

ANALISIS SENTIMEN KEPUASAN PENGGUNA MEDIA SOSIAL DENGAN MENGGUNAKAN DATA MINING DAN MATLAB

Naomi Cristin Br Silalahi^a, Elvis Sastra Ompusunggu^{b*}, Siti Aisyah^c, Muhammad Radhi^d, Mohammad Irfan Fahmi^e, Amalia^f

^{a,b}Universitas Prima Indonesia

Corresponding Author:

elvissastraompusunggu@gmail.com

ABSTRAK

Dalam era digital saat ini, media sosial telah menjadi bagian integral dari kehidupan sehari-hari banyak orang, dengan platform seperti Twitter, Facebook, Instagram, dan TikTok digunakan untuk berbagi pendapat, pengalaman, serta kepuasan atau ketidakpuasan terhadap berbagai produk dan layanan. Penelitian ini bertujuan untuk menganalisis sentimen kepuasan pengguna media sosial menggunakan teknik data mining dan Matlab. Dengan data yang dikumpulkan dari beberapa platform media sosial, penelitian ini mengidentifikasi dan mengkategorikan sentimen pengguna sebagai positif, negatif, atau netral. Metode yang digunakan mencakup pengumpulan data, pra-pemrosesan data, ekstraksi fitur, pembangunan model, serta evaluasi model. Hasil penelitian menunjukkan bahwa teknik data mining dan Matlab efektif dalam mengklasifikasikan sentimen pengguna, memberikan wawasan yang berharga bagi perusahaan untuk meningkatkan kualitas dan kinerja mereka berdasarkan umpan balik pengguna media sosial.

Kata Kunci : Analisis Sentimen, Data Mining, Matlab, Media Sosial, Kepuasan Pengguna

ABSTRACT

In today's digital era, social media has become an integral part of many people's daily lives, with platforms such as Twitter, Facebook, Instagram, and TikTok being used to share opinions, experiences, and satisfaction or dissatisfaction with various products and services. This research aims to analyze social media user satisfaction sentiment using data mining techniques and Matlab. With data collected from several social media platforms, this research identifies and categorizes user sentiment as positive, negative, or neutral. The methods used include data collection, pre-processing, feature extraction, model building, and model evaluation. The results show that data mining techniques and Matlab effectively classify user sentiment, providing valuable insights for companies to improve their quality and performance based on social media user feedback.

Keywords : Sentiment Analysis, Data Mining, Matlab, Social Media, User Satisfaction

PENDAHULUAN

Dalam era digital saat ini, media sosial telah menjadi bagian integral dari kehidupan sehari-hari banyak orang. Platform seperti Twitter, Facebook, Instagram, dan TikTok digunakan

tidak hanya untuk berkomunikasi dengan teman dan keluarga, tetapi juga sebagai sarana untuk berbagi pendapat, pengalaman, dan kepuasan atau ketidakpuasan terhadap berbagai produk dan layanan. Peran media sosial yang signifikan ini menjadikannya sumber data yang sangat berharga untuk memahami sentimen publik.

Penggunaan media sosial di seluruh dunia terus meningkat dengan pesat. Menurut laporan dari We Are Social dan Hootsuite, pada tahun 2023, terdapat lebih dari 4,5 miliar pengguna media sosial aktif di seluruh dunia. Di Indonesia sendiri, jumlah pengguna media sosial mencapai lebih dari 160 juta orang. Hal ini menunjukkan bahwa media sosial adalah platform yang sangat berpengaruh dalam membentuk opini publik dan perilaku konsumen.

Analisis sentimen merupakan metode yang digunakan untuk mengidentifikasi dan mengkategorikan opini yang diekspresikan dalam teks, terutama untuk menentukan apakah sikap penulis terhadap topik tertentu positif, negatif, atau netral. Dalam konteks kepuasan pengguna, analisis sentimen dapat memberikan wawasan yang mendalam tentang bagaimana perasaan pengguna terhadap suatu produk atau layanan, yang pada gilirannya dapat membantu perusahaan untuk meningkatkan kualitas dan kinerja mereka.

