



Analisis Sentimen Terhadap Ulasan Netizen Pada Aplikasi *CamScanner* Menggunakan Metode *Support Vector Machine* (SVM)

Wina Ayunda Sari^{1*}, Deni Arifianto², Agung Nilogiri³

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember¹³

Sistem Informasi, Fakultas Teknik, Universitas Muhammadiyah Jember²

Email: winaayundasari02@gmail.com^{1*}, deniarifianto@unmuhjember.ac.id², agungnilogiri@unmuhjember.ac.id³

ABSTRAK

CamScanner merupakan aplikasi pemindai dari sesuatu yang dapat menghasilkan gambar, maupun dokumen. Pada penelitian ini penulis melakukan analisis sentimen terhadap ulasan netizen yang terdapat pada komentar di *Google Play Store*, proses ini menerapkan metode *Support Vector Machine* (SVM). *Dataset* yang digunakan pada penelitian ini berjumlah 1315 data dan dibagi menjadi 2 bagian, 1000 sebagai proses *Cross Validation*, 315 sebagai *unseen data*. Data diklasifikasi berdasarkan kelas positif, netral, dan negatif. Hasil akurasi terbaik yang didapatkan pada penelitian ini yaitu sebesar 95,2%. Hasil pengujian menggunakan *unseen data test* menghasilkan nilai akurasi sebesar 81%. Proses *Confusion Matrix* pada presisi menghasilkan nilai sebesar 96%, *recall* sebesar 98%, dan *F-Measure* sebesar 98%.

Kata Kunci: Analisis Sentimen, *Google Play Store*, *Teks Mining*, SVM

ABSTRACT

CamScanner is a scanning application for something that can produce images or document. In this study the authors conducted a sentiment analysis of netizen review contained in comments on the *Google Play Store*, this process implements the *Support Vector Machine* (SVM) method. The dataset used in this study totaled 1315 data and was divided into 2 parts, 1000 as *Cross Validation* process, 315 as *unseen data*. Data is classified base on positive, neutral, and negative class. The best accuracy results obtained in this study were 95.2%. The test results using the *unseen data test* produce an accuracy value of 81%. The *Confusion Matrix* process on Precision produces a value of 96%, Recall of 98%, and *F-Measure* of 98%.

Keywords: Sentiment Analysis, *Google Play Store*, Text Mining, SVM

1. PENDAHULUAN

Smartphone merupakan teknologi yang sering digunakan oleh banyaknya masyarakat. Sering dijumpai bahwa *smartphone* merupakan sesuatu yang sangat krusial bagi semua kalangan masyarakat karena dapat disimpan kedalam teknologi masa kini. Mungkin sulit bagi seseorang untuk membawa *file-file* dokumen karena seseorang sering menerimanya dalam bentuk *hardcopy*, sehingga dapat merasa kesulitan untuk membawa berkas-berkasnya. Untuk mengatasi masalah ini, masyarakat bisa memanfaatkan aplikasi *CamScanner* yang tersedia di *Google Play Store* dalam mendigitalkan *file* manual tanpa harus mengantri di tempat *fotocopy* (Sholihah & Indriyanti, 2022). Pada penelitian ini melakukan analisis sentimen terhadap Ulasan pengguna *CamScanner* agar memudahkan saat melakukan pengolahan data.

Pemberian ulasan pada kolom komentar memiliki rating 1 sampai 5, dimana komentar ini diurutkan berdasarkan rating yang sesuai dengan isinya. Namun seringkali ditemukan rating yang tidak sesuai dengan ulasan. sehingga hal tersebut belum cukup menggambarkan kualitas dari aplikasi. Hal ini memungkinkan ulasan tersebut dapat mempengaruhi pengunjung untuk mengunduh aplikasi *CamScanner* (Potharaju dkk., 2017).

