

Perbandingan Metode Seleksi Fitur Pada Analisis Sentimen (Studi Kasus Opini PILKADA DKI 2017)

Edwar¹, I Gusti Agung Ngurah Rai Semadi², Muhamad Samsudin³, I Komang Dharmendra^{4,*}

¹ Manajemen Informatika; Institut Teknologi dan Bisnis STIKOM Bali; Jalan Raya Puputan No. 86 Renon, Denpasar, (0361) 244445; e-mail: edwar.ridwan@stikom-bali.ac.id.

^{2,3,4} Sistem Informasi; Institut Teknologi dan Bisnis STIKOM Bali; Jalan Raya Puputan No. 86 Renon, Denpasar, (0361) 244445; e-mail: ²raisemadi@stikom-bali.ac.id, ³samsudin@stikom-bali.ac.id, ⁴dharmendra@stikom-bali.ac.id

* Korespondensi: e-mail: dharmendra@stikom-bali.ac.id

Diterima: 29 Juni 2023 ; Review: 04 Juli 2023; Disetujui: 11 Juli 2023

Cara sitasi: Edwar, Semadi IGA, Samsudin M, Dharmendra IK. 2023. Perbandingan Metode Seleksi Fitur Pada Analisis Sentimen (Studi Kasus Opini PILKADA DKI 2017). Informatics for Educators and Professionals : Journal of informatics. Vol.8 (1) : 11 – 18.

Abstrak: Dalam analisis sentimen, seleksi fitur menjadi tahap yang penting karena mampu meningkatkan kinerja dan efisiensi model analisis sentimen, dan seleksi fitur juga membantu dalam mengurangi dimensi data yang kompleks, sehingga memungkinkan proses analisis menjadi lebih cepat dan efisien. Dimana dalam prosesnya menjadi tantangan dalam memilih fitur yang relevan karena jika salah memilih fitur maka bisa mengurangi tingkat akurasi dari model yang dibangun. Pada penelitian ini dilakukan analisis sentimen pada data tweet pada pemilihan Kepala Daerah DKI Jakarta pada tahun 2017 dengan menggunakan seleksi fitur TF-IDF yang dikombinasikan dengan *Recursive Feature Elimination (RFE)*, *Chi Square*, dan *Mutual Information*. Model-model dievaluasi menggunakan algoritma *Naïve Bayes Classification (NBC)* dan *Support Vector Machine (SVM)*. Dan untuk mengukur digunakan metrik evaluasi akurasi, presisi (*precision*), *recall*, dan *F1-score*. Hasil eksperimen menunjukkan bahwa kombinasi *TfidfVectorizer* dan *RFE* dalam model NBC mencapai akurasi tertinggi sebesar 71,1111% dan menunjukkan kinerja yang signifikan dalam presisi, recall, dan F1-Score. Hasil tersebut menunjukkan bahwa dalam analisis sentimen berbahasa Indonesia, penggunaan kombinasi seleksi model *TfidfVectorizer* dan *RFE* menunjukkan hasil yang baik, terutama jika diterapkan pada model NBC.

Kata kunci: Analisa Sentimen, Seleksi fitur, *Recursive Feature Elimination (RFE)*, *Chi Square*, *Mutual Information*

Abstract: In sentiment analysis, feature selection is an important stage as it can enhance the performance and efficiency of sentiment analysis models. Feature selection also aids in reducing the complexity of data dimensions, enabling faster and more efficient analysis processes. However, selecting relevant features poses a challenge, as choosing the wrong features can decrease the accuracy of the constructed model. This research focuses on sentiment analysis of tweet data related to the Jakarta Regional Head Election in 2017, employing feature selection techniques such as TF-IDF combined with Recursive Feature Elimination (RFE), Chi Square, and Mutual Information. The models were evaluated using the Naïve Bayes Classification (NBC) and Support Vector Machine (SVM) algorithms. Evaluation metrics, including accuracy, precision, recall, and F1-score, were utilized to measure performance. Experimental results demonstrate that the combination of TfidfVectorizer and RFE in the NBC model achieves the highest accuracy of 71.1111% and exhibits significant performance in terms of precision, recall, and F1-score. These findings indicate that the utilization of the combined selection model of TfidfVectorizer and

RFE yields promising results in Indonesian sentiment analysis, particularly when applied to the NBC model.

