Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS

https://doi.org/10.35870/jimik.v5i3.1024

Analisis Sentimen Naturalisasi Tim Nasional Indonesia U-23 di Era Shin Tae-yong Menggunakan Algoritma *Naïve Bayes* dan *K-Nearest Neighbors*

Dava Rizky Perwira Jaya [1*], Sri Lestari [2]

[1*,2] Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

Email: dafajaya92@gmail.com [1*], sri.lestari1203@gmail.com [2]

Histori Artikel:
Dikirim 31 Juli 2024; *Diterima dalam bentuk revisi* 11 Agustus 2024; *Diterima* 30 Agustus 2024; *Diterbitkan* 30 September 2024.

Abstrak

Dalam era Shin Tae-yong, naturalisasi pemain telah menjadi topik kontroversial dalam tim nasional sepak bola Indonesia U-23. Penelitian ini bertujuan untuk menganalisis sentimen publik terkait naturalisasi pemain dalam tim tersebut menggunakan dua algoritma klasifikasi, yaitu Naive Bayes dan K-Nearest Neighbor. Data sentimen diperoleh dari sumber-sumber berita, media sosial, dan forum diskusi online terkait pertandingan dan keputusan manajemen tim. Pertama, dilakukan pengolahan data, termasuk pembersihan teks, tokenisasi, dan pembobotan kata. Selanjutnya, model Naive Bayes dan KNN dilatih menggunakan dataset yang telah diproses. Hasil analisis sentimen diharapkan memberikan wawasan yang berharga tentang persepsi masyarakat terhadap naturalisasi pemain dalam tim nasional Indonesia U-23 di bawah kendali Shin Tae-yong, serta perbandingan efektivitas antara algoritma Naive Bayes dan KNN dalam klasifikasi sentimen. Penelitian ini diharapkan dapat memberikan pandangan yang lebih mendalam tentang dinamika dalam tim sepak bola nasional Indonesia serta kontribusinya terhadap permainan sepak bola Indonesia secara keseluruhan.

Kata Kunci: Data Mining; Sosial Media X; Naturalisasi Tim Nasional; Naïve Bayes; K-Nearest Neighbor.

Abstract

In the Shin Tae-yong era, the naturalization of players has become a controversial topic within the Indonesian U-23 national football team. This research aims to analyze public sentiment related to the naturalization of players in the team using two classification algorithms, namely Naive Bayes and K-Nearest Neighbor. Sentiment data is obtained from news sources, social media, and online discussion forums related to matches and team management decisions. First, data processing is carried out, including text cleaning, tokenization, and word weighting. Next, Naive Bayes and KNN models are trained using the processed dataset. The results of the sentiment analysis will provide valuable insight into the public's perception of the naturalization of players in the Indonesian U-23 national team under the control of Shin Tae-yong, as well as a comparison of the effectiveness between the Naive Bayes and KNN algorithms in sentiment classification. It is hoped that this research will provide a more in-depth view of the dynamics within the Indonesian national football team and its contribution to Indonesian football as a whole.

Keyword: Data Mining; Social Media X; National Team Naturalization; Naïve Bayes; K-Nearest Neighbor.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

1. Pendahuluan

Tim nasional sepak bola Indonesia U-23 telah mengalami sejumlah perubahan signifikan dalam beberapa tahun terakhir, salah satunya adalah penerapan kebijakan naturalisasi pemain asing. Kebijakan ini diterapkan dengan tujuan memperkuat tim agar dapat bersaing di tingkat internasional. Sejak tim berada di bawah kepemimpinan Shin Tae-yong, seorang pelatih asal Korea Selatan yang mulai menjabat pada akhir 2019, kebijakan ini semakin mendapatkan perhatian publik. Shin Tae-yong dikenal karena pendekatan strategis dan disiplin ketat yang diterapkannya dalam membangun tim.

