

MENGGALI OPINI PUBLIK: SENTIMEN TERHADAP KEBIJAKAN MAKAN SIANG GRATIS DENGAN SUPERVISED LEARNING

Dwija Wisnu Brata^{*1}, Welly Purnomo², Hariz Farisi³, Henry Kevin Marcelino Ratu⁴

^{1,2,3,4}Universitas Brawijaya, Malang
Email: ¹wisnubrata@ub.ac.id, ²welly@ub.ac.id, ³hariz_farisi@ub.ac.id, ⁴henrykevin@student.ub.ac.id
^{*}Penulis Korespondensi

(Naskah masuk: 12 Desember 2024, diterima untuk diterbitkan: 18 Juni 2025)

Abstrak

Di era informasi saat ini, opini publik menjadi aset penting dalam membentuk kebijakan sosial dan politik. Salah satu kebijakan yang sering menarik perhatian adalah program makan siang gratis yang ditujukan untuk meningkatkan kesejahteraan sosial. Kebijakan yang direncanakan bertujuan untuk memberikan manfaat bagi anak-anak di Indonesia agar dapat meningkatkan asupan gizi dan nutrisi, tetapi sering kali menimbulkan berbagai reaksi dari masyarakat, terutama di media sosial. Penelitian ini dilakukan untuk menggali sentimen berdasarkan opini tentang rencana kebijakan makan siang gratis oleh pemerintah Indonesia pada media sosial twitter. Total dataset yang dikumpulkan sebanyak 1359 yang telah di-*preprocessing*, terbagi atas 1000 sentimen negatif, dan 359 sentimen positif. Metode klasifikasi yang digunakan dan menjadi pembanding yaitu *Support Vector Machine* (SVM), *Naïve Bayes* (NB), dan *Random Forest* (RF). Hasil penelitian menunjukkan bahwa metode SVM memiliki tingkat akurasi yang lebih tinggi (85%) dibanding dua metode lainnya.

Kata kunci: *opini makan siang gratis, supervised learning, sentimen, opini media sosial*

EXPLORING PUBLIC OPINION: SENTIMENT TOWARDS FREE LUNCH POLICY PLAN WITH SUPERVISED LEARNING

Abstract

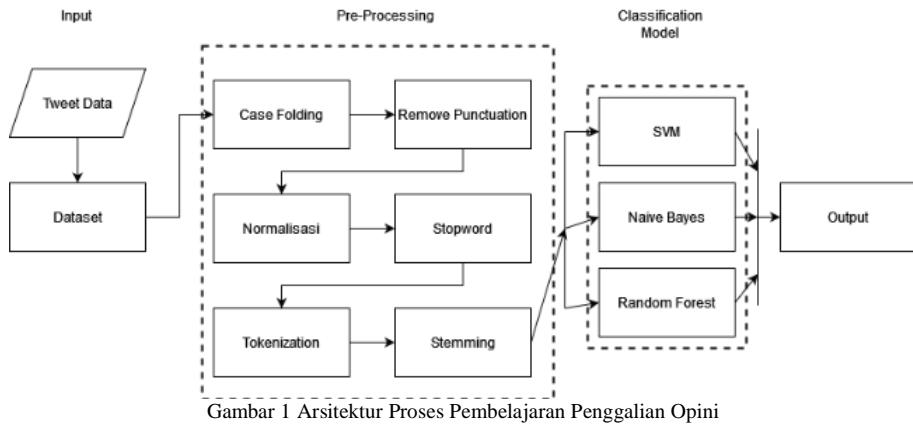
In today's information technology era, public opinion is an important asset in shaping social and political policies. One policy that often attracts attention is the free lunch program aimed at improving social welfare. The planned policy aims to benefit children in Indonesia in order to improve their nutritional intake, but it often generates various reactions from the public, especially on social media. This study was conducted to explore opinion-based sentiment about the free lunch policy plan by the Indonesian government on social media Twitter. The total dataset collected was 1359 that had been cleaned, divided into 1000 negative sentiments, and 359 positive sentiments. The classification methods used for comparison are Support Vector Machine (SVM), Naïve Bayes (NB), and Random Forest (RF). The results show that the SVM method has a higher accuracy rate (85%) than the other two methods.