Keywords: analisis sentimen, *feature selection*, *Recursive Feature Elimination (RFE)*, *Chi Square*, *Mutual Information*

1. Pendahuluan

Analisis sentimen adalah teknik yang digunakan untuk menganalisis dan mengidentifikasi sentimen atau opini yang terkandung dalam teks[1]. Salah satu area aplikasi penting dari analisis sentimen adalah dalam memahami sentimen terkait pemilu. Pada tahun 2017, pemilihan umum daerah DKI Jakarta menyediakan konteks yang menarik untuk menerapkan analisis sentimen pada *tweet* yang terkait dengan pemilihan tersebut[2].

Namun, dalam menerapkan analisis sentimen, penting untuk memperhatikan seleksi fitur. Seleksi fitur adalah proses pemilihan *subset* dari seluruh fitur yang tersedia dalam teks yang akan digunakan untuk analisis[3]. Dalam konteks analisis sentimen, fitur-fitur ini dapat berupa kata-kata, frasa, atau bahkan karakteristik linguistik tertentu.

Penerapan seleksi fitur dalam analisis sentimen bertujuan untuk mengidentifikasi fitur-fitur yang paling relevan dan memberikan kontribusi dalam memprediksi sentimen[4][5]. Dengan memilih fitur-fitur yang paling informatif, model analisis sentimen dapat menjadi lebih efisien dan akurat. Pentingnya seleksi fitur dalam analisis sentimen terletak pada kemampuannya untuk meningkatkan performa dan efisiensi model[6]. Dengan memilih fitur-fitur yang relevan dan informatif, model analisis sentimen dapat lebih fokus pada aspek penting dari teks dan menghindari kebisingan yang tidak relevan. Selain itu, seleksi fitur juga membantu dalam mengurangi dimensi data yang kompleks, sehingga memungkinkan proses analisis menjadi lebih cepat dan efisien.

Namun, kendala-kendala juga muncul dalam pemilihan seleksi fitur dalam analisis sentimen. Salah satu kendala utama adalah menentukan kriteria yang sesuai untuk memilih fitur-fitur yang paling relevan[7]. Selain itu, terdapat juga tantangan dalam menangani fitur-fitur yang ambigu atau memiliki banyak kemungkinan makna, yang dapat mempengaruhi akurasi model. Selain itu, perubahan konteks dan perubahan trend sentimen juga dapat menjadi kendala dalam pemilihan seleksi fitur yang efektif.

Dalam pemilihan seleksi fitur dalam analisis sentimen, salah satu pendekatan yang umum digunakan adalah menggunakan metode penghitungan *term frequency-inverse document frequency* (TF-IDF). Penggunaan TF-IDF sebagai alat bantu dalam seleksi fitur dapat membantu dalam mengidentifikasi kata-kata yang paling relevan dan memberikan kontribusi signifikan terhadap sentimen yang ingin diprediksi. Dengan menghitung skor TF-IDF untuk setiap kata dalam dataset, maka dapat menentukan kata-kata yang memiliki bobot tinggi dan mempengaruhi sentimen secara signifikan. Selain TF-IDF, dalam analisis sentimen juga digunakan seleksi fitur *Recursive Feature Elimination (RFE)*, *Chi Square*, dan *Mutual Information*.

Selain memahami berbagai metode seleksi fitur, penting juga untuk mengevaluasi kinerja model analisis sentimen yang dibangun. Untuk membangun model analisis sentimen menggunakan algoritma *Naive Bayes Classification* (NBC) dan *Support Vector Machine* (SVM). Untuk mengevaluasi kinerja model analisis sentimen yang dibangun, beberapa metrik evaluasi yang umum digunakan antara lain akurasi, presisi (*precision*), *recall*, dan *f1-score*[6]. Akurasi mengukur sejauh mana model mampu mengklasifikasikan teks dengan benar. Presisi mengukur sejauh mana model memberikan hasil positif yang relevan dari semua hasil positif yang diberikan. *Recall* mengukur sejauh mana model dapat menemukan kembali hasil positif yang relevan dari semua hasil positif yang sebenarnya ada. *F1-score* adalah ukuran rata-rata harmonik antara presisi dan *recall*.

