

Analisis Sentimen Terhadap Pembatalan Piala Dunia U-20 di Indonesia Menggunakan Algoritma Naïve Bayes dan Genetika

I Ketut Oning Pusparama¹, I Putu Gede Hendra Suputra², Gst. Ayu Vida Mastrika Giri³

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana Badung, Bali, Indonesia

¹iktoningpuspa21@gmail.com

²hendra.suputra@unud.ac.id

³vida@unud.ac.id

Abstrak

Pembatalan Piala Dunia U-20 di Indonesia menimbulkan berbagai reaksi masyarakat di media sosial, khususnya Twitter. Sebagian masyarakat menyayangkan keputusan tersebut karena dianggap merugikan Indonesia dalam kesempatan tampil di level dunia. Sebaliknya ada pihak yang mendukung pembatalan dengan alasan tertentu, seperti faktor politik atau prinsip nasional. Hal ini dapat dilakukan analisis sentimen terhadap permasalahan tersebut. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap pembatalan Piala Dunia U-20 dengan menerapkan algoritma Naïve Bayes dengan seleksi fitur Algoritma Genetika. Data penelitian diperoleh melalui proses crawling di Twitter pada periode 30 Maret hingga 5 April 2023 dengan kata kunci "Pembatalan Piala Dunia U-20 Indonesia" dan memperoleh data sebanyak 14.807 tweet. Penelitian ini menggunakan data sebanyak 2.000 dan dilakukan pelabelan data data tersebut bisa digunakan untuk pelatihan model klasifikasi. Tahap preprocessing meliputi case folding, filtering, tokenizing dan stemming. Tahap pembobotan data menggunakan TF-IDF mengukur tingkat kepentingan atau bobot suatu kata dalam sebuah dokumen. Penelitian ini menggunakan seleksi fitur genetika untuk memilih fitur (kata) terbaik dan mengurangi kata yang kurang relevan. Naive bayes merupakan algoritma klasifikasi berbasis probabilitas untuk memprediksi sentimen positif atau negatif. Proses klasifikasi dilakukan dengan dua skenario, yaitu tanpa seleksi fitur dan dengan seleksi fitur menggunakan Algoritma Genetika. Hasil penelitian menunjukkan bahwa model Naïve Bayes tanpa seleksi fitur memperoleh akurasi sebesar 71,5% dan mengalami peningkatan menjadi 76,8% menggunakan seleksi fitur genetika. Peningkatan juga terjadi pada nilai presisi, recall, dan F1-score, masing-masing mencapai sekitar 75,99%, 76%, dan 75,98%.

Keywords: klasifikasi, Naïve Bayes, genetika, piala dunia, twitter

1. Pendahuluan

Pada saat ini perkembangan dunia digital sudah semakin canggih. Hal ini menyebabkan perkembangan infromasi lebih mudah didapatkan melalui media sosial. Infromasi ini dapat berupa teks, foto ataupun video. Informasi yang paling sering di media sosial adalah teks. Teks ini biasanya berisi opini masyarakat berupa fakta, pujian, ujaran kebencian ataupun *hoax* yang dapat menimbulkan perdebatan di media sosial. Salah satu sosial media untuk melakukan diskusi adalah dengan penggunaan *twitter*.

Twitter adalah platform media sosial yang populer di mana pengguna dapat membuat, membaca, dan berinteraksi dengan pesan singkat yang disebut "tweets" [1]. *Tweet* dapat berupa sentimen seseorang baik opini positif ataupun negatif. *Twitter* mempunyai karakteristik yaitu dapat mengirim opini secara bebas tanpa batas, mencari berita terkini maupun lampau, membagikan *tweet* pengguna lain dan memberi komentar. Karena sifatnya yang terbuka, platform ini menjadi sumber data yang sangat

potensial untuk melakukan analisis sentimen guna mengetahui persepsi publik terhadap suatu peristiwa atau kebijakan.

