

Klasifikasi Kepuasan Mahasiswa Menggunakan Algoritma Support Vector Machine dan Metode Stemming Sastrawi

Ni Luh Ratniasih¹, Ni Wayan Ninik Jayanti², Ni Made Riska Aprilia³

Program Studi Sistem Informasi
Institut Teknologi dan Bisnis STIKOM Bali
Bali, Indonesia

e-mail: ¹ratni@stikom-bali.c.id, ²ninikjayanti01@gmail.com, ³maderiska23@gmail.com

Abstrak

Komentar hasil kepuasan mahasiswa dapat menghasilkan data baru dengan cara klasifikasi data menggunakan text mining. Hasil klasifikasi kepuasan mahasiswa diharapkan dapat membantu pengambilan kebijakan Perguruan Tinggi. Tujuan penelitian ini adalah memperoleh perbandingan hasil klasifikasi menggunakan algoritma Support Vector Machine, kemudian membandingkan hasil klasifikasi tersebut dengan menambahkan metode Stemming Sastrawi. Metode penelitian yang digunakan adalah Metode penelitian dalam penelitian ini terdiri dari beberapa tahap, mulai dari identifikasi masalah dan studi pustaka, kemudian pengumpulan data kepuasan pemangku kepentingan (Mahasiswa), kemudian preprocessing data, selanjutnya ekstraksi fitur agar dapat mempermudah klasifikasi menggunakan algoritma algoritma Support Vector Machine serta metode stemming sastrawi. Hasil penelitian menunjukkan hasil klasifikasi dengan algoritma Support Vector Machine menghasilkan klasifikasi yaitu komentar positif sejumlah 1648 dan negatif sejumlah 126. Kemudian hasil klasifikasi dengan menambahkan metode Stemming Sastrawi menghasilkan klasifikasi yaitu komentar positif sejumlah 1570 dan negatif sejumlah 204.

Kata kunci: Kepuasan Mahasiswa, Support Vector Machine, Metode Stemming Sastrawi

Abstract

Comments on the results of student satisfaction can produce new data by classifying data using text mining. It is hoped that the results of the classification of student satisfaction can assist in making higher education policies. The purpose of this study is to obtain a comparison of the classification results using the Support Vector Machine algorithm, then to compare the classification results by adding the Literary Stemming method. The research method used is the research method in this study consists of several stages, starting from problem identification and literature study, then collecting stakeholder satisfaction data (students), then data preprocessing, then feature extraction in order to facilitate classification using the Support Vector Machine algorithm. as well as literary stemming methods. The results showed that the results of the classification using the Support Vector Machine algorithm resulted in a classification of 1648 positive comments and 126 negative comments. Then the classification results by adding the Literary Stemming method resulted in a classification of 1570 positive comments and 204 negative comments.

Keywords: Student Satisfaction, Support Vector Machine, Literary Stemming Method.

1. Pendahuluan

Kepuasan pelanggan merupakan perasaan senang atau kecewa seseorang yang berasal dari perbandingan antara kesannya terhadap kinerja (hasil) suatu produk dan harapan – harapannya. Kepuasan mahasiswa dapat diartikan sikap positif mahasiswa terhadap pelayanan lembaga pendidikan tinggi karena adanya kesesuaian antara harapan dari pelayanan dibandingkan dengan kenyataan yang diterimanya. Pengukuran tingkat kepuasan mahasiswa merupakan hal yang wajib dilakukan oleh perguruan tinggi baik perguruan tinggi swasta maupun perguruan tinggi negeri. Hal ini dilakukan untuk mendapatkan umpan balik dan masukan bagi keperluan pengembangan dan implementasi strategi perguruan tinggi. Salah satu alat penguran yang digunakan untuk mengukur tingkat kepuasan mahasiswa adalah kuesioner.

