



**OPTIMASI METODE NAIVE BAYES PARTICLE SWARM OPTIMIZATION ANALISIS SENTIMEN
FORMULA E JAKARTA PADA TWITTER**

Donny Maulana¹, Hasim Budi Jatmiko², Nanang Tedi Kurniadi³

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

¹donny.maulana@pelitabangsa.ac.id, ²hasimbudijatmiko@gmail.com, ³nanang@pelitabangsa.ac.id

Abstract

The city of Jakarta plans to hold a Formula E racing event to promote electric cars as the vehicle of the future. The Covid-19 pandemic that hit Jakarta forced the plan to be postponed. The postponement caused a polemic in the community on social media due to the condition of Jakarta being hit by Covid-19 but the Jakarta city government still wants to hold Formula E by paying commitment money to the organizers which is not small. This difference of opinion on social media is used as material for sentiment analysis using the Naive Bayes classification method. The Naive Bayes method, which has a weakness in feature selection, is optimized by applying the Particle Swarm Optimization (PSO) feature selection. The results of the application of PSO optimization on the Naive Bayes method show an increase in performance with an accuracy value of 89.16%, precision 91.10%, recall 86.81% and AUC 0.690.

Keywords: Naive Bayes, Particle Swarm Optimization, Sentiment Analysis, Jakarta E-Prix.

Abstrak

Kota Jakarta berencana menyelenggarakan ajang balap Formula E untuk mempromosikan mobil listrik sebagai kendaraan masa depan. Pandemi Covid-19 yang melanda Jakarta membuat rencana tersebut harus ditunda. Penundaan tersebut menimbulkan polemik di masyarakat pada media sosial dikarenakan kondisi Jakarta yang sedang dilanda pandemi Covid-19 tetapi pemerintah kota Jakarta tetap ingin menyelenggarakan Formula E dengan membayar uang komitmen kepada penyelenggara yang jumlahnya tidak sedikit. Perbedaan pendapat pada media sosial ini dijadikan bahan untuk analisis sentimen menggunakan metode klasifikasi Naive Bayes. Metode Naive Bayes cukup yang memiliki kelemahan terhadap seleksi fitur dilakukan optimasi dengan menerapkan pemilihan fitur Particle Swarm Optimization (PSO). Hasil penerapan optimasi PSO pada metode Naive Bayes

menunjukkan adanya peningkatan performa dengan nilai akurasi 89,16%, precision 91,10%, recall, 86,81% dan AUC 0,690.

Kata kunci: naive bayes, particle swarm optimization, analisis sentimen, formula e jakarta.

1. Pendahuluan

Formula E atau yang memiliki nama resmi ABB FIA Formula E World Championship adalah sebuah ajang balap mobil kursi tunggal yang menggunakan mobil listrik [1]. Kejuaraan Formula E ini pertama kali di gagas pada tahun 2011 oleh presiden FIA Jean Todt dan Alejandro Agag yang juga sebagai pendiri dan CEO Formula E Holdings. Balapan Formula E pertama kali di gelar di Beijing, Tiongkok, pada bulan September 2014 dan berubah status menjadi kejuaraan dunia sejak tahun 2020.

Kota Jakarta resmi ditetapkan menjadi salah satu tuan rumah penyelenggaraan ajang balap Formula E yang akan dilaksanakan pada tanggal 4 Juni 2022 dengan slogan Jakarta E-Prix [2]. Penetapan Jakarta sebagai tuan rumah penyelenggaraan balapan Formula E ini melalui serangkaian proses yang panjang. Seharusnya Formula E Jakarta di gelar pada tahun 2020 tetapi ditunda akibat dari pandemi Covid-19 yang melanda kota Jakarta. Penundaan penyelenggaraan ini menimbulkan polemik dan menuai berbagai komentar dari masyarakat pada media sosial Twitter dikarenakan Pemerintah Provinsi Jakarta harus membayarkan sejumlah biaya komitmen [3] kepada Formula E Operations (FEO).

