is printed, it can be deleted by the operator. Must be justified immediately very dangerous in its use. After it's finished, it will be given 5 stars. Thank you	Negative	2	Negative
Cool this	Negative	5	Positive
Helpfull.. unfortunately I can't input non-sales related product operational costs	Negative	5	Positive
So far, very helpful and simple, still waiting, hopefully there will be other features.	Negative	5	Positive
Still often fails to connect to the customer input time server	Negative	3	Negative
My staff encounter a problem - automatic logout at the end of every transaction	Negative	3	Negative
Mantaaapppp, this is the application you are looking for. Look for small traders who have difficulty keeping books, very useful	Negative	5	Positive
Great POS application. I wish Qasir can upgrade more features which is Daily Report. Thanks Qasir. God bless you.	Negative	5	Positive
The application is very helpful for MSMEs in recording transactions and improving patterns and systems for documenting financial transactions. As MSME players, we are helped through this Qasir application, besides being free and not being charged anything for its features. This application was made by the nation's children, so we need to support it so that it can help MSMEs in Indonesia, especially in the category of micro and small MSMEs.	Negative	5	Positive
for my small shop this application is very useful	Negative	5	Positive

international number isn't yet supported	Negative	3	Negative
Complaints: 1. When I want to enter the lowest order, I can't select it because it says "start selling". 2. Not a complaint but a suggestion. summary per day, not just per month. the rest I think is good.	Negative	4	Negative
cool....very helpful	Negative	5	Positive
So far I am satisfied with the application. very helpful for SME activists.	Negative	5	Positive
Very helpful even though there are not many notes. But ok for start up	Negative	5	Positive
More needs to be introduced to the public, the application is very useful for MSMEs	Negative	4	Positive
4 first.... later if there is an open price feature, while waiting ah.....	Negative	4	Positive
very useful and helpful for small business owners.	Negative	5	Positive
very helpful for small business	Negative	5	Positive
very useful application, very cool...	Negative	5	Positive
Wow, this is an application made by the nation's children that helps MSME owners like me, honestly not inferior to other paid cashier systems. thank you qasir for presenting his best work for the progress of umkm. continue to be developed. thank you	Negative	5	Positive
Thank you, very helpful for MSMEs... go ahead with the Qasir team and continue to improve the features in this application, such as: compatibility with several bluetooth printers that are widely circulated in the market, especially the cheap ones to reduce the cost of MSMEs and logos on receipts.	Negative	5	Positive
After updating, I get an error, right?	Negative	3	Negative
Cannot set date on receipt	Negative	2	Negative
The slower, the more complicated, it's not clear	Negative	3	Negative
very good application, complete, suggestions to make it more complete add integration with several other applications (gojek, shopee, tokopedia) and for the debt feature, add auto-chat via whatsapp lsg to gt customers when it's due	Negative	4	Positive
Cool and worth the app that I'm using for my shop right now,, The only thing that bothers me is that the product menu is sorted alphabetically, it can't be customized by the user, I'm a bit confused if there are a lot of menus,, if this feature is available, I definitely won't will try to try another POS application..	Negative	4	Positive
the features are complicated, so it can't be fast when inputting orders and there is no feature to give a "name" for each order unless you have to input customer data one by one using a phone number??? hhh I'm really sorry I bought the yearly one, how can I get a refund???	Negative	1	Negative
Always having problems when I want to print.. even though I have connected the bluetooth.. today I can print... tomorrow I can't anymore.. please fix it again..	Negative	3	Negative
Printer support is very limited, for users of 80x80 paper printers it is very difficult, like the one I currently use with the Blueprint BP-LITE80D printer, sometimes printing doesn't want to print often, the printouts on 80x80 paper are messy, the cash drawer doesn't open automatically, and the paper does not automatically cut. It's too bad for the size of the paid application.	Negative	2	Negative