Setiap model seleksi fitur dengan kombinasi TF-IDF dan metode seleksi fitur *Recursive Feature Elimination (RFE)*, *Chi Square*, dan *Mutual Information*. Setelah itu, model NBC dan SVM dilatih menggunakan fitur-fitur yang telah dipilih. Evaluasi dilakukan dengan menghitung akurasi, presisi, *recall*, dan *f1-score* dari setiap model yang dibangun. Penelitian ini dilakukan untuk mengetahui bagaimana pengaruh kombinasi penggunaan metode seleksi fitur dalam analisis sentimen, terutama pada *tweet* berbahasa Indonesia

2. Metode Penelitian

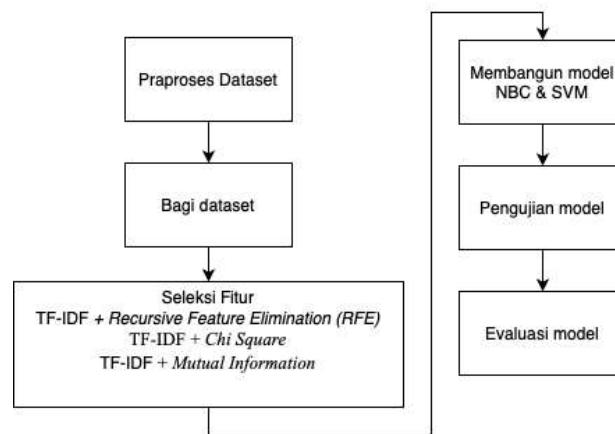
Penelitian dilakukan dengan menggunakan dataset dari penelitian yang telah dilakukan oleh Agnes Rossi Trisna Lestari, Rizal Setya Perdana, dan Mochammad Ali Fauzi [12]. Dataset berisi tweet opini masyarakat tentang kandidat calon gubernur pada Pemilu tahun 2017 yang diunggah secara publik.

Tabel 1 Penggunaan Data

Label	Jumlah Data
Positif	450
Negatif	450
Total	900

Sumber: Hasil Penelitian (2023)

Dataset yang digunakan adalah opini yang dibagikan secara publik dengan jumlah 900 data yang terbagi menjadi dua kelas, 450 data kelas positif dan 450 data kelas negatif, dan dataset berisi opini yang dipublikasikan secara publik pada platform Twitter. Jumlah data yang digunakan bisa dilihat pada tabel 1.



Sumber: Hasil Penelitian (2023)

Gambar 1. Alur Penelitian

2.1 Metode

Alur analisis yang dilakukan diawali dari melakukan praproses, hingga melakukan evaluasi model, seperti yang terlihat pada gambar 1.

1. Praproses Dataset

Dilakukan praproses dataset dengan mengubah semua karakter menjadi huruf kecil, menghilangkan karakter non-alphabet (angka, karakter khusus), menghapus emoticon, menghapus *hyperlink*, dan menghapus mention (@username), melakukan penghapusan *stopword* (kata yang sering muncul dalam sebuah teks namun tidak memiliki makna khusus atau relevan dalam konteks analisis teks) dan melakukan *stemming* menggunakan model Sastrawi[13]. Perbedaan data sebelum dilakukan praproses dan setelah melakukan praproses teks dapat dilihat pada tabel 2.

Tabel 2 Praproses tweet

Keterangan	kalimat
Tweet asli	Banyak akun kloning seolah2 pendukung #agussilvy mulai menyerang paslon #aniessandi dengan opini dan argumen membenaran..jangan terkecoh
Tweet hasil praproses	akun kloning dukung hastag serang paslon hastag opini argumen benar kecoh

Sumber: Hasil Penelitian (2023)

2. Bagi dataset menjadi data latih & data uji dataset akan dibagi menjadi data latih dan data uji, ratio data yang digunakan adalah 80:20. Tabel 3 Menampilkan jumlah data untuk data latih dan data uji.