Biasanya, metodologi yang digunakan untuk menganalisis data diperlukan saat melakukan analisis yaitu menggunakan Metode *Support Vector Machine* (SVM). Berbagai peneliti melaporkan bahwa metode SVM adalah metode yang sangat akurat dalam menganalisa teks penjabaran (Wahyudi & Kusumawardana, 2021). Penelitian yang dilakukan oleh (Sharma & Dey, 2012) telah membandingkan *Naive Bayes*, *Decision Tree*, *Maximum Entropy*, *Support Vector Machine*, *K-Nearest Neighbors*, *Winnow Classifier*, dan *Adaboost Classifier* menggunakan *dataset* pada *film review*, yang menghasilkan bahwa metode SVM memiliki performa yang lebih baik pada klasifikasi sentimen.

Penggunaan SVM dalam penelitian sudah umum digunakan. Salah satu kelebihan yang dimiliki yaitu dapat mengatasi berbagai permasalahan pada klasifikasi *text* bebas *linear*, jika dibandingkan, SVM lebih unggul dari model *machine learning* lainnya dalam hal nilai.

Berdasarkan permasalahan pada latar belakang yang telah dijelaskan sebelumnya, maka dari itu peneliti memutuskan menggunakan *Support Vector Machine* (SVM) yang diharapkan sesuai dengan pokok permasalahan. Oleh karena itu, penelitian ini bertujuan untuk mengetahui nilai klasifikasi lebih dalam terkait data dari hasil ulasan *CamScanner* di *Google Play Store*.

2. KAJIAN PUSTAKA

A. TF-IDF

TF-IDF merupakan metode yang digunakan dalam pencarian informasi untuk mengidentifikasi bobot berdasarkan satu kategori. Berikut rumus untuk menghitung *tf* yaitu:

$$tf_{t,d} = \frac{n_{t,d}}{N} \quad (1)$$

- $tf_{t,d}$: Frekuensi kemunculan kata dalam suatu dokumen
 $n_{t,d}$: Nilai istilah yang muncul
 N : Semua *term* pada dokumen

Berikut rumus persamaan *idf* sebagai berikut.

$$idf_d = \log \left(\frac{N_d}{df} \right) \quad (2)$$

- idf_d : Jumlah kemunculan suatu kata pada dokumen
 N_d : Banyaknya total dokumen
 df : Banyak dokumen yang mengandung *term*

berikut rumus pembobotan kata pada *tf-idf* sebagai berikut

$$tfidf_{t,d} = tf_{t,d} \times idf_d \quad (3)$$

- $tfidf_{t,d}$: *Term Frequency - Invers Document Frequency*
 $tf_{t,d}$: Nilai *tf*
 idf_d : Nilai *idf*

B. K-Fold Cross Validation

Teknik dalam melakukan verifikasi satu model terhadap satu *set* data dapat disebut *K-Fold Cross Validation*. Rumus pada *Cross Validation* sebagai berikut.

$$Data\ Testing = \frac{banyak\ data}{k-fold} \times 100 \quad (4)$$

Data yang digunakan dalam proses *Cross Validation* sebanyak 1000 data.

C. Support Vector Machine (SVM)

Support Vector Machine (SVM) yaitu Metode dari *Supervised Learning* untuk membuat prediksi baik secara kualitatif maupun kuantitatif (Fide dkk., 2021). Berikut langkah-langkah pada *Support Vector Machine* (SVM):

- Menentukan setiap kata dalam dokumen sehingga dimana kata tersebut sering muncul atau ulasan komentar yang ada pada *Google Play Store*
- Menentukan rumus yang ada dengan inisial awal dimana nilai $\alpha=0.5$, $C=1$, $\lambda=0.5$, $\epsilon = 0,001$ dan $\gamma = 0.5$
- Menghitung persamaan *matrix* menggunakan rumus sebagai berikut.

$$D_{ij} = y_i y_j (K(\rightarrow x_i \rightarrow x_j) + \lambda^2) \quad (5)$$

Keterangan:

