

# Analisis Sentimen Review Aplikasi PeduliLindungi Menggunakan Seleksi Fitur Information Gain Berbasis SVM

Dzulchan Abror

Fakultas Ilmu Komputer, Magister Teknik Informatika, Universitas Dian Nuswantoro Semarang  
Email: dzulchanabrор@gmail.com

## Abstrak

Aplikasi PeduliLindungi merupakan salah satu langkah pemerintah Indonesia dalam menekan penyebaran virus Covid 19. Aplikasi ini terdapat pada google play store yang didalamnya disediakan fasilitas rating dan ulasan bagi pengguna aplikasi. Fasilitas ini dapat kita manfaatkan untuk menggali beberapa informasi salah satunya dengan mengelompokkan ulasan menurut polaritas yang ada yaitu positif atau negative. Sehingga dapat diperoleh informasi pandangan masyarakat terhadap sebuah produk atau aplikasi tersebut, bidang penelitian ini dinamakan analisis sentimen. Dalam membangun sebuah model analisis sentimen terdapat suatu permasalahan dimensi fitur yang sangat tinggi, karena didalam teks mining memiliki ribuan fitur. Fitur tersebut kebanyakan irelevan dan tidak berpengaruh terhadap kelas, akibatnya tingkat akurasi model menurun. Solusinya diperlukan suatu metode yang dapat menyeleksi fitur-fitur teks tersebut. Pada penelitian kali ini diajukan metode sentimen analisis terhadap review aplikasi PeduliLindungi dengan metode SVM (Support Vector Machine) berbasis seleksi fitur information gain. Dataset diambil dari ulasan pengguna berjumlah 300 ulasan dengan metode penelitian membandingkan 2 model, yang pertama SVM tanpa seleksi fitur dengan hasil akurasi sebesar 74,33% dan model kedua SVM menggunakan seleksi fitur information gain dengan hasil akurasi sebesar 84%. Hal ini membuktikan bahwa dengan metode seleksi fitur dapat meningkatkan akurasi.

**Kata kunci :** *Pedulilindungi, Analisis Sentimen, Seleksi Fitur, Information Gain, SVM*

## Abstract

*The PeduliLindungi application is one of the steps taken by the Indonesian government to suppress the spread of the Covid 19 virus. This application is available on the Google Play Store in which rating and review facilities are provided for application users. We can use this facility to dig up some information, one of which is by grouping reviews according to the existing polarity, namely positive or negative. So that information can be obtained from the public's view of a product or application, this research field is called sentiment analysis. In building a sentiment analysis model there is a very high feature dimension problem, because in text mining it has thousands of features. Most of these features are irrelevant and have no effect on class, as a result the accuracy of the model decreases. The solution requires a method that can select the features of the text. This research proposes a sentiment analysis method for reviewing the PeduliLindungi application using the SVM (Support Vector Machine) method based on information gain feature selection. The dataset was taken from user reviews totaling 300 reviews with research methods comparing 2 models, the first SVM without feature selection with an accuracy of 74.33% and the second SVM model using information gain feature selection with an accuracy of 84%. This proves that the feature selection method can improve accuracy.*

**Keywords :** *Pedulilindungi, Sentiment Analysis, Feature Selection, Information Gain, SVM*

## 1. PENDAHULUAN

Pedulilindungi saat ini sangat penting serta bermanfaat bagi masyarakat dan pemerintah, karena setidaknya aplikasi ini memiliki 5 (lima) fitur yang berkaitan dengan pengendalian Covid 19, yaitu : memberi peringatan kepada pengguna tentang zona merah Covid 19, pengawasan pemerintah terhadap pergerakan