Tabel 3 Pembagian data latih dan data uji

Data	Jumlah Data
Data latih	720
Data Uji	180
Total	900

Sumber: Hasil Penelitian (2023)

3. Lakukan seleksi fitur

Data latih dan data uji yang telah dibentuk diterapkan seleksi fitur menggunakan beberapa kombinasi dengan TF-IDF untuk seleksi fitur yang lebih lanjut. TF-IDF adalah metode yang digunakan untuk mengukur sejauh mana sebuah kata penting dalam suatu dokumen dengan mempertimbangkan frekuensi kemunculan kata tersebut dalam dokumen tersebut dan seberapa umum kata tersebut dalam seluruh koleksi dokumen[8]. Dimana persamaan 1 menunjukkan bagaimana perhitungan menggunakan TF-IDF

$$tfidf(t, d, D) = tf(t, d) \times idf(t, D) \dots \dots \dots (1)$$

a) TF-IDF + *Recursive Feature Elimination (RFE)*: Metode *Recursive Feature Elimination (RFE)* merupakan pendekatan berbasis pemodelan yang digunakan untuk menghilangkan fitur secara rekursif berdasarkan pengaruhnya terhadap performa model[9]. Pada setiap iterasi, fitur dengan bobot terendah dieliminasi, dan model diperbarui. Proses ini berlanjut hingga mencapai jumlah fitur yang ditentukan. Persamaan 2 adalah persamaan untuk *Recursive Feature Elimination (RFE)*. Dengan menggunakan kombinasi TF-IDF dan RFE, dapat memilih subset fitur yang paling informatif dan relevan untuk analisis sentimen[9].

$$RFE = Select\ features\ recursively(Model, k) \dots \dots (2)$$

b) TF-IDF + *Chi Square*: Metode *Chi Square* dapat digunakan untuk mengevaluasi kepentingan statistik dari setiap fitur dalam hubungannya dengan label kelas[8]. Persamaan 3 adalah persamaan untuk penggunaan seleksi fitur *Chi Square*. Dengan menggunakan kombinasi TF-IDF dan *Chi Square*, dapat memilih fitur-fitur yang memiliki hubungan yang signifikan dengan sentimen yang ingin diprediksi[8].

$$\chi^2 = \sum \frac{(O-E)^2}{E} \dots \dots \dots (3)$$

c) TF-IDF + *Mutual Information*: Metode *Mutual Information* mengukur sejauh mana informasi mengenai satu variabel membantu dalam memperkirakan variabel lainnya. Dalam konteks analisis sentimen, *Mutual Information* dapat digunakan untuk mengevaluasi hubungan antara fitur-fitur dan sentimen. Dengan menggunakan kombinasi TF-IDF dan *Mutual Information*, dapat memilih fitur-fitur yang memiliki keterkaitan yang tinggi dengan sentimen yang ingin diprediksi.

4. Membangun model

Dilakukan permodelan menggunakan model NBC dan SVM, algoritma NBC adalah salah satu algoritma klasifikasi yang umum digunakan dalam analisis sentimen karena sederhana namun cukup efektif. Algoritma ini bekerja berdasarkan asumsi independensi fitur dan menggunakan prinsip probabilitas untuk memprediksi sentimen dari teks[10][11]. Persamaan 4 adalah persamaan untuk algoritma *Naïve Bayes Classification (NBC)*.

$$P(H|X) = \frac{P(X|H)}{P(X)} \cdot P(H) \dots \dots \dots (4)$$

Support Vector Machine (SVM) adalah algoritma yang berfungsi untuk memisahkan data dengan mencari *hyperplane* yang dapat memaksimalkan jarak antara kelas-kelas yang berbeda dalam ruang fitur[9][7]. SVM dapat diterapkan dalam konteks analisis sentimen dengan mengubah teks menjadi vektor fitur menggunakan metode seleksi fitur yang tepat. Persamaan 5 adalah persamaan untuk algoritma *Support Vector Machine*.

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(\log \frac{2l}{h} + 1 - \log \frac{n}{4})}{l}} \dots \dots \dots (5)$$

kemudian akan dilakukan prediksi menggunakan data latih dari setiap seleksi fitur yang diterapkan.

5. Pengujian model

Dari prediksi yang telah dibangun akan dilakukan pengujian menggunakan data uji yang telah dibentuk dari dataset yang digunakan.