Polemik penyelenggaraan ajang balap Formula E Jakarta ini menuai berbagai komentar dari masyarakat yang disampaikan pada media sosial salah satunya adalah Twitter. Twitter merupakan salah satu media sosial yang cukup populer di Indonesia [4] dan menjadi platform untuk mengungkapkan pendapat masyarakat terhadap kebijakan publik yang dilakukan oleh pemerintah. Kecenderungan masyarakat dalam mengungkapkan pendapat mereka tentang kebijakan publik yang dilakukan oleh pemerintah khususnya kebijakan penyelenggaraan ajang balap Formula E Jakarta pada Twitter ini dapat dijadikan sebagai bahan analisis sentimen untuk mengetahui pendapat masyarakat tentang Formula E Jakarta.

Analisis sentimen [5] merupakan teknik untuk mendapatkan informasi tentang sentimen bernilai positif, netral maupun negatif dari data teks. Analisis sentimen dilakukan untuk mengetahui sentimen masyarakat pengguna Twitter terhadap rencana penyelenggaraan Formula E Jakarta. Dalam analisis sentimen, metode klasifikasi yang banyak digunakan adalah klasifikasi dengan menggunakan algoritma Naive Bayes [6]–[8].

Metode Naive Bayes dalam pengklasifikasian mampu menghasilkan model akurasi yang baik meskipun dengan jumlah data yang sedikit [9]. Performa metode Naive Bayes meski memiliki akurasi yang baik namun masih dapat ditingkatkan dengan berbagai macam teknik optimasi seperti peningkatan bobot atribut, seleksi atribut, dan penerapan seleksi fitur [10]. Penerapan seleksi fitur dilakukan karena metode Naive Bayes sangat sensitif terhadap seleksi fitur [11]. Pemilihan fitur

dapat meningkatkan skalabilitas, efisiensi, dan akurasi dari metode Naive Bayes.

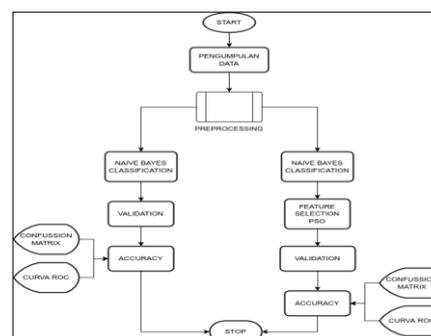
Beberapa penelitian sebelumnya [6], [12], [13] menerapkan Particle Swarm Optimization (PSO) pada metode klasifikasi dan mendapatkan hasil peningkatan akurasi dibandingkan tanpa optimasi dengan PSO. Salah satunya pada penelitian [12] menerapkan optimasi pada metode Naive Bayes dengan Particle Swarm Optimization (PSO) pada klasifikasi diagnosis penyakit diabetes Mellitus dan menunjukkan peningkatan hasil akurasi sebesar 82,58% dibanding metode Naive Bayes tanpa optimasi.

Pada penelitian [13] untuk meningkatkan performa dilakukan pengujian model Naive Bayes dengan PSO dengan beberapa parameter seperti jumlah partikel, jumlah iterasi, jumlah inertia weight dan didapatkan peningkatan hasil akurasi dari 89% menjadi 91,76%. PSO terbukti mampu untuk mengoptimasi akurasi dari metode Naive Bayes dan memecahkan masalah pada sentimen analisis.

Particle Swarm Optimization (PSO) adalah algoritma optimasi stokastik berbasis populasi yang dimotivasi oleh kecerdasan perilaku kolektif dari beberapa hewan seperti kawanan burung atau kumpulan ikan [14]. Terdapat beberapa parameter untuk melakukan optimasi dengan menggunakan PSO, di antaranya jumlah partikel yang menggambarkan kelompok data atau swarm, posisi vektor tiap partikel dalam kelompok data, arah gerakan dari partikel (velocity), learning rate, dan learning rate inertia.

