

Implementasi Perbandingan Metode Algoritma *Support Vector Machine* (SVM) DAN *Naïve Bayes* Untuk Menganalisa Pendapat Masyarakat Terkait *Cyberbullying* Diera Teknologi Digital Pada Aplikasi X

Ibnu Awaludin¹, Hadi Zakaria^{2*}

^{1,2}Teknik Informatika, Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ¹awaludinibnu@gmail.com, ^{2*}dosen00274@gmail.com

(* : coressponding author)

Abstrak – Di era teknologi digital, kemajuan teknologi telah mengubah cara orang berinteraksi, tetapi juga meningkatkan kasus cyberbullying, terutama di kalangan anak-anak dan remaja pengguna aplikasi X. Pelaku cyberbullying merasa aman dari konsekuensi langsung karena anonimitas internet. Yayasan Praktek Psikolog Indonesia (YAPI) prihatin terhadap masalah ini dan berencana mengembangkan aplikasi untuk menganalisis opini masyarakat tentang cyberbullying di media sosial X. Penelitian ini bertujuan untuk menganalisis pendapat masyarakat menggunakan Algoritma Support Vector Machine (SVM) dan Naïve Bayes, serta mengidentifikasi metode yang paling efisien dan akurat dalam hal kecepatan dan akurasi. Data dari platform X akan diambil menggunakan Python, diolah melalui preprocessing, dan disimpan di database MySQL. Penelitian ini diharapkan dapat menentukan metode yang lebih unggul antara SVM dan Naïve Bayes, serta memberikan pemahaman yang lebih baik tentang teknik preprocessing dalam analisis opini di media sosial, sehingga meningkatkan kualitas dan relevansi hasil analisis.

Kata Kunci: *Support Vector Machine*, Naïve Bayes, *Cyberbullying*, Teknologi Digital.

Abstract– *In the era of digital technology, technological advances have changed the way people interact, but have also increased cases of cyberbullying, especially among children and adolescents who use the X application. Cyberbullying perpetrators feel safe from direct consequences because of the anonymity of the internet. The Indonesian Psychological Practice Foundation (YAPI) is concerned about this problem and plans to develop an application to analyze public opinion about cyberbullying on social media X. This study aims to analyze public opinion using the Support Vector Machine (SVM) and Naïve Bayes Algorithms, and identify the most efficient and accurate methods in terms of speed and accuracy. Data from the X platform will be taken using Python, processed through preprocessing, and stored in a MySQL database. This study is expected to determine the superior method between SVM and Naïve Bayes, as well as provide a better understanding of preprocessing techniques in analyzing opinions on social media, thereby improving the quality and relevance of the analysis results.*

Keywords: *Support Vector Machine*, Naïve Bayes, *Cyberbullying*, *Digital Technology*.

1. PENDAHULUAN

Cyberbullying meningkat seiring kemajuan teknologi. Pengguna aplikasi X, terutama anak-anak dan remaja, sering menjadi korban intimidasi, pelecehan, dan penghinaan yang dilakukan melalui internet. Beberapa orang mengatakan bahwa kemajuan teknologi dan penggunaan aplikasi media sosial telah membuat cyberbullying menjadi lebih mudah. Internet membuat pelaku merasa aman tanpa khawatir akan akibatnya. Komentar ini sering dikirim melalui berbagai platform, salah satunya adalah aplikasi X, yang banyak digunakan masyarakat untuk menyampaikan pendapat mereka. Untuk mencegah kasus cyberbullying di era teknologi modern, sangat penting untuk menganalisis pendapat masyarakat ini. Untuk melakukan analisis ini, diperlukan teknik yang efektif untuk mengkategorikan opini saat ini.

Media sosial X, atau yang dulu dikenal sebagai Twitter, adalah salah satu sumber komunikasi global yang paling diminati. Hal ini dapat dilihat dari jumlah peningkatan pengguna X yang tercatat di seluruh dunia, salah satunya di Indonesia. Jumlah pengguna X di Indonesia mencapai 24 juta pada tahun 2023, menurut laporan Digital Hootsuite (We Are Social): Indonesian Report 2023 (Aulia, 2024). Pengguna X akan memberikan komentar atau kabar terbaru tentang hal-hal yang sedang menjadi topik utama di dunia. Topik-topik yang sering dikomentari oleh pengguna akan menimbulkan masalah atau topik trending di media sosial, terutama di X. (Supian et al., 2024)

Salah satu algoritma pembelajaran mesin yang paling banyak digunakan untuk tugas regresi dan klasifikasi adalah Support Vector Machine (SVM). SVM, yang dikembangkan oleh Vladimir Vapnik dan rekannya, mampu menangani data dengan dimensi tinggi dan masalah non-linear melalui penggunaan kernel. Ketika digunakan dalam analisis data ulasan, SVM dapat menjadi alat yang sangat efektif untuk mengklasifikasikan sentimen atau kategori lain dari teks yang singkat dan seringkali tidak terstruktur. Hyperplane optimal dipilih oleh SVM untuk memisahkan kelas-kelas dalam data pelatihan dengan margin maksimal. Batas keputusan yang memisahkan set data yang berbeda dalam ruang fitur disebut hyperplane.