- D_{ij} : Dokumen pada anggota *matrix* data ke- ij
 y_i : Label data atau kelas data ke- i
 y_j : Label data atau kelas data ke- j
 $K(\rightarrow x_i . \rightarrow x_j)$: Nilai fungsi kernel
 λ : Nilai turunan teoritis

- d. Menghitung nilai *error* menggunakan rumus untuk data ke- $n = 1,2,3,4,\dots,n$ dengan rumus diantaranya:

$$E_{D,i} = \sum_{j=1}^i \alpha_j D_{ij} \quad (6)$$

$$\delta \alpha_i = \min\{\max[\gamma(1 - E_i) - \alpha_i], C - \alpha_i\}$$

$$\alpha_i = \alpha_i + \delta \alpha_i$$

Keterangan:

- $E_{D,i}$: Nilai data *error* pada dokumen ke- i
 Γ : Laju pembelajaran
 $\delta \alpha_i$: *Delta alpha*
 $\text{Max}_{(i)} D_{ij}$: Nilai tertinggi diagonal matriks *hessian*.

- e. Menentukan nilai bias (b) dengan rumus sebagai berikut:

$$b = -\frac{1}{2}[w \cdot x^+ + w \cdot x^- + w \cdot x^0] \quad (7)$$

Keterangan:

- b : *Bias*
 w : Nilai *alpha* baru
 x^+ : Bobot positif
 x^- : Bobot negatif
 x^0 : Bobot netral

- f. Metode menghitung penilaian keputusan atau prakiraan (prediksi) dengan rumus sebagai berikut:

$$h(x) = \begin{cases} +x, & \text{if } w \cdot x + b \geq 0 \\ 0x, & \text{if } w \cdot x + b = 0 \\ -x, & \text{if } w \cdot x + b < 0 \end{cases} \quad (8)$$

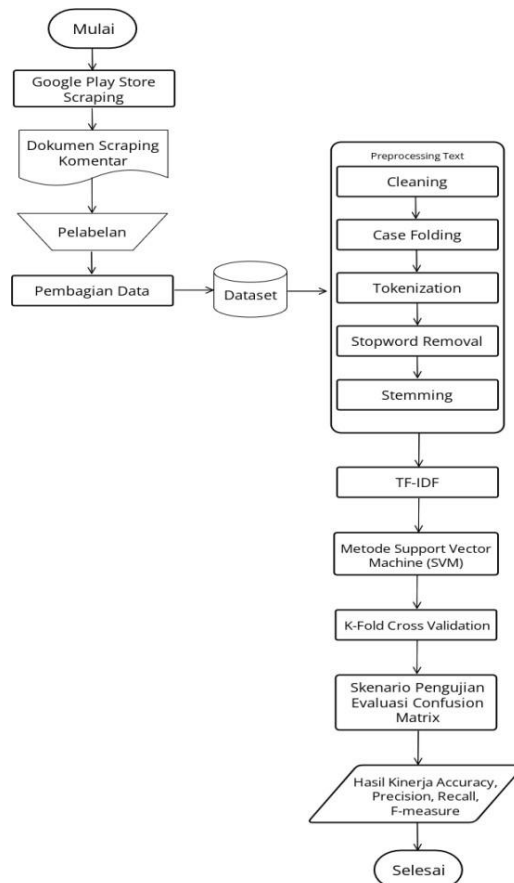
Keterangan:

- $h(x)$: Fungsi keputusan
 if : Perbandingan nilai logis dan perkiraan
 w : Nilai *alpha* baru
 x : Bobot nilai
 b : *Bias*

Selanjutnya untuk hasil fungsi pada $sign h(x)$ terhadap penilaian keputusan yang lebih dari sama dengan 0 maka hasilnya bernilai +1 yang merupakan sentimen “positif” dan sebaliknya, selanjutnya jika nilai sama dengan 0 maka hasilnya 0 yaitu “netral”, lalu jika hasil fungsi pada $sign h(x)$ dari perhitungan kurang dari 0 maka hasilnya -1 yaitu menghasilkan sentimen “negatif”.