2. Metode Penelitian

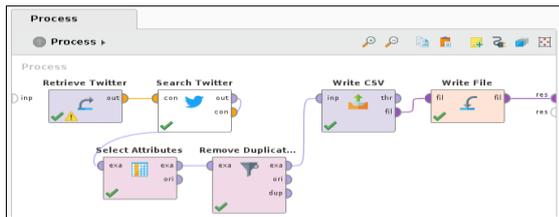
Penelitian ini akan melalui beberapa tahapan proses yang dijelaskan pada Gambar 1.



Gambar 1. Metode Penelitian

2.1. Pengumpulan Data

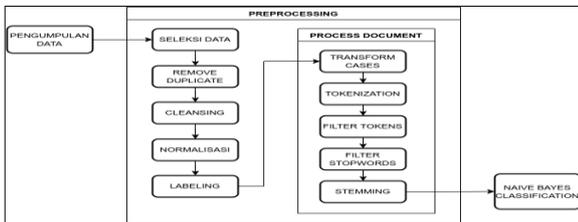
Pengumpulan data dilakukan dengan cara menarik atau crawling data komentar atau tweet menggunakan operator Search Twitter pada aplikasi RapidMiner Studio. Selanjutnya dilakukan pemilihan atribut menggunakan operator Select Attributes untuk memilih hanya data tweet yang diambil. Data yang diambil memiliki data yang duplikat sehingga diperlukan operator Remove Duplicate untuk menghapus data ganda. Data yang telah diambil kemudian di ekspor ke dalam berkas dengan format CSV.



Gambar 2. Proses Pengumpulan Data

2.2. Pengolahan Awal Data

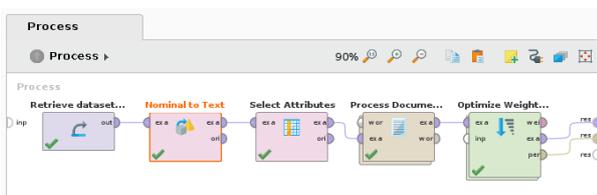
Pengolahan awal data dilakukan untuk membersihkan dataset dari data-data kotor atau noise yang tidak diperlukan dalam proses analisis sentimen. Untuk melakukan pengolahan awal data atau pre-processing dilakukan beberapa tahapan seperti yang dijelaskan pada Gambar 3.



Gambar 3. Proses Pengolahan Awal Data

2.3. Metode Yang Diusulkan

Pada penelitian ini metode yang diusulkan adalah klasifikasi sentimen menggunakan metode Naive Bayes dengan menerapkan feature selection Particle Swarm Optimization (PSO) untuk optimasi meningkatkan hasil akurasi dari metode Naive Bayes.



Gambar 4. Desain Pengujian Metode Naive Bayes Dengan PSO

2.4. Evaluasi Dan Validasi Hasil

Pada penelitian ini evaluasi yang dilakukan untuk mengetahui peningkatan akurasi dan kinerja dari metode Naive Bayes setelah diterapkan feature selection Particle Swarm Optimization (PSO). Sedangkan validasi hasil bertujuan untuk melihat perbedaan hasil akurasi dan kinerja dari metode atau model sebelum dan sesudah dilakukannya optimasi. Teknik validasi yang akan digunakan pada penelitian ini adalah Cross Validation sedangkan pengukuran akurasi akan menggunakan Confusion Matrix.

2.5. Pendukung Penelitian

Pada penelitian ini membutuhkan beberapa hal untuk menunjang jalannya penelitian yaitu:

2.5.1. Perangkat Lunak

Perangkat lunak yang digunakan dalam melakukan proses penelitian ini dijelaskan pada Tabel 1.