Metode Naive Bayes adalah metode statistik yang mengasumsikan independensi antara fitur-fitur dalam data dan menghitung kemungkinan untuk setiap kelas. Metode Naive Bayes diciptakan dengan menggunakan metode probabilistik dan statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes. Dengan kata lain, metode Naive Bayes menghitung kemungkinan bahwa setiap kelas keputusan benar jika di bawah kondisi tertentu. Menurut algoritma ini, atribut objek adalah independen. Probabilitas perkiraan akhir dihitung sebagai jumlah frekuensi dari tabel keputusan "master". Metode Naive Bayes digunakan untuk menghitung nilai akurasi, recall, dan ketepatan klasifikasi data. (Aponno, 2022).

Dengan menggunakan Algoritma Support Vector Machine (SVM) dan Naive Bayes, penelitian ini bertujuan untuk membandingkan kinerja kedua algoritma tersebut dalam mengklasifikasikan pendapat masyarakat terkait maraknya kasus cyberbullying di era teknologi digital. Dengan menggunakan dataset opini yang dikumpulkan dari aplikasi X, penelitian ini akan mengevaluasi seberapa baik SVM dan Naive Bayes dalam mengidentifikasi sentimen positif, negatif, dan netral dari masyarakat. Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang lebih baik mengenai metode mana yang lebih efektif untuk digunakan dalam analisis sentimen pada kasus ini, serta memberikan rekomendasi bagi pemerintah dan pihak terkait dalam menanggapi opini masyarakat secara lebih bijak dan berdasarkan data yang akurat.

Pada penelitian ini penulis akan merancang sebuah penelitian dengan judul, "IMPLEMENTASI PERBANDINGAN METODE ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN NAIVE BAYES UNTUK MENGANALISA PENDAPAT MASYARAKAT TERKAIT CYBERBULLYING DI ERA TEKNOLOGI DIGITAL PADA APLIKASI X (STUDI KASUS: YAYASAN PRAKTEK PSIKOLOG INDONESIA CABANG TANGERANG SELATAN)".

2. METODE

2.1 Metode Pengumpulan Data

Dalam metode penelitian untuk mendapatkan data dan informasi maka metode yang digunakan dalam proses pengumpulan data adalah sebagai berikut:

a. Pengamatan Langsung (Metode Observasi)

Dalam hal ini peneliti mengadakan penelitian langsung ke objek yang akan diteliti dengan mengadakan pengamatan terhadap sarana pemanfaatan teknologi yang dibutuhkan. Cara ini dilakukan dengan cara melakukan pemantauan langsung ke tempat objek yang ingin di teliti supaya penulis dapat mengetahui permasalahan yang ada.

b. Wawancara (Interview)

Pada langkah ini akan dilakukan analisis dan pengumpulan data yang dibutuhkan dalam pembangunan system, agar system yang dibangun sesuai dengan kebutuhan pengguna. Hal tersebut dilakukan dengan melakukan wawancara secara langsung untuk mengetahui bagaimana pelayanan yang dilakukan dan melakukan survey untuk melihat aspek kebutuhan.

c. Studi Pustaka (Studi Literature)

Peneliti melakukan studi pustaka karena dalam penelitian ini tidak terlepas dari buku-buku, jurnal, catatan, serta skripsi yang menjadi referensi untuk membantu melengkapi data-



data yang telah di dapat dan juga membantu dalam perancangan aplikasi tersebut. Cara ini dilakukan dengan cara merangkum informasi dari penelitian terdahulu yang dibutuhkan oleh penulis dalam memudahkan penelitian yang sedang dilakukan

2.2 Metode Pengembangan

Untuk pengembangan sistem, penulis menggunakan metode pengembangan perangkat lunak model air terjun (waterfall). Tahapan tahapan dalam pengembangan sistem aplikasi dengan metode waterfall adalah sebagai berikut :

a. Analisis Kebutuhan Perangkat Lunak

Proses pengumpulan kebutuhan dilakukan secara intensif untuk menspesifikasikan kebutuhan perangkat lunak agar dapat dipahami seperti apa yang dibutuhkan oleh user. Spesifikasi kebutuhan perangkat lunak pada tahap ini perlu di dokumentasikan.