So far, it's good for beginners and the cheapest. But unfortunately, I already joined the pro before September 22, so I couldn't try the business website. Time to wait 2023.	Negative	5	Positive
Wearing it is not relieved, the notification is a riot, via wa everything. Complicated.	Negative	1	Negative
Why don't wholesale prices appear on the business website? In the product catalog downloaded in excel form, wholesale prices are also not listed. Please update. What's so hard?	Negative	1	Negative
Why are you suddenly unable to do the transaction? Even though I've tried restarting it too	Negative	3	Negative
There are some problems with the new version.. some items can't be entered into the transaction.. has been clicked several times.. For example.. start a new transaction.. click on an item.. can't enter the next menu.. but other items can enter go to next menu	Negative	3	Negative
Is the qasir application in error? I was at the beginning of the opening of the shop in early March, it was safe why on Monday yesterday suddenly when the transaction couldn't click on the product, finally I was forced to have a manual transaction without an application	Negative	1	Negative
Errors !! The application always comes out when you want to update stock items & sometimes searching for product names also looks messy. I've tried synchronizing, uninstalling and reinstalling but still getting an error. Please fix it, because work is getting stuck	Negative	1	Negative
Hope the search, scan and pending order button on the bottom, not top.	Negative	4	Positive
A little input for Dev Qasir. The price cannot be filled with silver, for example, the selling price is 25,000, the base price is 21,551.72, the base price is added to tax and the service charge is 25,000. So what I mean is that the value that cannot be added is 72 silver, if silver cannot be input, there will be no balance. I hope it will be added soon, thank you	Negative	4	Negative
Please feature dark mode so that it's comfortable to see...	Negative	3	Negative
After updating, the error keeps getting worse. Loading not finished. Forced closed . So sorry for updating :(	Negative	2	Negative
It's too PRO application. Want this Pro, want it Pro. Even though the complete package is not very cheap. I am thinking to migrate to Moka or Maboo..	Negative	3	Negative
The features are various, but unfortunately many of these features are made half-heartedly, difficult to implement. The IT team's mindset is "The important thing is that it's fast, even if it doesn't solve the problem"	Negative	3	Negative
My shop made this application for more than 1 year and just extended it for the second year. Very disappointing, frequent errors, stock items often decrease and suddenly disappear. The brand of goods suddenly changed itself. Several times the application error. The call center phone is not picked up, the help menu in the old application is answered. I feel like it's a loss to extend this application and plan to move to another application	Negative	1	Negative
Good application. Sis suggest, give a feature for adding stock. So it can be monitored when the stock is added. So far, when the stock runs out, we'll edit it right away. So I don't know there is a change in the number of stock items. Thank you very much ðŸ™• ðŸ™•	Negative	5	Negative
Pretty good but the calculation of the purchase price is still manual... so some features are a bit less relevant and less useful...	Negative	2	Negative

Cool app	Negative	5	Positive
The mutation and stock don't match, I've checked the manual. Sometimes even different cellphone data takes a very long time to sync. Update: May 17, reports can't be accessed, for me, who checks daily income, is very disturbed by this problem, the inventory problem that is out of sync has not been completed, now reports can't be accessed. No recommendation	Negative	1	Negative
The app is good, it just has input. I want the operator to be unable to edit orders that have been saved The problem is that if you can, the data that comes out can be manipulated Thank you	Negative	4	Positive
Bad first impression. First time register and have to verify no. HP, the OTP code failed even though it was correct. Go to the Attendance Report menu - Click 'Start Subscribing' the problem is, the link is dead.	Negative	3	Negative
price negotiable function cannot be disabled	Negative	5	Positive
There is no trial feature. If you care about the user, please give it! â€¢login qasir, buy bundles, buy other features. Every time there is a change there is a different number contacting via chat from qasir. What are you doing? User data even spreads. Annoying â€¢the discount feature can only get 1 discount in total sales. Can't have multiple discounts or item discounts. Weird bundle feature, when you enter 1 item that has many variants, it becomes a separate item after setting the bundle. Can't choose the variant too. The stock database means messed up	Negative	3	Negative
App ok, just sometimes server error...	Negative	4	Positive
The stock changes frequently, sometimes it doesn't update, sometimes it goes missing. The trouble is when you want to make a sale transaction, you always say the stock is empty to the buyer. It's been 3 years complaining about the same thing. No change	Negative	2	Negative
So far the app is ok. But... The font is too small, the icon is too small, the thumbnail is too small. Great for eye pain. Hopefully the next one will be better in terms of the User Interface	Negative	5	Positive
The stock sucks. Not corrected.	Negative	2	Negative
Drop the star first. Because I like this error. Failed to load store keeps	Negative	1	Negative
Not once or twice for maintenance during peak hours, the application cannot be opened. The system is not reliable. Still have to write receipts on paper again. For hours the maintenance is not going right. Maintenance is scheduled at low hours like dawn, it won't hurt the user	Negative	1	Negative
The application after the update even got an error, couldn't log in to the application, I was opening a shop so it was hampered!	Negative	1	Negative
If you want to log in, you are asked to change the language, but still can't enter, ouch	Negative	1	Negative
Sometimes there is an error when closing the cashier even though you are only left to play other applications, causing delays in attendance at home, please fix the system again, thank you	Negative	3	Negative
The biggest stupidity is that when I first installed this app, I couldn't even enter the app because the language choice was NOTHING and NOT EXIT. The fastest solution I really hope. Edit: it can be used, the review will be updated after using the app.	Negative	3	Negative
There is no option to choose the size of the receipt. So you have to buy another printer	Negative	1	Negative