Sepak bola merupakan salah satu olahraga yang paling terkenal di dunia. Sepak bola sangat diminati oleh masyarakat dunia khususnya masyarakat Indonesia. Saat ini perkembangan sepak bola di Indonesia mengalami peningkatan yang lumayan signifikan. Hal ini dibuktikan dengan Indonesia menjadi runner-up di piala AFF 2022 dan dalam waktu dekat ini juga Piala Dunia U-20 akan diselenggarakan di Indonesia. Antusias Masyarakat sangat sangat positif menyambut event yang akan diadakan di Indonesia. Tetapi akibat isu yang beredar, Federasi Sepak Bola Internasional (FIFA) memutuskan untuk membatalkan penyelenggaraan piala dunia U-20 di Indonesia. Akibat pembatalan tersebut, sebagian masyarakat menganggap merugikan Indonesia atas kesempatan bagi pemain muda untuk tampil di level dunia. Sebaliknya ada pihak yang mendukung pembatalan dengan alasan tertentu, seperti faktor politik atau prinsip nasional. Hal ini dapat dilakukan analisis sentimen terhadap permasalahan tersebut.

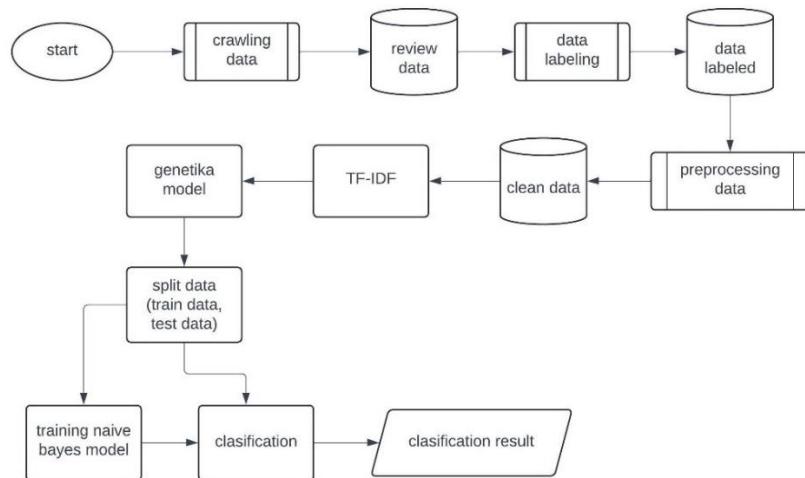
Analisis sentimen dapat mengungkap persepsi masyarakat terhadap kebijakan publik secara real-time melalui data media sosial. Dengan menggunakan pendekatan analisis sentimen, peneliti dapat menangkap kecenderungan sikap masyarakat, baik yang bersifat positif, negatif, maupun netral terhadap kebijakan tersebut [2]. Hasil analisis ini dapat dimanfaatkan oleh pemerintah, organisasi olahraga, maupun lembaga media untuk mengevaluasi kebijakan, merancang strategi komunikasi publik, serta memahami tingkat penerimaan masyarakat terhadap keputusan yang diambil.

Dalam berbagai penelitian sebelumnya, metode Naïve Bayes telah banyak digunakan dalam analisis sentimen karena kesederhanaan dan efisiensinya dalam melakukan klasifikasi teks Ciri khas dari Naïve Bayes adalah asumsi independensi antar fitur, yaitu menganggap bahwa setiap fitur atau kata dalam dokumen teks tidak saling bergantung satu sama lain dan menghasilkan model yang sederhana namun cukup efektif dalam banyak kasus, terutama dalam tugas klasifikasi berbasis teks seperti analisis sentimen [3]. Meskipun metode ini populer, asumsi independensi tersebut sering kali tidak terpenuhi dalam data teks yang kompleks, sehingga dapat menyebabkan penurunan akurasi klasifikasi.