b. Desain

Desain perangkat lunak adalah proses multi langkah yang fokus pada desain pembuatan program perangkat lunak termasuk struktur data, arsitektur perangkat lunak, representasi antar muka dan prosedur pengkodean. Tahap ini mentranslasi kebutuhan perangkat dari tahap analisis kebutuhan ke representasi desain agar dapat diimplementasikan program pada tahap selanjutnya. Desain perangkat lunak yang dihasilkan pada tahap ini juga perlu di dokumentasikan

c. Pembuatan Kode Program

Desain harus ditranslasikan kedalam program perangkat lunak. Hasil dari tahap ini adalah program komputer sesuai dengan desain yang telah dibuat pada tahap desain.

d. Pengujian

Pengujian fokus pada perangkat lunak dari segi logik dan fungsional serta memastikan bahwa semua bagian sudah diuji. Hal ini dilakukan untuk meminimalisir kesalahan (error) dan memastikan keluaran yang dihasilkan sesuai dengan yang diinginkan.

e. Pendukung (*support*) atau Pemeliharaan (*Maintenance*)

Tidak menutup kemungkinan sebuah perangkat lunak mengalami perubahan ketika sudah dikirimkan ke user. Perubahan bisa terjadi karena adanya kesalahan yang muncul dan tidak terdeteksi saat pengujian atau perangkat lunak harus beradaptasi dengan lingkungan baru. Tahap pendukung atau pemeliharaan dapat mengulangi proses pengembangan mulai dari analisis spesifikasi untuk perubahan perangkat lunak yang sudah ada, tapi tidak untuk membuat perangkat lunak baru

2.3 Algoritma Support Vector Machine (SVM)

SVM adalah teknik pembelajaran terawasi yang membuat pemetaan fitur input-output dari sekumpulan data pelatihan. Fungsi pemetaan ini dapat berupa fungsi klasifikasi atau regresi. SVM menggunakan ruang hipotesis berupa fungsi linier pada ruang fitur. SVM dilatih menggunakan algoritma pembelajaran berdasarkan teori optimasi dengan menerapkan bias pembelajaran yang berasal dari teori pembelajaran statistik. Meskipun teori SVM telah dikembangkan sejak tahun 1960an, teori ini diperkenalkan kembali oleh Vapnik, Boser, dan Guyon pada tahun 1992, dan SVM telah berkembang pesat sejak saat itu.

Teks	kurang	perhatian	keluarga	ini	ajar	kawal	terus	kelakuannya	kayak	tai	yang	sabar	pak	vincent	semangat	jadikan	pembelajaran
kurang perhatian keluarga ini	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
kurang ajar, kawal terus	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
kelakuannya kayak tai	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
yang sabar pak vincent	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0
semangat pak vincent, jadikan pembelajaran	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1

Gambar 1. Direkctory Kata

teks	sentimen
kurang perhatian keluarga ini	-1
kurang ajar, kawal terus	-1
kelakuannya kayak tai	-1
yang sabar pak vincent	1
semangat pak vincent, jadikan pembelajaran	1

Gambar 2. Pelabelan

2.4 Algoritma Naïve Bayes

Naïve Bayes merupakan teknik klasifikasi berdasarkan Teorema Bayes dengan asumsi independensi di antara para prediktor. Naive Bayes Classifier memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. *Naïve bayes* merupakan metode paling sederhana dari pengklasifikasian probabilitas, memiliki tingkat akurasi yang sangat tinggi ketika diaplikasikan pada database dengan big data. Selain itu model naïve bayes juga memiliki tingkat kesalahan yang sangat minimum dibandingkan dengan algoritma klasifikasi lainnya. Bayes adalah klasifikasi menggunakan metode probabilistik dan statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes Dengan kata lain, memprediksi masa depan berdasarkan data masa lalu Naive Bayes menghitung, untuk setiap kelas keputusan, probabilitas bahwa kelas keputusan itu benar jika diberikan vektor informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen Probabilitas dalam perkiraan akhir dihitung sebagai jumlah frekuensi dari tabel keputusan "master (Ernawati et al., 2023)

Rumus Bayes dalam (Muslehatin dkk, 2017) secara umum adalah sebagai berikut:

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)}$$

dengan :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Adapun aturan Bayes adalah sebagai berikut :