orang yang terkena Covid 19, mengunduh sertifikat vaksin, informasi hasil tes Covid 19 dan sebagai bukti untuk mengakses layanan publik.(KOMINFO/RDJS/YOY, 2021). PeduliLindungi telah tersedia di google play store, dimana pengguna tinggal mengunduh serta memasangnya kedalam smartphone. Pada Google play store disediakan fasilitas rating dan ulasan pengguna aplikasi, dimana pengguna dapat memberikan ulasannya secara bebas baik bernada positif, netral maupun negative. Dari fasilitas rating dan ulasan ini dapat kita analisis untuk menggali beberapa informasi yang terkandung didalamnya, salah satunya dengan cara mengelompokkan ulasan menurut polaritas yang ada yaitu positif, netral atau negative. Dari pengelompokkan ulasan tersebut dapat kita peroleh informasi fakta dan opini masyarakat terhadap sebuah produk, merek, layanan, politik, atau topik lainnya. Bidang penelitian ini dinamakan analisis sentimen dan termasuk kedalam bidang NLP (Natural Language Processing). Analisis sentimen dapat membantu sebuah perusahaan atau organisasi untuk mengetahui bagaimana orang melihat produk dan layanan mereka serta perubahan apa yang sekiranya diperlukan pada produk dan layanannya untuk meningkatkan produk dan layanannya tersebut demi kepuasan konsumen, yang imbasnya dapat meningkatkan profit. (Paliwal & Sunil Kumar Khatri, 2019). Untuk membangun model tersebut diperlukan sebuah algoritma klasifikasi, beberapa algoritma tersebut seperti naïve bayes, regresi linier, maksimum entropi, support vector machine (SVM) dan neural network (NN) telah banyak di usulkan oleh para peneliti. Abinash Tripathy, Ankit Agrawal dan Santanu Kumar Rath melakukan penelitian terhadap ulasan film yang terdiri dari 1000 ulasan negative dan 1000 ulasan positif, hasilnya kemudian dibandingkan antara metode Naïve Bayes dan SVM dimana diperoleh hasil tingkat akurasi 89,54% untuk metode Naïve Bayes dan 94,06 % untuk metode SVM. (Tripathy et al., 2015). Sneh Paliwal, Sunil Kumar Khatri, dan Mayank Sharma menganalisis sentiment terhadap produk “iPhoneX” dengan mengambil data set dari Twitter, Facebook dan Website menggunakan metode ANN menghasilkan akurasi sebesar 78%. (Paliwal & Sunil Kumar Khatri, 2019)

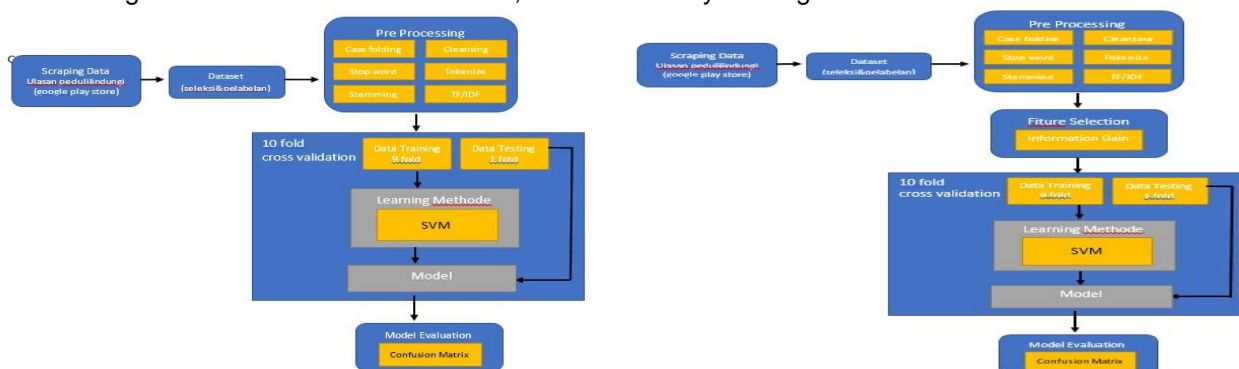
Ahmad Efriza Irsad, Yuita Arum Sari dan M. Alu Fauzi menganalisa klasifikasi teks twitter dengan metode seleksi fitur information gain menggunakan algoritma Naïve Bayes berdasarkan tempat tinggal pada kota Malang, dengan hasil akurasi sebesar 71,66%. (Irsad et al., 2019)

