

# Penggunaan Seleksi Fitur *Query Expansion Ranking* dan *Genetic Algorithm-Support Vector Machine* untuk Analisis Sentimen pada Aplikasi Perbankan Jenius

Haksar Haksar<sup>1\*</sup>, Siswanto Siswanto<sup>2</sup>, Nirwan Ilyas<sup>3</sup>

<sup>123</sup>Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Hasanuddin, Makassar, 60294, Indonesia

\* Corresponding author, email: [haksarh@gmail.com](mailto:haksarh@gmail.com)

## Abstract

*Jenius is a digital banking product from BTPN launched in 2016. In 2021, various opinions emerged on Twitter regarding cases of lost customer funds, necessitating sentiment analysis to understand public perception. This study used the Support Vector Machine (SVM) method with two feature selection approaches: Query Expansion Ranking (QER) and Genetic Algorithm (GA). The data used were 2,008 manually labeled tweets. The results showed that the QER-SVM method produced an accuracy of 87.81%, while the GA-SVM achieved an accuracy of 88.31% with improvements in precision and F-measure. Thus, the combination of Genetic Algorithm and SVM was more effective in classifying sentiment towards the Jenius application on Twitter.*

**Keywords:** *Genetic Algorithm, Jenius, Query Expansion Ranking, SVM, GA-SVM, QER-SVM.*

## Abstrak

Jenius merupakan produk perbankan digital dari BTPN yang diluncurkan pada tahun 2016. Pada tahun 2021, muncul berbagai opini di Twitter terkait kasus kehilangan dana nasabah, sehingga diperlukan analisis sentimen untuk memahami persepsi publik. Penelitian ini menggunakan metode Support Vector Machine (SVM) dengan dua pendekatan seleksi fitur, yaitu Query Expansion Ranking (QER) dan Genetic Algorithm (GA). Data yang digunakan adalah 2008 tweet yang telah dilabeli secara manual. Hasil menunjukkan bahwa metode QER-SVM menghasilkan akurasi 87,81%, sedangkan GA-SVM mencapai akurasi 88,31% dengan peningkatan pada precision dan F-measure. Dengan demikian, kombinasi Genetic Algorithm dan SVM lebih efektif dalam mengklasifikasikan sentimen terhadap aplikasi Jenius di Twitter.

**Kata kunci:** *Genetic Algorithm, Jenius, Query Expansion Ranking, SVM, GA-SVM, QER-SVM.*

## 1. Pendahuluan

Media internet memberikan dampak terhadap perkembangan sistem informasi di berbagai sektor, termasuk keuangan dan perbankan. Hal ini mendorong bank untuk mengembangkan teknologi yang dapat meningkatkan layanan perbankan yang diberikan kepada nasabah dan peminjam. Perkembangan teknologi informasi dan komunikasi serta internet menyebabkan munculnya aplikasi bisnis online yang salah satunya mendapat perhatian karena meluasnya penggunaannya pada layanan *mobile banking* atau BPR (Bank Perkreditan Rakyat) bank.

Produk perbankan digital bernama Jenius yang dikembangkan oleh bank BTPN. Diluncurkan pada tahun 2016, produk ini merupakan layanan perbankan digital yang

menyediakan opsi perbankan *online* yang cepat dan nyaman, termasuk pembuatan rekening, transfer bank, dan kliring tagihan. Semuanya terjadi di satu tempat, di ponsel Anda, dengan menggunakan sistem operasi Android maupun iOS. Namun, kehilangan uang nasabah di rekeningnya menjadi salah satu hal yang menimbulkan beragam opini tentang Jenius di beberapa *platform* media sosial, terutama *Twitter*. *Twitter* diketahui sebagai *platform* media sosial yang dapat memberikan data opini. Pengumpulan data opini tersebut dilakukan melalui *text mining*.

Teknik *text mining* menggunakan analisis sentimen eksploratif guna mencari tahu opini publik terkait topik tertentu yang diperoleh melalui kegiatan pengumpulan data. Kemajuan teknologi informasi membawa dampak terhadap perubahan interaksi manusia. Algoritma diketahui juga digunakan dalam analisis sentimen. Penggunaan algoritma dimaksudkan untuk memproses serta mengklasifikasikan data yang didapat. Klasifikasi data dalam analisis sentimen dapat dilakukan dengan menggunakan sejumlah metode algoritma seperti: *Naïve Bayes*, *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), *Decision Tree*, dan sebagainya [1].