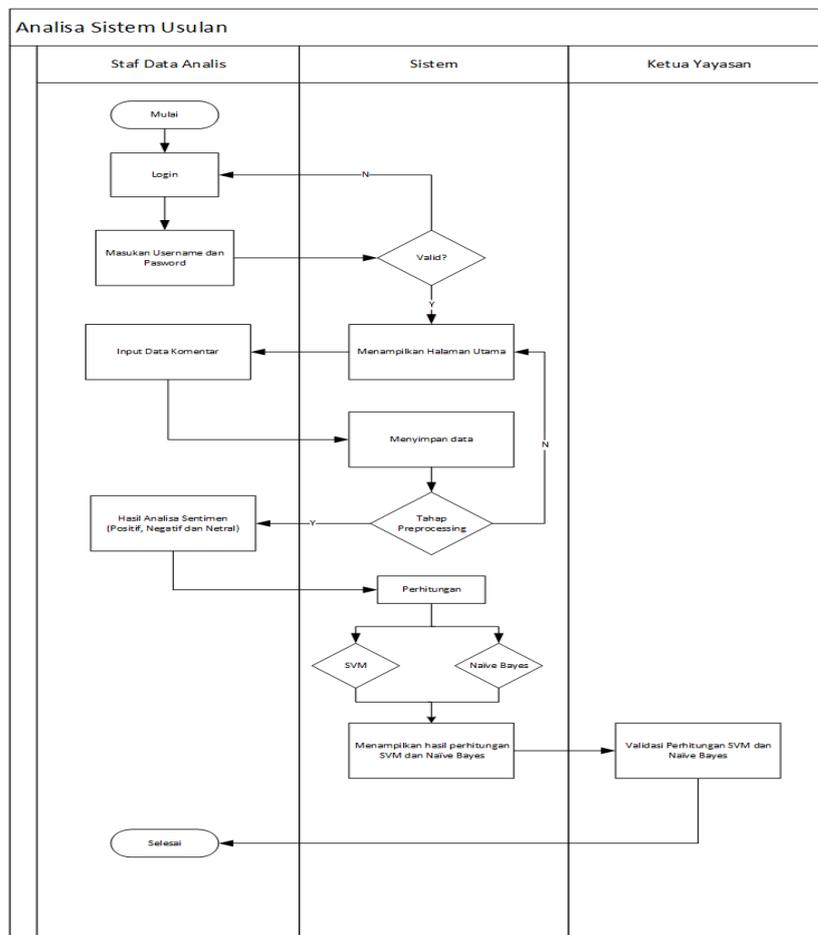
Jika $P(h1|x) < P(h2|x)$, maka x diklasifikasikan sebagai $h2$. Pernyataan $P(h1|x)$ mengindikasikan probabilitas hipotesis $h1$ berdasarkan kondisi x terjadi, begitu pula dengan $h2$. Sehingga didapat klasifikasi dari x sesuai dengan probabilitas terbesar diantara probabilitas x terhadap semua kelas.

3. ANALISA DAN PEMBAHASAN

Analisa sistem informasi merupakan penguraian suatu sistem informasi yang untuk kedalam bagian komponen-komponen dengan maksud untuk mengidentifikasi serta mengevaluasi permasalahan-permasalahan yang ada serta hambatan yang terjadi dan kebutuhan yang diharapkan sehingga dapat diusulkan menjadi perancangan sistem informasi (Puspitasari & Zakaria, 2023).

3.1 Analisa Sistem Usulan

Pada sistem ini diusulkan beberapa hal yang menjadi batasan masalah, yang akan diberikan solusi atau alternatif dengan maksud memperjelaskan tentang kebutuhan-kebutuhan yang harus dipenuhi oleh sistem yang dirancang. Berdasarkan hasil analisa tersebut, maka dibuat suatu kebutuhan dalam perancangan sebagai berikut :