Kebijakan naturalisasi di Timnas Indonesia U-23 memicu berbagai reaksi di masyarakat. Beberapa pihak mendukung langkah ini dengan alasan bahwa pemain naturalisasi dapat meningkatkan kualitas serta memberikan pengalaman yang diperlukan bagi tim nasional. Mereka berargumen bahwa pemain naturalisasi yang telah berkompetisi di liga luar negeri memiliki keunggulan dari sisi kemampuan teknis dan pengalaman bermain di panggung internasional. Namun, di sisi lain, terdapat kekhawatiran bahwa naturalisasi pemain asing akan mengurangi kesempatan bagi pemain lokal untuk berkembang dan tampil di kancah internasional. Selain itu, terdapat perdebatan mengenai bagaimana kebijakan ini dapat memengaruhi identitas nasional dan rasa kebersamaan dalam tim, terutama terkait dengan representasi Indonesia di level global.

Untuk memahami bagaimana pandangan masyarakat terhadap kebijakan ini, analisis sentimen digunakan sebagai pendekatan. Analisis sentimen merupakan salah satu metode dalam ilmu data yang mampu mengidentifikasi serta mengklasifikasikan opini publik yang tersebar melalui berbagai platform digital, seperti media sosial, forum diskusi, dan artikel berita. Menurut penelitian Alfandi Safira dan Hasan (2023), penerapan metode *Naive Bayes* dalam analisis sentimen terbukti efektif dalam memetakan opini masyarakat terhadap layanan *Paylater*, dengan hasil prediksi yang akurat untuk berbagai kategori sentimen (positif, negatif, dan netral). Penerapan metode ini dalam penelitian terkait kebijakan naturalisasi akan membantu dalam memahami persepsi publik secara lebih tepat.

Dalam penelitian ini, dua algoritma pembelajaran mesin, yaitu *Naive Bayes* dan *K-Nearest Neighbor* (*K-NN*), digunakan untuk melakukan analisis sentimen publik terhadap kebijakan naturalisasi. Algoritma *Naive Bayes* merupakan model probabilistik yang mengklasifikasikan teks berdasarkan probabilitas kemunculan kata-kata tertentu dalam kategori sentimen tertentu (Fikri et al., 2020). Algoritma ini dikenal efektif dalam menangani berbagai jenis data teks dan telah banyak digunakan dalam analisis sentimen. Sebagai contoh, penelitian Nurhuda et al. (2016) menunjukkan bahwa *Naive Bayes* sangat efektif dalam mengklasifikasikan opini publik terhadap calon presiden Indonesia pada Pemilu 2014, dengan hasil yang memuaskan dari sisi akurasi. Begitu pula, Mas Pintoko dan Muslim (2018) menemukan bahwa metode ini mampu menganalisis sentimen terhadap jasa transportasi online dengan baik.

Selain *Naive Bayes*, metode *K-Nearest Neighbor* (K-NN) juga digunakan dalam penelitian ini. K-NN merupakan algoritma berbasis jarak yang mengklasifikasikan teks berdasarkan kemiripannya dengan data yang sudah terklasifikasi sebelumnya (Asro'i & Februariyanti, 2022). Walaupun sederhana, metode ini dikenal mampu mengidentifikasi pola dalam data teks dengan baik. Deviyanto dan Wahyudi (2018) dalam penelitian mereka tentang sentimen pengguna Twitter terhadap kebijakan publik di Indonesia menunjukkan bahwa K-NN dapat digunakan dengan hasil yang cukup akurat dalam menganalisis opini publik. Ernawati dan Wati (2018) juga menemukan bahwa K-NN efektif digunakan dalam analisis ulasan pelanggan di sektor agen perjalanan.

Dengan menggunakan dua algoritma ini, penelitian bertujuan untuk memetakan persepsi publik terhadap kebijakan naturalisasi pemain dalam tim nasional Indonesia U-23 selama masa kepemimpinan Shin Tae-yong. Hasil penelitian diharapkan dapat memberikan panduan yang jelas bagi pembuat kebijakan, pelatih, dan pihak terkait lainnya dalam mengambil keputusan terkait pengembangan kebijakan ke depannya. Temuan ini sejalan dengan penelitian Muttaqin dan Kharisudin (2021), yang menunjukkan bahwa penggunaan algoritma *machine learning* dalam analisis sentimen mampu meningkatkan pengambilan keputusan yang lebih baik di berbagai bidang, termasuk olahraga. Penelitian ini juga diharapkan dapat menjadi rujukan penting bagi para pemangku

