

Analisis Perbandingan Algoritma *Machine Learning* Terhadap Sentimen Analis Pemindahan Ibu Kota Negara

Arif Rahman Hakim ^{1*}, Windu Gata ², Alda Zevana Putri Widodo ³, Oky Kurniawan ⁴, Arief Rama Syarif ⁵

^{1,2,3,4,5} Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia.

article info

Article history:

Received 1 July 2022

Received in revised form

20 November 2022

Accepted 12 February 2023

Available online April 2023

DOI:

<https://doi.org/10.35870/jti.k.v7i2.701>

Keywords:

Comparative Analysis;
Machine Learning Algorithm;
Analyst Sentiment; Relocation
of the State Capital.

abstract

The Indonesian nation was enlivened with news about the relocation of the State Capital (IKN). The government's plan to move IKN is based on Indonesia's vision and mission in 2045, namely advanced Indonesia. Twitter is one of the microblogging communication tools used to express opinions. Various algorithms have been used to analyze sentiment towards an opinion such as Support Vector Machine, Naive Bayes and Random Forest. This study aims to compare the performance of three classification algorithms, namely Support Vector Machine, Naïve Bayes and Random Fores. The highest accuracy results are using the Support Vector Machine algorithm and added with the Synthetic Minority Oversampling Technique Method (SMOTE) feature of 82.82%, Precision 79.34%, Recall 88.75%, 87.78% and ROC AUC 82.82%. Naive Bayes accuracy is 81.18%, Precision 84.89%, Recall 75.86%, 80.13% and ROC AUC 81.18%. and Random Forest accuracy of 79.55, precision 84.48%, recall 72.39%, 77.97% and ROC AUC 77.55%.

abstrak

Bangsa indonesia diramaikan dengan berita tentang pemindahan Ibu Kota Negara (IKN). Rencana pemerintah memindahkan IKN beralaskan pada visi misi Indonesia tahun 2045 yaitu Indonesia maju. Twitter menjadi salah satu alat komunikasi microblogging yang digunakan untuk menyampaikan opini. Berbagai algoritma telah digunakan untuk menganalisa sentimen terhadap suatu opini seperti Support Vector Machine, Naive Bayes dan Random Forest. Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma klasifikasi yaitu Support Vector Machine, Naïve Bayes dan Random Fores. Hasil akurasi tertinggi menggunakan algoritma Support Vector Machine dan di tambah feature Synthetic Minority Oversampling Technique Method (SMOTE) sebesar 82.82%, Presisi 79.34%, Recall 88.75%, 87,78% dan ROC AUC 82.82%. Akurasi Naive bayes sebesar 81.18%, Presisi 84.89%, Recall 75.86%, 80.13% dan ROC AUC 81.18%. dan akurasi Random Forest sebebar 79.55 Presisi 84.48%, Recall 72.39%, 77.97% dan ROC AUC 77.55%.

Corresponding author. Email: 14210167@nusamandiri.ac.id ^{1}.

© E-ISSN: 2580-1643.

Copyright @ 2023. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan RISET)
[\(http://creativecommons.org/licenses/by/4.0/\).](http://creativecommons.org/licenses/by/4.0/)

1. Latar Belakang

Tahun lalu, 26 Agustus 2019 bangsa indonesia diramaikan dengan berita tentang pemindahan Ibu Kota Negara (IKN) [1]. Meski rencana pemindahan IKN bukanlah rencana baru, mengingat hal ini pernah direncanakan oleh presiden pertama republik Indonesia bapak ir. Sockarno pada tanggal 17 Juli 1957 dengan memilih palangkaraya sebagai ibu kota negara. Pada masa orde baru, yakni tahun 1990-an, presiden kedua republik Indonesia juga merencanakan pemindahan IKN ke Jonggol Jawa Barat. Namun, pemindahan IKN baru serius digarap pada masa pemerintahan presiden Joko Widodo, hal ini terlihat dengan disahkannya RUU IKN menjadi UU IKN pada tanggal 18 Januari 2022 [3][4]. Rencana pemerintah memindahkan IKN beralaskan pada visi misi Indonesia tahun 2045 yaitu Indonesia maju, ekonomi Indonesia akan masuk 5 besar dunia 2045 (Nainggolan, 2022), pemindahan IKN akan mendorong pertumbuhan ekonomi yang merata termasuk di kawasan timur Indonesia (Aziz, 2022). Namun demikian pemindahan IKN juga akan beresiko pada tingginya pembiayaan pemindahan ibukota [1].