Text mining merupakan proses mendeteksi informasi atau sesuatu yang baru dan meneliti informasi besar. *Text mining* menganalisis text yang tidak terstruktur yang terkait satu sama lain dan yang berhubungan dengan prinsip dan aturan lain. Hasil yang diharapkan adalah pemahaman baru yang tidak

diketahui dan belum jelas sebelumnya. Proses perekapan hasil kuesioner yang berupa pendapat dilakukan dengan membaca satu persatu angket kuisisioner, karena kuisisioner yang berupa pendapat tidak memiliki batasan nilai pasti untuk masukan yang diberikan. Pengklasifikasian secara manual menghasilkan data yang akurat karena manusia dapat melakukan klasifikasi dengan tepat kalimat pada opini sesuai objek kata pada opini yang dimaksud. Melakukan penilaian atau analisis satu per-satu terhadap isian angket terbuka secara manual berdampak pada biaya (*cost*) yang tinggi, seperti waktu dan tenaga yang diperlukan dimana biaya ini berbanding lurus dengan jumlah kuisisioner yang dinilai.

Berdasarkan studi literatur terdapat penelitian sebelumnya yang dilakukan oleh Retno Sari tentang Analisis Sentimen dapat diketahui bahwa peneliti menggunakan metode KNN, namun pada penelitian tersebut tidak menjelaskan *tools* yang digunakan serta menggunakan satu metode saja [1]. Selanjutnya pada penelitian yang dilakukan oleh Ghulam Asrofi Buntoro tentang Analisis Sentimen dapat diketahui bahwa peneliti menggunakan metode SVM, namun tidak menjelaskan *tools* dalam analisisnya [2]. Penelitian analisis sentimen yang menggunakan algoritma Naïve Bayes sudah banyak dilakukan sebelumnya. Penelitian yang dilakukan oleh Samsir., dkk dalam analisis sentimen opini masyarakat tentang pembelajaran daring di masa pandemi COVID-19 menghasilkan 30% sentiment positif, 69% sentimen negatif, dan 1% netral [3]. Penelitian yang dilakukan oleh Safitri Juanita terkait dengan persepsi masyarakat terhadap pemilu 2019 pada media sosial twitter menyimpulkan persepsi negatif sebesar 52% jauh lebih besar dari persepsi positif 18% dan persepsi netral 31% lebih tinggi dari persepsi positif [4]. Berdasarkan penelitian sebelumnya dapat disimpulkan bahwa belum terdapat penelitian terkait kepuasan pemangku kepentingan (mahasiswa) menggunakan algoritma *Support Vector Machine* serta *Metode Stemming Sastrawi*.

2. Metode Penelitian

Metode penelitian terdiri dari beberapa tahap diantaranya tahap pertama dilakukan identifikasi masalah dan studi pustaka, tahap kedua pengumpulan data kepuasan pemangku kepentingan (Mahasiswa), tahap ketiga preprocessing data, tahap keempat adalah implementasi metode stemming sastrawi. Tahap kelima merupakan tahap pengujian dan evaluasi model. Tahap keenam adalah pengujian tingkat akurasi metode. Pada Gambar 1 menunjukkan tahapan atau alur penelitian ini.

Pada tahap identifikasi masalah dilakukan identifikasi tentang kepuasan pemangku kepentingan (Mahasiswa). Kemudian mencari teori-teori terkait *text mining* dan analisis sentimen serta mempelajari pustaka yang relevan terhadap masalah penelitian sejenis yang pernah dilakukan sebelumnya yang digunakan untuk menunjang penelitian ini.

Pengumpulan data diperoleh dari Direktorat Penjaminan Mutu Perguruan tinggi yang disebarakan setiap semester. Data yang digunakan dalam penelitian ini adalah hasil penyebaran kuesioner pada semester Ganjil 2021/2022 pada seluruh mahasiswa aktif. Jumlah pengisian kuesioner pada semester tersebut adalah 2558 responden. Pada Gambar 2 terdapat data kepuasan pemangku kepentingan (Mahasiswa), dimana pada kolom CL terdapat komentar para responden.

