

Analisis Sentimen Komentar Perplexity AI di X Tentang Pendidikan Menggunakan Support Vector Machine

Yoga Ardiansah*, Siti Monalisa, Fitriani Muttakin

Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}12050316847@students.uin-suska.ac.id, ²siti.monalisa@uin-suska.ac.id, ³fitrianimuttakin@uin-suska.ac.id

Email Penulis Korespondensi: 12050316847@students.uin-suska.ac.id

Submitted: 04/12/2024; Accepted: 29/12/2024; Published: 30/12/2024

Abstrak—Chatbot dengan Artificial Intelligence semakin populer dalam kehidupan sehari-hari. Karena kemampuannya untuk menalar dan menyampaikan informasi secara ekspresif, Artificial Intelligence (AI) yang menggunakan model Natural Language Processing (NLP) dapat berkomunikasi layaknya manusia. Pengguna menganggap salah satu chatbot ciptaan Perplexity yang menggunakan AI menarik karena dapat menunjukkan sumber informasi dengan tepat. Seiring berjalannya waktu dan bertambahnya jumlah pengguna Perplexity, analisis sentimen digunakan untuk mengukur kebahagiaan pengguna. Analisis sentimen ini berfungsi sebagai sumber data untuk penelitian ini, membantu dalam memahami bagaimana pengguna bereaksi terhadap media sosial X (Twitter). Metode Support Vector Machines (SVM) digunakan dalam penelitian ini, di mana SVM memaksimalkan jarak (margin) antara kelompok data untuk menentukan hyperplane yang ideal. Menurut penelitian tersebut, 90,11% responden menyatakan sentimen positif, 5,30% menyatakan opini negatif, dan 4,69% menyatakan sentimen netral. Dengan menggunakan rasio 80% data pelatihan dan 20% data uji, skor f1 mencapai 96%, dengan akurasi dan presisi masing-masing sebesar 92%.

Kata Kunci: Artificial Intelligence; Chatbot; Perplexity; Analisis Sentimen; Support Vector Machine

Abstract—Chatbots with Artificial Intelligence are increasingly popular in everyday life. Due to its ability to reason and convey information expressively, Artificial Intelligence (AI) using Natural Language Processing (NLP) models can communicate like humans. Users find one of Perplexity's AI chatbots interesting because it can pinpoint sources of information. As time goes by and the number of Perplexity users increases, sentiment analysis is used to measure user happiness. This sentiment analysis serves as the data source for this research, helping understand how users react to social media X (Twitter). The Support Vector Machines (SVM) method was used in this study, where SVM maximises the distance (margin) between data groups to determine the ideal hyperplane. According to the survey, 90.11% of respondents expressed positive sentiments, 5.30% expressed negative opinions, and 4.69% expressed neutral sentiments. Using a ratio of 80% training data and 20% test data, the f1 score reached 96%, with accuracy and precision of 92% each.

Keywords: Artificial Intelligence; Chatbot; Perplexity; Sentiment Analysis; Support Vector Machine

1. PENDAHULUAN

Komunikasi dan teknologi informasi telah berkembang karena kompleksitas sistem kecerdasan buatan (*artificial intelligence/AI*). Perplexity AI menggabungkan model bahasa yang sangat besar dengan kemampuan GPT-3 [1]. Perplexity AI dapat merespons pertanyaan dan menganalisis data dalam jumlah besar, memastikan Anda memberikan akses ke banyak informasi. Perplexity AI memberikan jawaban yang terverifikasi dan telah diteliti dengan baik kepada pengguna dalam bahasa akademis. Perplexity menawarkan versi gratis agar dapat diakses oleh setiap pengguna dan memberikan jawaban yang akurat dengan mengumpulkan informasi dari berbagai sumber seperti situs web dan jurnal akademis, Perplexity, versi gratis yang telah diperbarui pada tanggal 26 Juli 2023, didasarkan pada algoritma GPT-3 [2][3]. Perplexity AI beroperasi mirip dengan ChatGPT yang juga dibangun di atas model GPT OpenAI. Perplexity AI diperkirakan beroperasi pada proses pelatihan yang sama mengingat kerangka kerja GPT yang digunakan bersama, tetapi mungkin ada sedikit perbedaan hak milik karena diproduksi oleh perusahaan yang berbeda. Dataset Perplexity AI adalah yang terbaru pada saat penelitian ini dilakukan. Sebagai sebuah produk, Perplexity AI memberi merek pada dirinya sendiri di sekitar penelitian dan pertanyaan. Perbedaan dalam branding ini terlihat dalam struktur respons yang secara umum tampak tidak terlalu rumit secara percakapan jika dibandingkan dengan ChatGPT [4].

Kinerja Perplexity AI dalam dunia pendidikan telah membantu para siswa dan memberikan hasil yang sesuai selama beberapa tahun terakhir [5]. Perplexity AI secara khusus dirancang untuk memotivasi dan meningkatkan kemampuan menulis siswa. Dengan menggunakan chatbot AI, guru dapat membantu siswa mengembangkan tulisan yang lebih tepat dengan mengetahui kualitas tulisan, serta menawarkan situs yang menyenangkan dan menarik untuk melatih ketepatan dalam situasi nyata. Dengan demikian, integrasi teknologi ini sangat penting untuk memodernisasi dan meningkatkan keterampilan menulis [6]. Untuk mengetahui tingkat respon pengguna terhadap layanan ini, penelitian tentang Perplexity dan chatbot lain yang melayani pengguna menjadi sangat penting. Bisnis dapat menggunakan analisis sentimen untuk menentukan apakah pelanggan puas atau tidak dengan layanan chatbot [7].

