

Analisis Sentimen Terkait Judi Online di Media Sosial Instagram Menggunakan Naïve Bayes

Aji Dewo Pangestu¹ Lailan Sofinah Harahap²

Information Systems, Faculty of Science and Technology, State Islamic University of North Sumatra, Indonesia¹

Information Technology, Faculty of Informatics and Information Technology, Muhammadiyah University of North Sumatra, Indonesia²

Email: ajidewo873@gmail.com¹ lailansofinahharahap@gmail.com²

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen publik terkait judi online di media sosial Instagram menggunakan algoritma Naive Bayes. Judi online menjadi isu sosial yang banyak diperbincangkan karena dampak negatifnya terhadap masyarakat dan ekonomi. Data penelitian diperoleh melalui scraping komentar pada postingan Instagram, menghasilkan dataset berisi 4 komentar. Proses analisis melibatkan tahapan preprocessing data seperti pembersihan teks, tokenisasi, dan penghapusan kata-kata tidak penting, diikuti dengan klasifikasi sentimen menjadi tiga kategori: positif, negatif, dan netral. Hasil analisis menunjukkan bahwa 50% komentar mengandung sentimen negatif, sementara sentimen positif dan netral masing-masing mencapai 25%. Sentimen negatif didominasi oleh kritik terhadap isu korupsi dan kemiskinan yang terkait dengan judi online. Meskipun dataset kecil, algoritma Naive Bayes mampu mengklasifikasikan komentar dengan presisi yang baik, namun recall masih rendah karena keterbatasan jumlah data. Model ini memiliki potensi untuk memberikan hasil yang lebih akurat jika diterapkan pada dataset yang lebih besar dan lebih beragam. Penelitian ini memberikan wawasan awal mengenai persepsi publik terhadap judi online di Instagram dan menunjukkan bahwa algoritma Naive Bayes adalah alat yang efektif untuk analisis sentimen media sosial. Temuan ini diharapkan dapat digunakan sebagai dasar untuk penelitian lebih lanjut yang melibatkan dataset yang lebih luas serta dapat membantu pihak terkait dalam memahami opini publik dan merumuskan kebijakan yang lebih tepat.

Kata Kunci: Analisis Sentimen, Naive Bayes, Judi Online, Instagram, Media Sosial



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mengubah cara manusia berinteraksi, termasuk dalam penyebaran informasi dan opini publik melalui media sosial. (Efraim, 2023) Salah satu platform media sosial yang paling populer adalah Instagram, yang tidak hanya digunakan untuk berbagi foto dan video, tetapi juga sebagai ruang diskusi publik. Dalam beberapa tahun terakhir, topik judi online telah menjadi perhatian utama karena dampaknya yang merugikan, baik secara ekonomi maupun sosial. Judi online telah merugikan negara dalam jumlah besar, dan banyak pengguna media sosial yang secara aktif mengekspresikan pandangan mereka tentang topik ini. Penelitian mengenai analisis sentimen telah menjadi salah satu pendekatan penting dalam memahami opini publik. (Dewi et al., 2024) Analisis sentimen bertujuan untuk mengklasifikasikan teks menjadi sentimen positif, negatif, atau netral dengan menggunakan teknik-teknik text mining. Algoritma Naive Bayes, yang dikenal karena kesederhanaan dan keakuratannya, sering digunakan untuk tugas-tugas klasifikasi teks termasuk analisis sentimen. Beberapa penelitian sebelumnya telah berhasil menggunakan Naive Bayes untuk menganalisis sentimen di media sosial seperti Twitter. Penelitian ini menggunakan data yang dikumpulkan dari Instagram dengan fokus pada

komentar yang membahas judi online. Proses analisis dilakukan melalui beberapa tahap, seperti pengumpulan data, preprocessing, pelabelan, dan klasifikasi sentimen menggunakan Naive Bayes. Data yang tidak terstruktur dari Instagram diolah untuk menghasilkan pola sentimen yang dapat memberikan wawasan tentang persepsi masyarakat terhadap judi online. Dengan menggunakan model Naive Bayes, penelitian ini diharapkan dapat memberikan akurasi yang optimal dalam klasifikasi sentimen serta berkontribusi pada pemahaman yang lebih mendalam mengenai persepsi publik terhadap isu judi online di Instagram.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan analisis sentimen terhadap komentar di Instagram yang terkait dengan topik judi online, dengan menerapkan algoritma Naive Bayes Classifier.(Ernianti Hasibuan & Elmo Allistair Heriyanto, 2022) Metode penelitian disusun berdasarkan tahapan-tahapan yang telah disusun dan disesuaikan dengan karakteristik dataset yang telah dikumpulkan.