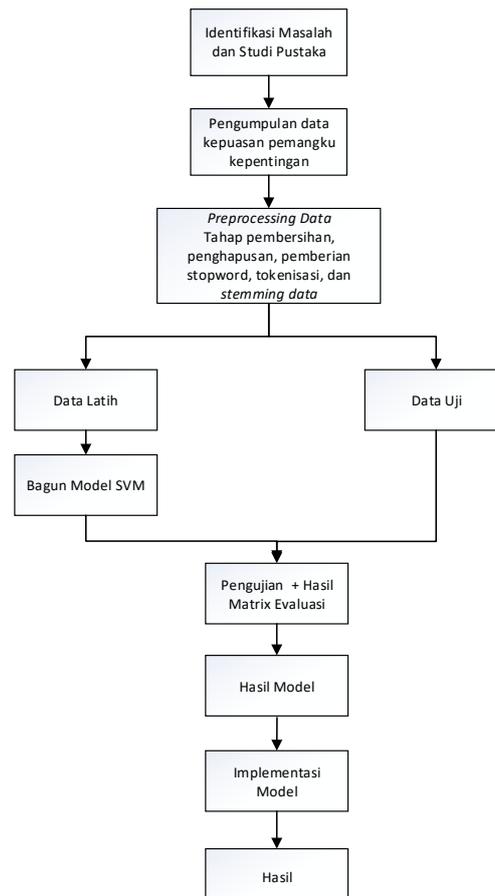
BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	
2	3	2	3	3	3	3	3	2	3	3	3	4	2	3	3	2	3	3	3	Kemudahan aksesibilitas fasilitas sarana dan prasarana yang mendukung proses belajar mengajar.
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Sudah memiliki jabatan fungsional yang dapat digunakan untuk perguruan tinggi
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	Pengelolaan administrasi untuk menunjang tridharma secara umum sudah baik
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Pelayanan terhadap dosen terkait tridharma perguruan tinggi sudah baik.
3	3	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Kegiatan belajar mengajar
3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	Pelayanan akademik
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	Sistem E-learning, SID & SINAK semakin baik dan lengkap
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	Sangat Baik
2	2	2	3	2	3	2	3	3	2	2	2	3	2	3	2	2	2	2	2	jempol untuk pusat pengembangan pembelajaran digital yang telah membuktikan kinerjanya saat pandemi. perlu ditingkatkan lagi model pembelajaran dan sistem manajemen pembelajaran.

Gambar 1. Hasil Kepuasan Mahasiswa

Tahap *preprocessing data* merupakan tahap seleksi data dan mengubahnya terstruktur. Terdapat beberapa proses dalam preprocessing data diataranya tahap Cleansing. Kalimat yang didapat biasanya masih terdapat noise, yaitu kesalahan acak atau varian dalam variable terukur kita harus menghilangkan noise tersebut. Tahap berikutnya *parsing* yaitu proses memecah dokumen menjadi sebuah kata dengan

melakukan analisa terhadap kumpulan kata dengan memisahkan kata tersebut dan menentukan struktur sintaksis dari tiap kata tersebut. Selanjutnya tahap normalisasi kalimat yang bertujuan untuk menormalkan kalimat sehingga kalimat gaul menjadi normal sehingga bahasa gaul tersebut dapat dikenali sebagai bahasa yang sesuai dengan KBBI. Proses *tokenization* digunakan dalam identifikasi kata – kata dan memecah kalimat menjadi istilah berdasarkan spasi dan tanda baca. Tahap terakhir dalam preprocessing yaitu *stemming* merubah kata imbuhan menjadi kata dasar [5].

Tahap selanjutnya, dari *hasil preprocessing data* dilakukan pembagian data menjadi data latih dan data uji. Tahap selanjutnya membangun model menggunakan algoritma SVM dan melakukan pengujian terhadap model yang sudah dibangun sehingga menghasilkan Matrix Evaluasi agar dapat imlementasi metode *stemming sastrawi*. Tahapan dalam metode penelitian digambarkan dalam bentuk bagan Gambar 2.



Gambar 2. Metode Penelitian

3. Hasil dan Pembahasan

Penelitian sebelumnya dengan pembahasan analisis sentimen terhadap layanan Indihome berdasarkan Twitter dengan metode klasifikasi *Support Vector Machine* (SVM) dengan tujuan untuk mendapatkan model klasifikasi sentimen menggunakan SVM, dan untuk mengetahui seberapa besar akurasi yang dihasilkan oleh metode SVM yang diterapkan pada analisis sentimen, serta untuk mengetahui seberapa puas pengguna layanan Indihome berdasarkan Twitter [6]. Berdasarkan penelitian tersebut klasifikasi sentiment pada penelitian ini dilakukan berdasarkan hasil Preprocessing Data dan pembangunan model dengan algoritma *Support Vector Machine* (SVM).