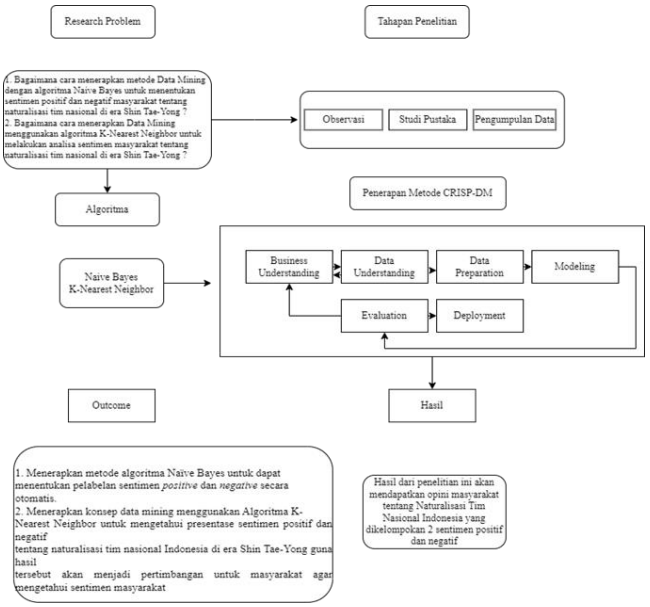
Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

kepentingan dalam dunia sepak bola Indonesia, terutama dalam menghadapi tantangan modernisasi dan globalisasi yang memengaruhi berbagai aspek olahraga. Hasilnya dapat digunakan sebagai bahan pertimbangan dalam merumuskan strategi ke depan yang lebih relevan dengan dinamika publik serta untuk memastikan bahwa kebijakan yang diambil mencerminkan harapan masyarakat luas.

2. Metode Penelitian

2.1 Penerapan Metodologi

Penelitian ini menggunakan dua algoritma pembelajaran mesin, yaitu *Naive Bayes* dan *K-Nearest Neighbors* (KNN), untuk klasifikasi sentimen. Data yang digunakan berasal dari platform Twitter, di mana tweet dengan tagar #NaturalisasiTimNasional diambil untuk dianalisis. Proses penelitian ini mengacu pada enam tahap metode *Cross Industry Standard for Data Mining* (CRISP-DM). Metode CRISP-DM dipilih karena merupakan standar yang sering digunakan dalam pengembangan dan pemecahan masalah berbasis data. Langkah-langkah utama dalam metode ini meliputi pengumpulan data, pembersihan data, pemilihan atribut, dan pemodelan data.



Gambar 1. Tahapan Penerapan Metodologi

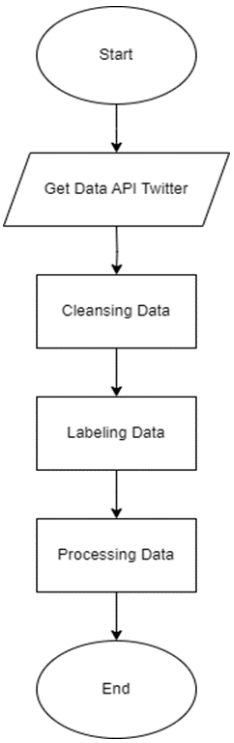
2.1.1 Proses Pengumpulan Data

Data dikumpulkan menggunakan *Application Programming Interface* (API) dari Twitter. Tweet yang dikumpulkan adalah tweet berbahasa Indonesia dengan kata kunci "Naturalisasi Tim Nasional Indonesia" yang diambil dalam jumlah terbatas, yaitu 5.000 tweet. Dari jumlah ini, sebanyak 2.344 tweet berhasil diolah melalui tahapan *text preprocessing*. Tahapan ini bertujuan untuk membersihkan data dari elemen-elemen yang tidak diperlukan, sehingga data siap digunakan dalam analisis sentimen. Proses pengumpulan data dapat dijelaskan sebagai berikut:

1) GET Data API Twitter: Data yang diambil merupakan opini pengguna Twitter terhadap keputusan naturalisasi pemain tim nasional.
2) Pembersihan Data (Cleansing Data): Proses ini melibatkan langkah-langkah pembersihan seperti *tokenisasi*, *case transformation*, *stopword removal*, dan *filtering*.
3) Pelabelan Data (Labeling Data): Setelah data dibersihkan, dilakukan pelabelan secara manual terhadap setiap tweet untuk menentukan sentimen.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

4) Pengolahan Data (Processing Data): Pengujian data dilakukan dengan matriks *confusion* untuk menilai akurasi algoritma KNN yang digunakan.