Keywords: *free lunch opinion, supervised learning, sentiment, social media opinion*

1. PENDAHULUAN

Di era informasi saat ini, opini publik menjadi aset penting dalam membentuk kebijakan sosial dan politik. Kebijakan yang baik senantiasa akan berdampak positif bagi yang melaksanakan kebijakan tersebut, hal ini terlihat dari aturan-aturan yang melindungi serta mengayomi. Pada tahun 2024, terdapat peristiwa lima tahunan yaitu pemilihan presiden Indonesia, di mana terdapat salah satu pasangan yang menjanjikan kebijakan tentang makan siang gratis. Pada tanggal 29 Februari 2024 CNN Indonesia melaporkan bahwa program makan siang gratis sudah mulai disimulasikan oleh kementerian ekonomi yaitu bapak Airlangga Hartarto. Hal ini

direncanakan menurut beliau akan menjadi sebuah kebijakan dimasa yang akan datang. Kebijakan yang direncanakan bertujuan untuk memberikan manfaat bagi anak-anak di Indonesia agar dapat meningkatkan asupan gizi dan nutrisi, tetapi sering kali menimbulkan berbagai reaksi dari masyarakat, terutama di media sosial. Opini yang muncul pada media sosial terdapat yang positif dan negatif. Pemahaman terkait sentimen yang muncul dalam tulisan yang bersifat opini pada media sosial terkadang bersifat ambigu (Ekawati, Sumadyo and Whidhiasih, 2022), sehingga dibutuhkan usaha lebih untuk mengidentifikasi opini yang ditulis termasuk sentimen positif atau negatif.



Gambar 1 Arsitektur Proses Pembelajaran Penggalian Opini

Proses identifikasi sentimen dapat dilakukan dengan analisis sentimen yaitu text analytics (Wira Yudha and Wahyudi, 2018). Proses analisis sentimen juga dilakukan oleh (Pamungkas and Kharisudin, 2021) untuk menganalisis sentimen masyarakat Indonesia terhadap pandemi covid-19. Peneliti membandingkan dua metode (SVM, dan Naive Bayes) dan memperoleh akurasi metode *Support Vector Machine* (SVM) 79.20%, dan Naive Bayes 62,10%. Metode *machine learning* juga (SVM, dan Naive Bayes) digunakan dalam menganalisis sentimen pada objek review dataset dan kanker payudara. Peneliti (Widiastuti, Rainarli and Dewi, 2017) dan (Rahat, Kahir and Masum, 2019) mendapatkan hasil prediksi terhadap metode yang digunakan yaitu *Support Vector Machine* (SVM) diatas 90% untuk akurasi.

Berdasarkan penelitian yang berkaitan dengan sentimen dengan metode yang digunakan, penelitian yang dilakukan bertujuan untuk mendapatkan hasil prediksi sentimen dari tiga metode yang dibandingkan yaitu SVM, Naive Bayes, dan Random Forest yang terbaik, serta membuat dataset dari opini publik terkait rencana kebijakan pemerintah Indonesia untuk makan siang gratis.

2. PENELITIAN TERDAHULU

(Chauhan, 2017) membandingkan beberapa metode machine learning yaitu Support Vector Machine, Naïve Bayes, Decision Tree, dan Neural Network. Data yang digunakan berasal dari hasil scraping Twitter dengan keyword “Donald Trump”. Hasilnya diperoleh tingkat akurasi SVM 64% dengan presisi 70,83%, Naïve Bayes memiliki tingkat akurasi sebesar 60.40% dengan presisi 67.25%, Decision Tree memiliki tingkat akurasi sebesar 55.67 % dengan presisi 60,53%, Neural Network memiliki akurasi sebesar 54.17% dengan presisi 48.86%.