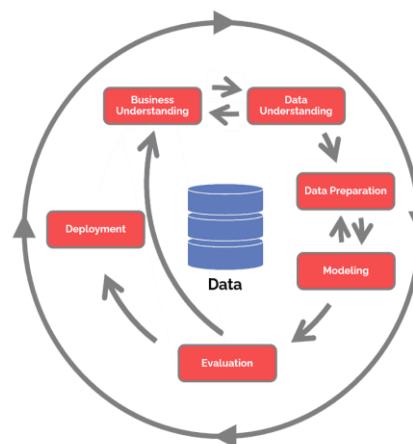
Sentimen adalah pandangan atau emosi seseorang yang tertulis dalam bahasa tertulis [4]. Analisis adalah suatu kegiatan penyelidikan untuk mengetahui fakta yang sebenarnya [4]. Sedangkan sentimen analisis adalah salah satu bidang penelitian yang bertujuan untuk menganalisa pendapat, perasaan, penilaian, sikap dan perasaan individu terhadap suatu peristiwa, topik, isu, politik, produk, jasa, organisasi, individu, dan properti [5]. Twitter adalah salah satu alat komunikasi *microblogging* populer bagi pengguna internet [6]. Selain itu twitter juga digunakan oleh pengguna dalam menyampaikan opininya [7]. Dalam tweet tersebut terdapat cuitan yang berupa opini masyarakat baik itu opini positif maupun opini negatif mengenai wacana pemindahan IKN. Hal yang sama telah dilakukan oleh Mas'udah (2020), penelitian tersebut mengambil data dari twitter menggunakan library *Twitterscrapper* dengan kata kunci "ibu kota". Data kemudian diklasifikasi dengan membandingkan 3 buah metode yaitu *Gaussian Naïve Bayes* (GNB), *Multinomial Naïve Bayes* (MNB) dan *Bernoulli Naïve Bayes* (BNB). Hasil akurasi tertinggi diperoleh sebanyak 68.10% menggunakan metode BNB dengan ditambah *feature Synthetic Minority*

Oversampling Technique (SMOTE) [8]. Pada penelitian Natasuwarna (2019), didapatkan dataset sebanyak 200 tweet dengan tagar Ibukota Palangkaraya, Indonesia Ibu Kotabaru, Yuk Usul Ibukota. Data tersebut di klasifikasi menggunakan metode *Naïve Bayes* dan tools *Rapidminer Studio*, didapatkan hasil akurasi rata rata sebanyak 89.86% [9]. Penelitian yang dilakukan Asri dkk (2021), algoritma *Support Vector Machine* (SVM) digunakan untuk analisis sentimen *twitter* dengan lima buah kata kunci yaitu ibu kota pindah, pemindahan ibu kota jakarta, pemindahan ibu kota, dan ibu kota jakarta. Penelitian tersebut mendapatkan hasil akurasi tertinggi sebesar 96,68% [10].

Pada penelitian sebelumnya masih menggunakan satu jenis algoritma untuk melakukan sentimen analis, belum ada perbandingan nilai akurasi dengan algoritma klasifikasi yang lain, maka dari itu penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma klasifikasi yaitu SVM, *Naïve Bayes* (NB) dan Random Forest (RF) dalam menganalisa sentimen yang dilontarkan masyarakat melalui media sosial *twitter* dengan topik ibu kota negara. Ketiga algoritma tersebut dievaluasi menggunakan teknik *cross validation* dengan melihat lima metrik evaluasi yaitu *Accururation*, *Pression*, *Recall*, *F1 score*, dan *ROC – AUC* [11][12].

2. Metode Penelitian

Dalam penelitian ini terdapat alur proses penelitian untuk dapat melakukan analisis sentimen menggunakan metode CRISP DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri 6 tahapan, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [13].



Gambar 1. Alur kerja CRISP DM

Pada gambar 1 penjabaran alur CRISP DM sebagai berikut [14]:

a) *Business Understanding*

Pada tahapan business understanding dilakukan langkah pendefinisian pengetahuan yang diinginkan dan diperoleh sebagai pertanyaan umum seperti tingkat proses pengambilan data melalui sosial media.

b) *Data Understanding*.

merupakan tahapan untuk mengumpulkan data dari twitter, mengidentifikasi kalimat yang tidak diperlukan yang digunakan sebagai data tweet dan data yang diperoleh serta menentukan penggunaan Jupyter Notebook dan library Twint sebagai alat crawling

c) *Data Preparation*.