Langkah-langkah dalam penelitian ini menggunakan analisis sentimen untuk mengevaluasi layanan Perplexity. Penelitian sentimen media sosial dalam akses chatbot menghasilkan informasi yang sangat baik dan respon yang cepat dan efektif [8]. Media sosial X digunakan dalam studi ini untuk menganalisis sentimen tweet yang berhubungan dengan interaksi dengan Perplexity AI. *Support Vector Machine* (SVM) merupakan langkah untuk menguji data. Pendekatan *Support Vector Machine* (SVM) dipilih karena dapat memberikan hasil dampak positif dalam penggalian opini juga dapat mengungguli metode kategorisasi teks yang bersifat linier [9].

Pada tahun 2024, penelitian yang dilakukan oleh Sagala dan Samuel menggunakan tiga dari algoritma *Support Vector Machines*. Hasil studi menjelaskan metode pada *Random Forest* dan *Support Vector Machine* (SVM) akurasi yang tertinggi. SVM mencatat validitas positif sebesar 86,80% dan validitas negatif sebesar 0,13%, dengan skor F1 mencapai 90%. Sementara itu, *Random Forest* mencatat data positif yang benar sebesar 87,43% dan data negatif yang benar sebesar 0,75% [10].

Pada Tahun 2023 Penelitian Tubishat dan kawan-kawan, Penelitian yang menggunakan data komentar media sosial Twitter, didapatkan hasil penelitian bahwa *Support Vector Machine* memiliki akurasi tertinggi yaitu 81% dan f1-score sebesar 0,77% sedangkan algoritma *Random Forest* akurasi 79% f1-score 0,72%, algoritma *Naive Bayes* akurasi 60% nilai f1-score 0,44% dan algoritma *KNN* akurasi sebesar 57% dan nilai f1-score 0,53% [11].

Pada tahun 2024, penelitian yang dilakukan oleh Laia dan rekan-rekan mengenai Jasa Ojek Online dengan menggunakan data Google Play Store dengan algoritma SVM, sentimen positif mencapai 40%. Selain itu, model SVM juga mencatat kinerja terbaik dengan nilai Precision dan F1-score memiliki nilai sebesar 0,93 dan 0,92 [12].

Pada Tahun 2022 Penelitian Yolanda dan kawan-kawan, Penelitian Pada Aplikasi Moodle dan Edmodo menggunakan sosial media twitter dengan menggunakan algoritma *Support Vector Machine* aplikasi Moodle nilai positif 67% dengan akurasi 84% nilai *f1-score* sebesar 87%, sedangkan Edmodo nilai positif 84% f1-score 87% [13].

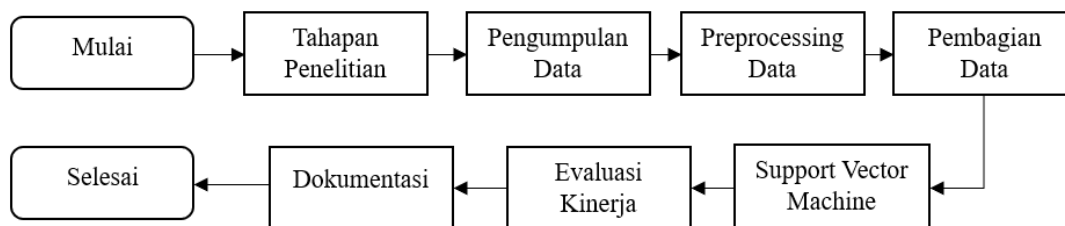
Pada Tahun 2024 Penelitian Kusman dan kawan-kawan, Penelitian terhadap data debat pemilihan Presiden 2024 Menggunakan Media Sosial X algoritma *Support Vector Machine* Pengujian Akurasi sebesar 52,5%, Hasil tersebut tidak memuaskan dengan optimasi dengan *Scaling* dan *oversampling* mendapatkan akurasi tertingginya 94% [14].

Dari penelitian ini, diharapkan bahwa studi mengenai *Chatbot AI*, khususnya pada aspek *Perplexity*, dapat memberikan kemajuan yang lebih signifikan. Penelitian ini diharapkan dapat menerima saran dari pengguna yang bertujuan untuk meningkatkan minat pengguna, serta memberikan manfaat yang nyata dalam memudahkan pengguna dalam memenuhi standar kualitas yang diharapkan.

2. METODOLOGI PENELITIAN

2.1 Tahap Penelitian

Analisis Sentimen adalah proses manajemen teks yang bersifat subjektif mulai dari menganalisis, memproses, meringkas, dan pemrosesan inferensial. Analisis sentimen adalah penggunaan metode klasifikasi pembelajaran mesin dan klasifikasi fitur yang mengkarakterisasi teks sehingga memberikan kesan emosi dan sentimen dengan cepat dan mudah. Berikut ini adalah Gambar 1 dari tahapan penelitian.



