

Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine

Faizal Fakhri Irfani

Informatika Universitas Amikom Yogyakarta

Mohamad Triyanto

Informatika Universitas Amikom Yogyakarta

Anggit Dwi Hartanto

Informatika Universitas Amikom Yogyakarta

Kusnawi

Informatika Universitas Amikom Yogyakarta

Abstrak: Analisis Sentiment Review Aplikasi Ruangguru Merupakan Proses Menganalisa, Memahami Dan Mengklasifikasikan Suatu Penilaian Yang Dikeluarkan Masyarakat Terhadap Entitas Aplikasi Ruangguru. Data Penelitian Ini Diambil Dari Website Google Play Store, Data Yang Diambil Yaitu Data Teks Review Dengan Jumlah 2000 Review. Data Diklasifikasikan Dengan Menggunakan Algoritma Support Vector Machine, Dan Dilakukan Pengujian Menggunakan Kombinasi Dari Pembagian Data Latih Dan Data Uji, Serta Menggunakan Sistem *K-Fold Cross-Validation*. Pengujian Menggunakan Kernel *Linear* Menghasilkan Akurasi Tertinggi Dengan Nilai 0.897. Kombinasi Data *Training* 60% Dan Data *Testing* 40% Menghasilkan Akurasi Tertinggi Sebesar 0.900. Pengujian Dengan Menggunakan Sistem *K-Fold*, Akurasi Tertinggi Berada Di Nilai *K-Fold* 6, 9, Dan 10 Dengan Nilai Akurasi Sebesar 0.902. Namun Pada *K-Fold* 10, Tingkat Presisi Nya Lebih Tinggi Dibanding Nilai *K-Fold* Lainnya Dengan Nilai Presisi Sebesar 0.903. Tingkat Akurasi Dalam Penelitian Ini Tinggi Berada Di Kisaran 90%. Hasil Dari Beberapa Pengujian Menunjukkan Bahwa Sentimen Masyarakat Terhadap Aplikasi Ruangguru Cenderung Positif.

Kata Kunci : Text Mining, Analisis Sentimen, Support Vector Machine (Svm), Ruangguru

Abstract: *Ruangguru Application Review Sentiment Analysis Is The Process Of Analyzing, Understanding And Classifying An Assessment Issued By The Public On Ruangguru Application Entities. The Research Data Was Taken From The Google Play Store Website, The Data Taken Were Review Text Data With A Total Of 2000 Reviews. Data Were Classified Using The Support Vector Machine Algorithm, And Testing Was Performed Using A Combination Of Training And Test Data Sharing, And Using The K-Fold Cross-Validation System. Testing Using A Linear Kernel Produces The Highest Accuracy With A Value Of 0.889. A Combination Of 60% Training Data And 40% Testing Data Produces The Highest Accuracy Of 0.900. Tests Using The K-Fold Cross-Validation System, The Highest Accuracy Is In The K-Fold Values 6, 9, And 10 With An Accuracy Value Of 0.902. But At K-Fold 10, The Level Of Precision Is Higher Than Other K-Fold Values With A Precision Value Of 0.903. The Level Of Accuracy In This Study Is High In The Range Of 90%. The Results Of Several Tests Indicate That Public Sentiment Towards Ruangguru Applications Tends To Be Positive.*

Keywords : *Text Mining, Sentiment Analysis, Support Vector Machine (Svm), Ruangguru*

1 Pendahuluan

Teknologi Berkembang Dengan Cepat Tidak Terkecuali Dalam Dunia Pendidikan, Peran Teknologi Dalam Pendidikan Salah Satunya Yaitu *Mobile Learning*, Penggunaan *Mobile Learning* Menjadi Salah Satu Media Pendukung Pembelajaran. Dengan *Mobile Learning* Maka Proses Belajar Menjadi Lebih Mudah, Seperti Waktu Yang Tidak Terikat, Tidak Terbatas Ruang Dan Waktu, Dan Dapat Diakses Dari Mana Saja.