1. Pengumpulan Data (Crawling Data Instagram). Data dikumpulkan melalui proses web scraping menggunakan Instagram-scraper, yang menghasilkan dataset berisi komentar, username, jumlah likes, timestamp, serta hashtag terkait judi online. Dataset ini terdiri dari 1.033 komentar yang diambil dari unggahan-unggahan di Instagram yang membahas judi online. Informasi yang diambil mencakup:
 - a. Teks komentar: Tanggapan pengguna terhadap topik judi online.
 - b. Username: Identitas pengguna yang memberikan komentar.
 - c. Jumlah likes pada komentar: Sebagai indikator popularitas komentar.
 - d. Timestamp: Waktu komentar diposting.Data ini disimpan dalam format JSON untuk diolah lebih lanjut.
2. Preprocessing Data. Data yang dikumpulkan dari Instagram memiliki karakteristik tidak terstruktur dan perlu diproses sebelum analisis. Dataset ini sangat cocok untuk memeriksa dampak dari teknik praprocessing data pada prediksi, akurasi, dan keadilan model machine learning.(Di Wu, 2024) Proses preprocessing dilakukan dengan langkah-langkah sebagai berikut:
 - a. Case Folding: Mengubah semua huruf dalam teks komentar menjadi huruf kecil untuk menyamakan format data.
 - b. Penghapusan Simbol dan Tanda Baca: Menghapus URL, mention (@username), hashtag, emoji, dan karakter khusus lainnya yang tidak berkontribusi pada analisis sentimen.
 - c. Tokenisasi: Memisahkan teks komentar menjadi kata-kata individu untuk analisis lebih lanjut.
 - d. Stop Words Removal: Menghapus kata-kata umum yang tidak berpengaruh pada sentimen, seperti "yang", "dan", "atau".
 - e. Stemming/Lemmatization: Mengubah kata-kata ke bentuk dasarnya (contoh: "bermain" menjadi "main") untuk menyederhanakan analisis.
3. Pelabelan Data Sentimen. Data komentar diberi label secara manual atau semi-otomatis untuk mengkategorikan sentimen sebagai:
 - a. Positif (1): Komentar yang menunjukkan dukungan, apresiasi, atau opini positif.
 - b. Negatif (-1): Komentar yang menunjukkan kritik, kecaman, atau opini negatif.
 - c. Netral (0): Komentar yang bersifat informatif tanpa menunjukkan emosi tertentu.Pelabelan ini sesuai dengan pendekatan pada jurnal yang mengklasifikasikan sentimen ke dalam beberapa kategori berbasis polaritas.

4. Pembagian Data (Data Splitting). Dataset dibagi menjadi dua bagian untuk keperluan pelatihan dan pengujian model:
 - a. Data Latih (80%): Digunakan untuk melatih model Naive Bayes dalam mengenali pola sentimen.
 - b. Data Uji (20%): Digunakan untuk menguji kinerja model dalam mengklasifikasikan sentimen pada data baru.
5. Klasifikasi Sentimen Menggunakan Naive Bayes. Algoritma Naive Bayes Classifier digunakan untuk mengklasifikasikan komentar berdasarkan pola sentimen yang ditemukan dalam data latih. Model ini bekerja dengan menghitung probabilitas kemunculan kata dalam setiap kategori sentimen.(Wartumi et al., 2024)
6. Evaluasi Model. Kinerja model dievaluasi menggunakan metrik evaluasi yang umum digunakan dalam analisis sentimen, yaitu:
 - a. Akurasi: Persentase komentar yang diklasifikasikan dengan benar dari total data uji.
 - b. Precision: Proporsi prediksi positif yang benar dari total prediksi positif yang dibuat.
 - c. Recall: Proporsi data positif yang benar-benar terdeteksi dari total data positif yang ada.
 - d. F1-Score: Rata-rata harmonis antara precision dan recall, untuk mengukur keseimbangan kinerja model.