Pembobotan TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode yang bertujuan memberikan nilai pada kata-kata yang sering digunakan [4]. Kombinasi Naïve Bayes dan TF-IDF telah diterapkan oleh beberapa peneliti, seperti penelitian oleh Rusdiaman dan Rosiyadi (2019) yang berhasil memperoleh akurasi sebesar 77,78% dalam analisis sentimen terhadap tokoh publik. Namun, hasil akurasi masih dapat ditingkatkan dengan menambahkan teknik seleksi fitur untuk mengeliminasi fitur yang tidak relevan. Salah satu pendekatan yang efektif untuk meningkatkan performa klasifikasi adalah seleksi fitur menggunakan algoritma Genetika. Genetika merupakan metode berbasis evolusi biologis yang mencari kombinasi fitur terbaik dengan proses seleksi, crossover, dan mutasi. Beberapa penelitian menunjukkan bahwa penerapan GA dapat meningkatkan akurasi Naïve Bayes secara signifikan. Misalnya, penelitian oleh Rafdi et al. (2021) menunjukkan peningkatan akurasi dari 60,26% menjadi 70,71% setelah menerapkan seleksi fitur genetika. Selain algoritma Genetika, metode lain seperti Information Gain, Chi-Square, dan Support Vector Machine (SVM) juga telah digunakan untuk analisis sentimen, namun GA lebih unggul dalam mencari subset fitur optimal secara heuristik dan adaptif.

2. Metode Penelitian

Penelitian ini dibagi menjadi beberapa tahapan. Tahapan-tahapan ini mulai dari tahap pengumpulan data, *data labelling, preprocessing*, TF-IDF, seleksi fitur, klasifikasi dan evaluasi hasil. Berikut ini adalah *flowchart* yang menunjukkan alur dari penelitian yang dilakukan.



Gambar 1. Alur Penelitian

2.1. Pengumpulan data

Data yang digunakan dalam penelitian ini adalah data sekunder. Data ini didapatkan melalui proses *crawling data* pada *twitter*. Data *twitter* di ambil pada tanggal 30 Maret 2023 sampai 05 April 2023 dengan kata kunci pencarian “Pembatalan Piala Dunia U-20 Indonesia”. Hasil dari proses *crawling* memperoleh data sebanyak 14.807 dengan 3 *data feature*.

Tabel 1. Contoh dataset

Datetime	Username	Text
2023-04-05	halintarbungkam	@ch_chotimah2 @Dennysiregar7 @erickthohir @jokowi Lupakan kegagalan tuan rumah piala dunia u-20 mari kita menata masa depan sepak bola seluruh Indonesia kedepan yang kelam di mata internasional.
2023-04-05	kirarachelsea	Gagalnya perhelatan piala dunia U-20 di Indonesia menjadikan suatu pengalaman yg mendewasakan kita semua.

Pada tabel 1 merupakan contoh data yang di peroleh melalui proses *crawling data*. Dataset yang diperoleh berisikan tanggal unggah *twitter*, *username* dan *text tweet*. Dataset belum memiliki label data positif dan negatif, sehingga pada proses selanjutnya dilakukan proses *labelling data*.

2.2. Data Labelling

Pada penelitian ini, data yang digunakan sebanyak 2.000 data sehingga dilakukan reduksi dari data yang diperoleh. Reduksi data dilakukan agar saat proses *labelling* tidak memakan waktu yang lama. Proses *labelling* dilakukan menerjemahkan data bahasa Indonesia ke bahasa Inggris. Hasil terjemahan data dapat dilihat pada tabel 2 berikut:

Tabel 2. Data yang sudah di terjemah

Text	Translate Text
Sepakbola tidak bisa dilepaskan dari urusan politik. Kegagalan menjadi tuan rumah Piala Dunia U-20 menjadi pelajaran untuk membenahi sepak	Football cannot be separated from political matters. The failure to host the U-20 World Cup is a lesson for

bola dan lebih luas lagi, menanggapi perkara persatuan bangsa.	improving football and more broadly, responding to matters of national unity.
--	---

Setelah dilakukan proses terjemah, dilakukan *labelling* terhadap data yang sudah diterjemahkan sebelumnya. Hasil *labelling* dapat dilihat pada tabel 3 berikut:

Tabel 3. Data yang sudah di label

Text	Label
Sepakbola tidak bisa dilepaskan dari urusan politik. Kegagalan menjadi tuan rumah Piala Dunia U-20 menjadi pelajaran untuk membenahi sepak bola dan lebih luas lagi, menanggapi perkara persatuan bangsa.	Positif

Setelah didapatkan hasil label pada data, dilakukan proses validasi secara manual oleh ahli bahasa universitas udayana.