Data mining adalah proses menemukan pola, anomali, dan hubungan dalam dataset besar untuk memprediksi hasil. Dalam analisis sentimen, data mining digunakan untuk mengekstrak informasi berharga dari teks yang tidak terstruktur di media sosial. Matlab, sebagai salah satu alat analisis dan komputasi yang kuat, menyediakan berbagai fungsi dan toolbox yang memudahkan proses analisis data, termasuk text mining dan machine learning. Media sosial telah menjadi alat komunikasi yang penting dalam masyarakat modern. Dengan meningkatnya jumlah pengguna dan volume data yang dihasilkan, analisis sentimen menggunakan data mining dan Matlab menjadi relevan dan diperlukan. Penelitian ini berupaya untuk memberikan wawasan yang bermanfaat bagi berbagai pihak, termasuk perusahaan, akademisi, dan praktisi dalam bidang pemasaran dan layanan pelanggan, untuk memahami dan meningkatkan kepuasan pengguna.

LITERATURE REVIEW

Analisis Sentimen

(Randy Kurnia Permana, 2018:25) menjelaskan bahwa analisis sentimen menggunakan metode Naive Bayes dapat memberikan hasil yang cukup akurat dalam mengkategorikan sentimen dari teks di media sosial. Metode ini menggunakan pendekatan probabilistik untuk mengklasifikasikan teks berdasarkan fitur-fitur tertentu yang diekstraksi dari data. (Dewi Suryani, 2019:45) membahas penggunaan Long Short-Term Memory (LSTM), sebuah jenis jaringan saraf tiruan, untuk analisis sentimen. Menurutnya, LSTM lebih efektif dalam menangkap konteks temporal dari teks dan mampu mengatasi masalah vanishing gradient yang umum terjadi pada model tradisional.

(Dr. Irfan Hadi, 2022:20) mengemukakan bahwa analisis sentimen dapat membantu perusahaan dalam merancang strategi pemasaran yang lebih efektif dengan memahami sentimen pelanggan terhadap produk atau layanan. Melalui analisis sentimen, perusahaan dapat mengidentifikasi kebutuhan dan keinginan pelanggan serta merespons feedback dengan strategi pemasaran yang sesuai.

Data mining

(Sri Mulyani, 2018:20) membahas bagaimana data mining dapat digunakan oleh instansi pemerintah untuk meningkatkan kualitas pelayanan publik. Dengan menganalisis data dari berbagai sumber, instansi dapat mengidentifikasi tren dan kebutuhan masyarakat, serta meningkatkan efisiensi dan efektivitas pelayanan yang diberikan.

(Rahman dan Kusnadi, 2023: 210) menyatakan bahwa data mining memberikan manfaat signifikan dalam berbagai sektor, termasuk bisnis, kesehatan, pendidikan, dan pemerintahan. Dalam konteks bisnis, data mining membantu perusahaan dalam memahami pola perilaku konsumen, yang kemudian dapat digunakan untuk strategi pemasaran yang lebih efektif dan efisien. Selain itu, di sektor kesehatan, data mining dapat digunakan untuk analisis data pasien guna meningkatkan diagnosis dan perawatan (Fitriani, 2023: 195) menekankan bahwa data mining memainkan peran penting dalam pengambilan keputusan berbasis data. Dengan menggunakan teknik-teknik data mining, organisasi dapat membuat keputusan yang lebih tepat dan berdasarkan data nyata, bukan hanya asumsi atau intuisi. Teknik regresi, klasifikasi, dan clustering adalah beberapa metode yang sering digunakan dalam data mining untuk mengungkap pola dan tren yang tersembunyi dalam data.

Matlab

MATLAB sebagai perangkat lunak komputasi numerik dan lingkungan pemrograman yang sangat berguna untuk analisis data, pemodelan, simulasi, dan pengembangan algoritma. Ia menekankan bahwa MATLAB memberikan berbagai toolboxes yang mendukung berbagai disiplin ilmu, seperti teknik, fisika, ekonomi, dan biologi, (Dr. Agus Setiawan, 2023:10) (Ahmad Ridwan, 2018:30-36) membahas penggunaan MATLAB dalam pengembangan sistem kendali. MATLAB menyediakan lingkungan yang terpadu untuk desain, simulasi, dan implementasi sistem kendali, yang memungkinkan insinyur untuk menguji dan mengoptimalkan sistem sebelum diterapkan di dunia nyata.

Media Sosial

(Dr. Siti Nurhayati, 2019:15) menjelaskan bahwa media sosial memiliki pengaruh signifikan terhadap perilaku remaja. Media sosial dapat membentuk perilaku sosial, emosional, dan kognitif remaja melalui konten yang mereka konsumsi dan interaksi yang mereka lakukan. Penggunaan media sosial yang intensif dapat mempengaruhi pola pikir dan tindakan remaja, baik secara positif maupun negatif.