Gambar 2. Flowchart Pengumpulan Data

2.2 Rancangan Penelitian

Dalam penelitian ini, metode CRISP-DM diadopsi karena struktur tahapan yang komprehensif dalam pemrosesan data. CRISP-DM merupakan pendekatan umum yang digunakan dalam proyek *data mining*, dengan fokus pada proses pengembangan data secara bertahap. Adapun enam tahapan CRISP-DM yang diterapkan dalam penelitian ini adalah sebagai berikut:

2.2.1 Pemahaman Bisnis (*Business Understanding*)

Langkah awal dalam metode CRISP-DM adalah memahami tujuan bisnis dari penelitian, yang dalam hal ini adalah menganalisis sentimen publik terhadap kebijakan naturalisasi pemain dalam Timnas Indonesia. Data yang dikumpulkan akan dianalisis untuk melihat persepsi publik terhadap kebijakan tersebut. Setelah pengumpulan data melalui API Twitter, data diproses menggunakan berbagai fitur, seperti *replace RT*, *replace link*, *replace hashtag*, *replace mention*, dan *remove duplicates*. Proses ini bertujuan untuk menghilangkan elemen yang tidak relevan, sehingga menghasilkan data yang siap dianalisis menggunakan algoritma KNN.

2.2.2 Pemahaman Data (*Data Understanding*)

Setelah pemahaman bisnis, tahap selanjutnya adalah memahami data yang digunakan. Data yang dikumpulkan dari Twitter berjumlah 2.145 tweet dengan dua atribut utama: teks tweet dan sentimen. Pengambilan data dilakukan dari tanggal 2 Juni 2022 hingga 9 Juni 2024. Setelah data diambil, dilakukan proses pembersihan data yang mencakup penghapusan duplikasi, link, serta simbol-simbol yang tidak diperlukan untuk analisis sentimen.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

2.2.3 Persiapan Data (*Data Preparation*)

Persiapan data adalah proses yang mencakup semua langkah yang diperlukan untuk membangun dataset yang siap digunakan untuk pemodelan klasifikasi. Ada tujuh tahap persiapan data, yaitu:

1) Pengumpulan Data: Data diambil dari Twitter dengan menggunakan kata kunci "naturalisasi tim nasional Indonesia".
2) Pembersihan Data: Data dibersihkan dari elemen-elemen yang tidak relevan, seperti karakter khusus, tautan, dan tanda baca.
3) Tokenisasi: Teks dipecah menjadi unit-unit yang lebih kecil (token) untuk memudahkan proses analisis.
4) Stemming dan Lematisasi: Pengurangan kata-kata menjadi bentuk dasarnya untuk mempermudah analisis. Stemming adalah pemangkasan akhiran kata, sedangkan lemmatisasi mengembalikan kata ke bentuk dasarnya.
5) Penghapusan *Noise*: Menghilangkan kata-kata atau simbol yang tidak relevan, seperti angka atau simbol berlebihan.
6) Normalisasi: Menghilangkan tanda baca atau karakter khusus tertentu untuk konsistensi data.
7) Pelabelan Sentimen: Data diberi label sentimen, baik secara manual maupun otomatis menggunakan kamus sentimen.

2.2.4 Pemodelan (*Modeling*)

Pada tahap ini, algoritma *Naive Bayes* dan KNN digunakan untuk klasifikasi sentimen. Kedua algoritma ini dipilih karena kemampuannya dalam menangani data teks yang besar. Tweet yang telah diproses akan dipisahkan menjadi dua atribut utama, yaitu teks tweet dan sentimen. Pemodelan dilakukan menggunakan perangkat lunak Rapidminer, yang digunakan untuk melakukan analisis dan evaluasi klasifikasi.