3.1 Preprocessing Data

Data diperoleh dalam format excel data hasil pengukuran kepuasan pemangku kepentingan (Mahasiswa) terdiri dari NIM, nilai harapan dan aktual, serta komentar. Pada penelitian ini yang digunakan sebagai data penelitian hanya NIM dan komentar mahasiswa seperti yang terlihat pada Tabel 1.

Tabel 1. Data Komentar Pemangku Kepentingan (Mahasiswa)

NO	NIM	KOMENTAR
1	140010021	Bagus
2	140010045	Selama ini kebetulan belum ada pengalaman mengecewakan yang saya alami. terima kasih
3	140010092	top
4	140010136	Mantap
5	140010139	Sudah Bagus
6	140010211	Tidak ada
7	140010221	Mantap
8	140010229	Tidak ada
9	140010287	semua sudah baik
10	140010306	Pemberian informasi perlu ditingkatkan

Proses *preprocessing data* dilakukan untuk menghilangkan noise dan memperjelas fitur serta membersihkan dengan menghapus atribut – atribut yang tidak diperlukan. Proses ini akan mengolah sebuah teks menjadi data yang mudah diterima oleh sistem saat proses utama dilakukan [6]. Tahap *preprocessing* diperlukan sebelum melakukan implementasi algoritma, dimana terdapat beberapa tahapan proses yang dilakukan antara lain :

- a. Import Data : mengambil data yang akan diproses dari tempat penyimpanan sehingga data dapat dibaca pada local Rapid Miner. Pada tahap ini akan terlihat data secara keseluruhan dan missing data yang perlu preprocessing. Pada proses *import data* dilakukan dengan mengambil data dari file excel namun kolom yang digunakan hanya kolom komentar.
- b. Replace Data : merupakan operator ini digunakan untuk menghilangkan beberapa karakter yang ada pada komentar. Karakter atau simbol yang dihilangkan seperti tanda titik (.), tanda seru (!), tanda tanya (?), dan lain – lain.
- c. Filter Examples : merupakan operator yang digunakan untuk menghilangkan baris – baris data yang kosong yang ditandai dengan simbol tanda tanya (?) (*missing attribute*).
- d. Remove Duplicates : merupakan operator yang digunakan untuk menghapus beberapa komentar yang sama untuk mempermudah proses analisis sentimen dan mengurangi waktu pelabelan komentar. Jumlah data komentar setelah diproses dengan operator ini adalah 2333 komentar. Tahap ini menghasilkan data yang sudah siap untuk diproses.
- e. Write CSV : merupakan operator untuk menyimpan data hasil terakhir dengan format data *.CSV.

5.2 Membangun Model

Data set yang dihasilkan dari tahap *preprocessing data* akan diberikan label secara manual dengan dua kategori yaitu “Positif” dan “Negatif”. Pada pembangunan model dilakukan tahap pembangunan model klasifikasi menggunakan algoritma SVM, dimana dalam proses pembangunan model dimulai dari membagi dataset, hingga mendapatkan model terbaik pada penggunaan algoritma SVM.

5.3 Pembagian Data

Pada tahap ini dilakukan pembagian data menjadi 2 kelompok yaitu data latih, dan data uji dengan ratio 80:20. Dimana data awal sebanyak 2558 komentar, setelah proses *preprocessing data* dan dihilangkan nilai yang kosong kemudian dilakukan sentimen menggunakan algoritma SVM jumlah data menjadi 1774 komentar.

5.4 Melakukan Implementasi Model SVM

Dilanjutkan dengan implementasi penggunaan *Support Vector Machine* (SVM) untuk melakukan klasifikasi pada data teks dengan menggunakan representasi TF-IDF sebagai pembobotan fitur. Dimana tahap proses implementasi adalah sebagai berikut:

1. Proses implementasi SVM dilakukan dengan kernel linear.
2. Dilanjutkan dengan melakukan pelatihan data latih pada implementasi *Support Vector Machine* (SVM) yang telah dibuat.
3. Proses dilakukan dengan memprediksi label pada data uji.
4. Hasil prediksi dan akurasi dari model tersebut akan dibandingkan dengan label asli pada data uji.