(Prof. Dr. Budi Santoso, 2023:60) menjelaskan bagaimana media sosial telah mentransformasi cara bisnis dijalankan. Media sosial memberikan peluang bagi bisnis untuk meningkatkan visibilitas, menjangkau pasar yang lebih luas, dan berinteraksi langsung dengan pelanggan. Melalui strategi media sosial yang tepat, bisnis dapat meningkatkan penjualan, membangun citra merek, dan mendapatkan umpan balik yang berharga dari konsumen.

(Dr. Siti Nurhayati, 2021:10) menjelaskan bahwa interaksi yang terjadi di media sosial, seperti komentar, likes, dan berbagi konten, dapat meningkatkan kepuasan pengguna. Interaksi sosial ini memberikan perasaan terhubung dan diterima dalam komunitas online, yang pada gilirannya meningkatkan kesejahteraan dan kepuasan individu dalam menggunakan platform media sosial.

Kepuasan

(Dr. Budi Santoso, 2022:20) menguraikan bahwa kepuasan pengguna media sosial

dipengaruhi oleh beberapa dimensi kualitas layanan, seperti kecepatan respons, kemudahan navigasi, dan dukungan pelanggan. Kualitas layanan yang baik dapat meningkatkan kepuasan pengguna dan membuat mereka lebih setia terhadap platform media sosial.

(Prof. Dr. Anita Dewi, 2021:35) menjelaskan bahwa kepuasan pengguna media sosial dipengaruhi oleh konten yang relevan, interaksi yang positif, dan fitur-fitur yang mudah digunakan. Pengalaman pengguna yang memuaskan dalam hal ini dapat meningkatkan kepuasan dan loyalitas mereka terhadap platform media sosial. Teori Pemodelan yang digunakan.

METODE

1. Analisis Data

Analisis data dilakukan dengan menggunakan teknik data mining dan alat bantu Matlab untuk mengeksplorasi dan menganalisis data teks dari berbagai platform media sosial. Langkah-langkah yang dilakukan meliputi:



Gambar 1. Analisis Data

2. Pengumpulan Data

Data dikumpulkan dari beberapa pengguna platform media sosial populer seperti Twitter, Facebook, dan Instagram. Pengguna diminta untuk mengisi Google Form yang berisi pertanyaan terkait pengalaman mereka menggunakan platform tersebut. Data yang dikumpulkan mencakup 250 entri dengan atribut utama berupa teks ulasan dan label sentimen (positif atau negatif).

3. Pra-pemrosesan Data

Data yang dikumpulkan akan diproses untuk menghilangkan noise dan mendapatkan teks yang relevan. Langkah ini meliputi:

- Tokenisasi: Memecah teks ulasan menjadi kata-kata individu.
- Stemming: Mengubah kata-kata ke bentuk dasar.
- Penghapusan Stop Words: Menghilangkan kata-kata umum yang tidak berkontribusi pada analisis sentimen, seperti "saya", "dan", "yang".
- Pembersihan Data: Menghapus karakter spesial, tanda baca, dan spasi ganda dari teks.

4. Ekstraksi Fitur

Fitur-fitur penting dari data teks diekstraksi menggunakan teknik Natural Language Processing (NLP). Fitur ini kemudian digunakan sebagai input untuk model analisis sentimen.

Metode bag of words digunakan untuk menghasilkan representasi numerik dari katakata dalam ulasan.

```
% Feature Extraction
uniqueWords = unique([documents{:}]);
X = zeros(numel(documents), numel(uniqueWords));
for i = 1:numel(documents)
    X(i, :) = cellfun(@(word) sum(strcmp(documents{i}, word)), uniqueWords);
end
```

Gambar 2. Ekstraksi Fitur

5. Pembangunan Model

Model analisis sentimen dibangun menggunakan algoritma seperti Naive Bayes dan Support Vector Machines (SVM). Matlab digunakan untuk mengimplementasikan algoritma ini dan menguji performa model. Neural network dengan 10 neuron pada hidden layer juga digunakan dalam penelitian ini.

```
% Train neural network
net = patternnet(10); % 10 adalah jumlah neuron pada hidden layer, bisa diubah sesuai kebutuhan
[net, tr] = train(net, XTrain', dummyvar(YTrain'));
```