Dalam membangun sebuah model klasifikasi teks, terdapat suatu permasalahan yaitu dimensi fitur yang sangat tinggi, karena didalam teks mining memiliki ribuan fitur dimana terdapat banyak fitur yang irelevan dan tidak berpengaruh terhadap kelas, yang dapat mengakibatkan menurunnya tingkat akurasi sebuah model. (Chen et al., 2009) Dalam hal data berdimensi tinggi, dimana mengandung atribut yang sangat besar dapat mempengaruhi biaya dan waktu komputasi serta dapat menurunkan tingkat akurasi pembelajaran. Untuk itulah diperlukan suatu metode yang dapat menyeleksi fitur-fitur dari teks tersebut, dimana metode ini diharapkan dapat mengurangi jumlah fitur dengan cara menyeleksi fitur yang memang berpengaruh terhadap kelas dan membuang fitur-fitur yang tidak berpengaruh terhadap kelas.

Beberapa algoritma seleksi fitur antara lain : Information Gain, Particle Swarm Optimazation (PSO), Genetic Algorithm, Chi-square, dll. Dari beberapa algoritma tersebut, Information Gain adalah metode yang paling sederhana dalam seleksi fitur dan banyak digunakan dalam pengkategorian teks, analisis data microarray dan analisis data gambar.(Chormunge & Jena, 2016). Untuk itu pada penelitian kali ini diajukan metode analisis sentimen terhadap ulasan aplikasi PeduliLindungi dengan menggunakan metode Support Vector Machine (SVM) berbasis seleksi fitur information gain agar dapat meningkatkan nilai akurasi.

**2. METODE PENELITIAN**

Pada penelitian ini akan dilakukan eksperimen terhadap 2 (dua) model analisis sentimen, model pertama menggunakan metode SVM tanpa dilakukan seleksi fitur dan model kedua menggunakan metode SVM dengan seleksi fitur Information Gain, desain modelnya sebagai berikut :



Gambar 1. Model 1: SVM tanpa seleksi fitur & Model 2 : SVM dengan Information Gain

## A. Sumber Data

Dataset diperoleh dari ulasan pengguna aplikasi PeduliLindungi pada Google Play Store, dengan cara melakukan proses Scraping data menggunakan aplikasi google\_play\_scraper pada tanggal 19 Maret 2022 dan diperoleh hasil scraping 2.000 ulasan pengguna.

## B. Pengolahan Data

Data hasil scraping kemudian disimpan kedalam file .csv selanjutnya dilakukan proses pengambilan baris data yang relevan dengan penelitian analisis sentimen. Yaitu kolom 'content' dan 'score' saja yang akan di ambil, kolom lainnya dihapus.

Setelah mendapatkan data 'content' dan 'score', dilakukan proses pelabelan dimana content dengan nilai score 1 – 3 di beri label "negative", content dengan nilai score 4 – 5 di beri label "positif". Setelah itu data diseleksi sehingga diperoleh dataset sebanyak 300 data, dimana terdiri dari 150 sentimen negative dan 150 sentimen positif. Untuk memastikan ke validan dalam proses pelabelan, 300 dataset tersebut juga divalidasi oleh seorang ahli Bahasa Indonesia.

## C. Preprocessing Data

Untuk mengolah data teks kedalam model machine learning, tidak bisa data teks hasil scraping atau download langsung kita masukkan ke dalam model. Teks dokumen harus kita proses terlebih dahulu melalui proses pembersihan data dan transformasi data, untuk merubah data tidak terstruktur dirubah ke data yang terstruktur.(Isnain et al., 2021).

Dari 300 dataset yang telah diberi label, dilakukan preprocessing data agar data siap untuk di seleksi fitur dan dimasukkan ke dalam model klasifikasi. Tahapan-tahapan dari preprocessing data :

- Case Folding : yaitu proses merubah semua teks kedalam huruf kecil
- Cleansing : yaitu proses menghilangkan tanda baca, angka, @, url, dll dari sebuah dokumen
- Stopword : menghilangkan kata-kata yang tidak mempunyai arti, seperti kata sambung, aku, dan, dll
- Tokenization : memotong kalimat menjadi sebuah kata-kata
- Steaming : merubah kata menjadi kata dasar, menghilangkan awalan dan akhiran.
- Feature extraction : menghitung jumlah kemunculan term atau kata, inverse dokumen frequency (idf) dan term weighting, serta mengubah teks menjadi vector.