Gambar 1. Tahapan Penelitian

2.2 Tahap Pengumpulan Data

Tahap awal studi ini data yang dikumpul berdasarkan komentar dari pengguna X (Twitter) di *Perplexity AI*, data yang dikumpulkan rentan dari tanggal 1 September 2023 hingga 1 Oktober 2024 sehingga diperoleh 2673. Setelah itu, Google Colab digunakan untuk memproses dan membersihkan data yang diperoleh. Kamus leksikal dari GitHub yang telah diverifikasi oleh pakar bahasa Indonesia digunakan untuk pelabelan data. Untuk memperoleh temuan dan akurasi pengujian dalam penelitian ini, proses pelabelan dilanjutkan dengan proses pemodelan menggunakan metode *Support Vector Machine* (SVM). Pada studi akan digunakan data media sosial X dengan pencarian tweet menggunakan “PERLENGKAPAN PENDIDIKAN”, “PERLENGKAPAN TUGAS”, dan “PERLENGKAPAN REFERENSI”.

Studi ini menunjukkan, sistem pembelajaran *Support Vector Machine* (SVM) dipelajari menggunakan metode pembelajaran yang berasal dari teori optimasi dan menggunakan hipotesis untuk fungsi linier dalam fitur berdimensi tinggi. [15]. Metode *Support Vector Machine* (SVM) memungkinkan penyelesaian masalah linier dengan melakukan transformasi matematis pada ruang pembelajaran melalui fungsi kernel. Konsep utama SVM merupakan menemukan hyperplane terbaik yang dapat memisahkan dua kelas yang telah ditentukan dengan jarak yang optimal [12].

2.3 Tahapan Pre-Processing Data

Pada tahapan pre-processing data beberapa tahapan yang perlu dilakukan untuk memastikan data memiliki kualitas yang baik dan bebas dari informasi yang tidak relevan. Data tersebut harus memiliki struktur yang jelas dan keakuratan yang tinggi sebelum memasuki tahap penelitian. Setelah Pre-processing Data selesai, data dapat digunakan untuk keperluan penelitian.

a. Data labelling

Tahap ini dilaksanakan secara manual dengan menambahkan label pada komentar sebagai positif, negatif, atau netral. Proses pelabelan untuk kategori netral dilakukan dengan mempertimbangkan informasi yang dianggap tidak signifikan terhadap konteks yang dibahas. Dalam sistem pelabelan ini, ditetapkan bahwa label 0 mewakili kategori netral, label 1 untuk positif, dan label 2 untuk negatif. Pendekatan ini bertujuan untuk memastikan bahwa setiap komentar diklasifikasikan secara akurat sesuai dengan sentimen yang terkandung di dalamnya. [16].

b. Data cleaning

Langkah awal dalam proses pengolahan data adalah melakukan pembersihan data (data cleaning) dengan menghapus semua simbol, tagar, nama pengguna, URL, dan komentar yang serupa dari teks. Proses ini bertujuan untuk menjamin kualitas data yang lebih baik dan mengeliminasi elemen-elemen yang tidak relevan yang dapat mengganggu analisis sentimen. Dengan demikian, teks yang tersisa akan lebih murni dan fokus, memungkinkan algoritma analisis untuk bekerja dengan lebih efektif saat mengidentifikasi sentimen pengguna.

c. Case folding

Tahap ini melibatkan proses case folding, semua pada karakter huruf dalam teks diubah keseluruhan menjadi huruf kecil. Proses ini penting karena sistem analisis sentimen akan menganggap huruf kapital dan huruf kecil sebagai entitas yang berbeda. Dengan menerapkan case folding, konsistensi dalam pemrosesan teks dapat dicapai, yang pada gilirannya meningkatkan akurasi analisis dan meminimalkan kesalahan dalam pengenalan kata-kata yang sama yang ditulis dengan cara berbeda. [17].

d. Tokenization

Tahap ini melibatkan proses tokenization, di mana data teks dipisahkan menjadi token-token yang terdiri dari kata, karakter khusus, atau tanda baca. Proses ini penting untuk analisis teks, karena memungkinkan pemrosesan lebih lanjut dengan memecah teks dibagi menjadi bagian-bagian yang lebih kecil dan lebih mudah diatur. Dalam penelitian ini, digunakan fungsi yang dirancang khusus untuk melakukan tokenization secara efisien, sehingga setiap token dapat dianalisis secara independen. Dengan pemisahan ini, analisis sentimen dapat dilakukan dengan lebih tepat, mengingat setiap elemen teks dapat dievaluasi berdasarkan konteks dan maknanya [18].

e. Stemming

Tahap ini digunakan dalam proses pengolahan bahasa alami untuk mereduksi kata-kata ke bentuk dasarnya. Dengan cara ini, berbagai variasi kata yang memiliki makna serupa dapat dianggap sebagai satu kesatuan. Dalam tahap ini, akhiran pada kata-kata dihilangkan, yang memungkinkan analisis dan pemrosesan data menjadi lebih efisien. Dengan melakukan stemming, model dapat lebih mudah memahami konteks dan hubungan antar kata, sehingga meningkatkan akurasi dalam pengolahan informasi dan analisis teks..