Tabel 1. Perangkat Lunak Yang Digunakan

Perangkat Lunak	Versi	Keterangan
Debian	11	Sistem operasi berjalan
LibreOffice Writer	7.2	Alat penulisan laporan
LibreOffice Calc	7.2	Alat pengolahan dataset
RapidMiner Studio	0	Alat digunakan pengolahan data
Draw.io	16.02	Alat diagram alur penelitian

2.5.2. Perangkat Keras

Perangkat keras yang digunakan dalam melakukan proses penelitian ini adalah laptop dengan spesifikasi seperti yang dijelaskan pada Tabel 2.

Tabel 2. Perangkat Keras Yang Digunakan

Spesifikasi Perangkat Keras	Keterangan
Prosesor	Intel Core i5-7200U 2.5GHz with Intel HD Graphics 620
RAM	4 GB DDR4
Memori	1000 GB HDD
Tipe Sistem	64 bit

3. Hasil dan Pembahasan

3.1. Hasil Pengumpulan Data

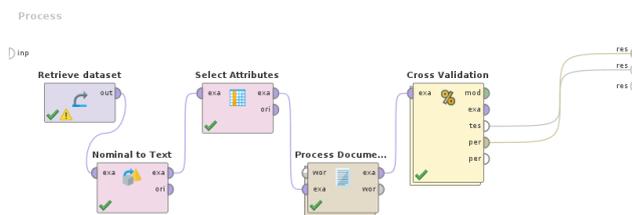
Data set yang digunakan pada penelitian ini berjumlah 629 data tweet yang terdiri dari tweet positif dan tweet negatif terhadap topik penyelenggaraan ajang balap Formula E Jakarta. Data tweet tersebut di ambil pada tanggal 01 Januari 2022 dengan proses crawling menggunakan aplikasi RapidMiner Studio dengan menggunakan kata kunci 'Formula E', 'Formula E Jakarta', 'Jakarta E-Prix', dan tagar #BalapanItuAduCepat.

Data set yang telah berhasil dikumpulkan kemudian dilakukan proses pengolahan awal data. Proses pengolahan awal data meliputi pemberian label. Proses pemberian label dilakukan untuk menentukan data tweet tersebut termasuk ke dalam tweet dengan sentimen positif atau tweet dengan sentimen negatif.

Setelah proses pemberian label dilakukan, tahap berikutnya adalah menghilangkan noise pada data sebelum melakukan implementasi algoritma. Pada tahapan ini dilakukan cleansing untuk menghilangkan atribut yang tidak diperlukan seperti URL, simbol RT, tagar, dan username.

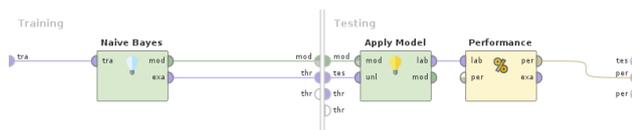
3.2. Pengujian Dataset Dengan Algoritma Naive Bayes

Pengujian pertama dilakukan dengan menguji dataset yang telah siap digunakan dengan menggunakan metode klasifikasi Naive Bayes tanpa optimasi. Proses pengujian menggunakan RapidMiner dengan merangkai beberapa proses yang dibutuhkan. Gambaran proses pengujian dijelaskan pada Gambar 5.



Gambar 5. Pengujian Dataset Dengan Algoritma Naive Bayes

Cross Validation memiliki dua sub proses yakni Training Subprocess dan Testing Subprocess. Pada Training Subprocess dilakukan untuk melatih model yang kemudian diterapkan ke Testing Subprocess. Performa dari sebuah model diukur ketika pada fase pengujian. Cross Validation pada pengujian dengan metode Naive Bayes tanpa optimasi diperlihatkan pada Gambar 6.