## D. Seleksi Fitur

Untuk melakukan seleksi fitur digunakan Information Gain, Merupakan salah satu teknik yang sederhana di dalam seleksi fitur, dimana pada metode information gain pada nominal menggunakan menggunakan teknik scoring serta teknik pembobotan atribut kontinu yang didiskretkan digunakan nilai dari maksimal entropi. Dalam mengkodekan suatu kelas entropi dipakai untuk menggambarkan sebuah nilai dari *information gain* serta menggambarkan seberapa banyak informasi yang dibutuhkan.

Pengukuran term atau kata pada information gain menggunakan cara dihitung bit informasi yang diperoleh dari kategori prediksi dimana ada atau tidak adanya kata dalam suatu dokumen.(Maulida et al., 2016)

Umumnya klasifikasi teks pada machine learning menggunakan banyak fitur, dimana digambarkan setiap record atau input sebagai vector  $x \in R^{1 \times F}$ , dimana F merupakan dimensi vector (banyaknya input), dan biasanya F bernilai besar, artinya model menjadi sangat kompleks, untuk mengurangi kompleksitas digunakanlah metode seleksi fitur. (Putra, 2019)

Seleksi fitur merupakan teknik yang dapat digunakan untuk mereduksi subset fitur yang tidak atau kecil pengaruhnya pada suatu kelas, dimana cara kerjanya dengan mengurangi fitur yang tidak relevan dan berlebihan yang dapat menyebabkan efek langsung bagi aplikasi. Hasil dari seleksi fitur dapat mempercepat pada proses dan dapat meningkatkan performa akurasi. Tugas utama didalam seleksi fitur adalah penentuan fitur mana yang akan dipilih dan dipakai untuk peramalan atribut-atribut fitur. (Djatna & Morimoto, 2008)

Tahapan ini dilakukan untuk model 2 yaitu model dengan metode SVM dengan seleksi fitur information gain. Dari data hasil preprocessing diatas, selanjutnya dilakukan seleksi fitur menggunakan algoritma Information Gain, yang tujuannya adalah untuk memperkecil dimensi dari fitur dataset, serta agar dihasilkan fitur yang memiliki tingkat relevan yang tinggi terhadap kelas dan diharapkan dapat meningkatkan performa akurasi dari model yang dibangun.

## E. Penentuan Data Training dan Data Test

Pada penelitian ini menggunakan 10 fold cross validation. Validasi silang adalah suatu teknik evaluasi kinerja suatu algoritma maupun model dengan menggunakan cara dataset yang ada dipecah dua menjadi data training serta data testing. Kemudian model akan dilatih oleh data latih dan diuji menggunakan data uji. Dataset yang dimiliki menjadi dasar dalam pemilihan jenis validasi silang. (Wibowo, 2017)

Salah satu alasan utama menggunakan cross validation dibanding dengan menggunakan validasi

konvensional atau split data adalah dataset yang ada tidak cukup untuk dibagi menjadi data training dan data test tanpa kehilangan pemodelan maupun performa pengujian yang signifikan. Dalam penelitian ini, metode yang adil dalam memprediksi model prediksi dengan pas adalah dengan memakai teknik cross validation yang merupakan teknik universal yang kokoh. (Wikipedia, 2020)

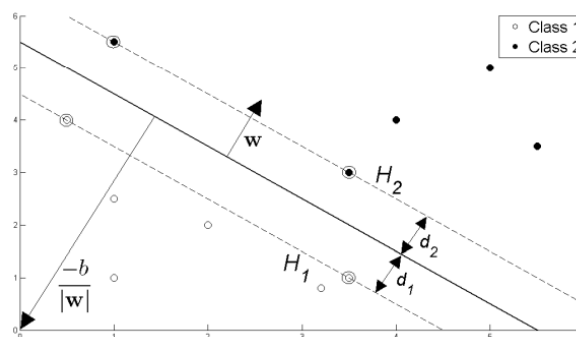
**F. Support Vector Machine**

Salah satu algoritma *classsifier* yang akhir-akhir ini banyak digunakan untuk melakukan berbagai keperluan pada model klasifikasi adalah Support Vector Machine (SVM), SVM dapat juga digunakan untuk regresi. Cara kerja SVM adalah membagi data menjadi dua class dengan sebuah *hyperplane*. *Hyperplane* tepat berada di tengah-tengah kedua class tersebut dengan jarak *d* ke titik data terdekat pada masing-masing class. Dimana *d* biasa disebut *margin*, dan titik-titik data yang tepat berada pada jarak *d* dari *hyperplane* adalah support vector. (Purnamawan, 2015)