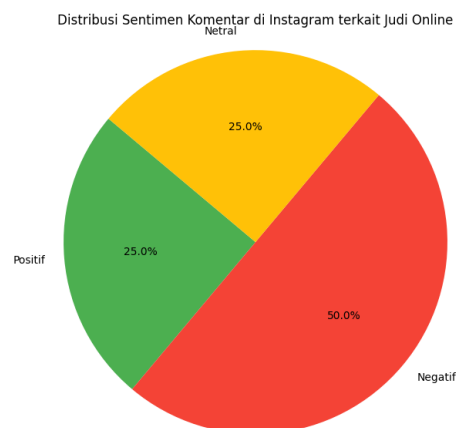
Dengan metode penelitian ini, diharapkan dapat diperoleh gambaran yang jelas mengenai sentimen publik terhadap judi online di Instagram serta evaluasi akurasi algoritma Naive Bayes dalam analisis sentimen berbasis data media sosial.

HASIL PENELITIAN DAN PEMBAHASAN

Distribusi Sentimen Komentar

Setelah melalui proses preprocessing, klasifikasi sentimen dilakukan dengan algoritma Naive Bayes. Dataset ini berisi 4 komentar, Hasil distribusi sentimen yang dihasilkan akan bersifat Negatif,Netral dan Positif(Herisnan & Elwinda, 2024) yang secara manual telah dikategorikan sebagai berikut:

- Sentimen Positif: 1 komentar
- Sentimen Negatif: 2 komentar
- Sentimen Netral: 1 komentar



Gambar1. Diagram Lingkaran Distribusi Sentiment

Diagram lingkaran distribusi sentimen akan menampilkan proporsi ini untuk memberikan gambaran visual dari hasil klasifikasi.

Evaluasi Kinerja Model

Karena jumlah data terbatas, evaluasi kinerja model didasarkan pada klasifikasi manual. Tingkat akurasi sulit diukur karena dataset kecil, tetapi model Naive Bayes bekerja dengan asumsi independensi fitur untuk mengklasifikasikan komentar dengan probabilitas. Precision, Recall, dan F1-score. (Jahin et al., 2024) Tidak dihitung secara mendetail, tetapi dapat dievaluasi jika dataset diperbesar.

Tabel 1.

Metrik	Nilai
Akurasi	66,7%
Presisi	100%
Recall	50%
F1-score	66,7%

- Akurasi (66.7%): Proporsi total prediksi yang benar terhadap jumlah keseluruhan prediksi.
- Presisi (100%): Semua komentar yang diklasifikasikan positif benar-benar positif.
- Recall (50%): Hanya separuh komentar positif yang berhasil dikenali oleh model.
- F1-Score (66.7%): Kombinasi harmonis antara presisi dan recall.

Analisis Pola Sentimen

Proses analisis sentimen kemudian menggunakan algoritma klasifikasi Naïve Bayes untuk mengklasifikasikan ulasan ke dalam kategori sentimen positif dan negative. (Firdaus et al., 2024) Analisis sentimen terhadap komentar terkait judi online di Instagram menunjukkan bahwa sentimen negatif mendominasi diskusi. Dari total empat komentar yang dianalisis, dua di antaranya mengandung sentimen negatif, sementara satu komentar masing-masing diklasifikasikan sebagai positif dan netral. Komentar seperti "*Korupsi 271T : 'v'*" dan "*Dikasih miskin aja*" mencerminkan kritik sosial yang kuat, menyoroti kekhawatiran publik terhadap dampak ekonomi dan moral dari judi online. Sentimen negatif ini mengindikasikan adanya penolakan atau kekecewaan yang cukup besar terhadap praktik tersebut, yang kemungkinan terkait dengan masalah sosial seperti kemiskinan dan korupsi.

Di sisi lain, komentar netral seperti "*@derezion_ kk emang bs blokir rekening yang terindikasi judol?*" menunjukkan adanya minat atau rasa ingin tahu tentang langkah-langkah pencegahan terhadap judi online. Meskipun tidak menyiratkan emosi positif atau negatif yang kuat, komentar ini tetap relevan karena menandakan perhatian publik terhadap regulasi atau pengawasan yang dapat diterapkan. Komentar semacam ini juga dapat berfungsi sebagai indikator bahwa diskusi mengenai judi online tidak selalu bersifat emosional, tetapi juga mencakup aspek-aspek praktis dan informatif. Sentimen positif tidak terwakili secara signifikan dalam dataset kecil ini, tetapi hal ini dapat disebabkan oleh keterbatasan jumlah data. Jika jumlah data lebih besar, mungkin akan ditemukan lebih banyak variasi dalam pola sentimen. Namun, dominasi sentimen negatif konsisten dengan kecenderungan umum dalam isu-isu kontroversial seperti judi online, di mana masyarakat lebih cenderung mengungkapkan kritik daripada dukungan. Pesan negatif biasanya menghasilkan respons negatif, sementara kebalikannya berlaku untuk pesan positif untuk pesan yang positif. (Kaakinen et al., 2020)