Metode klasifikasi SVM memungkinkan pembuatan prediksi kelas yang didasarkan kepada model yang didapat melalui hasil proses pelatihan. Metode SVM menghasilkan pola *hyperplane*, yaitu garis pemisah kelas. Berbeda dengan metode klasifikasi lainnya, SVM dapat menerapkan fitur yang tepat untuk melakukan diskriminasi linier pada data non-linier dan berdimensi tinggi [2].

Proses klasifikasi yang terdapat pada penelitian analisis sentimen dilakukan dengan menggolongkan teks menjadi sentimen positif dan juga sentimen negatif melalui perhitungan nilai dari setiap fitur. Penghapusan fitur-fitur yang ambigu dan berlebihan melalui pemilihan fitur dapat mempercepat dan mengoptimalkan proses perhitungan dalam sistem. Selain itu penggunaan seleksi fitur dapat meningkatkan akurasi proses klasifikasi yang dilakukan. Metode pemilihan fitur pada penelitian ini adalah melalui *Query Expansion Ranking* (QER). Pemilihan fitur dicapai melalui metode pemeringkatan ekstensi kueri, yang mempertimbangkan bobot atau skor fitur. Pada penelitian ini, penulis memadukan aplikasi Jenius dengan metode analisis sentimen *Genetic Algorithm-Support Vector Machine* (GA-SVM).

Terdapat beberapa penelitian yang menggunakan pendekatan seperti pada penelitian ini yaitu Ernawati et al. [3] yang menerapkan algoritma *Naïve Bayes* (NB) dengan fungsi envelope opsi algoritma genetika (GA) untuk menganalisis peringkat opini perusahaan *fashion online*. Akurasi dan AUC sebagai ukuran evaluasi meningkat secara signifikan. Peneliti sebelumnya Zainuddin dan Selamat [4] juga menggunakan SVM dalam penelitiannya dengan menggunakan banyak fitur dan SVM memberikan nilai rata-rata AUC lebih dari 0,8. Sementara itu, penelitian sebelumnya oleh Mentari et al. [5] tentang sikap *audiens* terhadap kurikulum 2013 menunjukkan peningkatan akurasi hasil dengan seleksi fitur QER menggunakan metode *K-nearest neighbour*, yaitu dari 93,63% dengan  $k = 1$  menjadi 96,36 % dengan rasio 50%.

Berdasarkan permasalahan yang muncul, maka penelitian ini dilakukan untuk memperoleh hasil klasifikasi sentimen pada aplikasi Jenius menggunakan seleksi fitur QER dengan metode SVM dan metode GA-SVM. Serta, memperoleh hasil perbandingan kinerja seleksi fitur QER menggunakan metode SVM dan metode GA-SVM pada analisis sentimen tentang aplikasi Jenius.

## 2. Material dan Metode

### 2.1. Pengumpulan Data

Data berupa *tweets* yang berkaitan dengan aplikasi Jenius dikumpulkan dari *Twitter* pada kurun waktu 1 Januari hingga 31 Desember 2021 menggunakan kata kunci “Bank Jenius”. Penarikan data didapat dengan memanfaatkan fasilitas *snscarpe* yang disimpan dalam format csv.

### 2.2. Term Frequency - Inverse Document Frequency

*Term Frequency-Inverse Document Frequency* (TF-IDF) dipahami sebagai proses transformasi data tekstual menjadi data numerik agar dapat dilakukan pembobotan terhadap setiap kata (fitur atau variabel). TF-IDF menjadi ukuran statistik pada saat evaluasi signifikansi sebuah kata dalam dokumen. Bobot kata diberikan berdasarkan frekuensi kemunculan kata dalam dokumen yang dianalisis. Kata yang muncul dengan frekuensi tinggi dalam satu dokumen akan memiliki bobot kata yang tinggi, sedangkan jika muncul dalam banyak dokumen, maka bobot kata yang dimiliki menjadi rendah. Persamaan (1) merupakan rumus dari TF-IDF [8].