Persamaan Hyperplane SVM dinyatakan dengan rumus sebagai berikut :

$$w \cdot x + b = 0$$

Dimana *w* merupakan normal dari *hyperplane*, dan  $\frac{b}{||w||}$  merupakan jarak hyperplane ke titik origin.



Gambar 2. Hyperplane membagi dua buah kelas

Dimana titik-titik data yang masuk ke dalam class 1 adalah titik-titik data yang memenuhi persamaan, sebagai berikut :

$$w \cdot x + b \leq -1$$

dan titik-titik data yang masuk ke dalam class 2 merupakan titik-titik data yang memenuhi persamaan, sebagai berikut :

$$w \cdot x + b \geq 1$$

Pada SVM, kernel memiliki properti yang menarik dimana disebut sebagai *kernel trik*, Kernel yang umumnya digunakan adalah *dot*, *radial*, *polynomial*, *neural*, *anova*, *epachnenikov*, *kombinasi gaussian*, *fungsi multikuadrik*, dapat di tuliskan persamaannya : (Kurniawati & Pardede, 2018)

$$k(x.y) = x.y$$

**G. Evaluasi Model**

Untuk mengevaluasi kinerja kedua model pada penelitian ini digunakan metode evaluasi confusion matrix. Dengan cara membandingkan hasil klasifikasi yang dilakukan oleh model 1 (satu) dan model 2 (dua) dengan hasil klasifikasi sebenarnya, yang selanjutnya akan diambil nilai dari akurasi, presisi dan recall.

Tabel 1. Confusion Matrix

Predicted Values	Actual Values	
	1 (Positive)	0 (Negative)
1 (positive)	TP (true positive)	FP (false positive)
0 (negative)	FN (false negative)	TN (true negative)

Dimana :

- a. TP : jumlah prediksi benar dari data positif
- b. TN : jumlah prediksi benar dari data negative
- c. FP : jumlah prediksi salah dari data negative (data negative diprediksi positif)
- d. FN : jumlah prediksi salah dari data positif (data positif diprediksi negative)

Berdasarkan keterangan tabel diatas, maka performace metrics untuk mengukur kinerja model dapat dihitung. Performa pengukuran yang banyak dipakai, adalah :

1. Akurasi : untuk menggambarkan seberapa akurat model jumlah predisksi yang benar. Akurasi adalah kecocokan data yang diprediksi dengan data sebenarnya. Persamaannya :

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

2. Presisi (Precision) : adalah penggambaran dari nilai prediksi positif atau tingkatan akurasi dari data hasil predisksi dengan data yang diminta pada sebuah model. Persamaannya di tuliskan sebagai berikut :

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

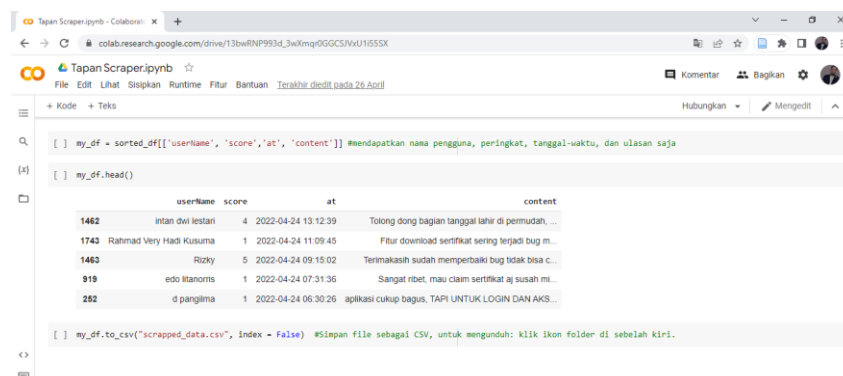
3. Recall atau Sensitivity : gambaran sensitivitas keberhasilan sebuah model didalam penemuan kembali informasi. Persamaannya adalah sebagai berikut : (Nugroho, 2019)