(Isnain et al., 2021) mengkaji persepsi masyarakat terhadap kebijakan *lockdown* melalui analisis sentimen di Twitter. Metode yang digunakan Support Vector Machine (SVM) dan ekstraksi fitur TF-IDF. Data yang digunakan sebanyak 2000 data yang berasal dari hasil scaping Twitter. Penelitian ini berhasil mengklasifikasikan sentimen menjadi positif

(68.75%) dan negatif (31.25%), serta model yang digunakan mampu menghasilkan akurasi sebesar 74%, presisi 75%, recall 92%, dan F1-Score 83%.

(Ardhani, Chamidah and Saifudin, 2021) melakukan penelitian tentang analisis Sentimen Pelaksanaan Kartu Prakerja yang diterapkan Pemerintah Indonesia. Penelitian yang dilakukan mengolah 500 tweet data dari periode Juli hingga Oktober 2020. Metode yang digunakan *Support Vector Machine* (SVM) dan kernel Radial Basis Function (RBF), serta memperoleh hasil sentimen sebesar 60% sentimen negatif, dan 40% sentimen positif. Tingkat akurasi dari model yang digunakan diperoleh 85,20%, sensitivitas 91,68%, spesifisitas 75.75%, prediksi sentimen negatif 85.03%; dan prediksi sentimen positif 86.04%.

3. METODE

Pada proses pembelajaran dalam mengklasifikasi sentimen dilakukan dalam 4 tahapan yaitu pengumpulan data raw, membentuk dataset, *preprocessing*, dan klasifikasi. Tahapan proses penelitian dapat terlihat pada gambar 1.

A. Dataset

Pengumpulan data dilakukan untuk membentuk sebuah dataset terkait data sentimen “makan gratis” yang akan digunakan sebagai input proses klasifikasi. Pengumpulan data dilakukan dengan melakukan scraping twitter berdasarkan keyword “makan siang gratis” dan hasilnya akan disimpan dalam format *.csv. Data raw dari hasil scraping diperoleh sebanyak 1400 data tweet dan dilakukan proses *cleaning* sehingga menjadi 1359 data tweet. Data tersebut terdiri dari 1000 data berlabel 0 (negatif) dan 350 berlabel 1(positif) (Padurariu and Breaban, 2019).

B. Machine Learning

Pada penelitian penggalian opini memanfaatkan metode *machine learning* dalam mengklasifikasi sentimen yang terdapat pada dataset. Terdapat tiga metode yang digunakan yaitu *Support Vector Machine* (SVM)(Pisner and Schnyer, 2020), *Naives Bayes* (NB), dan *Random Forest* (RF). Penerapan tiga metode dimaksudkan sebagai perbandingan hasil dari

input dataset yang disediakan, sehingga diketahui metode yang sesuai dengan dataset yang telah ada.

C. Experimental Scenario

Dalam penelitian yang dilakukan beberapa penggunaan model *machine learning* (SVM, *Naive Bayes*, dan *Random Forest*) dan beberapa *finetunning* dalam meningkatkan kinerja klasifikasi dalam mendekripsi kalimat-kalimat opini atau sentimen "makan siang gratis". Dataset terdiri dari kumpulan tweet yang diklasifikasikan sebagai sentimen negatif atau sentimen positif. Berbagai metode pra-pemrosesan teks, seperti *case folding*, *remove punctuation*, normalisasi, *stopword*, tokenisasi, dan *stemming*, diterapkan untuk menyiapkan data (Brata and Farisi, 2023). Data yang telah diproses ini kemudian digunakan untuk membuat vektor kata untuk melatih model *machine learning*. Untuk setiap percobaan 80% dari dataset dialokasikan untuk pelatihan dan 20% untuk pengujian, dengan pembagian tersebut bertujuan untuk memaksimalkan jumlah data yang digunakan dalam fase pelatihan.