$$W_{t,d} = tf_{t,d} \times idf_t = tf_{t,d} \times \log \frac{N}{df_t} \quad (1)$$

dengan

$W_{t,d}$  : Bobot *Term Frequency – Inverse Document Frequency*

$tf_{t,d}$  : Jumlah frekuensi kata

$idf_t$  : Jumlah *inverse* frekuensi dokumen tiap kata

$df_t$  : Jumlah frekuensi dokumen tiap kata

$N$  : Jumlah total dokumen

### 2.3. Query Expansion Ranking

*Query Expansion Ranking* (QER) dipahami sebagai metode perankingan dan digunakan dalam seleksi fitur atau variabel guna mereduksi kompleksitas komputasi tanpa mengurangi kualitas analisis sentimen. Berikut Persamaan (2) merupakan rumus dari QER [9].

$$Score_f = \left| \frac{p_f + q_f}{p_f - q_f} \right| \quad (2)$$

dengan:

$score_f$  : Nilai Query Expansion Ranking

$p_f$  : Peluang fitur  $f$  dalam dokumen kelas positif

$q_f$  : Peluang fitur  $f$  dalam dokumen kelas negatif

Nilai dari  $p_f$  dan  $q_f$  dapat diperoleh menggunakan rumus pada Persamaan (3) dan Persamaan (4).

$$p_f = \frac{df_+^f + 0.5}{n^+ + 1} \quad (3)$$

$$q_f = \frac{df_-^f + 0.5}{n^- + 0.5} \quad (4)$$

dengan:

$df_+^f$  : Jumlah dokumen positif yang mengandung fitur  $f$

$df_-^f$  : Jumlah dokumen negatif yang mengandung fitur  $f$

$n^+$  : Jumlah dokumen positif

$n^-$  : Jumlah dokumen negative

#### 2.4. Genetic Algorithm

Genetic Algorithm (GA), dikenal juga dengan Algoritma Genetika merupakan teknik pencarian dan optimasi mengikuti prinsip-prinsip evolusi dan genetika alam. Cara kerja dari seleksi fitur GA adalah dengan meniru evolusi alami dengan memodelkan populasi solusi yang dinamis. Anggota populasi disebut sebagai kromosom, kromosom terbaik diantara kromosom yang lain bertujuan memberikan solusi optimal dari fungsi objektif permasalahan optimasi [10]. Struktur yang akan diimplementasikan dalam Algoritma Genetika secara umum adalah sebagai berikut [11]:

1. Membangkitkan populasi awal
2. Seleksi kromosom dengan nilai *fitness*. Kromosom dengan *fitness value* yang tinggi akan terpilih untuk menjadi *parent*.
3. Melakukan *crossover* dan mutasi, dengan proses ini akan diperoleh hasil yang disebut *mutated child*.
4. Mengembalikan nilai *mutated child* ke populasi menggantikan kromosom terpilih pada langkah 2. Tahap ini hingga kriteria berhenti (generasi) terpenuhi.

## 2.5. Support Vector Machine

Support Vector Machine (SVM) adalah cara klasifikasi melalui *machine learning* (*supervised learning*) yang memberikan prediksi kelas atas dasar model atau pola yang dihasilkan melalui proses *training*. Klasifikasi mengacu kepada *hyperplane* atau garis pembatas (*decision boundary*) yang memisahkan suatu kelas dengan kelas lainnya [12]. Terdapat dua kelas pada persamaan optimasi SVM, yaitu klasifikasi *linear* serta klasifikasi *non-linear*. Persamaan (5) merupakan bentuk kelas atas dasar *hyperplane* optimal dalam kasus seperangkat data yang dipisahkan secara *non-linear* [7].

$$f(x_d) = \sum_{i=1}^n \alpha_i y_i K(x_i, x_d) + b \quad (5)$$

dan  $b$  diperoleh:

$$b = y_i - \frac{1}{2} \sum_{i=1}^n \alpha_i y_i K(x_i, x)$$

$K$  yaitu fungsi Kernel yang akan digunakan. Fungsi Kernel RBF diperlukan dalam pemetaan data non linier menjadi linier. Berikut Persamaan (6) merupakan fungsi kernel RBF [13].

$$K(x_i, x_j) = \exp \left( - \frac{(x_i - x_j)^T (x_i - x_j)}{2\gamma^2} \right) \quad (6)$$