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

### 3. HASIL DAN PEMBAHASAN

#### A. Sumber dan Pengolahan Data

Data ulasan pengguna aplikasi PeduliLindungi pada google play store diperoleh dengan menggunakan aplikasi google\_play\_scapper yang diambil pada tanggal 19 Maret 2022 dan diperoleh sejumlah 2000 ulasan. Proses pengambilan menggunakan google colab, dimana sebelumnya google\_play\_scapper di install terlebih dahulu ke dalam notebook kita. Kemudian data hasil scraper diseleksi hanya kolom user, score, waktu dan ulasan saja yang di ambil. Selanjutnya di simpan kedalam bentuk csv.



Gambar 3. Data hasil scraping

Selanjutnya dilakukan proses pengambilan kolom data yang relevan dengan penelitian analisis sentiment. Yaitu kolom 'content' dan 'score' saja yang akan di ambil, kolom lainnya dihapus.

Setelah mendapatkan data yang berisi kolom 'content' dan 'score', dilakukan proses pelabelan dimana content dengan nilai score 1 – 3 di beri label "negative", content dengan nilai score 4 – 5 di beri label "positif". Setelah itu data diseleksi sehingga diperoleh dataset sebanyak 300 data, dimana terdiri dari 150 sentimen negative dan 150 sentimen positif. Untuk memastikan ke validan dalam proses pelabelan, 300 dataset tersebut juga divalidasi oleh seorang ahli Bahasa Indonesia.

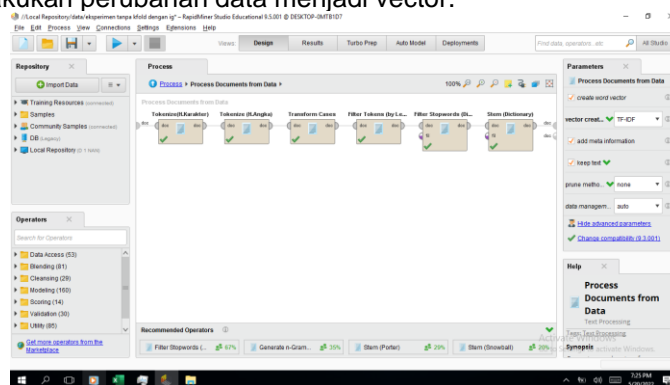
Tabel 2. Dataset sudah di seleksi baris dan diberi label

Sentimen	Ulasan
positif	Keren,,makin cepat ubah data sertifikat vaksin dengan chatbot ???• ???• ???• ???• teruskan berbuat cepat tanggap dalam segala hal jangan hanya untuk vaksin,, terimakasih
negatif	Ini aplikasi gunanya untuk apa, apakah dapat menghidupi keluarga, apakah dapat memberikan ekonomi lebih baik ??? Yg ada kok hidup tambah ribet cari rezeki tambah susah.

## B. Preprocessing Data

Data set yang telah terlabel dan tervalidasi, bentuk datanya masih belum terstruktur dimana didalamnya masih terdapat tanda baca, angka dan kata-kata yang tidak memiliki arti, agar menjadi data terstruktur dilakukan langkah preprocessing data dimana pada proses ini dilakukan pembersihan data dan transformasi data. Pada penelitian ini tahapan praprosesing data di lakukan menggunakan software Rapid Miner, data set berbentuk csv dilakukan proses pembacaan pada Rapid Miner selanjutnya dilakukan :

- Memotong kalimat menjadi kata-kata dan menghilangkan tanda baca/ punctuation character, menggunakan operator Tokenize dengan parameter mode : regular expression
- Memotong kalimat menjadi kata-kata dan menghilangkan angka, menggunakan operator Tokenize dengan parameter mode : non letters
- Merubah semua teks menjadi huruf kecil, menggunakan operator Transform Cases, dengan parameter transform to : lower case
- Menghapus kata yang kurang dari 4 huruf dan yang lebih dari 25 huruf, menggunakan operator Filter Tokens (by Length)
- Menghilangkan kata-kata yang tidak memiliki arti, seperti kata sambung dan lain-lain, menggunakan operator Filter Stopword (Dictionary) dengan cara membuat daftar kata-kata stopwords dalam Bahasa Indonesia yang dimasukkan ke dalam operator Filter Stopword (Dictionary).
- Merubah serta menghilangkan awalan dan akhiran kata, menggunakan operator Stem (Dictionary)
- Semua proses diatas dijadikan satu menjadi sub proses ke dalam operator Process Document from Data, selanjutnya dilakukan perubahan data menjadi vector.