6. Evaluasi model

Proses evaluasi model dilakukan dengan menghitung akurasi, presisi, recall, dan f1-score untuk setiap model dari setiap seleksi fitur yang diterapkan. Akurasi adalah rasio antara jumlah prediksi yang benar dengan jumlah total sampel. Persamaan 6 adalah persamaan untuk menghitung nilai akurasi.

$$\text{akurasi} = \frac{\text{truePositives} + \text{trueNegatives}}{\text{truePositives} + \text{trueNegatives} + \text{falsePositives} + \text{falseNegatives}} \dots\dots\dots (6)$$

Presisi menghitung rasio antara jumlah positif yang benar diprediksi dengan jumlah total prediksi positif. Persamaan 7 adalah persamaan untuk menghitung nilai presisi.

$$\text{presisi} = \frac{\text{truePositives}}{\text{truePositives} + \text{falsePositives}} \dots\dots\dots (7)$$

Recall menghitung rasio antara jumlah positif yang benar diprediksi dengan jumlah total sentimen positif. Persamaan 8 adalah persamaan untuk menghitung nilai *recall*.

$$\text{recall} = \frac{\text{truePositives}}{\text{truePositives} + \text{falseNegatives}} \dots\dots\dots (8)$$

F1-Score digunakan sebagai metrik evaluasi yang baik ketika ingin mempertimbangkan keseimbangan antara presisi dan *recall*. Persamaan 9 adalah persamaan untuk menghitung nilai *F1-Score*.

$$\text{F1 - score} = \frac{2 \cdot \text{truePositives}}{2 \cdot \text{truePositives} + \text{falsePositives} + \text{falseNegatives}} \dots\dots\dots (9)$$

3. Hasil dan Pembahasan

Dari setiap model yang dibangun dengan skenario kombinasi seleksi fitur dan algoritma NBC & SVM akan dievaluasi dengan metrik evaluasi akurasi, presisi, recall, dan F1-score untuk membandingkan performa dari setiap skenario yang dibangun.

Akurasi

Akurasi mengukur sejauh mana model klasifikasi mampu memprediksi dengan benar kelas sentimen pada keseluruhan dataset. Tabel 4 adalah nilai akurasi dari setiap model yang dibangun.

Tabel 4. Nilai akurasi

<i>akurasi</i>	SVM	NBC
TF-IDF + <i>RFE feature selector</i>	70.5556	71.1111
TF-IDF + <i>Chi Square</i>	60.0000	41.6667
TF-IDF + <i>mutual information</i>	70.5556	70.5556

Sumber: Hasil Penelitian (2023)

Presisi

Presisi mengukur sejauh mana model mampu mengidentifikasi dengan benar sentimen positif dari prediksi yang dilakukan. Tabel 5 adalah nilai presisi dari setiap model yang dibangun.

Tabel 5 Nilai Presisi

Presisi	SVM	NBC
TF-IDF + <i>RFE feature selector</i>	71.3767	71.1144
TF-IDF + <i>Chi Square</i>	73.0338	27.8978
TF-IDF + <i>mutual information</i>	70.5893	71.3767

Sumber: Hasil Penelitian (2023)

Recall

Recall, juga dikenal sebagai sensitivitas, mengukur sejauh mana model mampu menemukan dengan benar sentimen positif dari semua *instance* sentimen positif yang ada dalam dataset. Tabel 6 adalah nilai *recall* dari setiap model yang dibangun.

Tabel 6 Nilai Recall

Recall	SVM	NBC
TF-IDF + <i>RFE feature selector</i>	70.5556	71.1111
TF-IDF + <i>Chi Square</i>	60.0000	41.6667
TF-IDF + <i>mutual information</i>	70.5556	70.5556

Sumber: Hasil Penelitian (2023)

F1-score

F1-Score adalah ukuran rata-rata harmonik dari presisi dan *recall*. *F1-Score* memberikan pengukuran yang seimbang antara kedua metrik tersebut. Tabel 7 adalah nilai *F1-Score* dari setiap model yang dibangun.