2.3. *Preprocessing*

Pada penelitian ini, terdapat beberapa tahapan *pre-processing* seperti reduksi data, *case folding*, *filtering*, *tokenizing* dan *stemming*. *Pre-processing* ini dilakukan untuk menghilangkan missing value, *data redundant*, *outliers*, ataupun format data yang tidak sesuai dengan sistem [5].

- a. **Reduksi data**
Proses ini merupakan proses untuk mengurangi jumlah data agar data yang diinginkan sudah sesuai.
- b. **Case folding**
Proses ini merupakan proses untuk merubah keseluruhan teks yang ada dalam *dataset* menjadi huruf kecil sehingga semua teks memiliki bentuk yang sama.
- c. **Filtering**
Proses ini merupakan proses untuk menghilangkan simbol, angka dan kata yang kurang penting (stopwords) pada kalimat.
- d. **Tokenizing**
Proses ini merupakan proses untuk memecah teks menjadi bagian-bagian yang lebih kecil, seperti kata atau frasa.
- e. **Stemming**
Proses ini merupakan proses menghilangkan kata imbuhan awal maupun akhir dari sebuah teks.

2.4. TF-IDF

TF-IDF merupakan sebuah metode integrasi antara term frequency (TF) dan inverse document frequency (IDF) [6]. Term Frequency (TF) adalah proses menghitung jumlah frekuensi kata pada suatu dokumen atau data. Inverse Document Frequency (IDF) merupakan jumlah dokumen atau data yang berisikan kata yang dicari dalam dokumen atau data tersebut. TF-IDF digunakan untuk menghitung setiap kata pada dokumen dan memberikan bobot untuk merepresentasikan seberapa penting kata tersebut dalam dokumen tersebut. Bobot ini dihitung dengan mengalikan frekuensi kemunculan kata tersebut pada dokumen dengan nilai IDF dari kata tersebut. Metode ini digunakan untuk mengetahui seberapa sering sebuah kata muncul dalam suatu dokumen atau data. Pembobotan yang dilakukan agar fitur atau kata dapat di proses pada mesin klasifikasi.

2.5. Genetika

Genetika merupakan suatu metode heuristik yang dikembangkan berdasarkan prinsip genetika dan proses seleksi alamiah seperti pewarisan sifat, seleksi alam, mutasi gen dan kombinasi (*crossover*) pada teori Evolusi Darwin [7]. Metode optimasi dikembangkan oleh John Hollan sekitar tahun 1960-an

dan dipopulerkan pada tahun 1980-an oleh salah seorang mahasiswanya, David Goldberg. Algoritma Genetika ini digunakan untuk mencari solusi optimal dari permasalahan yang mempunyai banyak solusi. Teknik ini akan melakukan pencarian dari beberapa solusi yang diperoleh hingga mendapatkan solusi paling optimal sesuai dengan kriteria yang sudah ditentukan atau yang disebut sebagai nilai fitness. Berikut adalah proses yang dilakukan dalam algoritma genetika:

1. Pembentukan individu dengan menentukan nilai populasi awal. Individu dalam algoritma genetika disebut dengan kromosom. Kromosom terdiri dari Kumpulan gen dimana gen berisikan nilai biner, float ataupun kombinasi dari keduanya.
2. Crossover, yaitu proses kawin silang yang dilakukan oleh dua kromosom induk yang disebut dengan "parent" sehingga menghasilkan individu atau kromosom baru.
3. Mutasi, yaitu proses penukaran nilai gen dengan nilai intervensinya misalnya gen bernilai 0 akan menjadi 1. Proses mutasi ini dilakukan dengan probabilitas mutasi yang sudah ditentukan sebelumnya.
4. Nilai *fitness*, yaitu nilai yang menyatakan baik tidaknya suatu individu atau kromosom. Nilai fitness ini dijadikan acuan untuk menentukan nilai optimal pada algoritma genetika.
5. Seleksi, yaitu proses pemilihan individu atau kromosom terbaik berdasarkan nilai fitness yang ada pada individu atau kromosom tersebut. Individu yang memiliki nilai fitness tertinggi akan menjadi parent untuk proses mutasi selanjutnya.