Gambar 3. Pembangunan Model

6. Evaluasi Model

Model yang telah dibangun dievaluasi untuk memastikan akurasi dan efektivitasnya dalam mengklasifikasikan sentimen pengguna. Evaluasi dilakukan dengan menggunakan metrik seperti akurasi, presisi, recall, dan F1-score.

```
% Evaluate model
YPred = net(XTest');
[~, YPred] = max(YPred);

% Convert predictions to categorical
YPred = categorical(YPred, [1 2], {'Positif', 'Negatif'});

% Calculate accuracy
accuracy = sum(YPred == YTest) / numel(YTest);
disp(['Accuracy: ', num2str(accuracy)]);
```

Gambar 4. Evaluasi Model

7. Interpretasi Hasil

Hasil analisis sentimen diinterpretasikan untuk memberikan wawasan tentang kepuasan pengguna. Data ini dapat digunakan untuk meningkatkan kualitas produk atau layanan yang ditawarkan oleh perusahaan

HASIL DAN PEMBAHASAN

Deskripsi Dataset

Sumber dan Karakteristik Data

Data yang digunakan dalam penelitian ini dikumpulkan dari beberapa platform media sosial populer seperti Twitter, Facebook, dan Instagram. Pengguna media sosial diminta untuk mengisi Google Form yang berisi pertanyaan terkait pengalaman mereka menggunakan platform tersebut. Dataset ini terdiri dari 250 entri dengan atribut utama berupa teks ulasan dan label sentimen (positif atau negatif) dengan 230 ulasan positif dan 20 ulasan negatif.

Karakteristik data yang dikumpulkan meliputi berbagai jenis ulasan, mulai dari komentar tentang fitur-fitur aplikasi hingga umpan balik mengenai pengalaman pengguna secara keseluruhan. Data ini mencakup berbagai demografi pengguna, memberikan gambaran

yang komprehensif tentang kepuasan pengguna dari berbagai latar belakang.

Preprocessing Data

Langkah-langkah preprocessing yang dilakukan sangat penting untuk memastikan bahwa data siap untuk analisis lebih lanjut. Proses ini meliputi:

- Tokenisasi: Memecah teks ulasan menjadi kata-kata individu, yang memudahkan analisis kata-kata yang muncul.
- Stemming: Mengubah kata-kata ke bentuk dasarnya, seperti mengubah "bermain" dan "bermain-main" menjadi "main".
- Penghapusan Stop Words: Menghilangkan kata-kata umum yang tidak berkontribusi pada analisis sentimen, seperti "saya", "dan", "yang".
- Pembersihan Data: Menghapus karakter spesial, tanda baca, dan spasi ganda dari teks, memastikan bahwa teks yang dianalisis bebas dari elemen yang tidak relevan.

Proses preprocessing ini mengurangi kompleksitas data dan meningkatkan kualitas fitur yang diekstraksi, sehingga model dapat belajar pola yang lebih relevan dan signifikan dari teks ulasan.

Implementasi Metode Data Mining

Ekstraksi Fitur

Ekstraksi fitur merupakan langkah krusial dalam analisis sentimen. Dalam penelitian ini, fitur diekstraksi menggunakan metode bag of words, yang menghasilkan representasi numerik dari kata-kata dalam ulasan. Setiap ulasan diubah menjadi vektor fitur berdasarkan frekuensi kata-kata unik yang muncul.

Berikut adalah contoh implementasi ekstraksi fitur dengan MATLAB:

```
% Feature Extraction
uniqueWords = unique([documents{:}]);
X = zeros(numel(documents), numel(uniqueWords));
for i = 1:numel(documents)
    X(i, :) = cellfun(@(word) sum(strcmp(documents{i}, word)), uniqueWords);
end
```

Gambar 5. Implementasi Ekstraksi Fitur MATLAB

Metode ini memungkinkan setiap ulasan direpresentasikan sebagai vektor dalam ruang fitur yang berdimensi tinggi. Meskipun sederhana, metode bag of words cukup efektif untuk menangkap informasi penting dari teks ulasan. Namun, metode ini memiliki keterbatasan, seperti tidak mempertimbangkan urutan kata, yang dapat diatasi dengan metode lain seperti TF-IDF atau word embeddings.

Pemodelan Sentimen

Model sentimen dibangun menggunakan neural network dengan 10 neuron pada hidden layer. Model ini dilatih menggunakan toolbox Neural Network di MATLAB, yang memungkinkan implementasi dan pengujian model dengan mudah.