Langkah ini memiliki banyak proses seperti pembersihan data yang telah didapatkan, format ulang data, mengeleminasi jumlah data, dan lain lain.

d) *Modeling*.

Langkah ini adalah representasi komputasi hasil pengamatan dan merupakan hasil identifikasi pola yang terkandung pada data.

e) *Evaluation*.

Langkah ini adalah langkah untuk memastikan nilai dari model yang telah ditentukan pada tahap modeling sebelumnya

f) *Deployment*.

Langkah ini merupakan langkah terakhir dari pendekatan CRISP DM, langkah ini merupakan hasil yang didapatkan dari seluruh tahapan sebelumnya yang kita gunakan.

3. Hasil dan Pembahasan

Business Understanding

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *SVM*, *NB* dan *RF* dalam melakukan klasifikasi sentimen analisis pemindahan IKN. Data dibagi menjadi dua opini sentimen negatif dan sentimen positif menggunakan library *Textblob*. Selain itu algoritma juga akan dibandingkan dengan menambahkan feature *Synthetic Minority Oversampling Technique* (SMOTE).

Data Understanding

Tahapan ini adalah tahapan pemahaman data yang

akan digunakan. Penelitian ini menggunakan data *tweet* dari sosial media *twitter* dengan kata kunci “IKN” menggunakan *Jupyter Notebook* dan library *Twint*. Data diambil sebanyak 1500 data dari mulai tanggal 01-09-2021 sampai dengan tanggal 09-05-2022.

date	username	tweet
0 2022-05-09 06:59:19	gusuran76	Dulu ada seorang bilang begini, dibilangnya be...
1 2022-05-09 06:58:17	syadiah_d	@SahronihO @machbeach Saran saya kalau punya...
2 2022-05-09 06:57:53	mimakhayla	Pemerintah resmi membentuk TIM Transisi Pemind...
3 2022-05-09 06:56:46	Mah_Mood2	Muncul Petisi Sejumlah Tokoh Penting Tolak IKN...
4 2022-05-09 06:54:23	EdiKeceput	@Sangkuriang5551 Waduh ... Jadi sumber dana IK...
...
1501 2022-05-08 07:53:39	kalalo_ftri	Jokowi Mania: Penolak IKN Nusantara Belum Move...
1502 2022-05-08 07:53:21	ralebdotcom	IKN Sarung https://t.co/5IPh2TgHTF
1503 2022-05-08 07:52:47	Sastidevi2	@machbeach Rakyat g perlu ikn
1504 2022-05-08 07:51:16	kalalo_ftri	[HOAKS] WNA China Calon Penghuni IKN Sudah Ber...
1505 2022-05-08 07:50:52	JokoHar68301204	Proyek IKN Baru Disebut hanya Ambisi dan Akal...

1506 rows x 3 columns

Gambar 2. Preview Tweet IKN

Data Preparation

Tahap ini adalah tahap pemberian label sentimen negatif atau positif dan pembersihan data atau data *preprocessing*. Adapun tahapan pembersihan data pada penelitian ini adalah sebagai berikut

a) *Case Folding*

Case folding merupakan proses merubah semua huruf, kata atau kalimat menjadi huruf kecil [15]. Proses ini menghilangkan inkonsistensi penggunaan huruf kapital.

b) *Cleaning*

Cleaning merupakan proses pembersihan data dari noise [16]. Pada penelitian ini pembersihan noise meliputi pembersihan karakter “RT”, emoji, url, username, duplikat data, duplikat huruf, *number*, duplikat spasi, tanda baca, tanda *hashtag*, duplikat data dan data kosong,

c) *Tokenizing*

Tokenizing merupakan proses memotong kalimat atau *string input* berdasarkan tiap kata penyusun kalimat [17]

d) *Stopword*

Stopword adalah proses penghapusan atau penghilangan kata tidak penting seperti di, ke, dan , yang, dan lain lain [18].

e) *Stemming*

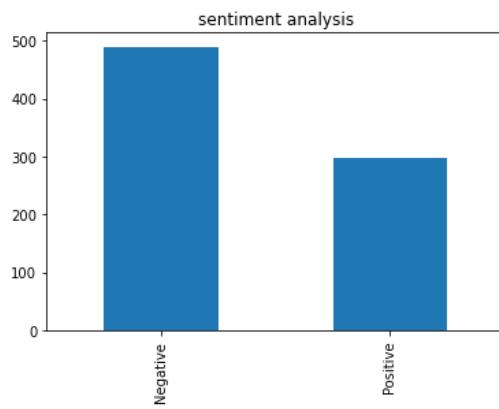
Stemming adalah proses penghilangan kata yang mempunyai imbuhan menjadi bentuk kata dasar [19].