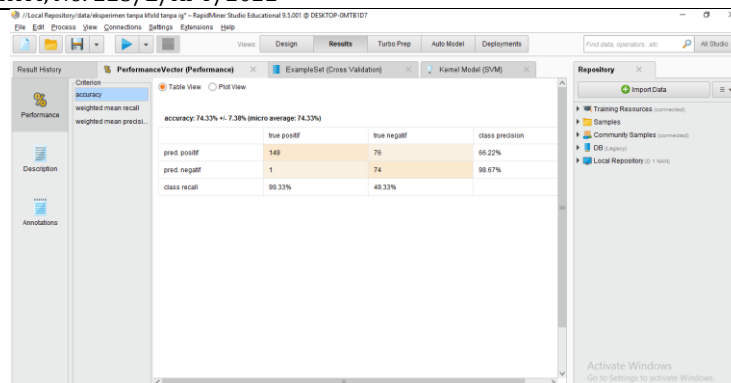


Gambar 4. Preprocessing data

Sampai pada tahap ini diperoleh data set dengan jumlah fitur sebesar 1.841 fitur dan data set telah siap untuk dimasukkan ke dalam 2 (dua) model analisis sentimen yang telah direncanakan.

## C. Eksperimen Model pertama

Data set hasil preprocessing yang berjumlah 1.841 fitur dimasukkan kedalam model 1, yaitu model yang tidak menggunakan seleksi fitur, dan dihasilkan performa dengan nilai akurasi sebesar 74,33 %, precision : 82,67 % dan recall : 74,33 %.

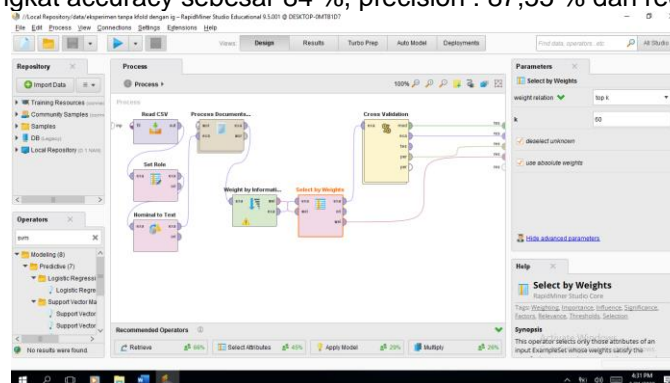


Gambar 5. Hasil performa model pertama

#### D. Eksperimen Model Kedua

Pada model kedua, setelah tahapan preprosesing dilakukan dilanjutkan ke tahapan seleksi fitur menggunakan Information Gain. Information Gain akan melakukan proses pembobotan dari setiap fitur yang dihasilkan pada tahap preprocessing data, dimana jumlah fitur nya sebesar 1.481 fitur. Setelah dilakukan pembobotan, fitur akan diurutkan berdasarkan bobot dari nilai terbesar ke nilai yang terkecil.

Selanjutnya dilakukan proses pemilihan fitur dengan nilai bobot terbesar sehingga dihasilkan 60 fitur. Pada Rapid Miner proses seleksi fitur dan pembobotan information gain menggunakan operator Weight by Information Gain, dengan setting parameters : Sort Weights dan Sort Direction menggunakan Descending. Untuk memilih 60 fitur teratas berdasarkan bobot digunakan operator Select by Weights, dengan setting paramaters weight relation : top k dan nilai k : 60. Fitur ini dimasukkan kedalam model sehingga menghasilkan performa tingkat accuracy sebesar 84 %, precision : 87,55 % dan recall : 84 %.