Many features were locked behind paywall, multiple tracking attempts, notification offer are spammy, even their CS are giving you offer from WhatsApp, was muting one of them and then get contacted by other CS smh. Not again.	Negative	2	Negative
Product entry from backoffice via excel can generate double data. My advice (1) please provide the option to select multiple data entries at once so that they can be modified, edited or deleted at the same time (2) The tile product display option displays the price, on my device it is not visible except the list display (3) The product purchase option is below the number 1 is placed outside, without having to press the edit button. Meanwhile, the experience for using it for a week. The rest are satisfied	Negative	4	Positive
Each several times a disturbance at a critical moment. Data is often error, save orders suddenly appear	Negative	3	Negative
Unfortunately, the operator can sell the product below the capital price, you have to check the operator often	Negative	3	Negative
why is it always after the update that it doesn't immediately sync to another device after updating the stock and having to update it in the settings. after this update it's also slow	Negative	1	Negative
can't open, sip. I've bought a lot of features too.	Negative	1	Negative
Because it's been smooth again, the 5 stars are back again. Btw, if there is a repair plan, do it in the early morning and with notification from a few days in advance. Thanks. Addendum: thanks for fixing the 100% discount issue. Previously, such transactions could not be completed.	Negative	5	Positive
A little disappointed because the latest update can't save additional records.	Negative	4	Positive
I intend to subscribe to Qasir because I want to integrate Grabfood, but after I subscribed I just received information that I have to make categories and products from the Qasir application which Grabfood will follow. And promos and ads will be deleted and we will reset it again. There should be information at the beginning like this, suitable for those who want to open a new store, not one that is already running. Had to uninstall.	Negative	1	Negative
Please search transaction history not only by note number, but also customer name.... For receipts so that it can be more custom...	Negative	1	Negative
These apps really make it easier for me to manage my distributor business. Forward etalastic!	Positive	5	Positive
The new version is better than the one that hasn't been updated, so the discount feature can be synchronized without manual input to make it easier	Positive	5	Positive
The apps are really cool	Positive	5	Positive
Helpful enough	Positive	4	Positive
Very helpful for my financial report,	Positive	5	Positive
The app is good. Only after the update it keeps crashing. I hope the next update will be good.	Positive	4	Positive
If the purchase could choose the debt option would be better. There is only cash option	Positive	4	Positive
Simple. Powerful. Free	Positive	5	Positive
love it	Positive	5	Positive
the most suitable application for me, of the many applications available, only this is the right one	Positive	5	Positive

Good, can be equipped with product photos and displays if the category is the same.	Positive	4	Positive
The application is good, just personalization is a bit difficult.. maybe it can be made easier..	Positive	5	Positive
Nice app... Thanks...	Positive	5	Positive
This application is good, it is online and free, but there are a few suggestions that I would like to add, --- 1. Please give each operator a password setting, so that each person has a different password, so that reports between operators do not have to be tampered with, and the supervisor has the right to reset if for example the operator forgets the password, also the total documents are confidential, ... 2. Please give the feature so that I can export to excel/pdf maybe in 1 transaction, and daily when it closes that day you can also know that it can be done in total with employees.... , and monthly reports (this has been set to start the bookkeeping on what date because every business is different, the closing date of the book) so that you know the profit and loss.... just keep it simple, only those that are sold capital....., sm like other cashier applications, that way the bookkeeping will be more detailed... I'm waiting for the latest update so that this application becomes the top plg	Positive	4	Positive
Great.... Hopefully it will be free always, very helpful for SMEs. My suggestion is to support more cashier printers	Positive	4	Positive
Very good, hopefully in the future there will be additional unit features	Positive	5	Positive
nice	Positive	5	Positive
Qasir is very helpful in recording, hopefully next time you can see the total sales per day. Continued success Qasir to help MSMEs.	Positive	5	Positive
The apps are really easy to use.. I want to try to print the receipt.. Can I use the Eppos mobile printer brand? Please answer yes	Positive	5	Positive
Woooow, amazing. I tried the demo, the results are interesting, neater and more organized.	Positive	5	Positive
Nice app	Positive	5	Positive
Pretty helpful, just confused when you enter the item category, where can you edit the category?	Positive	3	Positive
Cool app, really helpful	Positive	5	Positive
Very helpful for MSMEs .. no more ads really sip.	Positive	5	Positive
The best app for checking stock	Positive	5	Positive
Easy to use	Positive	5	Positive
nice	Positive	5	Positive
An application that is very helpful for those who have a business, so you can supervise too.. it's really good	Positive	5	Positive
OK app have	Positive	5	Positive
Very good application, hopefully next time there will be a menu to export pdf or excel for a list of stock items	Positive	5	Positive
Very much help full... Continues to develop qasir, still waiting for other improvements... Can you help with recording business assets, right?	Positive	4	Positive
Great. Haven't tried the bluetooth printer yet	Positive	5	Positive
So far it is good.	Positive	4	Positive