## 2.6. Klasifikasi

Klasifikasi merupakan proses mengelompokkan data ke dalam kelas atau label yang memiliki atribut yang sama. Merujuk pada Prasetyo [14], klasifikasi merupakan proses pengelompokan objek data ke dalam kelas tertentu berdasarkan kelas-kelas yang ada. Klasifikasi dalam analisis sentimen dilakukan dengan mengelompokkan teks menjadi sentimen positif serta sentimen negatif melalui perhitungan nilai setiap fitur. Kualitas hasil klasifikasi dapat dinilai dan dievaluasi berdasarkan beberapa ukuran [15]:

### 1. Accuracy

$$accuracy = \frac{TP + FN}{(TP + TN + FP + FN)} \quad (7)$$

### 2. Precision

$$precision = \frac{TP}{TP + FP} \quad (8)$$

### 3. Sensitivity

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (9)$$

### 4. F-Measure

$$F_1 = \frac{2 * \text{precision} * \text{Sensitivity}}{\text{precision} + \text{Sensitivity}} \quad (10)$$

dengan:

*TP* : True Positive

*TN* : True Negative

*FP* : False Positive

*FN* : False Negative

## 3. Hasil dan Diskusi

### 3.1. Deskripsi Data

Proses *crawling* yang dilakukan pada *platform* media sosial *Twitter* memberikan sejumlah data yang dibutuhkan dalam penelitian. *Tweet* yang dikumpulkan ialah *tweet* dengan menggunakan kata kunci “Bank Jenius” yang diunggah pada kurun waktu 1 Januari hingga 31 Desember 2021. Data yang didapat berjumlah 9896 *tweet* dalam bahasa Indonesia yang kemudian diklasifikasikan secara manual menjadi sentimen positif serta sentimen negatif. Setelah dilakukan klasifikasi manual terhadap 9896 data *tweet*, hanya 2008 data *tweet* yang masuk ke dalam klasifikasi, sementara 7888 data *tweet* lainnya merupakan data *noise* sehingga harus diabaikan.

*Preprocessing* data teks dilakukan terhadap data *tweet* yang telah dikumpulkan. *Preprocessing* dilakukan melalui serangkaian tahap berupa *normalization*, *Tokenizing*, *stemming*, dan *stopword removal*.

### 3.2. Pembobotan Kata *Term Frequency-Inverse Document Frequency*

Perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF) dilakukan setelah melalui proses tahap awal yaitu *preprocessing*. Perhitungan TF-IDF dilakukan guna memperoleh nilai bobot setiap kata dasar yang didapat saat *tokenizing*. Dalam penelitian ini menampilkan kata dasar sebanyak 3427 kata dari 2008 data *tweet*. Perhitungan nilai TF-IDF dilakukan menggunakan persamaan (1), Tabel 1 merupakan hasil perhitungan TF-IDF pada dokumen 1.

**Tabel 1.** Term Frequency-Inverse Document Frequency.

No	Kata	$tf_{t,d}$	$df_t$	$idf_t$	$W_{t,d}$
1	Jarang	0.0625	15	2.3588	0.1474
2	Pakai	0.0625	624	0.7397	0.0462
3	Jenius	0.0625	2008	0.2322	0.0145
4	Kecuali	0.0625	11	2.4935	0.1558
5	Transaksi	0.0625	126	1.4345	0.0897
6	Visa	0.0625	74	1.6657	0.0141
⋮	⋮	⋮	⋮	⋮	⋮
3425	Zakat	0.00	1	3.5349	0.00
3426	Zaman	0.00	1	3.5349	0.00
3427	Zuriya	0.00	1	3.5349	0.00

### 3.3. Query Expansion Ranking

Query Expansion Ranking (QER) merupakan metode seleksi yang akan memilih fitur atau variabel atas dasar bobot atau nilai yang dimiliki fitur tersebut. Kata atau *term* terpilih nantinya merupakan kata dengan bobot atau nilai yang paling tinggi. Nilai probabilitas masing-masing fitur pada sentimen positif dan sentimen negatif dihitung dengan Persamaan (3) serta Persamaan (4) yang digunakan dalam perhitungan *score* QER pada Persamaan (2).