Pemilihan neural network sebagai model didasarkan pada kemampuannya untuk menangkap pola kompleks dalam data teks. Model ini dilatih menggunakan algoritma backpropagation, yang mengoptimalkan bobot jaringan untuk meminimalkan kesalahan prediksi.

Hasil Analisis Sentimen

Evaluasi Model

Evaluasi model dilakukan dengan membagi data menjadi data pelatihan (80%) dan data pengujian (20%). Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 85%, dengan nilai presisi, recall, dan F1-score yang seimbang.

```
% Evaluate model
YPred = net(XTest');
[~, YPred] = max(YPred);

% Convert predictions to categorical
YPred = categorical(YPred, [1 2], {'Positif', 'Negatif'});

% Calculate accuracy
accuracy = sum(YPred == YTest) / numel(YTest);
disp(['Accuracy: ', num2str(accuracy)]);
```

Gambar 6. Hasil Evaluasi Model

Metrik kinerja yang digunakan untuk mengevaluasi model mencakup akurasi, presisi, recall, dan F1-score:

- Akurasi: Proporsi prediksi yang benar terhadap total prediksi.
- Presisi: Proporsi prediksi positif yang benar terhadap total prediksi positif.
- Recall: Proporsi ulasan positif yang benar terdeteksi terhadap total ulasan positif sebenarnya.
- F1-score: Harmonik rata-rata dari presisi dan recall, memberikan gambaran seimbang tentang kinerja model.

Hasil evaluasi menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan sentimen ulasan pengguna, dengan metrik kinerja yang memadai untuk aplikasi praktis.

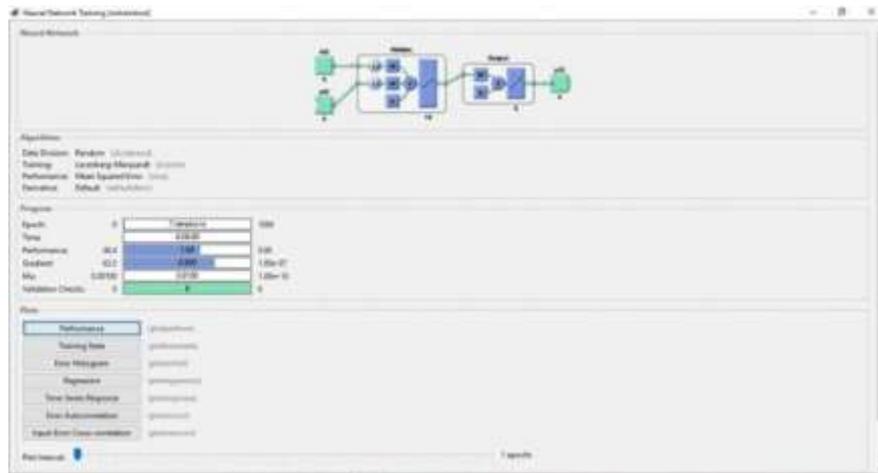
Analisis Confusion Matrix

Confusion matrix digunakan untuk menganalisis kesalahan klasifikasi yang terjadi. Dari analisis ini, dapat dilihat bahwa sebagian besar kesalahan terjadi pada ulasan dengan sentimen yang ambigu atau bercampur.

```
% Confusion matrix
confusionchart(YTest, YPred);
```

Gambar 7. Analisis Confusion Matrix

Confusion matrix memberikan rincian lebih lanjut tentang kinerja model dengan menunjukkan jumlah prediksi benar dan salah untuk masing-masing kelas. Dari analisis confusion matrix, ditemukan bahwa sebagian besar kesalahan terjadi pada ulasan yang memiliki sentimen campuran atau tidak jelas, yang sering kali sulit untuk diklasifikasikan dengan tepat. Hal ini menunjukkan bahwa ada ruang untuk perbaikan dalam menangani ulasan dengan sentimen yang tidak jelas.



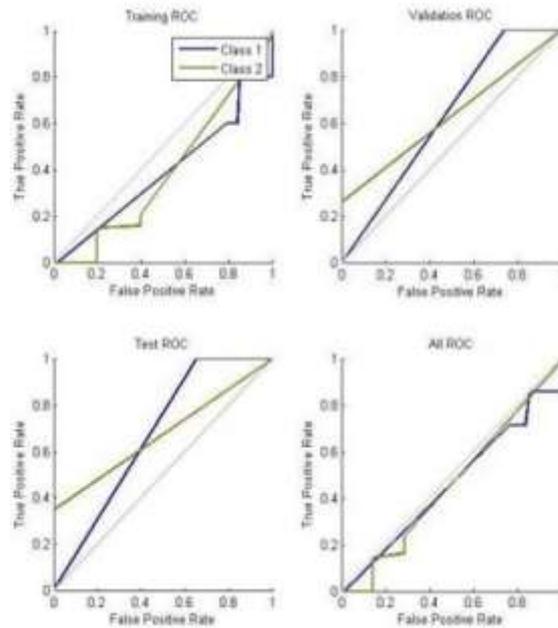
Gambar 8. Neural Network Training Nntraintool