Salah Satu Aplikasi *Mobile Learning* Yang Sedang *Tren* Dikalangan Masyarakat Saat Ini Ialah Ruangguru. Ruang Guru Adalah Aplikasi Belajar Online Yang Dibangun Oleh Adamas Belva Syah Devara Dan Muhammad Iman Usman Pada Tahun 2014. Ruangguru Menghubungkan Siswa Dengan Pengajar Yang Tepat Untuk Membantu Siswa Dalam Mempelajari Ilmu Yang Baru Serta Mendapatkan Solusi Belajar Diluar Sekolah. Aplikasi Ruangguru Dapat Diunduh Di Google Play Store Maupun App Store Ios. Dengan Lahirnya Inovasi Tersebut, Banyak Mengalir Pro-Kontra Dari Masyarakat Dengan Adanya Ruangguru Di Dunia Pendidikan. Berbagai Komentar Bermunculan Di Kolom *Review* Google Play Store. Untuk Mengetahui Penilaian Suatu Aplikasi *Mobile Learning* Dapat Menggunakan Analisis Sentimen. Dari Hasil Yang Diperoleh Dari Analisis Sentiment Dapat Diketahui Sentiment Positif Dan Negatif Dari Para Pengguna. Hasil Dari Analisis Ini Dapat Digunakan Sebagai Salah Satu Alat Untuk Bahan Pertimbangan Perusahaan Dalam Mengambil Sebuah Keputusan.

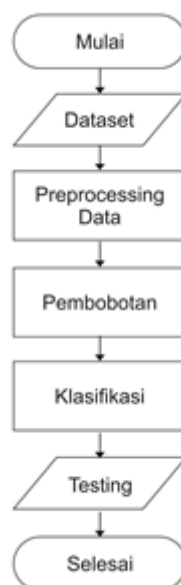
2 Tinjauan Pustaka

Support Vector Machine (Svm) Diperkenalkan Oleh Vapnik Pada Tahun 1992. Svm Merupakan Metode Pembelajaran Linier Yang Menemukan *Hyperline Optimal* Untuk Memisahkan Dua Kelas (Positif Dan Negatif). Svm Memiliki Kemampuan Dalam Menerapkan Pemisah Input Non-Linier Berdimensi Tinggi Yang Membuat Svm Menjadi Istimewa Dengan Menggunakan Fungsi Kernelnya. Terdapat Beberapa Kernel Svm Yang Dapat Digunakan, Namun Didalam Suatu Penelitian, Kernel Yang Sering Digunakan Ialah Linear, Polynomial, Radial Basis Function (Rbf). Kernel-Kernel Tersebut Memiliki Tingkat Akurasi Yang Berbeda-Beda Dalam Pengimplementasiannya Di Suatu Penelitian, Sehingga Banyak Peneliti Yang Sering Membandingkan Kernel Mana Yang Terbaik Guna Meningkatkan Akurasi Dari Hasil Penelitiannya.

Dari Penelitian Yang Telah Dilakukan Resky Rayvano Moningka Dan Syahfitri Kartika Lidya, Dimana Didalam Penelitian Resky Rayvano Moningka , Metode Yang Digunakan Adalah Svm Dan Maxent. Didalam Penelitian Syahfitri Kartika Lidya, Metode Yang Digunakan Adalah Svm Dan Knn. Dari Hasil Kedua Penelitian Menunjukkan Bahwa Akurasi Metode Svm Lebih Tinggi Dibandingkan Metode Lainnya. Oleh Karena Itu Didalam Penelitian Ini Penulis Melakukan Penelitian Analisis Sentimen Untuk Mengetahui Sentimen Serta Penilaian Dari Masyarakat Terhadap Adanya Aplikasi Ruangguru Menggunakan Metode Svm.

3 Metode Penelitian

Terdapat Beberapa Tahapan Yang Dilakukan Pada Penelitian. Alur Tahapan Tersebut Dijelaskan Pada Gambar 1.



Gambar 1. Langkah Penelitian

Mengumpulkan Data Review

Data Review Diambil Menggunakan Teknik *Crawling*. Pada Proses Ini Dilakukan Pengembangan Program Sederhana Berbasis Python Untuk Proses *Crawling* Data. Data Diambil Dari Website Google Play Store, Data Yang Diambil Yaitu Data Teks Review Dengan Jumlah 2000 Review Dan Dari Rentang Waktu 5 November 2019 – 9 November 2019.