Proses crossover, mutasi dan seleksi akan terus berulang hingga mencapai nilai threshold (ambang batas) atau jumlah generasi maksimum yang ditentukan.

2.6. Naïve Bayes

Naïve Bayes merupakan salah satu algoritma klasifikasi yang paling sederhana dan efisien dalam bidang pembelajaran mesin, khususnya dalam pemrosesan bahasa alami (Natural Language Processing) [8]. Metode ini memprediksi kelas berdasarkan kemungkinan kelas yang tersedia dari data latih yang diberikan. Teori bayes sangat berguna untuk penyediaan perhitungan kemungkinan posterior, dimana kemungkinan posterior adalah kemungkinan suatu peristiwa yang akan terjadi setelah semua bukti dan informasi sudah di perhitungkan. Teori bayes dirumuskan pada persamaan 1 berikut:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Dengan:

A = kelas yang sudah diketahui

B = kelas yang belum diketahui

$P(A|B)$ = kemungkinan posterior dari A bersyarat B sudah terjadi

$P(B|A)$ = kemungkinan posterior dari B bersyarat A sudah terjadi

$P(A)$ = kemungkinan prior dari A

$P(B)$ = kemungkinan prior dari B

2.7. Confusion Matrix

Confusion matrix merupakan sebuah metode untuk mengukur kinerja dari suatu metode klasifikasi. *Confusion matrix* menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [9]. Pengukuran ini dilakukan untuk melihat dan menggambarkan seberapa efektif sistem dalam mengklasifikasi data.

1. Akurasi adalah nilai berapa banyak jumlah hasil TP dan TN dibandingkan jumlah semua hasil. Nilai akurasi dapat dihitung dengan persamaan (2):

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

2. Presisi adalah pengukuran seberapa baik hasil klasifikasi dan nilai yang didapat pada rasio TP dengan semua deteksi TP dan FP, dengan persamaan (3):

$$Presisi = \frac{TP}{TP+FP} \quad (3)$$

3. *Recall* adalah pengukuran seberapa baik hasil klasifikasi dan nilai yang didapat pada rasio hasil deteksi TP dari kasus positif yang seharusnya, dengan persamaan (4):

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

4. F-1 score adalah nilai rata-rata antara *precision* dan *recall* yang diberi bobot. F-1 score digunakan sebagai acuan jika jumlah data dari FP dan FN tidak mendekati. Nilai F-1 score dapat dihitung dengan persamaan (5):

$$F - 1 Score = 2 \times \frac{presisi \times recall}{presisi + recall} \quad (5)$$

3. Hasil dan Pembahasan

3.1. Preprocessing

Pada tahap ini, dilakukan beberapa proses *preprocessing* yaitu *case folding*, *cleaning*, *tokenizing*, *stopword removal* dan *stemming*. Proses ini dilakukan agar nantinya data bersifat bersih dan stabil sehingga dapat masuk ke proses selanjutnya. Hasil dari proses *preprocessing* dapat dilihat pada tabel 4 berikut:

Tabel 4. Data sebelum dan sesudah di *preprocessing*

Sebelum	Sesudah
Tetap semangat anak-anak bangsa	semangat anak anak bangsa
Kecewa Piala Dunia U-20 di Indonesia Batal, Suporter Timnas: Tandai Mereka di 2024, Selamat Tertawa di Atas Penderitaan	kecewa piala dunia indoensia batal supporter timnas tanda selamat tertawa derita