Proses pelabelan sentimen dilakukan setelah data dilakukan proses *case folding* dan *cleaning*. Pelabelan

sentimen menggunakan *library Textblob*. *Textblob* merupakan *library python* yang menyediakan API untuk menganalisis *Natural Language Processing* (NLP) salah satunya adalah analis sentimen [20].

	tweet	label
0	bilang bilang benci kalimantan rasain aja deh ...	Negative
1	saran otak peke bodoh pilih tolol jgn bangga...	Negative
2	perintah resmi bentuk tim transisi pindah ikn ...	Negative
3	muncul petisi tokoh tolak ikn ruhut cari popul...	Positive
4	sumber dana ikn sumbang sukarela masyarakat d...	Positive
...
781	coba yg udh nnton spoiler meh poin	Positive
782	celamitan aja embat dana haji utk haji bikin h...	Positive
783	jis hutang perintah pusat tp apbn dana pulih e...	Positive
784	ralyat g ikn	Negative
785	proyek ikn ambisi akalakalan jokowi rocky geru...	Positive

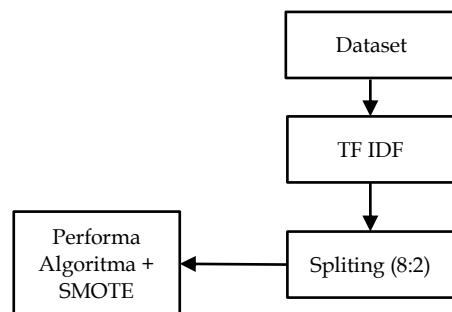
Gambar 3. Data dataset dengan label sentimen



Gambar 4. Grafik perbandingan sentimen

Modeling

Pemodelan menggunakan algoritma *SVM*, Hasil dari algoritma *SVM* akan dibandingkan dengan algoritma *NB* dan algoritma *RF*. Sebelum dilakukan perbandingan, data akan dilakukan pembobotan kata menggunakan dengan *TF-IDF* (*Term Frequency - Inverse Document Frequency*). Perbandingan kinerja juga akan dibandingkan dengan ditambahkan nya feature *Synthetic Minority Oversampling Technique Method* (SMOTE). Setelah dilakukan pemodelan data kemudian dipecah menjadi dua bagian yaitu untuk data latih sebanyak 80% dan data uji sebanyak 20%.



Gambar 5. Diagram alur modelling

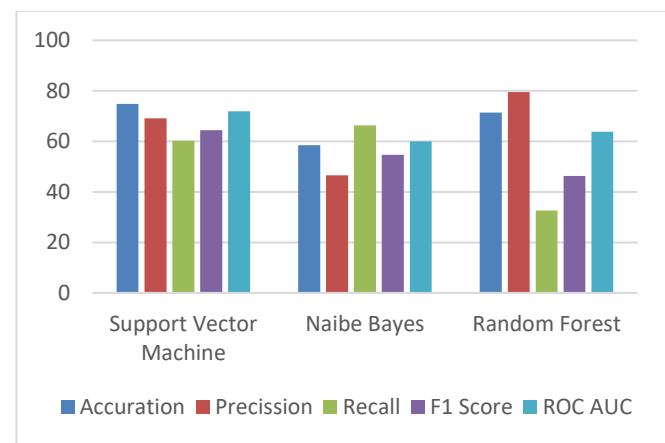
Evaluation

Tahap ini merupakan tahap *validasi* keakuratan dari sebuah model algoritma. *Validasi* akan digunakan untuk mendapatkan hasil prediksi dari model yang ada, kemudian hasil validasi ini dibandingkan dengan model yang lain, ini adalah langkah terpenting dalam proses membangun model (Saputra, 2020).