Gambar 5. Sentimen analisis model kedua

#### 4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan diatas dapat ditarik kesimpulan bahwa pada penelitian sentimen analisis terhadap review aplikasi PeduliLindungi menggunakan seleksi fitur Information Gain berbasis SVM ini terbukti dapat meningkatkan performa model, dimana model tanpa seleksi fitur menghasilkan nilai akurasi 74,33% dan model menggunakan seleksi fitur menghasilkan nilai akurasi 84%, dengan nilai peningkatan sebesar 9,67 %. Selain itu, penerapan seleksi fitur dengan information gain juga dapat meningkatkan nilai precision suatu model, dimana hasilkan nilai presicion model pertama adalah 82,67 % setelah diterapkan seleksi fitur information gain meningkat menjadi 87,55 %.

#### REFERENSI

- Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications*, 36(3 PART 1), 5432–5435. <https://doi.org/10.1016/j.eswa.2008.06.054>
- Chormunge, S., & Jena, S. (2016). Efficient feature subset selection algorithm for high dimensional data. *International Journal of Electrical and Computer Engineering*, 6(4), 1880–1888. <https://doi.org/10.11591/ijece.v6i4.9800>
- Djatna, T., & Morimoto, Y. (2008). Perbandingan Stabilitas Algoritma Seleksi Fitur Menggunakan Transformasi Ranking Normal. *Jurnal Ilmiah Ilmu Komputer*, 6(2), 245006.

- Irsad, A. E., Sari, Y. A., & Fauzi, M. A. (2019). Seleksi Fitur Information Gain untuk Klasifikasi Informasi Tempat Tinggal di Kota Malang Berdasarkan Tweet Menggunakan Metode Naïve Bayes dan Pembobotan TF-IDF-CF. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(5), 4907–4913.
- Isnain, A. R., Sakti, A. I., Alita, D., Marga, N. S., Kunci, K., Analisis, S., & Machine, S. V. (2021). *JAKARTA MENGGUNAKAN ALGORITMA SVM Media sosial menjadikan masyarakat mengalami pergeseran perilaku baik budaya , etika dan norma yang ada , sehingga mereka dapat mengeluarkan opini - opini yang mereka miliki . Opini merupakan suatu pendapat dari pemikiran.* 2(1), 31–37.
- KOMINFO/RDJS/YOY. (2021). *Ini Manfaat Aplikasi PeduliLindungi yang Belum Banyak Diketahui.*
- Kurniawati, I., & Pardede, H. F. (2018). Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of SVM-Based Sentiment Analysis. *2018 International Conference on Information Technology Systems and Innovation, ICITSI 2018 - Proceedings*, 1–5. <https://doi.org/10.1109/ICITSI.2018.8695953>
- Maulida, I., Suyatno, A., Rahmania Hatta, H., & Mulawarman, U. (2016). Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain. *JSM STMIK Mikroskil*, 17(2), 249–258.
- Nugroho, K. S. (2019). *Confusion Matrix untuk Evaluasi Model pada Supervised Learning.*
- Paliwal, S., & , Sunil Kumar Khatri, and M. S. (2019). Sentiment Analysis and Prediction Using Neural Networks. *Springer Nature Singapore Pte Ltd*, 458–470.
- Purnamawan, I. K. (2015). yang mempunyai keunggulan dapat mengolah data berdimensi tinggi, tanpa mengalami penurunan performa yang signifikan. SVM sekarang ini semakin banyak dipergunakan. Pada. *Information and Software Technology*, 12, 173–180.
- Putra, J. W. G. (2019). Pengenalan Konsep Pembelajaran Mesin dan Deep Learning. In *Computational Linguistics and Natural Language Processing Laboratory* (1.4, Vol. 4).
- Tripathy, A., Agrawal, A., & Rath, S. K. (2015). Classification of Sentimental Reviews Using Machine Learning Techniques. *Procedia Computer Science*, 57, 821–829. <https://doi.org/10.1016/j.procs.2015.07.523>
- Wibowo, A. (2017). *10 FOLD-CROSS VALIDATION.*
- Wikipedia. (2020). *Validasi- Silang (statistik).*