3.2. Pembobotan Data

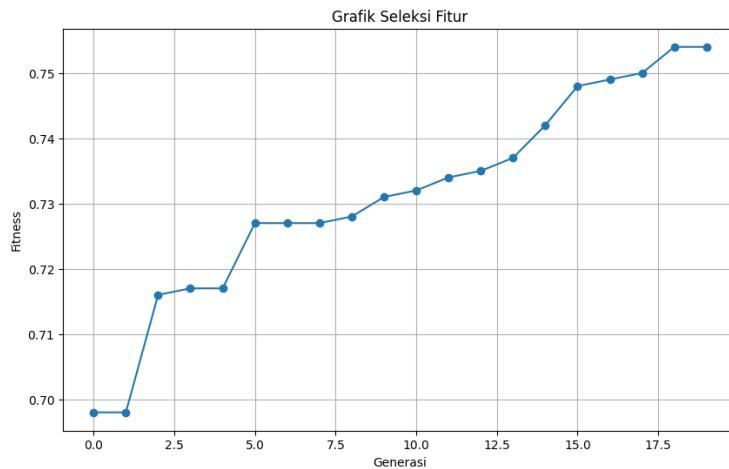
Pada tahap ini, dilakukan beberapa proses pembobotan data dengan metode TF-IDF. Proses perhitungan TF-IDF ini mendapat fitur-fitur sejumlah 2.612 fitur (kata). Kata-kata ini akan diproses lebih lanjut pada tahap pelatihan model. Hasil dari pembobotan data dengan 5 kata dengan nilai TF-IDF tertinggi dapat dilihat pada tabel 5 berikut:

Tabel 5. 5 Nilai Tertinggi TF-IDF

No	Fitur	Nilai TF-IDF
1	degradasi	0.414
2	rusuh	0.374
3	tentu	0.374
4	kompetisi	0.361
5	transformasi	0.342

3.3. Klasifikasi

Terdapat 2 skenario klasifikasi, yaitu dengan menggunakan seleksi fitur dan tanpa seleksi fitur. Seleksi fitur dilakukan dengan algoritma genetika. Terdapat 3 tuning *hyperparameter* yang digunakan yaitu: populasi awal, *crossover* dan *mutation*. Nilai populasi awal yang digunakan sebesar 50, *crossover* sebesar 0,5 dengan metode *Two-Point crossover* dan *mutation* sebesar 0,2 dengan metode *Flip Bit Mutation*. Pada proses inisiasi populasi awal, ditentukan jumlah individu sebanyak 50. Dalam populasi terdapat 50 individu dengan gen (solusi) yang berbeda. Pembentukan individu dilakukan secara acak berdasarkan kamus yang sudah dibentuk dalam proses TF-IDF. Individu-individu pada populasi ini akan mengalami proses yang sama seperti generasi sebelumnya yaitu proses evaluasi nilai fitness, seleksi, *crossover* dan mutasi. Proses ini akan menghasilkan kromosom-kromosom baru untuk generasi yang selanjutnya. Proses ini akan berulang sampai sejumlah generasi yang telah ditetapkan yaitu sebanyak 20 kali. Hasil dari proses seleksi fitur dapat dilihat pada gambar 1 berikut:



Dapat dilihat pada gambar 1 bahwa terjadi peningkatan nilai *fitness* pada setiap generasi. Nilai *fitness* tertinggi didapat pada generasi ke-20 yaitu 0,755. Hal ini menandakan bahwa saat proses seleksi fitur sudah menemukan kombinasi cukup optimal, sehingga perbaikan tambahan semakin kecil. Nilai *fitness* yang tinggi juga menandakan bahwa kualitas model semakin baik.

Setelah dilakukan proses seleksi fitur selesai, dilakukan evaluasi menggunakan *confusion matrix*. Perbandingan hasil evaluasi klasifikasi dengan seleksi fitur dan tanpa seleksi fitur ditunjukkan pada tabel 6 berikut:

Tabel 6. Perbandingan Evaluasi Model

Model	Akurasi	Presisi	Recall	F1-score
Tanpa Seleksi Fitur	71,5%	71,49%	71,5%	71,49%
Dengan Seleksi Fitur	76,8%	75,99%	76%	71,98%

Pada tabel 6 dapat dilihat bahwa klasifikasi tanpa seleksi fitur hanya memperoleh akurasi sebesar 71,5%, presisi sebesar 71,49%, recall sebesar 71,5% dan F1-score sebesar 71,49% sedangkan pengklasifikasian sentimen menggunakan algoritma Naive Bayes dan seleksi fitur menggunakan algoritma Genetika memperoleh nilai evaluasi terbaik pada iterasi ke-20 dengan akurasi 76,8%, presisi sebesar 75,99%, recall sebesar 76% dan F-1 score sebesar 75,98%.