Pengujian dilakukan dengan menggunakan *environment Jupyter Notebook*, Pada model pertama menggunakan *cross validation* dengan nilai $k = 10$. Algoritma *SVM* mendapatkan skor *accuracy* 74.80%, *precision* 69.11%, *recall* 60.20%, *F1 score* 64.38%, ROC-AUC 71.95%. Perbandingan kinerja algoritma dapat dilihat pada tabel berikut.

Tabel 1. Tabel hasil perbandingan kinerja model

Algoritma	Accururation	Precision	Recall	F1 Score	ROC AUC
Support Vector Machine	74.80	69.11	60.20	64.38	71.95
Naive Bayes	58.52	46.57	66.33	54.72	60.05
Random Forest	71.37	79.50	32.66	46.30	63.77

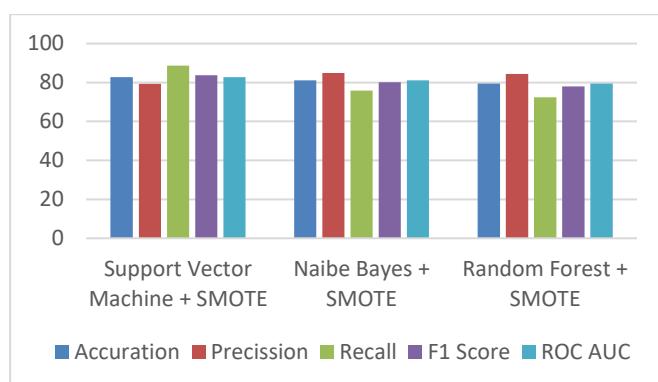


Gambar 6. Grafik perbandingan kinerja algoritma

Pengujian selanjutnya masih dengan menggunakan *environment Jupyter Notebook* namun dengan menambahkan teknik SMOTE pada data label target agar data target menjadi seimbang. Dengan menambahkan SMOTE didapatkan hasil *accuration* 82.82%, *precision* 79.34%, *recall* 88.75%, *F1 Score* 83.78%, *ROC AUC* 82.82%. Perbandingan dengan algoritma lain dapat dilihat pada tabel berikut.

Tabel 2. Tabel perbandingan kinerja model + SMOTE

Algoritma	Accurration	Precision	Recall	F1 Score	ROC AUC
Support Vector	82.82	79.34	88.75	83.78	82.82
Machine + SMOTE					
Naive Bayes	81.18	84.89	75.86	80.13	81.18
Bayes + SMOTE					
Random Forest	79.55	84.48	72.39	77.97	79.55
Forest + SMOTE					



Gambar 7. Grafik perbandingan kinerja algoritma + SMOTE

Deployment

Hasil dari evaluasi pengujian model antara algoritma *SVM* dengan model algoritma *NB*, dan algoritma *RF* dan perbandingan algoritma yang ditambahkan dengan feature Synthetic Minority Over-sampling Technique Methode (SMOTE), didapati hasil pengujian model tertinggi dari seluruh hasil pengujian algoritma adalah model algoritma *SVM* yang ditambahkan feature SMOTE. Proses Deployment menggunakan *environment jupiter Notebook*.

4. Kesimpulan

Dari hasil pengujian yang sudah dilakukan dapat disimpulkan dari total 1506 data *tweet* setelah dilakukan proses preparation, data berkurang menjadi 786 data yang terdiri dari 489 sentimen negatif dan 297 sentimen positif. Algoritma *SVM* memiliki nilai akurasi paling tinggi sebesar 74.80% untuk data tanpa SMOTE dan 82.82% untuk data yang telah ditambahkan SMOTE. Terdapat peningkatan nilai akurasi setelah ditambahkan teknik SMOTE pada algoritma *SVM* maupun *RF*.

5. Daftar Pustaka

- [1] Silalahi, S.A.F., 2019. Dampak Ekonomi dan Resiko Pemindahan Ibu Kota Negara. *Info Singkat*, 11(16), pp.19-24. Available at: <http://puslit.dpr.go.id>.
- [2] Nainggolan, E.U. 2022. *Urgensi Pemindahan Ibu Kota Negara*. Available at: <https://www.djkn.kemenkeu.go.id/kanwil-kalbar/baca-artikel/14671/Urgensi-Pemindahan-Ibu-Kota-Negara.html> (Accessed: 29 June 2022).
- [3] Aziz, A. and Fauziah, F., 2022. Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 6(1), pp.115-125. DOI: <http://dx.doi.org/10.30645/j-sakti.v6i1.430>.
- [4] Sucayyo, N., Kurniati, I. and Harvit, K., 2022. Analisis Sentimen Masyarakat Terhadap UU Cipta Kerja Pada Media Sosial Twitter. *JRIS: Jurnal Rekayasa Informasi Swadharma*, 2(1), pp.63-70. DOI: <https://doi.org/10.56486/jris.vol2no1.167>.