Maka diperoleh nilai  $p_f$ ,  $q_f$  dan *score* QER pada fitur berjumlah 2212 kemudian dilakukan perankingan setiap fitur berdasarkan nilai *score* QER yang diperoleh. Tabel 2 merupakan hasil *ranking* dan fitur yang digunakan pada proses *genetic algorithm*.

**Tabel 2.** Dataset Query Expansion Ranking

No	Fitur	$p_f$	$q_f$	$score_f$
1	Bank	0.999472	1	3786.8788
2	Jenius	0.999472	1	3786.8788
3	Rekening	0.09019	0.089967	808.4381
4	Aplikasi	0.117616	0.118229	384.8095
5	Mobile	0.007911	0.008008	165.5736
6	Tahan	0.010021	0.009892	153.8445
7	Untung	0.021624	0.022138	85.14016
⋮	⋮	⋮	⋮	⋮
2210	Emosi	0.000527	0.014602	1.074947

No	Fitur	$p_f$	$q_f$	$score_f$
2211	Kecewa	0.000527	0.01837	1.059119
2212	Parah	0.000527	0.021196	1.051035

### 3.4. Genetic Algorithm

Seleksi fitur GA menerapkan konsep evolusi alam untuk mendapatkan fitur terbaik yang akan menjadi himpunan solusi optimal. Dalam penelitian ini, kromosom merepresentasikan fitur, yang dikodekan sebagai *string biner* dari 0 atau 1, kromosom yang bernilai 0 berarti bahwa fitur tersebut dihilangkan sedangkan kromosom bernilai 1 menunjukkan bahwa fitur tersebut terpilih.

Pada penelitian ini akan dibangkitkan sebuah populasi yang berukuran 50 kromosom sampai iterasi konvergen yaitu iterasi yang menghasilkan akurasi lebih baik. Kemudian diambil 2 nilai *fitness* terbaik yang diperoleh dari akurasi kromosom dengan pembagian data latih dan data uji sebesar 80:20 dengan menggunakan metode klasifikasi SVM yang akan menjadi *parent*.

*Crossover* dilakukan pada kromosom *parent* dengan *crossover rate* 0.5 dan diperoleh kromosom *child* dimana dari 2212 data dilakukan *crossover* sebanyak 1106. Kemudian mutasi dilakukan untuk mempertahankan keanekaragaman gen dalam kromosom dengan *mutated rate* 0.005. Setelah proses mutasi selesai, maka kedua kromosom *mutated child*  $K_1$  dan  $K_2$  dikembalikan ke populasi sebelumnya dengan menggantikan posisi kromosom *parent*  $K_1$  dan  $K_2$ . Proses tersebut dilakukan hingga akurasi mengalami peningkatan dari akurasi sebelumnya dan tidak mengalami peningkatan lagi. Pada Tabel 3 disajikan dari 2212 fitur yang ada, diperoleh sebanyak 2182 fitur dengan nilai akurasi sebesar 88.3%.

**Tabel 3.** Dataset Genetic Algorithm

No	Fitur	Genetic Algorithm
1	Bank	1
2	Jenius	1
3	Rekening	1
4	Aplikasi	1
5	Mobile	1
6	Tahan	1
7	Untung	1
⋮	⋮	⋮
2180	Emosi	1
2181	Kecewa	1
2182	Parah	1

### 3.5. Klasifikasi *Support Vector Machine*

Berdasarkan hasil dari tahap *preprocessing* dan pembobotan TF-IDF, maka dilanjutkan ketahap analisis klasifikasi menggunakan *Support Vector Machine* (SVM). Klasifikasi SVM menggunakan data *training* dalam pembentukan model klasifikasi serta data *testing* guna memvalidasi model klasifikasi yang terbentuk menggunakan rasio 80:20 dimana sebanyak 1606 data latih dan 402 data uji. Model SVM kernel RBF menjadi model klasifikasi terbaik yang digunakan untuk data *non-linear separable*, terlihat dalam persamaan (5) dan persamaan (6).

#### 3.5.1. *Confusion Matrix* dengan Seleksi Fitur *Query Expansion Ranking*

Tahap pengujian ini dilakukan klasifikasi menggunakan dataset QER pada Tabel 2. Berikut Tabel 4 yang menunjukkan *confusion matrix* dari metode seleksi fitur QER.