D. Receiver Operating Characteristic (ROC) dan Area Under the Curve (AUC)

Proses yang dilakukan dalam mengevaluasi kinerja model klasifikasi biner digunakan ROC dan AUC. ROC adalah kurva yang menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai nilai ambang batas keputusan. Kurva ini memvisualisasikan kemampuan model untuk membedakan antara dua kelas (positif dan negatif) berdasarkan hasil prediksi. AUC merupakan ukuran tunggal yang diperoleh dari luas di bawah kurva ROC, yang memberikan ringkasan kinerja keseluruhan model. Nilai AUC berkisar antara 0 hingga 1, dengan nilai mendekati 1 menunjukkan kinerja model yang sangat baik dalam membedakan kelas. Evaluasi menggunakan ROC dan AUC menjadi penting dalam text mining karena memungkinkan identifikasi performa model secara menyeluruh tanpa terikat pada nilai ambang tertentu, yang sangat relevan ketika model diterapkan pada data teks yang kompleks dan bervariasi (M and M.N, 2015; Liu et al., 2020).

E. Evaluasi

Hasil model prediksi yang dihasilkan oleh metode SVM, NB, dan RF akan dievaluasi kinerjanya dengan menganalisis presisi (Prec), *recall* (Rec), *f1-score* (f1) dan akurasi (Acc) (Tharwat, 2021; Brata, Purnomo and Nofandi, 2024). Data analisis dari evaluasi kinerja antar metode akan dibandingkan, sehingga dapat diketahui metode yang memiliki kinerja diatas 80%.

4. HASIL DAN PEMBAHASAN

Pada proses persiapan dataset, konten yang digunakan berasal dari situs Twitter. Dataset yang

digunakan dalam penelitian ini terdiri dari 359 tweet positif dan 1000 tweet negatif yang telah dilatih menggunakan pengklasifikasi. Hal ini menghasilkan total 1359 titik data yang telah dilabeli secara sentimen untuk pengklasifikasi. Ilustrasi hasil pelabelan pada data twitter ditunjukkan pada Tabel 1.

Tabel 1. Data Tweet Hasil Pelabelan

| No | Data Tweet | Label |
|----|--|---------|
| 1 | Anggota Dewan Pakar TKN Prabowo Gibran Soedrajad Djiwandono mengatakan bahwa makan siang gratis sudah dipikirkan pentahapannya. Hal ini disampaikannya saat menjadi narasumber di salah satu acara televisi sebagaimana yang kami kutip pada Sabtu (30/3/2024) | Positif |
| 2 | @HisyamMochtar Yang ngamuk itu kan si IQ 58 penikmat bansom dan pengemis makan siang gratis? @geloraco ga ngaruh tetap gasspoll sambil ngantri makan siang gratis bansom dan serangan fajar 20K ~ 50K tiap pemilu | Negatif |
| 3 | | Negatif |

Raw dataset yang telah dikumpulkan diolah untuk didapatkan data yang valid agar dapat diproses oleh metode *machine learning*, seperti *case folding*, normalisasi, seperti terlihat pada Tabel 2.

Tabel 2. Hasil Olah Raw Data

| No | Raw Data Tweet | Normalisasi | Label |
|----|--|--|---------|
| 1 | Anggota Dewan Pakar TKN Prabowo Gibran Soedrajad Djiwandono mengatakan bahwa makan siang gratis sudah dipikirkan pentahapannya. Hal ini disampaikannya saat menjadi narasumber di salah satu acara televisi sebagaimana yang kami kutip pada Sabtu (30/3/2024) | anggota dewan pakar tkn prabowo gibran soedrajad djiwandono mengatakan bahwa makan siang gratis sudah dipikirkan pentahapannya. Hal ini disampaikannya saat menjadi narasumber di salah satu acara televisi sebagaimana yang kami kutip pada sabtu 3032024 | Positif |
| 2 | @HisyamMochtar Yang ngamuk itu kan si iq 58 penikmat bansom dan pengemis makan siang gratis? @geloraco ga ngaruh tetap gasspoll sambil ngantri makan siang gratis bansom dan serangan fajar 20 ribu 50 ribu tiap pemilu | uangnya ada itu yang ngamuk itu kan si iq 58 penikmat bansom dan pengemis makan siang gratis tidak ngaruh tetap gasspoll sambil ngantri makan siang gratis bansom dan serangan fajar 20 ribu 50 ribu tiap pemilu | Negatif |
| 3 | | | Negatif |