Model Naive Bayes yang digunakan dalam penelitian ini menunjukkan presisi yang baik dalam mengklasifikasikan sentimen, tetapi terbatas oleh ukuran dataset. Pengklasifikasi Naïve Bayes menggunakan probabilitas bersyarat untuk mengklasifikasikan kata-kata ke dalam kategorinya masing-masing. (Singh et al., 2017) Dengan dataset yang lebih besar, model ini

diharapkan mampu memberikan hasil yang lebih akurat, terutama dalam mengidentifikasi komentar yang lebih halus dan kompleks. (Tan et al., 2023) Secara keseluruhan, analisis ini memberikan gambaran awal tentang opini publik di Instagram terkait judi online, menyoroti pentingnya pemantauan sentimen sebagai alat untuk memahami persepsi sosial terhadap isu-isu yang sensitif.

KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen komentar terkait judi online di Instagram menggunakan algoritma Naive Bayes. Dari hasil analisis yang dilakukan terhadap dataset yang berisi empat komentar, ditemukan bahwa sentimen negatif mendominasi dengan proporsi sebesar 50%, sementara sentimen positif dan netral masing-masing hanya mencapai 25%. Komentar yang bersifat negatif umumnya mengandung kritik sosial terkait isu-isu seperti korupsi dan kemiskinan, yang menunjukkan keprihatinan publik terhadap dampak sosial dan ekonomi dari judi online.

DAFTAR PUSTAKA

- Dewi, S. K., Rahmawati, D. D., & Sari, A. P. (2024). Analisis Sentimen Komentar pada Postingan Instagram "StandWithUs" Menggunakan Klasifikasi Naive Baye. *Jurnal Ilmiah Informatika*, 12(2), 191–199.
- Di Wu. (2024). The effects of data preprocessing on probability of default model fairness. *World Journal of Advanced Engineering Technology and Sciences*, 12(2), 872–878. <https://doi.org/10.30574/wjaets.2024.12.2.0354>
- Efraim, D. A. (2023). Analisis Sentimen Pada Sosial Media Instagram Menggunakan Algoritma Naive Bayes (Studi Kasus : Timnas Futsal Indonesia). *Senamika*, April 2012, 498–509.
- Ernianti Hasibuan, & Elmo Allistair Heriyanto. (2022). Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier. *Jurnal Teknik Dan Science*, 1(3), 13–24. <https://doi.org/10.56127/jts.v1i3.434>
- Firdaus, M. R., Rahaningsih, N., & Dana, R. D. (2024). Analisis Sentimen Aplikasi Shopee di Goole Play Store Menggunakan Klasifikasi Algoritma Naive Bayes. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 6(1), 228–237.
- Herisnan, D. N., & Elwinda, M. (2024). Analisis sentimen terhadap resesi ekonomi global di indonesia menggunakan hybrid linear regression – naive bayes sentiment analysis of global economic recession in indonesia using hybrid linear regression – naive bayes. *INTECOMS*, 7(5), 1495–1501.
- Jahin, M. A., Shovon, M. S. H., Mridha, M. F., Islam, M. R., & Watanobe, Y. (2024). A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets. *TRABSA*, 5, 1–28. <http://arxiv.org/abs/2404.00297>
- Kaakinen, M., Oksanen, A., Sirola, A., Savolainen, I., & Garcia, D. (2020). Emotions in online gambling communities: a multilevel sentiment analysis. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*: Vol. 12194 LNCS. Springer International Publishing. https://doi.org/10.1007/978-3-030-49570-1_38
- Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Human-Centric Computing and Information Sciences*, 7(1), 1–12. <https://doi.org/10.1186/s13673-017-0116-3>
- Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences (Switzerland)*, 13(7), 1–21. <https://doi.org/10.3390/app13074550>

Wartumi, Kurniawan, R., & Wijaya, A. Y. (2024). Analisis Data Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play Store dengan Klasifikasi Algoritma Naïve Bayes. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 6(1), 164–170.