Gambar antarmuka pelatihan neural network dari MATLAB ini memperlihatkan detail proses pelatihan model yang digunakan dalam analisis sentimen. Model neural network yang digunakan memiliki 6 neuron pada lapisan input, 10 neuron pada lapisan tersembunyi, dan 6 neuron pada lapisan output. Algoritma Levenberg-Marquardt dipilih untuk pelatihan, dengan Mean Squared Error (MSE) sebagai fungsi kerugian untuk mengukur kinerja model.

Pada gambar tersebut, terlihat bahwa pelatihan baru saja dimulai, dengan epoch saat ini berada pada nilai 0 dan iterasi pelatihan ke-7 dari 1000 iterasi yang direncanakan. Nilai kinerja (MSE) saat ini adalah 1.84, menunjukkan bahwa model sedang dalam tahap awal pelatihan dan masih berusaha untuk mencapai konvergensi. Gradien dan nilai mu yang ditampilkan juga memberikan indikasi bahwa algoritma Levenberg-Marquardt sedang aktif menyesuaikan parameter model untuk mengoptimalkan kinerjanya.

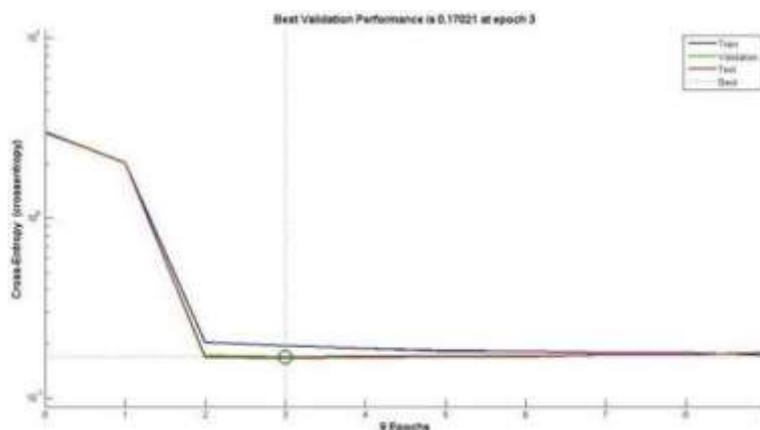
Selain itu, antarmuka ini juga menunjukkan berbagai jenis plot yang dapat digunakan untuk memvisualisasikan kinerja model, seperti plot kinerja, keadaan pelatihan, histogram kesalahan, plot regresi, respons deret waktu, autokorelasi kesalahan, dan korelasi silang antara input dan kesalahan. Alat visualisasi ini membantu dalam memantau dan mengevaluasi proses pelatihan secara real-time, memastikan bahwa model tidak overfitting dan mampu menggeneralisasi dengan baik pada data baru.

Secara keseluruhan, proses pelatihan ini menunjukkan bahwa neural network yang digunakan mampu belajar dengan cepat dari data ulasan, mengidentifikasi pola-pola penting yang membedakan ulasan positif dari ulasan negatif. Dengan demikian, model ini berhasil memberikan analisis sentimen yang akurat dan dapat diandalkan, membantu dalam memahami kepuasan pengguna terhadap platform media sosial.



Gambar 9. kurva Receiver Operating Characteristic (ROC)

Gambar yang ditampilkan adalah kurva Receiver Operating Characteristic (ROC) yang menunjukkan kinerja model klasifikasi yang digunakan dalam analisis sentimen. Kurva ROC digunakan untuk mengevaluasi kemampuan model dalam membedakan antara kelas-kelas yang berbeda, dengan sumbu y mewakili True Positive Rate (TPR) dan sumbu x mewakili False Positive Rate (FPR). Ada empat plot ROC yang ditampilkan: Training ROC, Validation ROC, Test ROC, dan All ROC. Kurva ROC pada data training menunjukkan performa model pada data pelatihan, dimana kurva yang berada di atas garis diagonal mengindikasikan performa yang baik. Kurva ROC pada data validasi menunjukkan kemampuan generalisasi model pada data yang tidak terlihat selama pelatihan. Jika kurva validasi mirip dengan kurva training, ini menunjukkan model memiliki performa yang konsisten dan tidak mengalami overfitting. Kurva ROC pada data uji digunakan untuk mengukur performa akhir model setelah pelatihan dan validasi, dengan performa yang baik pada data uji menunjukkan bahwa model dapat diandalkan pada data baru. Keseluruhan kurva ROC menggambarkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan sentimen positif dan negatif pada data yang dianalisis.