f. TF-IDF

Pada tahap TF-IDF, pembobotan kata dilakukan. Setiap kata dalam dokumen memiliki berat yang berbeda dengan menggunakan frekuensi kata dan frekuensi dokumen balik (TF-IDF). Perhitungan TF-IDF memiliki beberapa rumus seperti Rumus 1. IDF (Inverse Document Frequency), Rumus 2 TF-IDF dan Rumus 3 Normalisasi Vektor [19], sebagai berikut:

$$idf(t) = \ln \frac{1+n}{1+df(t)} + 1 \quad (1)$$

$$tf-idf(t, d) = tf(t, d) \times idf(t) \quad (2)$$

$$v_{norm} = \frac{v}{||v||_2} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \quad (3)$$

Pada Rumus 1, rumus *IDF* dinyatakan seberapa spesifik atau unik kata tersebut. Di sini, *n* adalah jumlah total dokumen, sementara *df(t)* adalah jumlah dokumen yang mengandung kata *t*. Semakin sedikit dokumen yang mengandung kata tersebut, semakin tinggi nilai *IDF*-nya, yang menunjukkan bahwa kata itu lebih relevan dan signifikan. Pada Rumus 2, rumus *TF-IDF* dihitung dengan mengalikan frekuensi kata dalam dokumen, *tf(t,d)* dengan *IDF*, sehingga menghasilkan nilai *tf-idf(t,d) = tf(t,d) × idf(t)*. Semakin sedikit dokumen yang mengandung kata tersebut, semakin tinggi nilai *IDF*-nya, yang menunjukkan bahwa kata itu lebih relevan dan signifikan. dan Pada Rumus 3, normalisasi Vektor ini memungkinkan kita untuk membandingkan vektor - vektor yang merepresentasikan dokumen dengan lebih efektif, sehingga meningkatkan kemampuan analisis dan pengambilan informasi dari teks.

2.4 Pembagian Data

Pengujian data melibatkan proses pemisahan data menjadi dua kategori: 80% untuk data pelatihan (training) dan 20% untuk data pengujian (testing). Pembagian ini merupakan praktik umum dalam pembelajaran mesin, bertujuan agar model dapat melakukan generalisasi dengan baik pada data baru. Dalam skema ini, 80% dari total dataset digunakan untuk melatih model, sementara 20% sisanya digunakan untuk menguji kinerja model tersebut. Dengan menggunakan fungsi kernel, pendekatan SVM memungkinkan modifikasi matematis pada ruang pembelajaran, sehingga dapat menangani masalah linier dengan lebih efektif [20].

- a. Data Training (data latih) Data ini dimanfaatkan untuk melatih model. Dalam fase ini, algoritma belajar dari fitur dan label yang ada dalam data pelatihan, mengidentifikasi pola, dan mengoptimalkan parameter model. Model akan berusaha menemukan hubungan antara fitur input dan output yang diinginkan berdasarkan data ini.
- b. Data Testing (data uji) Data pengujian digunakan untuk mengevaluasi kinerja model setelah dilatih. Dengan memanfaatkan data yang belum pernah dilihat oleh model sebelumnya, kita dapat mengevaluasi seberapa baik model dapat menggeneralisasi dan melakukan prediksi pada data baru. Kinerja ini biasanya diukur dengan metrik seperti akurasi, *precision*, *recall*, dan *F1-score*.

2.5 Support Vector Machine

Pengguna algoritma SVM digunakan untuk mencari hyperplane terbaik yang dapat memisahkan dua kelas dengan jarak yang paling optimal. hyperplane dengan margin tertinggi. Margin didefinisikan sebagai jarak antara titik data terdekat dengan hyperplane. Titik yang paling dekat dengan hyperplane adalah support vector [15].

$$w \cdot x + b = 0 \quad (4)$$

Pada Rumus 4, Hyperplane berfungsi sebagai pemisah antara dua kelas dalam ruang fitur multi dimensi. Dalam konteks ini, vektor bobot w menentukan arah dan kemiringan hyperplane, sementara bias b menggeser hyperplane dari titik asal. Dengan kata lain, hyperplane membagi ruang fitur menjadi dua bagian: satu untuk setiap kelas. Titik-titik data yang terletak di satu sisi hyperplane akan memiliki nilai $w \cdot x + b > 0$ (misalnya, kelas positif), sedangkan titik-titik di sisi lainnya akan memiliki nilai $w \cdot x + b < 0$ (kelas negatif). Tujuan dari SVM adalah untuk menemukan hyperplane yang tidak hanya memisahkan kedua kelas tersebut, tetapi juga memaksimalkan margin berarti memperbesar jarak antara hyperplane dan titik data terdekat dari setiap kelas.

2.6 Evaluasi Model

Tahapan evaluasi metode Support Vector Machine dalam penelitian ini dilakukan dengan menggunakan Confusion Matrix dan K-Fold Cross Validation. Berikut Penjelasan Confusion Matrix dan K-Fold Cross Validation pada metode *Support Vector Machine* [9].