Tabel 3. Confusion Matrix Model

| Actual | Prediksi | | ROC |
|--------|----------|---------|---------|
| | Positif | Negatif | |
| SVM | Positif | 21(TP) | 12(FP) |
| | Negatif | 15(FN) | 88(TN) |
| NB | Positif | 5(TP) | 0(FP) |
| | Negatif | 31(FN) | 100(TN) |
| RF | Positif | 17(TP) | 10(FP) |
| | Negatif | 19(FN) | 90(TN) |

Tabel 3 menjelaskan kemampuan model yang baik secara keseluruhan untuk membedakan antara kelas positif dan negatif ditunjukkan oleh skor ROC AUC sebesar 0,82. Dengan presisi 0,8549, model meminimalkan *false positive* dengan sangat berhasil dalam menemukan situasi positif ketika memprediksi mereka. Meskipun ada sedikit *trade-off* antara presisi dan *recall*, model ini berhasil mendeteksi sebagian besar kejadian positif yang sebenarnya, sebagaimana dibuktikan dengan *recall* sebesar 0,8509. Kedua ukuran ini diimbangi dengan *F1-score* sebesar 0,8361, yang memberikan ringkasan kinerja model.

Terdapat 88 kejadian *true negative*, 21 kejadian *true positive*, 12 kejadian *false positive*, dan 15 kejadian *false negative*, berdasarkan *confusion matrix*. Hal ini menyiratkan bahwa kinerja model yang baik, masih terdapat beberapa kesalahan klasifikasi masih terjadi.

Tabel 4. Hasil Klasifikasi Model

| Metode | Prec | Rec | F1-Score | Accuracy |
|--------|----------|----------|----------|----------|
| SVM | 0,854906 | 0,850909 | 0,836127 | 0,850909 |
| NB | 0,821178 | 0,763636 | 0,685178 | 0,763636 |
| RF | 0,801465 | 0,810909 | 0,794603 | 0,810909 |

Dalam analisis Tabel 4, kedua algoritma memperoleh tingkat akurasi keseluruhan yang sangat baik yaitu lebih dari 80% (SVM, dan RF) sedangkan NB memperoleh dibawah 80%. Sejumlah penelitian sebelumnya (A. Saepulrohman, S. Saepudin and D. Gustian, 2021) juga dapat diperkuat dengan hasil eksperimen ini. Hasil perbandingan antara Naïve Bayes dan SVM, yang menunjukkan bahwa SVM lebih unggul dalam proses analisis sentimen, berbeda dengan hasil penelitian sebelumnya (Ardhani, Chamidah and Saifudin, 2021; Isnain et al., 2021).

5. KESIMPULAN

Tingkat akurasi dari analisis yang dilakukan pada penelitian ini untuk meramalkan data twitter tentang “menggali opini publik: sentimen terhadap kebijakan makan siang gratis dengan supervised learning” diperoleh metode *Support Vector Machine* (SVM) lebih tinggi akurasinya dibanding dua model lainnya, sebesar 85% sedangkan *Naïve Bayes* 76.36%, *Random Forest* 81.09%. Sikap positif dan negatif terhadap data tweet dari media sosial Twitter adalah variabel input yang digunakan. *Naïve Bayes* dan *Support Vector Machines* adalah pilihan yang sangat layak untuk memprediksi data Twitter, menghasilkan nilai prediksi positif dan negatif, berdasarkan kemampuan identifikasi (akurasi).

Secara keseluruhan proses yang ada bahwa model yang dapat mengidentifikasi sentimen dalam dataset opini “makan siang gratis” media sosial (tweet) adalah *Support Vector Machine* (SVM) dibanding dua model lainnya.