5.5 Matrix Evaluasi

Pada tahap ini dilakukan proses evaluasi model *Support Vector Machine* (SVM) menggunakan *confusion matrix* dan *classification report*. *Confusion matrix* digunakan untuk menghitung jumlah prediksi yang benar atau salah dari model, sedangkan *classification report* memberikan informasi mengenai *precision*, *recall*, dan *f1-score* dari model untuk setiap kelas yang ada pada data uji. Hasil *Confusion matrix* dapat dilihat pada Gambar 3. Hasil tingkat akurasi implementasi metode *Support Vector Machine* (SVM) sebesar 95%.

	Algoritma	Akurasi	Precision	Recall	F1-score
0	SVM (Support Vector Machine)	0.95	0.96	0.95	0.951971

Gambar 3. Hasil *Confusion matrix*

5.6 Hasil Klasifikasi dengan Algoritma *Support Vector Machine* (SVM) dan Metode *Stemming Sastrawi*

Hasil implementasi model menggunakan metode *Support Vector Machine* (SVM) akan digunakan untuk memperoleh hasil perbandingan penerapan metode *Stemming Sastrawi*. Hasil implementasi tanpa menggunakan metode *Stemming Sastrawi* menunjukkan hasil komentar positif sebanyak 1648 dan komentar negatif sebanyak 126. Kemudian hasil implementasi menggunakan metode *Stemming Sastrawi* menunjukkan hasil komentar positif sebanyak 1570 dan komentar negatif sebanyak 204.

```
df_clean_valuecount
✓ 0.0s
Positif    1648
Negatif    126
Name: sentimen, dtype: int64
```

Gambar 4. Tanpa *Stemming Sastrawi*

```
df_clean_valuecount
Positif    1570
Negatif    204
Name: sentimen, dtype: int64
```

Gambar 5. *Stemming Sastrawi*

4. Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa :

- Hasil klasifikasi sentiment dengan algoritma *Support Vector Machine* menghasilkan klasifikasi yaitu komentar positif sejumlah 1648 dan negatif sejumlah 126. Kemudian hasil klasifikasi sentiment dengan menambahkan metode *Stemming Sastrawi* menghasilkan klasifikasi yaitu komentar positif sejumlah 1570 dan negatif sejumlah 204.
- Berdasarkan hasil klasifikasi tersebut dapat ditarik kesimpulan bahwa, mahasiswa lebih banyak berkomentar positif dibandingkan komentar negatif.

Daftar Pustaka

- Sari, R. (2020). Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma KNearest Neighbor (K-Nn). *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1), 10–17. <https://doi.org/10.31294/evolusi.v8i1.7371>.
- Buntoro, Ghulam Asrofi. 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal*. Vol. 2. No. 1. Hal. 32-41.
- Samsir, A., Verawardina, U., Edi, F., Watrianthos, R. 2021. Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*. Vol. 5, No. 1. Page 157-163.
- Juanita, S. 2020. Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 3, p. 552.

-
- [5] Rasenda, R., H. Lubis, and R. Ridwan 2020. "Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter," J. MEDIA Inform. BUDIDARMA, doi: 10.30865/mib.v4i2.2051.
- [6] Rian Tineges. 2020. "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)". Jurnal Media Informatika Budidarma. Vol. 4. No. 3. Page 650-658.
- [7] Santosa B. 2007. Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi. Graha Ilmu Yogyakarta.
- [8] Gunawan, Billy, Helen Sasty Pratiwi, Enda Esyudha Pratama. 2018. Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. Jurnal Edukasi dan Penelitian Informatika. Vol. 4. No. 2. Hal. 113-118.
- [9] Hidayah, N., Hutagalung, S. S., & Hermawan, D. 2019. Analisis peran stakeholder dalam pengembangan wisata talang air peninggalan kolonial Belanda di Kelurahan Pajaresuk Kabupaten Pringsewu. Jurnal Ilmu Administrasi Publik. Vol. 7. No. 1. Hal. 55-71. doi:10.31289/publika.v7i1.2179.
- [10] Han, J., Kamber, M. & Pei, J. 2012. Data Mining Concepts and Techniques. 3rd ed. Waltham: Morgan Kaufmann Publishers.
- [11] Juanita, S. 2020. Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. J. MEDIA Inform. BUDIDARMA, vol. 4, no. 3, p. 552.
- [12] Junaedi, Hartanto., Herman, Budianto. 2011. Data Transformation Pada Data Mining. Prosiding Konferensi nasional "Inovasi dalam Desain dan Teknologi". IDEaTech 2011.