3. METODE PENELITIAN

Tahap penelitian dari proses awal sampai akhir ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

Tahap awal setelah menentukan topik yaitu mengidentifikasi permasalahan. Tahapan identifikasi masalah yaitu penegasan terhadap batasan masalah sehingga tidak keluar dari tujuan. Pada identifikasi masalah ini mengupas berisi ulasan pengguna pada aplikasi *CamScanner* di *Google Play Store*, yang berupa komentar positif, netral, negatif dan berbagai masalah dalam melakukan analisis sentimen.

A. Pengumpulan Data

Pengumpulan data ini berasal dari ulasan yang ada pada komentar pengguna *CamScanner* di *Google Play Store*. Proses pengambilan data tersebut memakai teknik *Scraping* dan menggunakan bahasa pemrograman *Python*. Jumlah data yang diambil pada periode Januari sampai Oktober 2022 sebanyak 1315 data.

B. Preprocessing Data

Langkah-langkah yang harus dilakukan selama proses *Preprocessing Text* diantaranya:

1. *Cleaning*

Pada tahap ini yaitu melakukan menghapus tanda baca yang tidak konsisten pada setiap jenis penulisan yang tidak relevan.

2. *Case Folding*

Pada tahap ini yaitu perubahan kata menjadi huruf kecil penuh atau disebut *lowercase*.

3. *Tokenization*

Pada tahap ini yaitu memecah teks menjadi sebuah kalimat lalu menjadikan kata-kata.

4. *Stopword Removal*

Pada tahap ini yaitu menghilangkan kata-kata yang tidak perlu berdasarkan pustaka yang ada di *library*.

5. Stemming

Pada tahap ini yaitu proses pengubahan kata berimbuhan menjadi kata dasar atau kata baku.

4. HASIL DAN PEMBAHASAN

Berikut hasil nilai akurasi yang didapatkan pada *Support Vector Machine* (SVM) dapat dijelaskan sebagai berikut:

Tabel 1. Hasil Seluruh Langkah Uji

<i>K-Fold Cross</i>	Langkah Uji	<i>Support Vector Machine</i> (SVM) (<i>Data Training</i>)
2-fold	Langkah Uji 1	93,4%
	Langkah Uji 2	89,4%
4-fold	Langkah Uji 1	92,8%
	Langkah Uji 2	94,8%
	Langkah Uji 3	92,4%
	Langkah Uji 4	90,4%
5-fold	Langkah Uji 1	93%
	Langkah Uji 2	94,5%
	Langkah Uji 3	93,5%
	Langkah Uji 4	93,5%
	Langkah Uji 5	89,5%
8-fold	Langkah Uji 1	92,8%
	Langkah Uji 2	94,4%
	Langkah Uji 3	94,4%
	Langkah Uji 4	93,6%
	Langkah Uji 5	90,4%
	Langkah Uji 6	94,4%
	Langkah Uji 7	95,2%
	Langkah Uji 8	86,4%
10-fold	Langkah Uji 1	95%
	Langkah Uji 2	92%
	Langkah Uji 3	94%
	Langkah Uji 4	94%
	Langkah Uji 5	95%
	Langkah Uji 6	92%
	Langkah Uji 7	92%
	Langkah Uji 8	95%
	Langkah Uji 9	94%
	Langkah Uji 10	86%

Berdasarkan hasil Langkah Uji yang sudah dipaparkan diatas, maka Langkah Uji terbaik yaitu:

- 1) Langkah Uji terbaik pada 8-fold Langkah Uji ke 7 dengan nilai 95,2%.
- 2) Langkah Uji terendah pada 10-fold Langkah Uji ke 10 dengan nilai 86%.