- a. Confusion Matrix pada Metode SVM adalah alat evaluasi yang digunakan untuk menilai kinerja model SVM dengan membandingkan prediksi model dengan label yang sebenarnya. Confusion matrix membantu dalam mengidentifikasi kesalahan klasifikasi spesifik yang dilakukan oleh model SVM, seperti apakah model sering salah mengklasifikasikan positif sebagai negatif. Confusion Matrix Komponen utama seperti *True Positif* (TP), *False Positif* (FP), *False Negatif* (FN), *True Negatif* (TN). Dan memiliki nilai metrik evaluasi seperti *Precision*, *Recall*, dan *F1-Score*.
- b. K-Fold Cross Validation pada Metode SVM adalah teknik yang digunakan untuk mengevaluasi model SVM dengan membagi dataset menjadi K bagian (fold) untuk melatih dan menguji model secara bergantian. Proses pengerjaan Dataset dibagi menjadi K subset yang sama, setiap kali menggunakan K-1 fold untuk pelatihan dan 1 fold untuk pengujian. Nilai K sering dipilih berdasarkan ukuran dataset. Pada umumnya, K=5 atau K=10 digunakan, tetapi dapat disesuaikan berdasarkan kompleksitas data dan model. Setelah semua iterasi, metrik evaluasi (seperti akurasi, precision, recall, dan F1-score) dihitung sebagai rata-rata dari K iterasi. Ini memberikan wawasan yang lebih mendalam tentang kinerja model SVM di seluruh dataset.

2.7 Dokumentasi

Pada tahap Dokumentasi, yang merupakan fase akhir dari penelitian, tujuan utamanya adalah menyajikan visualisasi dalam bentuk word cloud. Word cloud ini berfungsi untuk menggambarkan kata-kata yang paling sering muncul berdasarkan analisis sentimen yang telah dilakukan. Visualisasi ini memungkinkan peneliti untuk dengan mudah mengidentifikasi istilah-istilah kunci dan pola yang dominan dalam komentar pengguna, serta memberikan gambaran yang jelas mengenai fokus perhatian dan persepsi pengguna terhadap Perplexity AI.

3. HASIL DAN PEMBAHASAN

Pada tahapan Hasil dan Pembahasan, akan dibahas hasil analisis sentimen yang dilakukan terhadap komentar pengguna Perplexity AI di media sosial X dalam konteks pendidikan. Bagian ini akan menguraikan alur pengerjaan penelitian secara sistematis. Penelitian ini mengimplementasikan Algoritma Support Vector Machine (SVM) untuk menganalisis sentimen, disertai dengan pengujian menggunakan Confusion Matrix dan K-Fold Cross Validation. Selanjutnya, dalam tahapan dokumentasi, akan ditampilkan frekuensi kemunculan kata-kata yang relevan serta visualisasi dalam bentuk word cloud (awan kata). Hasil analisis ini diharapkan dapat memberikan wawasan yang mendalam mengenai persepsi pengguna terhadap Perplexity AI dan kontribusinya dalam bidang pendidikan.

3.1 Data Dasar

Data yang diperoleh dari media sosial X dikumpulkan dengan menggunakan teknik *scraping* yang diterapkan melalui bahasa pemrograman Python. Jumlah komentar yang berhasil dikumpulkan mencapai 2.673. Hasil dari proses pencarian yang telah difilter mencakup informasi mengenai tanggal, nama pengguna, dan teks komentar pengguna

terkait Perplexity AI dalam konteks pendidikan. Berikut disajikan data dasar mengenai Perplexity AI dalam Tabel 1, yang menunjukkan informasi sebelum dilakukan proses pengolahan lebih lanjut.

Tabel 1. Data Dasar

No	Tanggal	Username	Text
1.	Fri Aug 25, 11:13:50 AM	driedtrees	@collegemenfess Udahnya masuk ke perplexity. Nanti di perplexity dia bakalan ngasih penjelasan lebih detail + resourceny (bentuknya kaya dapus jadi kepike banget)
2.	Mon Aug 21, 4:48:46 AM	eudaimonophile	@raquewlla @fellexandro Jlek bnget perplexity kasih sumber linkedin
....
2040.	Sun Sep 22, 4:20:44 AM	zmx	@zmx perplexity telusur sumber percaya sistem lelang, iklan senang sistem cari proses ai dasar kece fakta saing terap file tanda tangan teman fakta interpretasi

Tabel 1 menyajikan data dasar yang diperoleh dari media sosial X (Twitter) terkait Perplexity AI. Kolom-kolom dalam tabel mencakup nomor urut, tanggal, nama pengguna, dan isi teks komentar. Data ini berfungsi sebagai sumber informasi awal yang akan dianalisis untuk mengidentifikasi sentimen pengguna terhadap Perplexity AI. Analisis ini bertujuan untuk memahami persepsi pengguna dan memberikan wawasan mengenai dampak Perplexity AI dalam konteks pendidikan.