DAFTAR PUSTAKA

- A. SAEPULROHMAN, S. SAEPUDIN AND D. GUSTIAN, 2021. Analisis Sentimen Kepuasan Pengguna Aplikasi WhatsApp Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine privat maupun grup. In: *is The Best Accounting Information Systems and Information Technology Business Enterprise this is link for OJS us.*
- ARDHANI, B.A., CHAMIDAH, N. AND SAIFUDIN, T., 2021. Sentiment Analysis Towards Kartu Prakerja Using Text Mining with Support Vector Machine and Radial Basis Function Kernel. *Journal of Information Systems Engineering and Business Intelligence*, [online] 7(2), p.119. <https://doi.org/10.20473/jisebi.7.2.119-128>.
- BRATA, D.W. AND FARISI, H., 2023. Etalase Online Pedagang Produk UMKM dan Fashion Di Wilayah Kota Malang dalam Penerapan UMKM Berbasis Teknologi Menggunakan Information Retrieval. *Jurnal Ilmiah Teknologi Informasi Asia*, 17(2), p.109. <https://doi.org/10.32815/jitika.v17i2.897>.
- BRATA, D.W., PURNOMO, W. AND NOFANDI, A., 2024. Extracting Customer Reviews of Restaurants to Explore Service Aspects on Google Review and Tripadvisor as Factors for Quality Improvement. *Jurnal Ilmiah Teknologi Infomasi Asia*.
- CHAUHAN, P., 2017. Sentiment Analysis: A Comparative Study of Supervised Machine Learning Algorithms Using Rapid miner. *International Journal for Research in Applied Science and Engineering Technology*, [online] V(XI), pp.80–89. <https://doi.org/10.22214/ijraset.2017.11011>.
- EKAWATI, I., SUMADYO, M. AND WHIDHIASIH, R.N., 2022. Analisa Sentimen Masyarakat Terhadap Kebijakan Pemerintah Selama Pandemi Covid-19 Menggunakan Support Vector Machine dan Naïve Bayes. *JREC (Journal of Electrical and Electronics) ISSN*, .
- ISNAIN, A.R., SAKTI, A.I., ALITA, D. AND MARGA, N.S., 2021. Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jurnal Data Mining dan Sistem Informasi*, 2(1), p.31. <https://doi.org/10.33365/jdmsi.v2i1.1021>.
- LIU, C., ZHONG, Q., AO, X., SUN, L., LIN, W., FENG, J., HE, Q. AND TANG, J., 2020.

- Fraud Transactions Detection via Behavior Tree with Local Intention Calibration. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. [online] New York, NY, USA: ACM. pp.3035–3043. <https://doi.org/10.1145/3394486.3403354>.
- M, H. and M.N, S., 2015. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), pp.01–11. <https://doi.org/10.5121/ijdkp.2015.5201>.
- PADURARIU, C. and BREABAN, M.E., 2019. Dealing with data imbalance in text classification. In: *Procedia Computer Science*. Elsevier B.V. pp.736–745. <https://doi.org/10.1016/j.procs.2019.09.229>.
- PAMUNGKAS, F.S. and KHARISUDIN, I., 2021. Analisis Sentimen dengan SVM. [online] 4, pp.628–634. Available at: <https://journal.unnes.ac.id/sju/index.php/p_risma/>.
- PISNER, D.A. and SCHNYER, D.M., 2020. Support vector machine. In: *Machine Learning*. Elsevier. pp.101–121.
- <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>.
- RAHAT, A.M., KAHIR, A. and MASUM, A.K.M., 2019. Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset. In: *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*. IEEE. pp.266–270. <https://doi.org/10.1109/SMART46866.2019.9117512>.
- THARWAT, A., 2021. Classification assessment methods. *Applied Computing and Informatics*, 17(1), pp.168–192. <https://doi.org/10.1016/j.aci.2018.08.003>.
- WIDIASTUTI, N.I., RAINARLI, E. and DEWI, K.E., 2017. Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen. *JURNAL INFOTEL*, 9(4), p.416. <https://doi.org/10.20895/infotel.v9i4.312>.
- WIRA YUDHA, S. and WAHYUDI, M., 2018. *Komparasi Algoritma Klasifikasi Untuk Analisis Sentimen Review Film Berbahasa Asing. Sistem Informasi Dan Keamanan Siber (SEINASI-KESI)* Jakarta-Indonesia

Halaman ini sengaja dikosongkan