Gambar 6. Cross Validation Naive Bayes

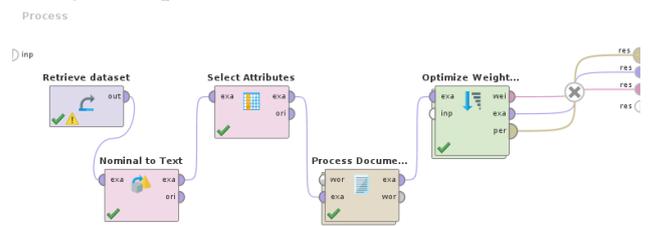
Proses pengujian gambar 5. dilakukan dengan teknik cross validation dengan meragamkan jumlah folds pada parameter untuk mendapatkan

hasil akurasi terbaik. Hasil pengujian dijelaskan pada Tabel 3.

Tabel 3. Hasil Pengujian Algoritma Naive Bayes

Number of Folds	Accuracy	Precision	Recall	AUC
2	78,59%	80,87%	74,75%	0,723
3	79,60%	81,67%	76,27%	0,500
4	79,09%	80,08%	77,79%	0,602
5	81,10%	81,57%	80,29%	0,541
6	81,36%	83,06%	78,80%	0,671
7	81,35%	83,48%	79,30%	0,579
8	81,14%	81,65%	80,96%	0,594
9	80,88%	81,43%	79,90%	0,580
10	81,37%	81,62%	81,54%	0,526

Pengujian berikutnya dilakukan dengan melakukan perubahan pada pengujian pertama yakni dengan menambahkan pemilihan fitur Particle Swarm Optimization (PSO). Proses pengujian dataset menggunakan metode Naive Bayes dengan optimasi Particle Swarm Optimization (PSO) ditunjukkan pada Gambar 7.

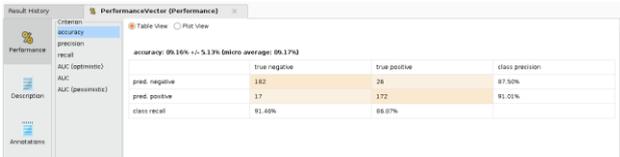


Gambar 7. Pengujian Dataset Dengan Algoritma Naive Bayes Dan Particle Swarm Optimization (PSO)

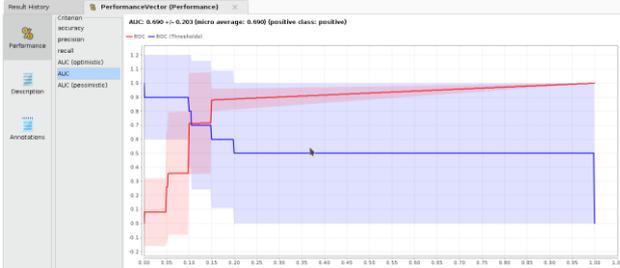
Dari pengujian yang dilakukan didapatkan nilai parameter terbaik Particle Swarm Optimization (PSO) untuk metode klasifikasi Naive Bayes seperti yang diperlihatkan pada Tabel 4.

Parameter	Nilai
Population Size	10
Inertia Weight	0,6
Global Best Weight	0,3
Local Best Weight	0,6

Berdasarkan nilai parameter pada Tabel 4. hasil pengujian Naive Bayes dengan optimasi menggunakan Parameter Swarm Optimization (PSO) menunjukkan peningkatan performa dengan mendapatkan nilai akurasi 89,16%, precision 91,10%, recall 86,81%, dan AUC 0,690. Nilai Confusion Matrix dijelaskan pada Gambar 8. dan kurva ROC pada Gambar 9.



Gambar 8. Confusion Matrix Naive Bayes PSO



Gambar 9. Kurva ROC Naive Bayes PSO

3.3. Pembahasan Hasil Pengujian

Performa dari metode klasifikasi yang digunakan dalam pengujian dataset ditunjukkan dengan seberapa besar nilai akurasi. Nilai akurasi merupakan nilai yang menunjukkan tingkat kedekatan nilai prediksi dengan nilai yang sebenarnya. Berdasarkan pengujian gambar 5. dan 7. dapat dilihat bahwa ada perubahan nilai performa dari pengujian setelah menerapkan pemilihan fitur Particle Swarm Optimization (PSO). Perbandingan dari hasil kedua pengujian dijelaskan pada Tabel 5.