3.2 Pre –Processing

Pada tahap ini, data dasar yang digunakan akan melalui proses penghapusan bagian yang tidak relevan, seperti tanda baca dan tanda makna yang lain, emotikon, dan penggunaan singkatan, sehingga data yang diproses menjadi lebih mudah dipahami oleh pengguna. Tahap pre-processing dilakukan untuk memastikan data yang digunakan valid dan siap untuk analisis yang akan dilakukan selanjutnya. Tabel 2 di bawah ini menunjukkan hasil dari proses pre-processing yang mencakup stemming.

Tabel 2. Pre-Processing

No	stemming	text
1.	['udahnya', 'masuk', 'perplexity', 'perplexity', 'masih', 'jelas', 'detail', 'resourceny', 'bentuk', 'kayak', 'dapus', 'kepike', 'banget']	udahnya masuk perplexity perplexity masih jelas detail resourceny bentuk kayak dapus kepike banget.
2.	['jelek', 'banget', 'Perplexity', 'kasih', 'sumber', 'linkedin']	jelek banget perplexity kasih sumber linkedin
....
2040.	['perplexity', 'telusur', 'sumber', 'percaya', 'sistem', 'lelang', 'iklan', 'senang', 'sistem', 'cari', 'proses', 'ai', 'dasar', 'kece', 'fakta', 'saing', 'terap', 'file', 'tanda', 'tangan', 'teman', 'fakta', 'interpretasi']	perplexity telusur sumber percaya sistem lelang iklan senang sistem cari proses ai dasar kece fakta saing terap file tanda tangan teman fakta interpretasi

Pada Tabel 2. Pre-Processing yang mencakup Stemming dari data yang dikumpulkan. Setiap entri menunjukkan kata-kata yang telah diproses untuk memperbaiki kualitas analisis sentimen. Proses ini penting untuk memastikan bahwa kata-kata dengan makna yang sama dapat dikelompokkan, sehingga analisis menjadi lebih akurat.

3.3 Labelling Data

Pada tahap ini, dilakukan pelabelan menggunakan pustaka TextBlob. Nilai yang dihasilkan akan mengindikasikan kategori sentimen data, yaitu positif, negatif, dan netral. Nilai 1 merepresentasikan sentimen positif, nilai 0 merepresentasikan sentimen netral, dan nilai -1 merepresentasikan sentimen negatif. Hasil pelabelan ini ditampilkan dalam Tabel 3.

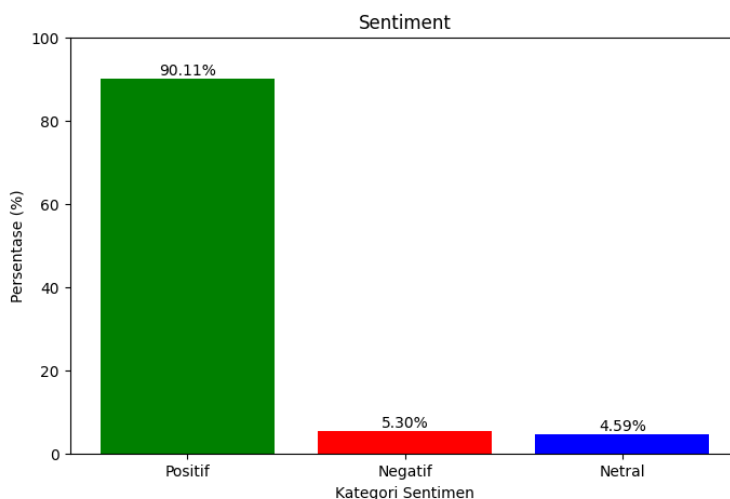
Tabel 3. Labelling Data

No	Text	Sentiment
1.	udahnya masuk perplexity perplexity asih jelas detail resourceny bentuk kayak dapus kepike banget	3 Positif
2.	jelek banget perplexity kasih sumber linkedin	-1 Negatif
....

No	Text	Sentiment
2040.	perplexity telusur sumber percaya sistem lelang iklan senang sistem cari proses ai dasar kece fakta saing terap file tanda tangan teman fakta interpretasi	5 Positif

Pada Tabel 3. Labelling ini menunjukkan hasil pelabelan sentimen dari data yang telah dibersihkan. Setiap entri mencantumkan teks asli dan hasil pelabelan sentimen, yaitu positif, negatif, atau netral. Hasil ini akan digunakan untuk analisis lebih lanjut dan pemodelan menggunakan algoritma SVM.

Setelah memperoleh hasil dari proses labelling data, tahapan selanjutnya adalah melakukan pengelompokan sentimen ke dalam kategori positif, negatif, dan netral. Awalnya, terdapat 2.673 data, namun setelah dilakukan pembersihan, jumlah data yang tersisa menjadi 2.040. pelabelan manual oleh ahli bahasa pakar pembersihan, jumlah komentar dengan sentimen positif mencapai 1.667, sedangkan komentar dengan sentimen negatif berjumlah 98, dan komentar dengan sentimen netral sebanyak 85 berikut ditampilkan Gambar 2. Hasil Bagan Sentimen.



Gambar 2. Hasil Bagan Sentimen

Gambar 2 ini menggambarkan distribusi sentimen yang dihasilkan dari analisis. Bagan ini menunjukkan persentase komentar yang termasuk dalam kategori positif menunjukkan 90,11%, negatif menunjukkan 5,30%, dan netral 4,59% menunjukkan bahwa sangat sedikit data yang memiliki sentimen netral, memberikan gambaran umum tentang bagaimana pengguna merasakan Perplexity AI.