Selanjutnya yaitu melakukan pengujian data menggunakan *unseen data test* dimana masing-masing Langkah Uji menggunakan 315 *unseen data*. Data tersebut dikelompokkan menggunakan *K-Fold Cross Validation* dengan jumlah setiap sentimennya yaitu 235 positif, 66 negatif, dan 14 netral. Berikut hasil pengujian *unseen data* dan sebagai berikut:

Tabel 2. Hasil Seluruh Langkah Uji

<i>K-Fold Cross</i>	Langkah Uji	<i>Support Vector Machine</i> (SVM) (<i>Hasil Unseen Data Test</i>)
2-fold	Langkah Uji 1	80,9%
	Langkah Uji 2	76,8%
	Langkah Uji 1	80,6%
	Langkah Uji 2	82,5%

4-fold	Langkah Uji 3	80,6%
	Langkah Uji 4	80,3%
5-fold	Langkah Uji 1	81,3%
	Langkah Uji 2	83,2%
	Langkah Uji 3	81%
	Langkah Uji 4	81%
	Langkah Uji 5	79,7%
8-fold	Langkah Uji 1	82,5%
	Langkah Uji 2	81,6%
	Langkah Uji 3	83,8%
	Langkah Uji 4	81,6%
	Langkah Uji 5	82,2%
	Langkah Uji 6	81,6%
	Langkah Uji 7	81%
	Langkah Uji 8	81,9%
10-fold	Langkah Uji 1	82,2%
	Langkah Uji 2	81,6%
	Langkah Uji 3	82,9%
	Langkah Uji 4	83,2%
	Langkah Uji 5	81%
	Langkah Uji 6	82,5%
	Langkah Uji 7	81,6%
	Langkah Uji 8	82,2%
	Langkah Uji 9	80%
	Langkah Uji 10	82,2%

Berdasarkan hasil *unseen data test* diatas, dijelaskan bahwa:

- 1) Hasil *unseen data test* lebih rendah dari hasil *data training*.
- 2) Hasil *unseen data* lebih rendah karena data yang di uji tidak sebanyak *data training*, dan terutama pada *unseen data* diambil dari total data yang telah di *crawling* sehingga *unseen data* digunakan sebagai data uji.

Berdasarkan hasil Langkah Uji data yang sudah dilakukan, selanjutnya diperoleh rekapitulasi hasil data *training* dan *unseen data test* yang dapat dipaparkan sebagai berikut.

Tabel 3. Hasil Rekapitulasi

K – Fold Cross	Langkah Uji	Support Vector Machine (SVM)	
		Data Training	Data Unseen
2-fold	Langkah Uji 1	93,4%	80,9%
	Langkah Uji 2	89,4%	76,8%
4-fold	Langkah Uji 1	92,8%	80,6%
	Langkah Uji 2	94,8%	82,5%
	Langkah Uji 3	92,4%	80,6%
	Langkah Uji 4	90,4%	80,3%
5-fold	Langkah Uji 1	93%	81,3%
	Langkah Uji 2	94,5%	83,2%
	Langkah Uji 3	93,5%	81%
	Langkah Uji 4	93,5%	81%
	Langkah Uji 5	89,5%	79,7%
8-fold	Langkah Uji 1	92,8%	82,5%
	Langkah Uji 2	94,4%	81,6%
	Langkah Uji 3	94,4%	83,8%
	Langkah Uji 4	93,6%	81,6%
	Langkah Uji 5	90,4%	82,2%
	Langkah Uji 6	94,4%	81,6%
	Langkah Uji 7	95,2%	81%
	Langkah Uji 8	86,4%	81,9%
10-fold	Langkah Uji 1	95%	82,2%
	Langkah Uji 2	92%	81,6%
	Langkah Uji 3	94%	82,9%
	Langkah Uji 4	94%	83,2%

Langkah Uji 5	95%	81%
Langkah Uji 6	92%	82,5%
Langkah Uji 7	92%	81,6%
Langkah Uji 8	95%	82,2%
Langkah Uji 9	94%	80%
Langkah Uji 10	86%	82,2%

A. Pengujian Evaluasi *Confusion Matrix*

Setelah mendapatkan nilai akurasi yang sudah di olah sebelumnya, selanjutnya melakukan proses Evaluasi *Confusion Matrix* yaitu bagaimana mencari hasil Akurasi, Presisi, *Recall*, dan *F-Measure*.