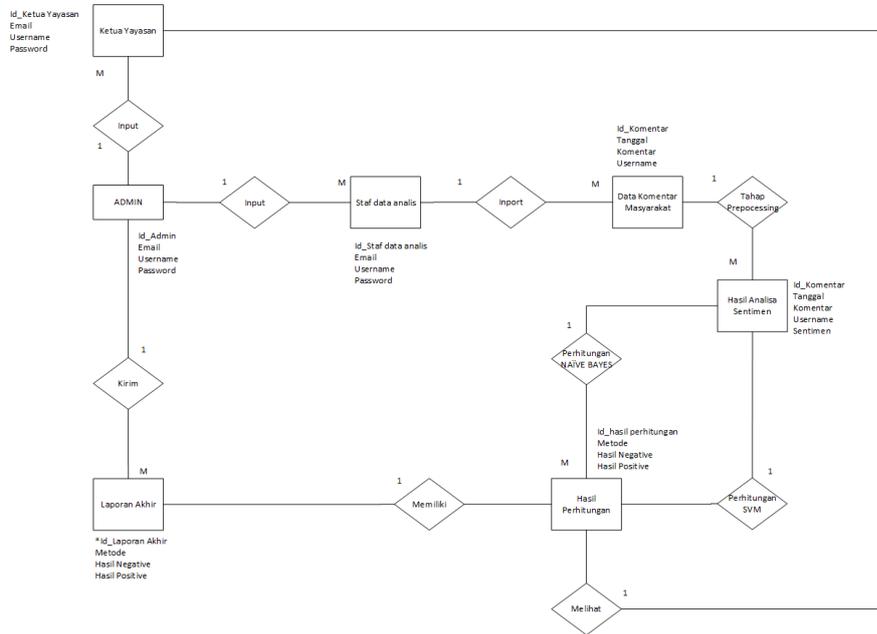


Gambar 3. Sistem Usulan

3.2 Perancangan Basis Data

Perancangan basis data merupakan penggambaran penyusunan data agar mempermudah dalam proses pengambilan keputusan. Rancangan ini mengidentifikasi komponen-komponen system informasi yang dirancang secara rinci.

3.2.1 Entity Relation Diagram (ERD)

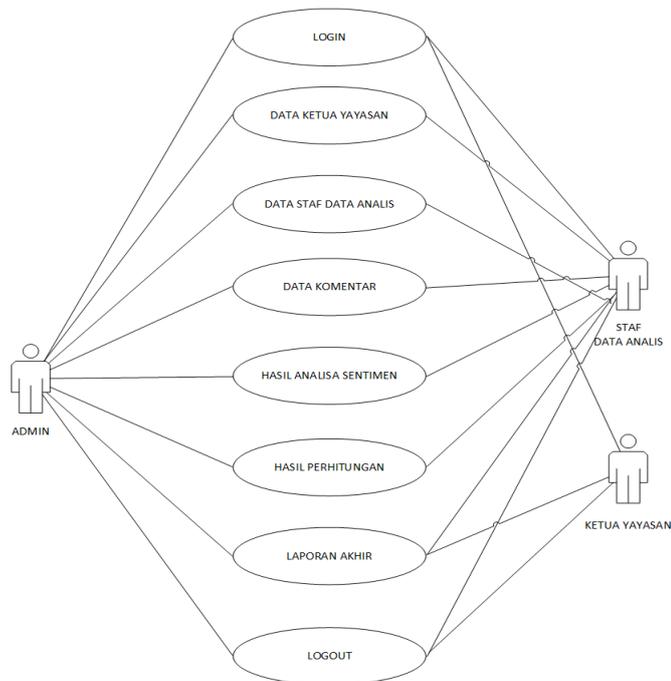


Gambar 4. Entity Relation Diagram (ERD)

3.3 Perancangan Unified Modeling Language (UML)

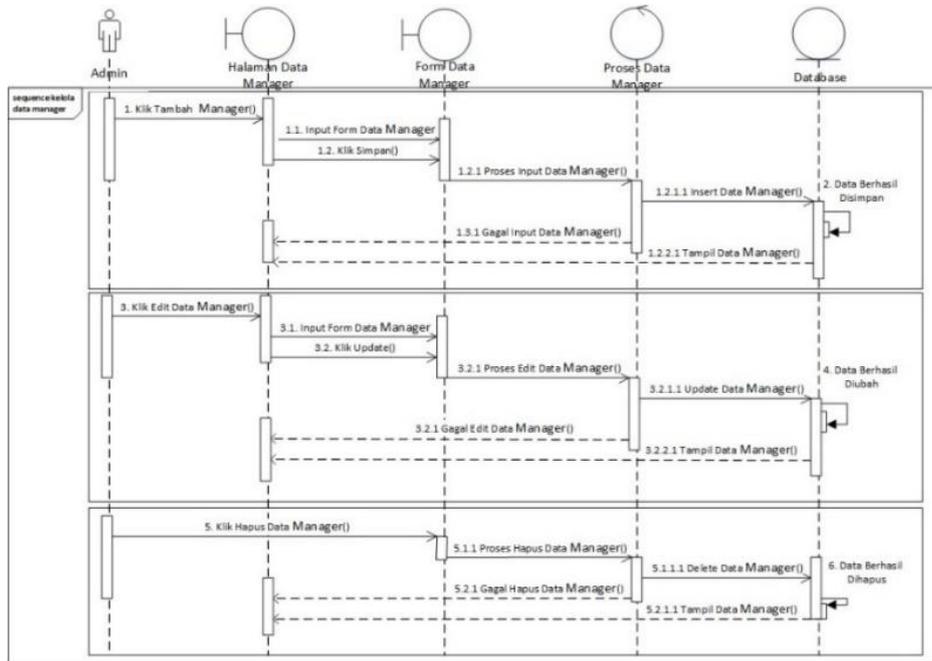
Perancangan Pengembangan perangkat dengan menggunakan metode permodelan secara visual untuk sarana perancangan sistem berorientasi objek, atau definisi UML yaitu sebagai suatu bahasa yang sudah menjadi standar pada visualisasi, perancangan dan juga pendokumentasian sistem software. Berikut perancangan dari aplikasi pelayanan terhadap pelanggan.