4. Kesimpulan

Berdasarkan penelitian yang sudah dilakukan, dapat dilihat bahwa seleksi fitur dengan algoritma genetika dapat meningkatkan evaluasi model pada data *twitter*. Seleksi fitur dengan algoritma genetika nilai populasi awal sebesar 50, *crossover* sebesar 0,5 dengan metode Two-Point Crossover, dan *mutation* sebesar 0,2 metode Flip Bit Mutation efektif meningkatkan akurasi, presisi, *recall* dan F1-score. Klasifikasi menggunakan *naïve bayes* tanpa seleksi fitur memperoleh akurasi sebesar 71,5%, presisi sebesar 71,49%, *recall* sebesar 71,5% dan F-1 score sebesar 71,49% sedangkan pengklasifikasian sentimen menggunakan algoritma Naive Bayes dan seleksi fitur menggunakan algoritma Genetika memperoleh nilai evaluasi terbaik pada generasi ke-20 dengan akurasi 76,8%, presisi sebesar 75,99%, *recall* sebesar 76% dan F-1 score sebesar 75,98%. Keaikan akurasi menggunakan seleksi fitur meningkatkan peforma klasifikasi pada data *twitter* sebesar 6-7%.

Referensi

- [1] Yusupa, A., and Tarigan, V. 2024. Perbandingan Algoritma Machine Learning dalam Analisis Sentimen Mobil Listrik di Indonesia pada Media Sosial Twitter/X JIP (Jurnal Informatika Polinema), vol. 10, no. 4, pp. 479–490.
- [2] Yanuar Laik, A. A., Nabilla, A., Diah, A., Sumanto, Indra, A., and Arya, Y. 2025. Komparasi Naive Bayes dan SVM untuk Analisis Sentimen Pada E-Commerce Seller Center, Jurnal Sains dan Teknologi (JSIT), vol. 5, no. 3, pp. 199–208.
- [3] Setiawan, D., Umar, N., and M. Adnan Nur. 2024. Optimasi Ekstraksi Fitur untuk Meningkatkan Akurasi Naïve Bayes dalam Analisis Sentimen Objek Wisata Bulukumba, Sistemasi: Jurnal Sistem Informasi, vol. 13, no. 5, pp. 2209–2221. Apriani, R., and Gustian, D. 2019. Analisis Sentimen dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia, vol. 6, no. 1, pp. 54-62.
- [4] Jiang, Z., Gao, B., He, Y., Han, Y., Doyle, P., and Zhu, Q. 2021. Text Classification Using Novel Term Weighting Scheme-Based Improved TF-IDF for Internet Media Reports, Mathematical Problems in Engineering, vol. 2021.
- [5] Wulan, S. T., Bettiza, M., and Hayaty, N. 2017. Optimasi Seleksi Fitur Klasifikasi Naïve Bayes Menggunaan Algoritma Genetika Untuk Prediksi Risiko Kredit Konsumen.
- [6] Ginting, A., Purba, D., Sinaga, L. M., and Sagala, M. 2025. Analisis Sentimen Masyarakat Twitter terhadap Emas Digital Menggunakan Algoritma Naïve Bayes, KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer), pp. 8–14, 2025.
- [7] Normati, D., and Prayogi S. A. 2021. Implementasi Naïve Bayes Clasifier dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter, vol. 5, no. 2, pp. 697-711.
- [8] Rusdiaman, D., and Rosiyadi, D. 2019. Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine. CESS (Journal of Computer Engineering System and Science), vol. 4, no. 2, pp. 230-235.
- [9] Rafdi, A., Mawengkang, H., and Efendi, S. 2021. Sentiment Analysis Using Naive Bayes Algorithm with Feature Selection Particle Swarm Optimization (PSO) and Genetic Algorithm, International Journal of Advances in Data and Information Systems, vol. 2, no. 2, pp. 96–104.