3.4 TF– IDF

Pada tahapan TF-IDF, fitur diekstrak dari teks dengan mengubah input teks menjadi representasi numerik. Teknik TF-IDF memberikan bobot pada setiap term dalam korpus, di mana bobot TF-IDF untuk setiap term dihitung dengan mengalikan nilai Term Frequency (TF) dan Inverse Document Frequency (IDF). Berikut adalah hasil dari perhitungan TF-IDF yang ditampilkan pada Tabel 4 berikut.

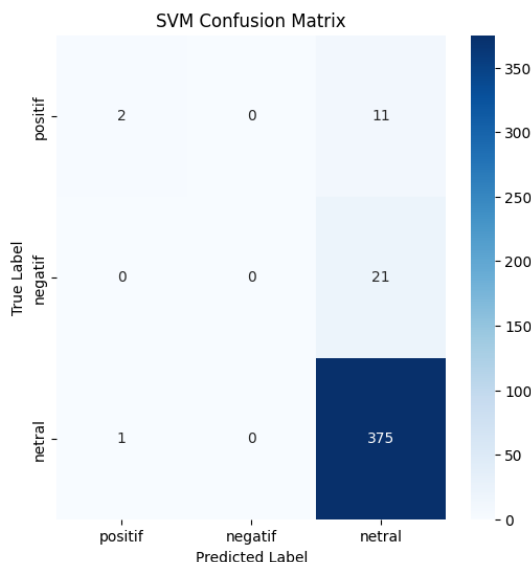
Tabel 4. TF-IDF

No	Ketentuan					
	abad	abadi	abai	abal	...	zuckerberg
1	0.0	0.0	0.0	0.0	...	0.0
2	0.0	0.0	0.0	0.0	...	0.0
3	0.0	0.0	0.0	0.0	...	0.0
4	0.0	0.0	0.0	0.0	...	0.0
...
2040	0.0	0.0	0.0	0.0	...	0.0

Pada Tabel 4, disajikan hasil perhitungan untuk setiap istilah yang ada dalam data. Nilai-nilai ini digunakan untuk menentukan bobot setiap kata dalam analisis, yang penting untuk memahami kata-kata mana yang paling berpengaruh dalam sentimen yang diungkapkan. Dengan bobot yang ditentukan oleh TF-IDF, kita dapat mengidentifikasi istilah yang memiliki dampak signifikan dalam konteks analisis sentimen.

3.5 Evaluasi Kinerja

Setelah melalui beberapa tahapan sebelumnya, tahap ini melibatkan evaluasi kinerja model menggunakan algoritma Support Vector Machine (SVM). Hasil tahapan ini akan dianalisis menggunakan confusion matrix serta mengevaluasi akurasi pada nilai, termasuk *Precision*, *Recall*, *F1-score*, dan *Support*. Selanjutnya, pengujian lebih lanjut akan dilakukan dengan membandingkan hasil k-fold cross-validation yang diulang sebanyak 10 kali iterasi.



Gambar 3. SVM Confusion Matrix

Pada Gambar 3, Confusion Matrix SVM untuk model *Support Vector Machine* (SVM) menunjukkan kinerja model dalam mengklasifikasikan data ke dalam tiga kategori: positif, negatif, dan netral. Dari analisis ini, terdapat 2 contoh yang diklasifikasikan sebagai positif dengan benar (true positif), sementara tidak ada contoh negatif yang salah diklasifikasikan sebagai positif (false positif). Namun, model mengalami kesalahan dalam mengklasifikasikan 11 contoh positif sebagai netral (false negatif). Dalam kategori negatif, model berhasil mengidentifikasi 21 contoh dengan benar (true negatif). Kategori netral menunjukkan performa terbaik dengan 375 contoh yang diklasifikasikan dengan akurat. Secara keseluruhan, matriks ini mencerminkan efektivitas model SVM, serta menunjukkan bahwa meskipun model cukup baik dalam mengenali kategori netral, masih terdapat tantangan dalam mengklasifikasikan kategori positif dan negatif secara akurat.

Tabel 5. Hasil SVM Confusion Matrix

	Precision	Recall	f1-score	support
Negatif	0.00	0.00	0.00	14
Netral	0.00	0.00	0.00	18
Positif	0.92	1.00	0.96	378
accuracy			0.92	410
macro avg	0.31	0.33	0.32	410
weighted avg	0.85	0.92	0.68	410
SVM Accuracy :91.95%				

Tabel 5 Hasil SVM Confusion Matrix di atas menyajikan metrik evaluasi untuk model *Support Vector Machine* (SVM), termasuk precision, recall, f1-score, dan support untuk masing-masing kategori: negatif, netral, dan positif. Hasil dari Tabel 5 Confusion Matrix menunjukkan bahwa akurasi model *Support Vector Machine* (SVM) mencapai 91,95%. Selain itu, rata-rata nilai *Precision*, *Recall*, dan *F1-score* juga menunjukkan kinerja yang tinggi dalam klasifikasi sentimen untuk semua kategori.