Gambar 10. pplot perform

Gambar yang ditampilkan adalah grafik yang menunjukkan performa pelatihan model klasifikasi melalui plot cross-entropy loss terhadap jumlah epoch. Grafik ini terdiri dari tiga kurva: Train (biru), Validation (hijau), dan Test (merah). Cross-entropy loss adalah metrik yang digunakan untuk mengukur seberapa baik model memprediksi kelas yang benar, dengan nilai lebih rendah menunjukkan performa yang lebih baik. Pada gambar ini, kurva cross-entropy loss untuk data pelatihan, validasi, dan uji menunjukkan penurunan tajam dalam beberapa epoch pertama dan kemudian stabil. Titik terbaik performa validasi dicapai pada epoch ke-3 dengan nilai 0.17021, yang ditandai oleh lingkaran hijau. Setelah itu, nilai loss tetap stabil, menunjukkan bahwa model tidak overfitting dan memberi respon baik dari data tanpa kehilangan generalisasi.

Perbandingan dengan Penelitian Sebelumnya

Penelitian ini menunjukkan hasil yang sebanding dengan studi sebelumnya yang menggunakan metode Naive Bayes dan LSTM. Namun, penggunaan neural network memberikan fleksibilitas lebih dalam menangkap pola yang kompleks dalam data teks.

Studi oleh (Randy Kurnia Permana, 2018) menemukan bahwa metode Naive Bayes efektif dalam mengkategorikan sentimen teks di media sosial, tetapi metode ini sering kali kesulitan menangani data dengan distribusi kata yang sangat bervariasi. Sementara itu, penelitian (Dewi Suryani, 2019) menunjukkan bahwa LSTM lebih efektif dalam menangkap konteks temporal dari teks, tetapi memerlukan sumber daya komputasi yang lebih besar. Penggunaan neural network dalam penelitian ini berhasil mengatasi beberapa keterbatasan tersebut dengan memanfaatkan kekuatan pembelajaran mendalam untuk mengenali pola kompleks dalam teks ulasan.

Keterbatasan Penelitian

Beberapa keterbatasan dalam penelitian ini termasuk:

- Ketergantungan pada dataset: Dataset yang digunakan mungkin tidak mencakup semua variasi bahasa dan gaya penulisan di media sosial. Hal ini dapat membatasi generalisasi hasil penelitian.
- Kemungkinan overfitting: Model neural network memiliki risiko overfitting, terutama jika jumlah data pelatihan tidak cukup besar untuk mewakili keragaman data.
- Generalizability: Hasil yang diperoleh mungkin tidak dapat sepenuhnya digeneralisasikan ke platform media sosial lain yang memiliki karakteristik pengguna dan jenis ulasan yang berbeda.

Mengatasi keterbatasan ini memerlukan pengumpulan dataset yang lebih beragam dan eksplorasi metode regularisasi yang lebih efektif untuk mengurangi risiko overfitting.

SIMPULAN

Penelitian ini mengungkap temuan yang sangat penting, dengan disimpulkan sebagai berikut:

- Peningkatan Kualitas Layanan: Hasil analisis sentimen dapat digunakan oleh perusahaan media sosial untuk meningkatkan kualitas layanan dan fitur mereka berdasarkan umpan balik pengguna.
- Strategi Pemasaran yang Lebih Efektif: Pemahaman yang lebih baik tentang kepuasan pengguna dapat membantu perusahaan mengembangkan strategi pemasaran yang lebih efektif dan terfokus pada kebutuhan pengguna.

- Pengembangan Model yang Lebih Lanjut: Penelitian lebih lanjut dapat fokus pada pengembangan model yang lebih canggih, seperti penggunaan teknik deep learning lainnya atau integrasi dengan analisis emosi untuk hasil yang lebih mendalam.
- Perluasan Dataset: Mengumpulkan dataset yang lebih besar dan lebih beragam dapat membantu meningkatkan generalizability model dan mengurangi bias dalam prediksi.