It's good, but it's still a bit 'slow' to enter orders, that's the amount of money	Positive	5	Positive
This application is simple and very easy to use.. very useful for my online shop	Positive	5	Positive
I think there's an error to input product code. And it will be better with additional features to manage cash flow with cash drawer and also to manage customers.	Positive	5	Positive
Very good, I want to ask, how to connect to a good printer, what kind of printer is it, what brand is it?	Positive	5	Positive
Nice and simple.. Can you add the customer's name.. So you don't get confused when recaping Thanks	Positive	5	Positive
super helpful.. simple and complete..	Positive	5	Positive
There is still something wrong with the cashier's calculation during the discount application section. Apart from that, so far so good.	Positive	4	Positive
Of the many applications, in my opinion this is the best, the latest version the features are getting really good... awesome!!!!	Positive	5	Positive
Actually, it's really good, but for an online shop that has stores in several marketplaces, it's a bit difficult to record the transaction history because it can't include the right information in the sales section. If possible, add an option for information in the sales section, it's very helpful	Positive	4	Positive
So far so good, if possible Sis, please add a barcode scanner using a camera that fits right into the product input to make it easier without any additional tools!	Positive	4	Positive
The better! Hook!	Positive	5	Positive
The features are like mokapos which is paid but free, good	Positive	5	Positive
good...	Positive	5	Positive
From the free cashier application. This application is best. Please add header and footer yak mimin cute hihihi...	Positive	5	Positive
The app is really good	Positive	5	Positive
Very cool, easy to use	Positive	5	Positive
Brilliant and easy to use!	Positive	5	Positive
This is a fairly good PoS app. The basic features are available for free and the additional features can be purchased separately for cheap prices.	Positive	5	Positive
Ok help	Positive	5	Positive
delete honest post & review from your customer? please consider using this apps	Positive	1	Negative
Edited: after 2 days of bugs, I can finally use the website features, even though I can't connect to Instagram (add shop), it's okay. The export stock feature from excel to the system also makes it easier, maybe add a column for the picture to make it easier for customers and sellers, so you don't have to re-enter photos one by one. Old comment: I feel like I lost paying 438 thousand to get the website, but it just keeps crushing, it took a long time to fix it.	Positive	5	Positive
Great, easy to use	Positive	4	Positive
God willing, a blessing for those who made it.. Helping us is a small umkm.. :)	Positive	5	Positive
If indeed this apk can be free, your services are eternal, bro...	Positive	5	Positive

Using the free one, there are still many limitations. If you can, the free ones will also have more features. Let more people be helped by this application. Just a hope.	Positive	5	Positive
It's so easy. Thanks Qasir	Positive	5	Positive
The application is very good, user friendly, easy to understand and use, and has very complete features. But unfortunately on my device why can't landscape display, even though if you can it's already the best, anyway, please make the developer more compatible for other devices so that it can display landscape. Thank you	Positive	5	Positive
I've updated, how come I still can't process the transaction history, right? At 6 o'clock I want to finish work, I need to do the bookkeeping	Positive	1	Negative
So far it really helped me in terms of stock items.	Positive	5	Positive
Good help...	Positive	5	Positive
Good application... I've been using it for a long time.. But when I reported a bug in the application after updating, the Qasir team didn't even understand..	Positive	2	Negative
Best app pos until now	Positive	5	Positive
The app is good, sadly it's only available on Android. When is it on iOS?	Positive	3	Positive
Nice application, easy to use, and cheaper...	Positive	5	Positive
easy application and very helpful	Positive	5	Positive
It was a really good app and simple to use. I subscribed for pro version and get the grab merchant integration but as well as you guys need to know, the grab option group feature was not available in this version. So be aware if your store really need those options.	Positive	3	Positive
System-wise, the apps are pretty easy to use. But for additional features (syncing to grabfood) I don't really recommend. Should have worked together with grab data can be synchronized with the system. But this is manual input with a menu of hundreds. Plus the data was lost and input again manually. I've paid and followed all the steps but until now I can't do it for almost 1 year. I have complained many times, just told me to take screenshots without any improvement. I've asked for a refund and it's not served.	Positive	3	Positive
Similar to EA Games, yes, a lot of paid DLC. I've bought a 6 million qasir device, but I still have to buy an additional set, if you want to print a receipt, you have to pay 18k first, where have you subscribed for another year to the premium one. Sorry, it's better to just book a stall. The practice of making money is annoying, corporate capitalist.	Positive	1	Negative
It's been getting more and more limited for its free features. Even for recording transactions, the most basic feature has been removed from the free account. Just as soon as you pay, everything doesn't have to be free. Local developers are indeed more capitalist than global developers.	Positive	1	Negative
Ok, I can do it.. please maintain it to make it run more smoothly, thank you	Positive	4	Positive
3 stars first. There are still many shortcomings. The selling price and capital of each outlet cannot be different. Logos can't be different either. On Android there is no inventory menu like on the website. Capital is not automatically replaced by the new price at purchase. Hopefully in the future updates will add these features. Thank you	Positive	3	Positive