2.2.5 Evaluasi (*Evaluation*)

Tahap evaluasi bertujuan untuk mengukur akurasi dan efektivitas model yang telah dibangun. Evaluasi dilakukan menggunakan matriks *confusion* untuk menilai seberapa baik model KNN dan *Naive Bayes* dalam mengklasifikasikan data. Evaluasi juga menentukan apakah model dapat digunakan atau perlu diulang dari awal jika hasilnya tidak sesuai dengan rencana awal.

2.2.6 Penyebaran (*Deployment*)

Tahap terakhir dalam metode CRISP-DM adalah penyebaran hasil penelitian. Hasil yang diperoleh dari pemodelan dan evaluasi akan disusun menjadi laporan yang menjelaskan pengetahuan yang diperoleh dari proses analisis data. Laporan ini diharapkan dapat memberikan panduan bagi pengambil keputusan dalam merumuskan kebijakan atau strategi berdasarkan hasil analisis sentimen yang dilakukan.

3. Hasil dan Pembahasan

3.1 Hasil

3.1.1 Pemahaman Bisnis (*Business Understanding*)

Perkembangan *artificial intelligence* yang cepat banyak mempengaruhi berbagai aspek kehidupan masyarakat seperti seperti dibidang industri, sosial, culture, dan sebagainya. Walaupun *artificial intelligence* dinilian sangat berpengaruh positif bagi masyarakat namun tidak sedikit masyarakat yang menilai negatif dari *artificial intelligence*. Dengan demikian maka akan dilakukan sentimen terhadap *artificial intelligence*.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

3.1.2 Pemahaman Data (Data *Understanding*)

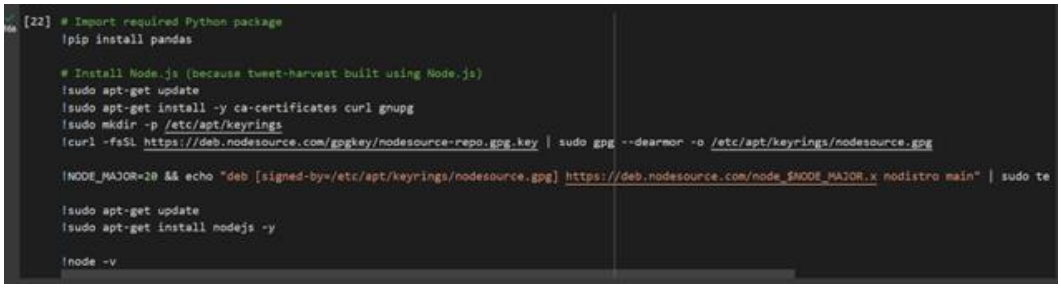
Pada tahapan ini maka peneliti mencoba untuk memahami data yang akan digunakan Data dari twitter (X) yang telah di ambil menggunakan pemprograman python berjumlah 5000 tweet dengan dua atribut yang akan dilakukan dengan cara pemilihan atribut menjadi dua atribut yaitu tweet dan sentimen. Data yang didapatkan dalam penelitian ini diperoleh tanggal 1 Januari 2022 sampai 30 Juni 2024.

3.1.3 Data Preparation

Pada tahap data preparation dilakukan beberapa tahapan antara lain sebagai berikut :

1) Pengumpulan Data

Untuk pengambilan data menggunakan pemprograman python dengan melalui google colab. Untuk tahapannya antara lain yaitu diawali instalasi package python, mengambil token dari twitter (X), lalu crawling data berdasarkan text, dan periode tanggal, dan yang terakhir data di save. Untuk instalasi package python bisa dilihat gambar dibawah ini.



Gambar 3. Instalasi Package Phyton

Setelah melalui package python dapat terinstall dengan baik, selanjutnya mengambil token dari akun twitter (X), lalu setelahnya memasukan token seperti gambar dibawah ini.