Tabel 5. Perbandingan Hasil Pengujian

Pengujian	Accuracy	Precision	Recall	AUC
Naive Bayes	81,37%	81,62%	81,54%	0,580
Naive Bayes + PSO	89,16%	91,10%	86,81%	0,690

4. Kesimpulan

Berdasarkan penelitian yang dilakukan dapat diambil kesimpulan bahwa:

Metode klasifikasi Naive Bayes memiliki hasil dan kinerja yang cukup baik tetapi masih belum optimal. Metode Naive Bayes tanpa optimasi mendapatkan nilai akurasi sebesar 81,27%, precision 81,62%, recall 81,54% dan AUC 0,526. Pemilihan fitur Particle Swarm Optimization (PSO) dapat meningkatkan hasil dan kinerja dari metode klasifikasi Naive Bayes. Setelah diterapkan PSO metode Naive Bayes mengalami peningkatan performa akurasi menjadi 89,16%, precision 91,10%, recall 86,81% dan AUC 0,690. Parameter PSO yang berpengaruh adalah population size, inertia weight, global best weight, dan local best weight.

Nilai indikator pada pemilihan fitur berpengaruh terhadap hasil dan kinerja metode klasifikasi Naive Bayes. Pemilihan parameter yang optimal

meningkatkan performa metode Naive Bayes seperti akurasi menjadi 89,16%, precision 91,10%, recall 86,81%, dan AUC 0,690.

Mengombinasikan metode klasifikasi Naive Bayes dengan pemilihan fitur yang lain untuk mendapatkan hasil dan kinerja yang maksimal.

Meragamkan nilai indikator untuk mendapatkan hasil optimal dari pemilihan fitur Particle Swarm Optimization (PSO).

Menggunakan metode klasifikasi selain Naive Bayes

Ucapan Terima Kasih

Peneliti mengucapkan terima kasih kepada Dinas Komunikasi dan Data DKI Jakarta dan Universitas Pelita Bangsa yang telah mendukung dan membantu dalam memberikan data maupun informasi untuk tujuan penelitian ini.

Referensi

- [1] "Formula E," Wikipedia bahasa Indonesia, ensiklopedia bebas. 2 Desember 2021. Diakses: 20 Desember 2021. [Daring]. Tersedia pada: https://id.wikipedia.org/w/index.php?title=Formula_E&oldid=19497301
- [2] K. C. Media, "Jakarta Resmi Jadi Tuan Rumah Formula E 2022," Kompas.com, 16 Oktober 2021. <https://otomotif.kompas.com/read/2021/10/16/110452815/jakarta-resmi-jadi-tuan-rumah-formula-e-2022> (diakses 20 Desember 2021).
- [3] "FEO Pastikan Jakarta Bayar Commitment Fee Formula E Sama dengan Kota Negara Lain." <https://megapolitan.kompas.com/read/2021/11/25/05371921/feo-pastikan-jakarta-bayar-commitment-fee-formula-e-sama-dengan-kota> (diakses 27 Desember 2021).
- [4] P. KOMINFO, "Indonesia Peringkat Lima Pengguna Twitter," Website Resmi Kementerian Komunikasi dan Informatika RI. http://content/detail/2366/%20indonesia-peringkat-lima-penggunatwitter/0/sorotan_media (diakses 20 Desember 2021).
- [5] B. Liu, "Sentiment Analysis and Opinion Mining," Synth. Lect. Hum. Lang. Technol., vol. 5, no. 1, hlm. 1–167, Mei 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.
- [6] Y. Cahyono, "Analisis Sentiment pada Sosial Media Twitter Menggunakan Naive Bayes Classifier dengan Feature Selection Particle Swarm Optimization (PSO)." *SIGMA - Jurnal Teknologi Pelita Bangsa*