As a Qasir Pro user, I was disappointed again. I thought paying Qasir Pro already got all the features, apparently not. There is a new feature that usually appears in POS restaurants, namely table settings. Apparently asked to pay again Rp. 15 thousand per month to use this feature. For those who want to subscribe to Qasir Pro, it might be considered carefully.	Positive	1	Negative
Had been trying here and there, finally manage to found an application wherein can add variants onto my items! Thank you very much.	Positive	5	Positive
Received the verification sms. Waited for hours. Repeatedly resend, after hours there was a barrage of sms. Why didn't you send the email verification, it's rubbish	Positive	1	Negative
Honestly, why did you delete the review, please fix the performance, please. I don't have clear information, it's complicated, I don't accept input. Better choose another application	Positive	1	Negative
Very helpful for my business. Continued success qasir	Positive	5	Positive
The app is good, easy to use and quite detailed. Suggestions, please make it integrated with go food, shopee food, traveloka eats and maxim food too	Positive	5	Positive
Hopefully soon there will be a shipping label feature that can only be printed to make it more effective	Positive	4	Positive
The application is very good, simple and has many features that can make business run smoother. But can you add "nite mode"	Positive	5	Positive
Good, how do you enlarge the product image?	Positive	5	Positive
Thanks for the prompt response Qasir team!	Positive	5	Positive
Connect to the printer can only be bluetooth, even though in the menu there is via wifi and lan, but can't be used,	Positive	1	Negative
For all those who don't know, this is the most correct application for business people. MSMEs, Entrepreneurs, the main businesswoman. Everything has complete features, we happen to have been using it for almost 2 years. Great help, please appreciate all the team. The struggle is extraordinary.. Hopefully all the Qasir team will be more successful, the more meritorious for the country that has helped us!! Qasir is getting more and more appraised, never cynical and makes traders don't have to worry	Positive	5	Positive
It would be nice if added features: 1. A special button to open the cash drawer like the next app without the need to print a receipt 2. The Bluetooth keyboard shortcut feature 3. The button to add the number of product items should be placed directly on the transaction page so that it is faster to add products	Positive	4	Positive
very good and easy application for use	Positive	5	Positive
good help	Positive	5	Positive
Very helpful for us small traders because it is free, the features are also quite complete. Nice!	Positive	5	Positive
Stocks that come out sometimes don't automatically count out	Positive	2	Negative
Very helpful for my business, very easy to use	Positive	5	Positive
maintenance without advance notice is very annoying !!	Positive	1	Negative
Maintenance at operational time very often	Positive	1	Negative
Ugly times are often a problem. Want to recap data even maintenance. Opening sync takes too long	Positive	2	Negative

The CSM is so annoying, they contact me personally by whatsapp just to announce a promo that they have. Why don't you just do it in the app? Its been a couple month and I'm going to uninstall, such a waste actually. I paid some features here.	Positive	2	Negative
The features are ok, it's just that the payment qris is too full in size, so the qris reaches the panel causing it to be difficult to scan, if the size is reduced a little the qris so that there is a limit around the qris so that it is easy to scan (using advan tab 8 elite, recommendation from qasir) . Hopefully it can be implemented. Thank you	Positive	5	Positive
I requested an activation code, many times over the past 3 days, no sms came in. Try sending the activation code via WA, it's easier.	Positive	1	Negative
As soon as possible so that transactions can be made via a computer, so that it will be easier, thank you	Positive	5	Positive
I think the more you update it, the worse the Qasir apps are... It's good at first, now most of the updates are even more confusing... Those who think this is me or someone else, right?	Positive	3	Positive