3.3.1 Usecase Diagram



Gambar 5. Use Case Diagram

3.3.2 SequenceDiagram



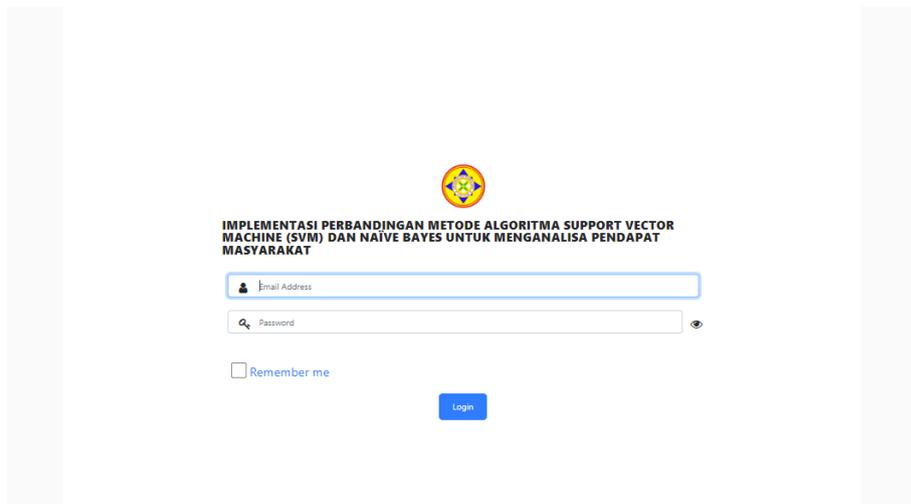
Gambar 6. *Squence Diagram*

4. IMPLEMENTASI

a. Implementasi Antarmuka

Implementasi antarmuka dari sistem aplikasi ini dibuat menggunakan bahasa pemrograman PHP, dalam implementasi antarmuka ini terdapat beberapa halaman menjelaskan proses penggunaan sistem aplikasi ini adalah sebagai berikut:

a. Halaman Login



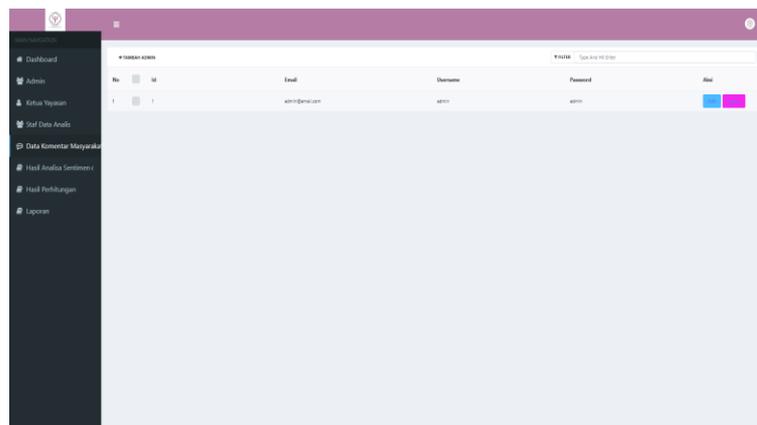
Gambar 5. *Halaman Login*

b. Halaman Dashboard



Gambar 6. Halaman *Dasboard*

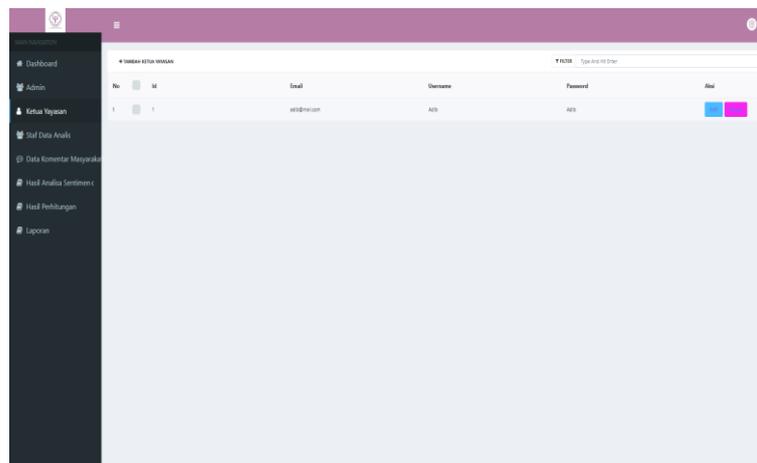
c. Halaman Admin



No	Aktif	Email	Username	Password	Role
1	<input type="checkbox"/>	admin@gmail.com	admin	admin	Admin

Gambar 7. Halaman Admin

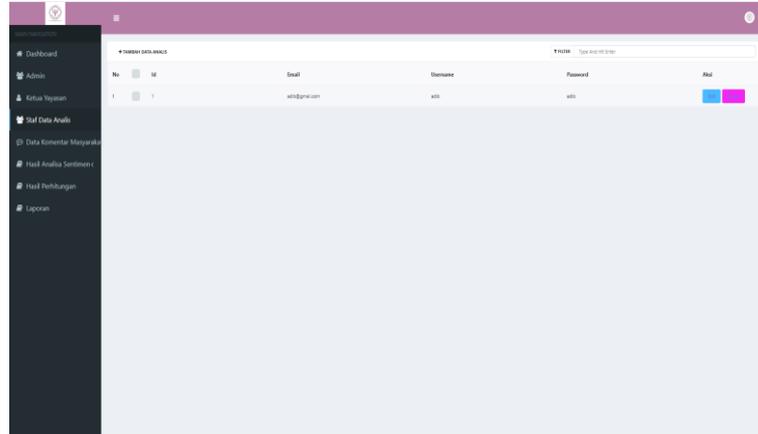
d. Halaman Ketua Yayasan



No	Aktif	Email	Username	Password	Role
1	<input type="checkbox"/>	ketua@yayasan.com	ketua	ketua	Ketua Yayasan

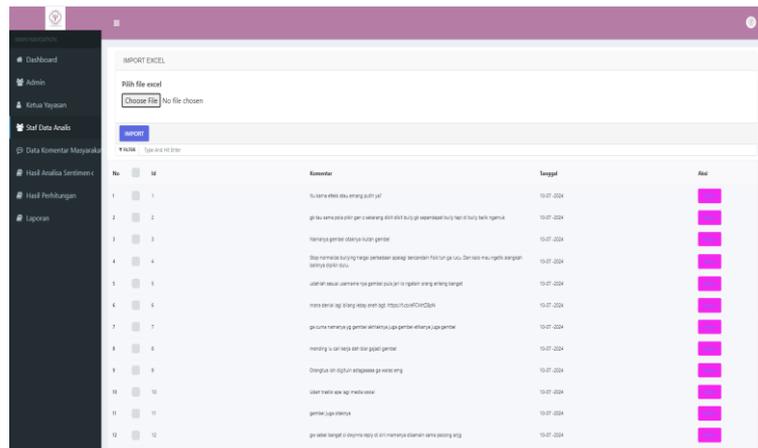
Gambar 8. Manager

e. Halaman Staf Data Analisis



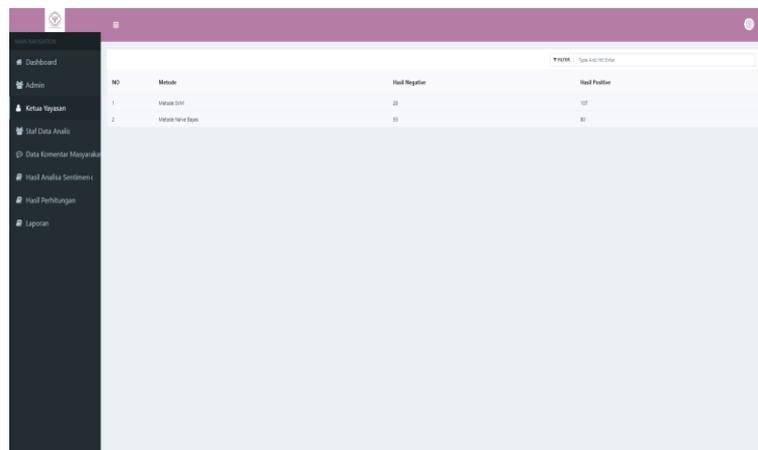
Gambar 9. Staf Data Analisis

f. Halaman Data Komentar



Gambar 10. Halaman Data Komentar

g. Halaman Data Hasil



Gambar 11. Halaman Data Hasil

h. Halaman Laporan Akhir

Laporan Hasil Klasifikasi
Tanggal: 10-Jul-2024

Laporan : Algoritma Naive Bayes & SVM

NO	Metode	Hasil Negative	Hasil Positive
1	Metode SVM	28	107
2	Metode Naive Bayes	55	80