Tabel 4 *Confusion Matrix* Seleksi Fitur *Query Expansion Ranking*

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	211	10
Negatif	39	142

Kinerja model klasifikasi berdasarkan *confusion matrix* dalam Tabel 4 didapat melalui perhitungan dengan menggunakan Persamaan (7) sampai Persamaan (10) sebagai berikut:

$$\begin{aligned} \text{Accuracy} &= \frac{211 + 142}{211 + 142 + 39 + 10} \\ &= 0.8781 \approx 87.81\% \end{aligned}$$

$$\begin{aligned} \text{Precision} &= \frac{211}{211 + 39} \\ &= 0.844 \approx 84.40\% \end{aligned}$$

$$\begin{aligned} \text{Sensitivity} &= \frac{211}{211 + 10} \\ &= 0.9547 \approx 95.47\% \end{aligned}$$

$$\begin{aligned} F - \text{Measure} &= 2 \times \frac{0.844 \times 0.95475}{0.844 + 0.95475} \\ &= 0.8959 \approx 89.59\% \end{aligned}$$

### 3.5.2. Confusion Matrix Seleksi Fitur Query Expansion Ranking dan Genetic Algorithm

Tahap pengujian ini dilakukan klasifikasi menggunakan dataset GA pada Tabel 3. Berikut Tabel 5 menunjukkan *confusion matrix* dengan seleksi fitur QER dan GA.

**Tabel 5.** *Confusion Matrix* seleksi fitur QER dan Genetic Algorithm

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	211	10
Negatif	37	144

Kinerja model klasifikasi berdasarkan *confusion matrix* dalam Tabel 5 didapat melalui pehitungan dengan menggunakan Persamaan (7) sampai Persamaan (10) sebagai berikut:

$$Accuracy = \frac{211 + 144}{211 + 144 + 37 + 10}$$

$$= 0.8831 \approx 88.31\%$$

$$Precision = \frac{211}{211 + 37}$$

$$= 0.8508 \approx 85.08\%$$

$$Sensitivity = \frac{211}{211 + 10}$$

$$= 0.9547 \approx 95.47\%$$

$$F - Measure = 2 \times \frac{0.8508 \times 0.9547}{0.8508 + 0.9547}$$

$$= 0.8998 \approx 89.98\%$$

### 4.7. Perbandingan Support Vector Machine dan Seleksi Fitur

Perbandingan metode SVM tanpa menggunakan seleksi fitur dengan metode SVM yang menggunakan seleksi fitur QER dan GA dalam penelitian ini memberikan pertimbangan berdasarkan hasil ketepatan klasifikasi yang terbaik. Tabel 6 menunjukkan perbandingan kualitas klasifikasi.

**Tabel 6.** Persentasi Perbandingan Support Vector Machine dan seleksi fitur

Metode	<i>Accuracy</i>	<i>Precision</i>	<i>Sensitivity</i>	<i>F-Measure</i>
SVM	78.36	79.39	81.90	80.62
QER + SVM	87.81	84.40	95.48	89.59
QER + GA-SVM	88.31	85.08	95.48	89.98

Tabel 6 menunjukkan perbandingan ketepatan klasifikasi data sehingga kemudian dapat disimpulkan bahwa secara keseluruhan kinerja metode SVM melalui penggunaan seleksi fitur QER dan GA lebih baik jika dibandingkan dengan tanpa penggunaan seleksi fitur dalam mengukur ketepatan klasifikasi mengenai opini masyarakat terhadap aplikasi Jenius. Perbandingan tersebut dilihat dari nilai *Accuracy*, *Precision* dan *F-measure* menunjukkan aplikasi metode SVM bersama seleksi fitur QER dan GA menghasilkan ketepatan klasifikasi yang lebih baik.

#### **4. Kesimpulan**

Hasil klasifikasi sentimen terhadap aplikasi Jenius di *Twitter* menggunakan *Support Vector Machine* kernel *Radial Basis Function* melalui seleksi fitur *Query Expansion Ranking* ialah 947 *tweet* sentimen positif serta 1061 *tweet* sentimen negatif dengan nilai *Accuracy* 87.81%, *Precision* 84.40%, *Sensitivity* 95.48% dan *F-Measure* 89.59%. Sedangkan, penggunaan seleksi fitur *Query Expansion Ranking* dan *Genetic Algorithm-Support Vector Machine* memberikan hasil *Accuracy* model klasifikasi 88.31%, *Precision* 85.08%, *Sensitivity* 95.48% dan *F-Measure* 89.98%. Berdasarkan perhitungan kinerja tersebut, dapat disimpulkan bahwa metode seleksi fitur *Query Expansion ranking* dengan *Genetic Algorithm-Support Vector Machine* lebih baik dalam mengklasifikasikan sentimen terhadap aplikasi Jenius di *Twitter*.