Tabel 6. Persentase K-Fold Cross Validations

	percentage
K Fold 1	89,76%
K Fold 2	89,76%
K Fold 3	89,27%
K Fold 4	90,24%
K Fold 5	87,32%
K Fold 6	90,73%
K Fold 7	86,27%
K Fold 8	93,14%
K Fold 9	89,71%
K Fold 10	84,86%

- doi: 10.1016/j.dld.2024.02.019.
- [6] C. U. Lubis, B. Ichsanda, dan R. Hz, “Perplexity AI on the Writing Efficiency of EFL Students in Higher Education: Students’ Insights,” vol. 8, no. 1, hal. 167–178, 2024, doi: 10.29240/ef.v8i1.
 - [7] D. Baidoo-Anu dan L. O. Ansah, “Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning,” *J. AI*, vol. 7, no. 1, hal. 52–62, 2023, doi: 10.2139/ssrn.4337484.
 - [8] S. Arora, S. Mahapatra, A. Jadav, M. Barla, dan N. Mallick, “Temporal and Sentimental Analysis of Customer Reviews,” in *2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2024, hal. 520–525. doi: 10.1109/Confluence60223.2024.10463473.
 - [9] R. Wahyudi dkk., “Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, hal. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
 - [10] G. J. Sagala dan Y. T. Samuel, “Sentiment Analysis on ChatGPT App Reviews on Google Play Store Using Random Forest Algorithm, Support Vector Machine and Naïve Bayes,” *Int. J. Eng. Bus. Soc. Sci.*, vol. 2, no. 04, hal. 1194–1204, 2024, doi: 10.58451/ijebss.v2i04.148.
 - [11] M. Tubishat, F. Al-Obeidat, dan A. Shuhaiber, “Sentiment Analysis of Using ChatGPT in Education,” *2023 Int. Conf. Smart Appl. Commun. Networking, SmartNets 2023*, hal. 1–7, 2023, doi: 10.1109/SmartNets58706.2023.10215977.
 - [12] Y. B. Laia Sunneng Sandino; Sumihar, Yo’el Pieter; Budiati, Haeni, “Implementasi Library Textblob dan Metode Support Vector Machine Pada Analisis Sentimen Pelanggan Terhadap Jasa Transportasi Online,” *Build. Informatics, Technol. Sci.*, no. Vol 6 No 1 (2024): June 2024, hal. 1–10, 2024, doi: 10.47065/bits.v6i1.5090.
 - [13] N. Yolanda, I. H. Santi, dan D. F. H. Permadi, “Analisis Sentimen Analisis Sentimen Popularitas Aplikasi Moodle dan Edmodo Menggunakan Algoritma Support Vector Machine,” *J. Algoritm.*, vol. 3, no. 1, hal. 48–59, 2022, doi: 10.35957/algoritm.v3i1.3313.
 - [14] V. Nava, M. Kusman, V. Metayani, dan O. Karnalim, “Prediksi Analisis Sentimen Data Debat Pemilihan Presiden 2024 Menggunakan Support Vector Machine (SVM) Prediction of Sentiment Analysis of 2024 Presidential Election Debate Data Using Support Vector Machine (SVM),” vol. 16, hal. 1–5, 2024, doi: 10.35891/explorit.v16i1.4887.
 - [15] A. R. Isnain, A. I. Sakti, D. Alita, dan N. S. Marga, “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm,” *J. Data Min. Dan Sist. Inf.*, vol. 2, no. 1, hal. 31–37, 2021, doi: 10.33365/JDMSI.V2I1.1021.
 - [16] M. L. Nugraha dan E. B. Setiawan, “Bank Central Asia (BBCA) Stock Price Sentiment Analysis On Twitter Data Using Neural Convolutional Network (CNN) And Bidirectional Long Short-Term Memory (BI-LSTM),” *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 3, hal. 936–947, 2023.
 - [17] N. P. Wahyuningtyas, D. E. Ratnawati, dan N. Y. Setiawan, “Root Cause Analysis (RCA) berbasis Sentimen menggunakan Metode K-Nearest Neighbor (K-NN)(Studi Kasus: Pengunjung Kolam Renang Brawijaya),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, hal. 2515–2520, 2023.
 - [18] A. O. Permana dan S. Saepudin, “Perbandingan algoritma k-nearest neighbor dan naïve bayes pada aplikasi shopee,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 1, hal. 25–32, 2023, doi: 10.37859/coscitech.v4i1.4474.
 - [19] S. Chamira, “Implementasi Metode Text Mining Frequency-Invers Document Frequency (Tf-Idf) Untuk Monitoring Diskusi Online,” *J. Informatics, Electr. Electron. Eng.*, vol. 1, no. 3, hal. 97–102, 2022, doi: 10.47065/jieee.v1i3.353.
 - [20] M. Aulia dan A. Hermawan, “Analisis Perbandingan Algoritma SVM, Naïve Bayes, dan Perceptron untuk Analisis Sentimen Ulasan Produk Tokopedia,” *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 4, hal. 1850, 2023, doi: 10.30865/mib.v7i4.6839.