Preprocessing Data

Pada Dasarnya, Dataset Yang Diperoleh Dari Proses *Text Mining* Memiliki Struktur Yang Sembarang Dan Tidak Beraturan. Oleh Karena Itu, Sebelum Dataset Dimasukan Kedalam Model, Data Terlebih Dahulu Melalui Tahap *Preprocessing Data*. Tahapan Ini Dilakukan Untuk Membersihkan Data Dari *Noise* Dan Mengubah Data Menjadi Data Yang Terstruktur. Tahap *Preprocessing* Meliputi :

1. *Case Folding* Ialah Tahapan Untuk Menyeragamkan Seluruh Teks Yang Akan Dimasukan Kedalam Model Menjadi Huruf Kecil Semua (*Lowercase*).
2. *Tokenize* Yaitu Proses Pemecahan Kata Pada Kalimat *Review*, Pemisahan Kata Dalam Kalimat Umumnya Menggunakan Karakter Spasi, Maka Karakter Spasi Diandalkan Untuk Proses Tokenisasi Ini.
3. *Cleansing* Yaitu Proses Menghilangkan *Noise* Yang Berupa *Emoticon* Dan Karakter Yang Kurang Penting Dalam Kalimat *Review*.
4. *Filtering* Yaitu Proses Ini Dilakukan Untuk Menghapus Data Yang Salah, Tidak Lengkap Atau Kesalahan Tipografi.
5. *Translation* Yaitu Proses Menerjemahkan Kata Yang Menggunakan Bahasa Asing Menjadi Kata Yang Berbahasa Indonesia
6. *Stopword Removal* Yaitu Proses Menghilangkan Kata Yang Termasuk Kedalam Kategori *Stopword*. *Stopword* Adalah Kata Yang Sering Muncul Namun Dianggap Tidak Memiliki Arti.
7. *Stemming* Yaitu Proses Menemukan Kata Dasar Dengan Menghilangkan Semua Imbuhan Yang Menempel Pada Kata.

Pembobotan Term

Dataset Yang Sudah Melewati Proses *Preprocessing Data* Selanjutnya Dilakukan Pembobotan Guna Mendapatkan Suatu Nilai Agar Bisa Diklasifikasikan. Pada Tahap Ini Pembobotan Menggunakan Metode *Term Frequency-Inverse Document Frequency* Atau *Tf-Idf*.

$$w_{ij} = tf_{ij} \times \log\left(\frac{D}{df_j}\right) + 1 \quad (1)$$

Dimana W_{ij} Merupakan Hasil Pembobotan Dari Kata/Term Terhadap Dokumen Yang Ada, Tf_{ij} Adalah Jumlah Frekuensi Munculnya Kata/Term Dalam Dokumen, D Adalah Jumlah Keseluruhan Dari Dokumen Yang Ada Didalam Dataset Dan Df_j Adalah Jumlah Dokumen Yang Mengandung Kata Yang Dimaksud.

Klasifikasi

Setelah Proses Preprocessing Dan Pembobotan, Proses Selanjutnya Adalah Klasifikasi Data. Metode Klasifikasi Yang Digunakan Yaitu Menggunakan Metode *Support Vector Machine*.

Berikut Merupakan Rumus Dari Klasifikasi Svm:

$$\min = \frac{1}{2} |w| \quad (2)$$

$$S.T \ Y_i(X_i.W+B)-1 \geq 0$$

Dimana $(X_i .W+B) \geq 1$ untuk Kelas 1, Dan $(X_i .W+B) \leq -1$ Untuk Kelas 2, X_i Adalah *Data Set*, Y_i Adalah *Output* Dari Data X_i , Dan W, B Adalah Parameter Yang Dicari Nilainya.

Testing

Dari Hasil Klasifikasi, Selanjutnya Dilakukan Pengujian. Pengujian Dilakukan Dengan Beberapa Ujicoba. Untuk Menghitung Performa Dari Model Dilakukan Pengujian Pertama Berdasarkan Kernel Yang Digunakan. Pengujian Kedua Menggunakan Kombinasi Dari Pembagian Data Latih Dan Data Uji, Serta Pengujian Ketiga Menggunakan Sistem *K-Fold Cross-Validation*. Didalam Metode *K-Folds Cross-Validation*, Data Yang Telah Dilakukan Pembobotan *Term* Dibagi Menjadi K Bagian, Dimana K Adalah Ukuran Dari *Term*. Dilakukan K Buah Percobaan Dalam Satu Set Percobaan Dengan Menggunakan Satu Bagian Sebagai Data Uji.