#### **Daftar Pustaka**

- [1] Fikri, M. I., Sabrila, T. S., & Azhar, Y. Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA Jurnal*, 10(2), 71–76, 2020. <https://doi.org/10.32664/smatika.v10i02.455>.
- [2] Haddi, E., Liu, X., & Shi, Y. The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 17, 26–32, 2013. <https://doi.org/10.1016/j.procs.2013.05.005>.
- [3] Ernawati, S., Yulia, E. R., Frieyadie, & Samudi. Implementation of The Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies. *Proceedings of the 6th International Conference on Cyber and IT Service Management (CITSM)*, 1–5, 2018. <https://doi.org/10.1109/CITSM.2018.8674286>.
- [4] Zainuddin, N., & Selamat, A. Sentiment Analysis using Support Vector Machine. *Proceedings of the International Conference on Computer, Communications, and Control Technology (I4CT)*, 333–337, 2014. <https://doi.org/10.1109/I4CT.2014.6914200>.
- [5] Mentari, N. D., Fauzi, M. A., & Muflikhah, L. Analisis Sentimen Kurikulum 2013 pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(8), 2739–2743, 2018.

- [6] Mulyani, S., Thamrin, S. A., & Siswanto, S. Analisis Sentimen Masyarakat pada Kebijakan Vaksinasi Covid-19 di Twitter Menggunakan Metode Support Vector Machine dengan Kernel Radial Basis Function Berbasis Fitur Leksikon. *Jambura Journal of Probability and Statistics*, 3(2), 110–119, 2022. <https://doi.org/10.34312/jjps.v3i2.16663>
- [7] Rezki, N., Thamrin, S. A., & Siswanto, S. Sentiment Analysis of Merdeka Belajar Kampus Merdeka Policy Using Support Vector Machine with Word2Vec. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 17(1), 481–486, 2023. <https://doi.org/10.30598/barekengvol17iss1pp0481-0486>
- [8] Dadgar, S. M. H., Araghi, M. S., & Farahani, M. M. A Novel Text Mining Approach Based on TF-IDF and Support Vector Machine for News Classification. *Proceedings of the IEEE International Conference on Engineering and Technology (ICETECH)*, 112–116, 2016. <https://doi.org/10.1109/ICETECH.2016.7569223>
- [9] Fanissa, S., Fauzi, M. A., & Adinugroho, S. Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(8), 2766–2770, 2018.
- [10] Ladha, L., & Deepa, T. Feature Selection Methods and Algorithms. *International Journal on Computer Science and Engineering*, 3(5), 1787–1797, 2011.
- [11] Thiang, Kurniawan, R., & Ferdinando, H. Implementasi Algoritma Genetika pada Mikrokontroler MCS51 untuk Mencari Rute Terpendek. *Proceedings of Seminar of Intelligent Technology and Its Applications (SITIA)*, Institut Teknologi Sepuluh Nopember, 2001.
- [12] Novantirani, A., Sabariah, M. K., & Effendy, V. Analisis Sentimen pada Twitter Mengenai Penggunaan Transportasi Umum Darat dalam Kota dengan Metode Support Vector Machine. *e-Proceeding of Engineering*, 2(1), 1177–1183, 2015.
- [13] Hsu, S. Developing a Scale for Teacher Integration of Information and Communication Technology in Grades 1–9. *Journal of Computer Assisted Learning*, 26(3), 175–189, 2010. <https://doi.org/10.1111/j.1365-2729.2010.00348.x>
- [14] Prasetyo, E. *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: CV Andi Offset, 2012.
- [15] Fadil, M., Islamiyati, A., & Thamrin, S. A. Classification of Nutritional Status in Toddlers using the Support Vector Machine Method. *Communications in Mathematical Biology and Neuroscience*, 2025, 2025. <https://doi.org/10.28919/cmbn/9126>