Tabel 7 Nilai *F1-Score*

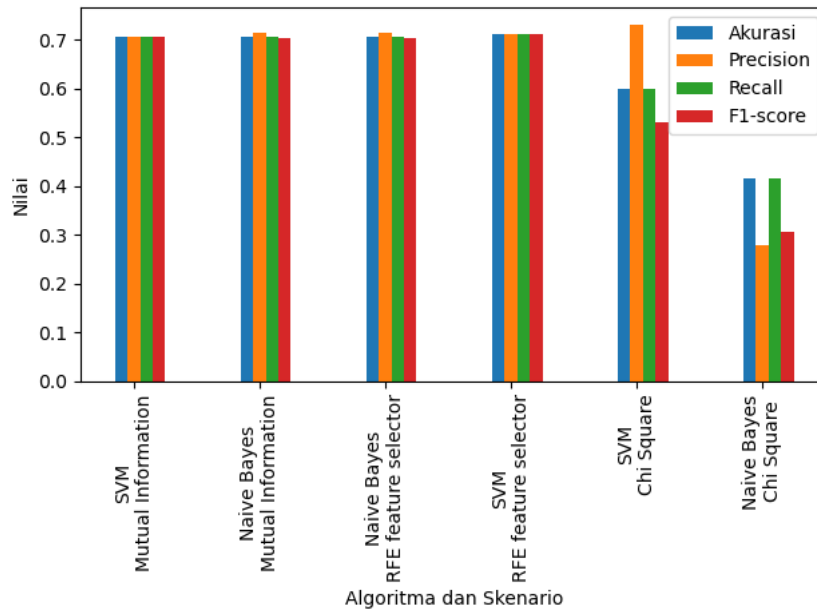
F1-score	SVM	NBC
TF-IDF + <i>RFE feature selector</i>	70.3222	71.1040
TF-IDF + <i>Chi Square</i>	53.0764	30.6158
TF-IDF + <i>mutual information</i>	70.5528	70.3222

Sumber: Hasil Penelitian (2023)

Tabel 4 menunjukkan nilai akurasi pada model algoritma NBC dengan kombinasi seleksi fitur TF-IDF dan *RFE feature selector* memiliki nilai akurasi tertinggi sebesar 71.1111 %, dan pada Nilai *F1-Score* juga menghasilkan nilai tertinggi sebesar 71.1040 %. Pada presisi adalah 71.1144 % dan *recall* sebesar 71.1111 %.

Kombinasi seleksi fitur TF-IDF dan *Chi Square* pada model NBC menghasilkan nilai terendah pada semua matrix evaluasi, dengan nilai akurasi sebesar 41.6667 %, presisi 27.8978 %, *recall* 41.6667 %, dan *F1-score* sebesar 30.6158 %. Gambar 2 menunjukkan grafik matrix

evaluasi pada algoritma SVM & NBC pada skenario seleksi fitur *RFE feature selector*, *Chi Square*, dan *mutual information*



Sumber: Hasil Penelitian (2023)

Gambar 2. Perbandingan Hasil Matrix Evaluasi

4. Kesimpulan

Dalam pengujian dengan matrix evaluasi, diketahui bahwa kombinasi seleksi fitur TF-IDF dan RFE feature selector pada model algoritma NBC menghasilkan nilai akurasi tertinggi sebesar 71.1111%. Hal ini menunjukkan bahwa metode seleksi fitur tersebut dapat memberikan performa yang baik dalam memprediksi sentimen. Selain itu, penggunaan kombinasi seleksi fitur juga memberikan hasil yang baik pada matrix evaluasi presisi, recall, dan F1-Score. Hal ini menunjukkan bahwa model yang menggunakan metode seleksi fitur dapat memberikan hasil yang konsisten dan seimbang dalam mengklasifikasikan sentimen. Dengan demikian, penerapan seleksi fitur dalam analisis sentimen menggunakan kombinasi seleksi fitur, khususnya TF-IDF dan RFE feature selector, dapat meningkatkan kualitas hasil analisis dengan memberikan akurasi yang lebih tinggi dan konsistensi dalam evaluasi presisi, recall, dan F1-Score.

Ucapan Terima Kasih

Penelitian ini berhasil diselesaikan berkat dukungan, bantuan, saran, dan masukan yang diberikan oleh berbagai pihak. Penulis sangat bersyukur dan mengucapkan terima kasih kepada semua yang telah membantu dalam menyelesaikan penelitian ini. Penulis menyadari bahwa penelitian ini masih memiliki kekurangan, dan dengan demikian, penulis mengharapkan kritik dan saran dari pembaca untuk melakukan perbaikan pada penelitian ini.