4 Hasil Dan Pembahasan

4.1 Dataset

Pada Penelitian Ini, Dataset Diperoleh Dari *Review* Aplikasi Ruangguru Di Google Play Store. Dari Proses *Crawling* Data, Didapatkan 2000 Data *Review*. Kemudian Data Tersebut Dibagi Menjadi 2 Untuk Data Latih Dan Data Uji. Didalam Data Latih Dilakukan Pelabelan Secara Manual Dimana Kalimat Positif Diberi Label 1 Dan Kalimat Negatif Diberi Label 0.

Tabel 1. Pembagian Dataset (Data Latih Dan Data Uji)

Prosentase	Data Latih	Data Uji
50% - 50%	500	500
60% - 40%	600	400
70% - 30%	700	300
80% - 20%	800	200
90% - 10 %	900	100

4.2 Pengujian

Pengujian Dilakukan Menggunakan Beberapa Aspek Seperti:

4.2.1 Pengujian Berdasarkan Kernel Yang Digunakan

Pengujian Ini Menggunakan Semua Data Latih Dan Data Uji Serta Kernel Svm Yang Berbeda-Beda. Kernel Svm Yang Diuji ialah *Linear*, *Radial Basis Function (Rbf)* , Dan *Polynomial*.

Tabel 2. Hasil Pengujian Berdasarkan Kernel

Kernel	Recall	Precision	Accuracy
Linear	0.981	0.905	0.897
Rbf	0.994	0.866	0.865
Polynomial	1.000	0.845	0.845

Dari Tabel 2 Memerlihatkan Bahwa Hasil Pengujian Menggunakan Kernel *Linear* Menghasilkan Akurasi Tertinggi Dengan Nilai 0.897.

4.2.2 Pengujian Berdasarkan Kombinasi Pembagian Data Latih Dan Data Uji

Pengujian Kedua Menggunakan Pembagian Data Seperti Yang Terdapat Di Tabel 1. Kernel Yang Akan Digunakan ialah *Linear*.

Tabel 3. Hasil Pengujian Berdasarkan Kombinasi Dataset

Data	Recall	Precision	Accuracy
50% - 50%	0.988	0.897	0.896
60% - 40%	0.990	0.901	0.900
70% - 30%	0.983	0.890	0.884
80% - 20%	0.987	0.897	0.894
90% - 10 %	0.980	0.902	0.892

Pada Tabel 3 Dapat Disimpulkan Bahwa Kombinasi Antara Data Latih 60% Dan Data Uji 40% Mendapatkan Hasil Akurasi Tertinggi Sebesar 0.900.

4.2.3 Pengujian Berdasarkan Sistem *Cross Validation* (Nilai K-Fold)

Pengujian Ketiga Menggunakan Nilai *Fold* Sebagai Parameter. Nilai *Fold* Yang Digunakan ialah 2-10 Fold. Pengujian Ini Menggunakan Kombinasi *Dataset* 60-40 Dan Kernel *Linear* Yang Di Pengujian Sebelumnya Memiliki Akurasi Tertinggi.

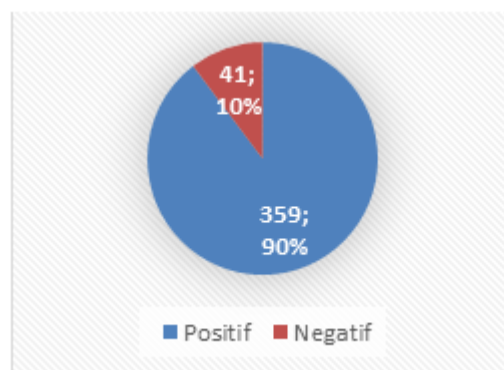
Tabel 4. Hasil Pengujian Berdasarkan Nilai *Fold*

Nilai K-Fold	Recall	Precision	Accuracy
2	0.990	0.887	0.887
3	0.990	0.892	0.892
4	0.992	0.899	0.900
5	0.992	0.901	0.900
6	0.990	0.902	0.902
7	0.988	0.900	0.898
8	0.990	0.899	0.898
9	0.990	0.902	0.902
10	0.990	0.903	0.902

Dari Tabel 4 Diperoleh Bahwa Akurasi Tertinggi Berada Di Nilai *K-Fold* 6, 9, Dan 10 Dengan Nilai Akurasi Sebesar 0.902. Namun Pada *K-Fold* 10, Tingkat Presisi Nya Lebih Tinggi Dibanding Nilai *K-Fold* Lainnya Dengan Nilai Presisi Sebesar 0.903.