- Swarm Optimization dan Term Frequency,” *J. Inform. Univ. Pamulang*, vol. 2, no. 1, hlm. 14, Mar 2017, doi: 10.32493/informatika.v2i1.1500.
- [7] S. Fanissa, M. A. Fauzi, dan S. Adinugroho, “Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking,” hlm. 5.
- [8] M. A. Fauzi, A. Z. Arifin, dan S. C. Gosaria, “Indonesian News Classification Using Naive Bayes and Two-Phase Feature Selection Model,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 8, no. 3, hlm. 610, Des 2017, doi: 10.11591/ijeecs.v8.i3.pp610-615.
- [9] G. Feng, J. Guo, B.-Y. Jing, dan T. Sun, “Feature subset selection using naive Bayes for text classification,” *Pattern Recognit. Lett.*, vol. 65, hlm. 109–115, Nov 2015, doi: 10.1016/j.patrec.2015.07.028.
- [10] F. Pramono, Didi Rosiyadi, dan Windu Gata, “Integrasi N-gram, Information Gain, Particle Swarm Optimization di Naive Bayes untuk Optimasi Sentimen Google Classroom,” *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 3, no. 3, hlm. 383–388, Des 2019, doi: 10.29207/resti.v3i3.1119.
- [11] J. Chen, H. Huang, S. Tian, dan Y. Qu, “Feature selection for text classification with Naive Bayes,” *Expert Syst. Appl.*, vol. 36, no. 3, hlm. 5432–5435, Apr 2009, doi: 10.1016/j.eswa.2008.06.054.
- [12] D. S. Susilawati dan D. Riana, “Optimization the Naive Bayes Classifier Method to diagnose diabetes Mellitus,” *IAIC Trans. Sustain. Digit. Innov. ITSDI*, vol. 1, no. 1, hlm. 78–86, Apr 2021, doi: 10.34306/itsdi.v1i1.21.
- [13] S. Khomsah, “Naive Bayes Classifier Optimization on Sentiment Analysis of Hotel Reviews,” *J. Penelit. Pos Dan Inform.*, vol. 10, no. 2, hlm. 157, Des 2020, doi: 10.17933/jppi.2020.100206.
- [14] R. Eberhart dan J. Kennedy, “A new optimizer using particle swarm theory,” dalam *MHS’95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, Nagoya, Japan, 1995, hlm. 39–43. doi: 10.1109/MHS.1995.494215.
- [15] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, hlm. 50, Jun 2018, doi: 10.35314/isi.v3i1.335.
- [16] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, dan W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, hlm. 115, Jul 2020, doi: 10.33365/jti.v14i2.679.
- [17] V. Chandani dan R. S. Wahono, “Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film,” *J. Intell. Syst.*, vol. 1, no. 1, hlm. 5, 2015.
- [18] Y. Cahyono dan S. Saprudin, “Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes Classifier dengan Seleksi Feature Chi Squared Statistic,” *J. Inform. Univ. Pamulang*, vol. 4, no. 3, hlm. 87, Sep 2019, doi: 10.32493/informatika.v4i3.3186.
- [19] M. Taboada, “Sentiment Analysis: An Overview from Linguistics,” *Annu. Rev. Linguist.*, vol. 2, no. 1, hlm. 325–347, Jan 2016, doi: 10.1146/annurev-linguistics-011415-040518.
- [20] A. M. S. M. Kom S. Kom, M. Kom DAN Adam Puspabhuana, S. Kom, *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group.
- [21] D. T. Larose, “Discovering Knowledge in Data 2nd Edition,” *Discov. Knowl. Data*, hlm. 15, 2014.
- [22] “Twitter,” *Wikipedia bahasa Indonesia, ensiklopedia bebas*. 30 November 2021. Diakses: 20 Desember 2021. [Daring]. Tersedia pada: <https://id.wikipedia.org/w/index.php?title=Twitter&oldid=19484193>
- [23] A. Taufiq, “Analisis Sentimen Terhadap Pemindahan Ibu Kota Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization,” hlm. 7, 202