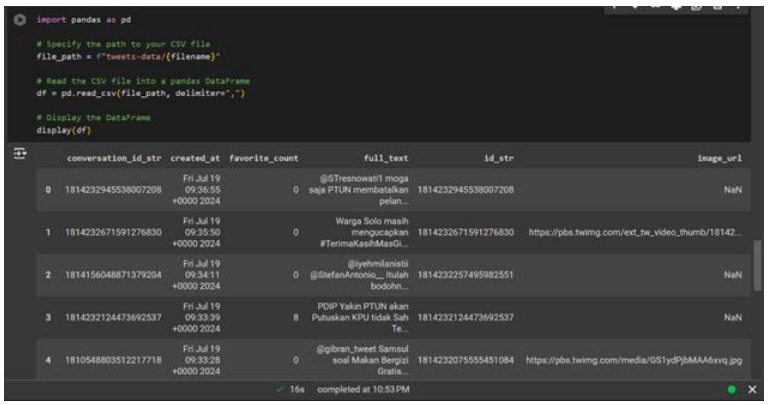
Gambar 4. Memasukan Token

Setelah memasukan token dan dapat terautentifikasi dengan baik maka selanjutnya tinggal memasukan keywoard *artificial intelligence*, limit yang dicari yaitu 5000 data dengan bentuk text, dan periode yang dicari yaitu dari 1 januari 2022 sampai 30 juni 2024.

Gambar 5. Proses Crawling Data

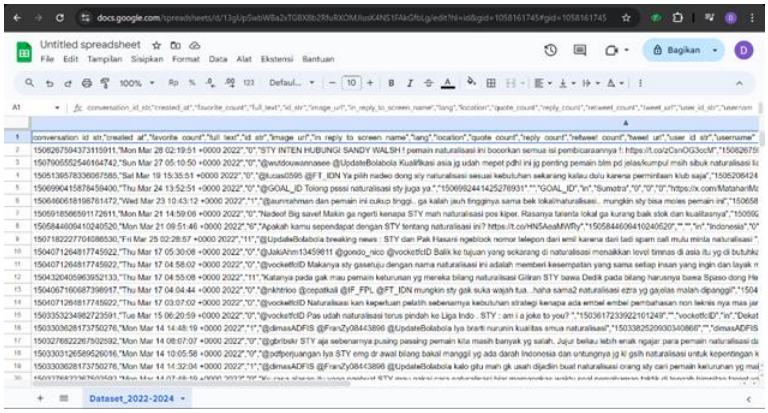
Setelah proses crawling data, file yang sudah siap tinggal di save dan di export.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E - ISSN: 2723-7079 | P - ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024



Gambar 6. Proses Export Data

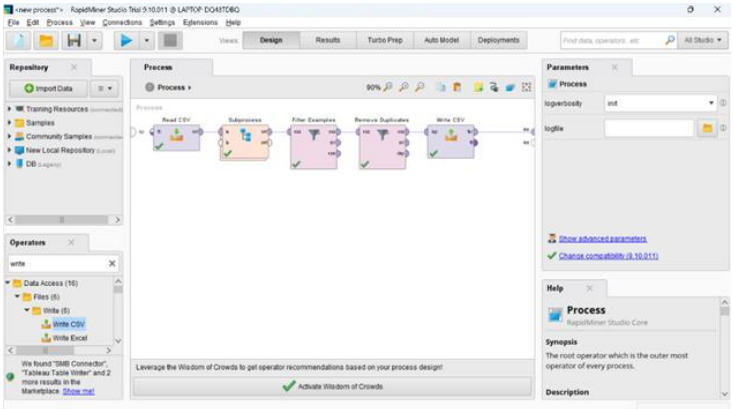
Setelah dapat di export data bisa didapatkan dengan jumlah 5000 data.