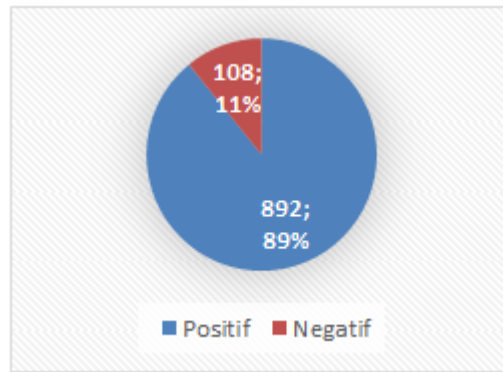
4.3 Evaluasi

Dari Tahap Pengujian Yang Dilakukan Dengan Menggunakan Kombinasi Pembagian Data Latih Sebesar 60% Dan Data Uji Sebesar 40% Dihasilkan Sebuah Sentimen, Dimana 400 Data Yang Diuji Berhasil Diklasifikasikan Menjadi 359 *Review* Positif Dan 41 *Review* Negatif.



Gambar 2. Hasil Pengujian 60-40

Selanjutnya Dilakukan Juga Pengujian Menggunakan Seluruh Data Latih Dan Data Uji. Dari Pengujian Tersebut Diperoleh Sebuah Sentimen, Dimana Dari 1000 Data Yang Diuji Berhasil Diklasifikasikan Menjadi 892 *Review* Positif Dan 108 *Review* Negatif.



Gambar 3. Hasil Pengujian Seluruh Data

5 Kesimpulan

Berdasarkan Analisa Dan Hasil Beberapa Pengujian Tersebut, Dapat Diperoleh Kesimpulan Bahwa:

1. Kernel *Linear* Memberikan Nilai Akurasi Yang Lebih Tinggi Dibandingkan Dengan Kernel *Rbf* Dan *Polynomial*.
2. Tingkat Akurasi Dalam Penelitian Ini Tinggi Berada Di Kisaran 90%.
3. Semakin Besar Data Latih Yang Digunakan Tidak Selalu Meningkatkan Nilai Akurasi. Hasil Pengujian Kombinasi Data Pada Data 60-40 Dan 90-10 Menunjukkan Jikalau Akurasi Yang Didapat 60-40 Lebih Tinggi Dibandingkan Data 90-10.
4. Pengujian Menggunakan *Cross Validation* Guna Menghitung Performa Tidak Cukup Mendapatkan Perbedaan Hasil Yang Signifikan Namun Semakin Tinggi Nilai *K-Fold* Cenderung Semakin Meningkatkan Akurasinya, Dimana Akurasi Tertinggi Dengan Pengujian Ini Didapat Dari Nilai *K-Fold* = 10.
5. Hasil Dari Beberapa Pengujian Menunjukkan Bahwa Sentimen Masyarakat Terhadap Aplikasi Ruanguru Cenderung Positif.

6 Daftar Pustaka

- Buntoro, G. A. (2016). Analisis Sentimen Hatespeech Pad Twitter Dengan Metode Naive Bayes Clasifier Dan Support Vector Machine. *Jurnal Dinamika Informatika*, 4-5.
- Fadholi Fat Haranto, B. W. (2019). Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet. *15*, 171-175.
- I Made Astra, U. D. (2012). Aplikasi Mobile Learning Fisika Dengan Menggunakan Adobe Flash. *Jurnal Pendidikan Dan Kebudayaan*, *18*, 1-3.

- Ramadhan Wp, A. N. (2017). Analisis Sentimen Menggunakan Support Vector Machine Dan Maximum Entropy. *E-Proceeding Of Engineering*, 4, 2389-2395.
- Resky Rayvano Moningka, D. B. (2018). Dentifikasi Kebutuhan Dasar Di Tempat Evakuasi Sementara Pasca Erupsi Merapi Dengan Sentiment Analisis Dan Support Vector Machine. *Telematika*, 15, 77-86.
- Syahfitri Kartika Lidya, O. S. (2015). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (Svm) Dan K-Nearest Neighbor (K-Nn). 1-8.
- Tri Listyorini, A. W. (2013). Perancangan Mobile Learning Mata Kuliah Sistem Operasi Berbasis Android. *Simetris*, 1-2.