Gambar 12. Halaman Laporan Akhir

5. KESIMPULAN

Hasil dari penelitian ini menunjukkan Algoritma Support Vector Machine (SVM) cenderung menawarkan akurasi yang lebih tinggi dalam analisis opini karena kemampuannya dalam menangani data dengan margin yang jelas dan memisahkan kelas yang berbeda dengan baik, terutama pada dataset yang kompleks dan tidak seimbang. Namun, proses pelatihan SVM sering kali memerlukan waktu yang lebih lama, terutama pada dataset besar, sehingga efisiensi waktu bisa menjadi kendala, di sisi lain, Algoritma Naïve Bayes dikenal lebih cepat dan efisien dalam memproses data, karena asumsi independensi antar fitur yang membuat perhitungannya lebih sederhana. Meski begitu, akurasinya dapat menurun apabila ada korelasi yang kuat antara fitur-fitur dalam dataset, karena asumsi independensi ini seringkali tidak sesuai dengan data nyata. Jika kecepatan analisis menjadi prioritas, terutama dalam skenario di mana respons cepat sangat dibutuhkan, Naïve Bayes bisa menjadi pilihan yang lebih baik meskipun dengan akurasi yang sedikit lebih rendah. Namun, jika tujuan utama adalah mencapai akurasi yang maksimal dalam identifikasi opini, terutama pada kasus yang kompleks dan data yang tidak seimbang, maka SVM lebih unggul, meskipun dengan konsekuensi waktu pemrosesan yang lebih lama.

Secara keseluruhan, penelitian ini mengindikasikan bahwa tidak ada algoritma yang secara absolut lebih unggul dalam semua aspek. Pilihan terbaik bergantung pada kebutuhan spesifik: apakah yang lebih penting adalah akurasi atau kecepatan. Penelitian ini juga menyoroti pentingnya teknik preprocessing yang tepat dalam meningkatkan kualitas analisis opini di media sosial.

REFERENCES

- Ernawati, A., Sari, A. O., Sofyan, S. N., Iqbal, M., & Wijaya, R. F. W. (2023). Implementasi Algoritma Naïve Bayes dalam Menganalisis Sentimen Review Pengguna Tokopedia pada Produk Kesehatan. *Bulletin of Infile:///G:/My Drive/TARGET 2024/REFERENSI/Sistem Informasi Aplikasi Penentuan Jurusan Yang Sesuai Dengan Minat Menggunakan Pendekatan Fuzzy Sugeno.pdf* *Information Technology (BIT)*, 4(4), 533–543. <https://doi.org/10.47065/bit.v4i4.1090>
- Fatmawati, F., & Narti, N. (2022). Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring. *JTIM: Jurnal Teknologi Informasi dan Multimedia*, 4(1), 1–12. <https://doi.org/10.35746/jtim.v4i1.196>
- Fitriani, E. (2020). Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *Sistemasi*, 9(1), 103. <https://doi.org/10.32520/stmsi.v9i1.596>
- Haqmanullah Pambudi, R., & Darma Setiawan, B. (2018). Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(7), 2637–2643. <http://j-ptiik.ub.ac.id>
- Ika, N., Kalingara, P., Pratiwi, O. N., & Anggana, H. D. (2021). Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi Jne Dan J & T Express Menggunakan Metode Naïve Bayes Sentiment Analysis



- Review Customer of Jne and J & T Express Expedition Services Using Naïve Bayes Method. *e-Proceeding of Engineering*, 8(5), 9035–9048.
- Prast, T. P., Zakaria, H., & Wiliantoro, P. (2022). Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier. *OKTAL : Jurnal Ilmu Komputer dan Sains*, 1(06), 573–582. <https://journal.mediapublikasi.id/index.php/oktal/article/view/427>
- Puspitasari, A., & Zakaria, H. (2023). *Sistem Informasi Aplikasi Penentuan Jurusan Yang Sesuai Dengan Minat Menggunakan Pendekatan Fuzzy Sugeno (Studi Kasus : SMK Fadilah)*. 1(3), 630–642.