Berikut hasil Presentase *Accuracy* yaitu:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Akurasi = \frac{114 + 5}{114 + 5 + 4 + 2} \times 100\% = 95,2\%$$

Berikut yaitu cara mencari nilai *Precision*:

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

$$Presisi = \frac{114}{114 + 4} = 0,9661 = 96\%$$

Berikut yaitu cara mencari nilai *Recall*:

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

$$Recall = \frac{114}{114 + 2} = 0,98276 = 98\%$$

Berikut cara mencari nilai *F-Measure*:

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$-Measure = 2 \times \frac{0,96 \times 0,98}{0,96 + 0,98}$$

5. KESIMPULAN

A. Kesimpulan

Kesimpulan yang dapat diberikan selama proses penelitian mengenai analisis sentimen terhadap ulasan netizen pada aplikasi *CamScanner* menggunakan metode *Support Vector Machine (SVM)* yaitu:

1. Data penelitian ini digunakan sebanyak 1315 dengan ulasan positif 1138 data, ulasan negatif 137 data, dan ulasan netral sebanyak 40 data. Selanjutnya data diambil sebanyak 315 ulasan yang terdiri dari 235 positif, 66 negatif, dan 14 netral. Sehingga data dalam proses penelitian yang digunakan yaitu 1000 data sebagai *training* dan 315 sebagai *data unseen*.
2. Pada proses *K-Fold Cross Validation* hasil akurasi terbaik yaitu pada 8-fold Langkah Uji 7 sebesar 95,2% dan nilai akurasi terendah berada pada 10-fold Langkah Uji 10 sebesar 86%. Selanjutnya melakukan pengujian *unseen data* yang menghasilkan akurasi sebesar 81% pada Langkah Uji ke 7 dengan *fold-8*.

3. Hasil *Confusion Matrix* untuk mencari akurasi yang menghasilkan nilai sebesar 95,2%, Hasil presisi sebesar 96%, Hasil *recall* sebesar 98%, dan hasil *F-Measure* dengan nilai sebesar 98%. Artinya, metode *Support Vector Machine* (SVM) inilah merupakan salah satu metode terbaik dalam pengujian data.

B. Saran

Setelah melakukan penelitian, penulis dapat memberikan saran sehingga dapat meningkatkan kualitas dan pengembangan pada penelitian yang akan datang yaitu:

1. Dapat menambahkan data yang lebih banyak lagi supaya nilai akurasi yang didapatkan bisa bertambah.
2. Menggunakan metode yang lainnya untuk dibandingkan dengan metode *Support Vector Machine* (SVM).

6. DAFTAR PUSTAKA

- Fide, S., Suparti, & Sudarno. (2021). Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (SVM) Dan Asosiasi. *Jurnal Gaussian*, 10(3), 346–358.
- Potharaju, R., Rahman, M., & Carbnar, B. (2017). A Longitudinal Study of Google Play. *IEEE Transactions on Computational Social Systems*, 4(3), 135–149.
- Sharma, A., & Dey, S. (2012). A comparative study of selection and machine learning techniques for sentiment analysis. *Conference: Proceedings of the 2012 ACM Research in Applied Computation Symposium*, 1–7.
- Sholihah, R., & Indriyanti, A. D. (2022). Analisis Kepuasan Pengguna Aplikasi Camscanner Menggunakan Metode Technology Acceptance Model (TAM) dan End-User Computing Satisfaction (EUCS). *Journal of Emerging Information System and Business Intelligence (JEISBI)*, 3(3), 102–109.
- Wahyudi, R., & Kusumawardana, G. (2021). Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *Jurnal Informatika*, 8(2).