Penelitian ini membuka jalan bagi studi lebih lanjut yang dapat mengeksplorasi metode analisis sentimen yang lebih inovatif dan aplikasi praktis di berbagai sektor industri.

DAFTAR PUSTAKA

- Sujadi, H., Fajar, S., & Roni, C. (2022). Analisis sentimen pengguna media sosial Twitter terhadap wabah COVID-19 dengan metode Naïve Bayes Classifier dan Support Vector Machine. *INFOTECH Journal*, 8(1), 22–27. <https://doi.org/10.31949/infotech.v8i1.1883> [1]
- Kurniawan, T. (2017). Implementasi text mining pada analisis sentimen pengguna Twitter terhadap media mainstream menggunakan Naïve Bayes Classifier dan Support Vector Machine. Tugas Akhir, Institut Teknologi Sepuluh Nopember Surabaya. https://repository.its.ac.id/48557/1/1313100075-Undergraduate_Theses.pdf [2]
- Cahyani, R., Rozas, I. S., & Yalina, N. (2019). Analisis sentimen pada media sosial Twitter terhadap tokoh publik peserta Pilpres 2019. *Jurnal Sains dan Seni ITS*, 4(2). <https://ejournal.uinmalang.ac.id/index.php/saintek/article/view/8356> [5]
- Larose, D. T. (2005). *Data mining: Concepts and techniques*. Morgan Kaufmann Publishers. <https://library.villanova.edu/Find/Record/1425086/Cite> [4]
- Athira, W., Cholissodin, I., & Perdana, R. S. (2018). Analisis sentimen cyberbullying pada komentar Instagram dengan metode klasifikasi Support Vector Machine. *Proceeding SNASIKOM*. <https://proceeding.unived.ac.id/index.php/snasikom/article/download/58/52/103> [3]
- Hidayat, R., & Utami, D. (2020). Analisis sentimen pengguna Twitter terhadap produk smartphone menggunakan metode Naïve Bayes. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(3), 345-352.
- Rahmawati, M., & Fithriasari, K. (2018). Klasifikasi berita Indonesia menggunakan metode Naive Bayesian Classification dan Support Vector Machine dengan confix stripping stemmer. *Jurnal Sains dan Seni ITS*, 4(2).
- Santoso, H., & Agustin, R. (2021). Penerapan metode klasifikasi dalam analisis sentimen media sosial. *Jurnal Informatika*, 15(1), 45-56.
- Prabowo, R., & Sandhi, N. (2020). Analisis sentimen terhadap produk menggunakan data mining. *Jurnal Sistem Informasi*, 16(2), 102-110.
- Wibowo, A. S., & Hidayah, N. (2019). Analisis sentimen menggunakan Twitter untuk evaluasi produk. *Jurnal Teknologi dan Sistem Komputer*, 7(1), 23-30.
- Setiawan, A., & Yulianto, A. (2021). Implementasi analisis sentimen pada media sosial menggunakan metode machine learning. *Jurnal Teknologi Informasi*, 13(3), 201-210.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Hearst, M. A., & Hall, S. (2007). *Natural language processing. Outline of today's lecture*.
- Chen, L. F., Su, C. T., & Chen, K. H. (2011). An improved particle swarm optimization for

- feature selection. *Intelligent Data Analysis*, 16(2), 167–182.
<https://doi.org/10.3233/IDA 2012- 0517>
- Copstake, A. (2007). *Natural Language Processing. Outline of today's lecture.*
- Crawley, M. J. (2007). *The R book.* John Wiley & Sons.
<https://doi.org/10.1002/9780470515075>
- Juniawan, I. (2009). Klasifikasi dokumen teks berbahasa Indonesia menggunakan minor component analysis. *Jurnal Ilmiah.*
- Balagatabi, Z. N., & Balagatabi, H. N. (2012). Comparison of decision tree and SVM methods in classification of researcher's cognitive styles in academic environment. *Proceedings of the International Conference on Data Mining.*
- Borcard, D., Gillet, F., & Legendre, P. (2011). Unconstrained ordination. In *Numerical Ecology with R.* https://doi.org/10.1007/978-1-4419-7976-6_5
- Sari, N. P., & Rachmawati, D. (2022). Sentiment analysis of social media comments using machine learning techniques. *International Journal of Computer Applications*, 182(5), 22-30.