- [5] Hakim, I., Nugroho, A., Sukmana, S.H. and Gata, W., 2020. Sentimen Analisis Stay Home menggunakan metode klasifikasi Naive Bayes, Support Vector Machine, dan k-Nearest Neighbor. *Paradigma-Jurnal Komputer dan Informatika*, 22(2020), pp.169-174. DOI: <https://doi.org/10.31294/p.v22i2.8237>.
- [6] Muhibin, D. and Wibowo, A., 2020. Perbandingan kinerja algoritma Support Vector Machine dan K-Nearest Neighbor terhadap analisis sentimen kebijakan new normal. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 5(2), pp.153-159. DOI: <https://doi.org/10.30998/string.v5i2.6715>.
- [7] Wiratama Putra, T., Triayudi, A. and Andrianingsih, A. 2022. Analisis Sentimen Pembelajaran Daring Menggunakan Metode Naïve Bayes, KNN, dan Decision Tree, *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 6(1), pp. 20–26. DOI: <https://doi.org/10.35870/jtik.v6i1.368>
- [8] Mas' udah, E., Wahyuni, E.D. and Anjani, A., 2020. Analisis sentimen: Pemindahan ibu kota Indonesia pada twitter. *Jurnal Informatika dan Sistem Informasi*, 1(2), pp.397-401.
- [9] Natasuwarna, A.P., 2019, December. Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi Naive Bayes. In *SENSITIf: Seminar Nasional Sistem Informasi dan Teknologi Informasi* (pp. 47-53).
- [10] Arsi, P. and Waluyo, R., 2021. Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), p.147. DOI: <https://doi.org/10.25126/jtiik.0813944>.
- [11] Pratiwi, B.P., Handayani, A.S. and Sarjana, S., 2020. Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix. *Jurnal Informatika Upgris*, 6(2). DOI: <https://doi.org/10.26877/jiu.v6i2.6552>.
- [12] Wahyuni, E.S., 2016. Penerapan metode seleksi fitur untuk meningkatkan hasil diagnosis kanker payudara. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 7(1), pp.283-294. DOI: <https://doi.org/10.24176/simet.v7i1.516>.
- [13] Sukarna, R.H. and Ansori, Y., 2022. Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Ilmiah Sains dan Teknologi*, 6(1), pp.50-61. DOI: <https://doi.org/10.47080/saintek.v6i1.1467>.
- [14] Saputra, D.D. 2020. *Otomasi Pengkategorian Terhadap Keluhan Pelanggan Pada Twitter Operator Telekomunikasi XL Axiata di Indonesia*. Universitas Nusa Mandiri.
- [15] Gunawan, B., Sastypratiwi, H. and Pratama, E.E., 2018. Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 4(2), pp.113-118. DOI: <https://doi.org/10.26418/jp.v4i2.27526>.
- [16] Buntoro, G.A., 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, 2(1). DOI: <https://doi.org/10.31284/j.integer.2017.v2i1.95>.
- [17] Pristiyanti, R.I., Fauzi, M.A. and Muflikhah, L. 2018. Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(3), pp. 1179–1186. Available at: <http://j-ptiik.ub.ac.id>.
- [18] Septian, J.A., Fachrudin, T.M. and Nugroho, A., 2019. Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *INSYST: Journal of Intelligent System and Computation*, 1(1), pp.43-49. DOI: <https://doi.org/10.52985/insyst.v1i1.36>.

- [19] Santosa, A., Purnamasari, I. and Mayasari, R., 2022. Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 6(1), pp.81-93. DOI: <http://dx.doi.org/10.30645/j-sakti.v6i1.427>.
- [20] Fauziyyah, A.K., 2020. Analisis sentimen pandemi Covid19 pada streaming Twitter dengan text mining Python. *Jurnal Ilmiah SINUS*, 18(2), pp.31-42. DOI: <https://doi.org/10.30646/sinus.v18i2.491>.