Referensi

- [1] I. K. Dharmendra, N. N. U. Januhari, I. P. Ramayasa, and I. M. A. W. Putra, "Uji Komparasi Analisis sentimen Pada Opini Alumni Terhadap Perguruan Tinggi," *Jurnal Teknik Informatika UNIKA Santo Thomas*, pp. 1–6, May 2022, doi: 10.54367/jtiust.v7i1.1748.
- [2] G. A. Buntoro, "Analisis sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," vol. 1, no. 1, pp. 32–41, 2017.
- [3] T. Desyani, A. Saifudin, and Y. Yulianti, "Feature Selection Based on Naive Bayes for Caesarean Section Prediction," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 879, no. 1, p. 012091, Jul. 2020, doi: 10.1088/1757-899X/879/1/012091.
- [4] A. K. Fauziyyah, "ANALISIS SENTIMEN PANDEMI COVID19 PADA STREAMING TWITTER DENGAN TEXT MINING PYTHON," *Jurnal Ilmiah SINUS*, vol. 18, no. 2, Art. no. 2, Jul. 2020, doi: 10.30646/sinus.v18i2.491.

- [5] P. H. Prastyo, I. Ardiyanto, and R. Hidayat, "A Review of Feature Selection Techniques in Analisis sentimen Using Filter, Wrapper, or Hybrid Methods," in *2020 6th International Conference on Science and Technology (ICST)*, Sep. 2020, pp. 1–6. doi: 10.1109/ICST50505.2020.9732885.
- [6] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based analisis sentimen of online product reviews with a novel term weighting and feature selection approach," *Information Processing & Management*, vol. 58, no. 5, p. 102656, Sep. 2021, doi: 10.1016/j.ipm.2021.102656.
- [7] "KOMPARASI FITUR SELEKSI PADA ALGORITMA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN REVIEW | Arifin | Jurnal Informatika." <https://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/868/936> (accessed May 21, 2021).
- [8] U. I. Larasati, M. A. Muslim, R. Arifudin, and A. Alamsyah, "Improve the Accuracy of Support Vector Machine Using Chi Square Statistic and Term Frequency Inverse Document Frequency on Movie Review Analisis sentimen," *Scientific Journal of Informatics*, vol. 6, no. 1, Art. no. 1, May 2019, doi: 10.15294/sji.v6i1.14244.
- [9] N. S. Mohd Nafis and S. Awang, "An Enhanced Hybrid Feature Selection Technique Using Term Frequency-Inverse Document Frequency and Support Vector Machine-Recursive Feature Elimination for Sentiment Classification," *IEEE Access*, vol. 9, pp. 52177–52192, 2021, doi: 10.1109/ACCESS.2021.3069001.
- [10] W. Ningsih, B. Budiman, and I. Umami, "Implementasi Algoritma Naïve Bayes Untuk Menentukan Calon Penerima Beasiswa Di SMK YPM 14 Sumobito Jombang," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 4, no. 2, Art. no. 2, Jul. 2022, doi: 10.47233/jteksis.v4i2.570.
- [11] "Klasifikasi Menggunakan Naïve Bayes Dan K-Nearest Neighbor Pada Manajemen Layanan Teknologi Informasi | Jurnal Teknologi Dan Sistem Informasi Bisnis." <http://jurnal.unidha.ac.id/index.php/jteksis/article/view/121> (accessed Jun. 11, 2023).
- [12] "Analisis sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji | Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer." <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/627> (accessed Apr. 30, 2023).
- [13] "sastrawi/sastrawi: High quality stemmer library for Indonesian Language (Bahasa)." <https://github.com/sastrawi/sastrawi> (accessed Jan. 09, 2022).
- [14] E. Retnoningsih, D. Diyah, P. Utami, and D. P. Utami, "Penerapan Knowledge Management pada Perguruan Tinggi ... PENERAPAN KNOWLEDGE MANAGEMENT PADA PERGURUAN TINGGI (STUDI KASUS AMIK BSI PURWOKERTO)".
- [15] E. Retnoningsih, "Seminar Nasional Teknologi Informasi dan Multimedia 2015 MENGUKUR TINGKAT KEPUASAN PENGGUNAAN LEARNING MANAGEMENT SYSTEM DALAM KNOWLEDGE SHARING," pp. 6–8, 2015.