Gambar 7. Jumlah Data Yang Didapat

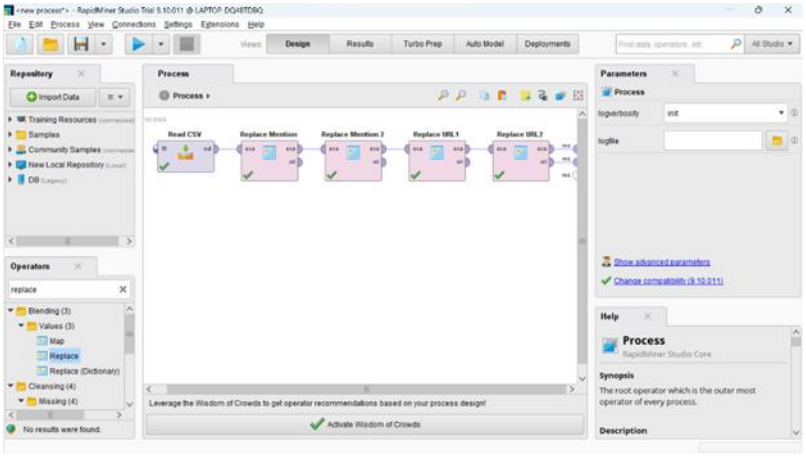
2) *Cleansing* Data

Pada tahapan ini peneliti akan melakukan proses Cleansing dengan menggunakan rapidminer sebelum melakukan pemberian label pada sentimen, agar dataset pada file csv terbaru terhindar dari duplikasi data dan tanda yang tidak diperlukan. Pada proses ini akan dilakukan pembersihan dari berbagai noise seperti menghilangkan link URL, username, retweet, digit angka dan karakter. Setelah selesai menjalankan proses Cleansing dengan menggunakan rapidminer. Awalnya terdapat 5000 data tweet pada file csv, kemudian setelah melewati proses Cleansing maka data tweet berkurang menjadi 2344 data.

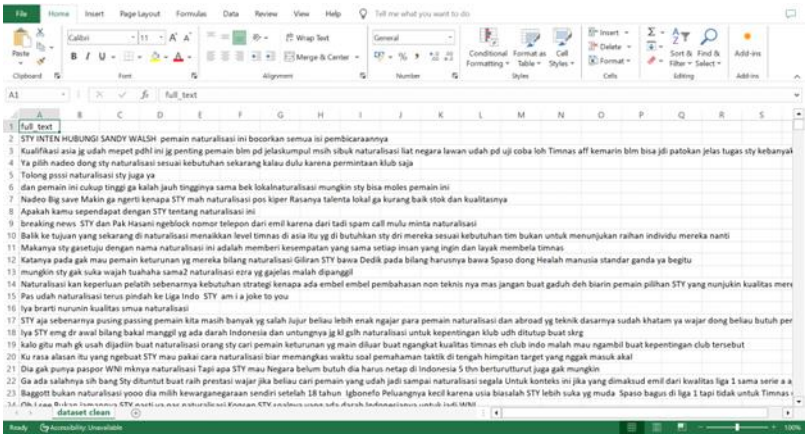


Gambar 8 Model Proses *Cleansing*

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024



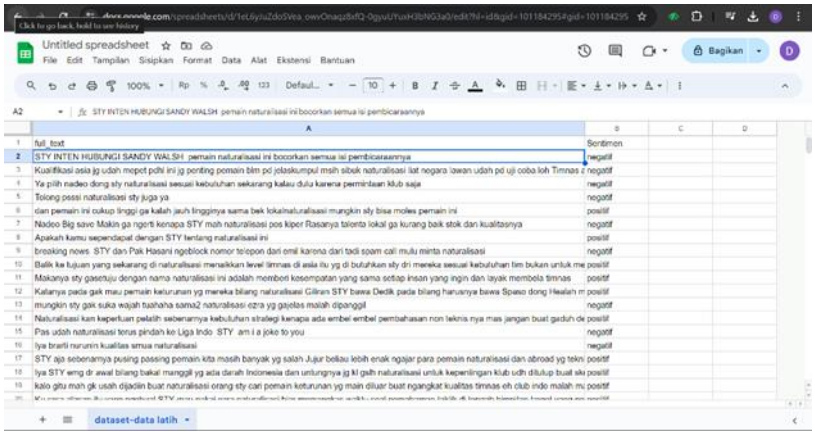
Gambar 9. Model Subproses *Cleansing*



Gambar 10. Hasil *Cleansing* Data

3) *Labeling*

Kemudian tahap selanjutnya adalah melakukan klasifikasi atau penentuan sentimen berdasarkan masing-masing tweet. Selama proses penentuan sentimen berlangsung, disini peneliti melakukan pemberian sentimen secara manual. Berikut ini data yang sudah di beri label.

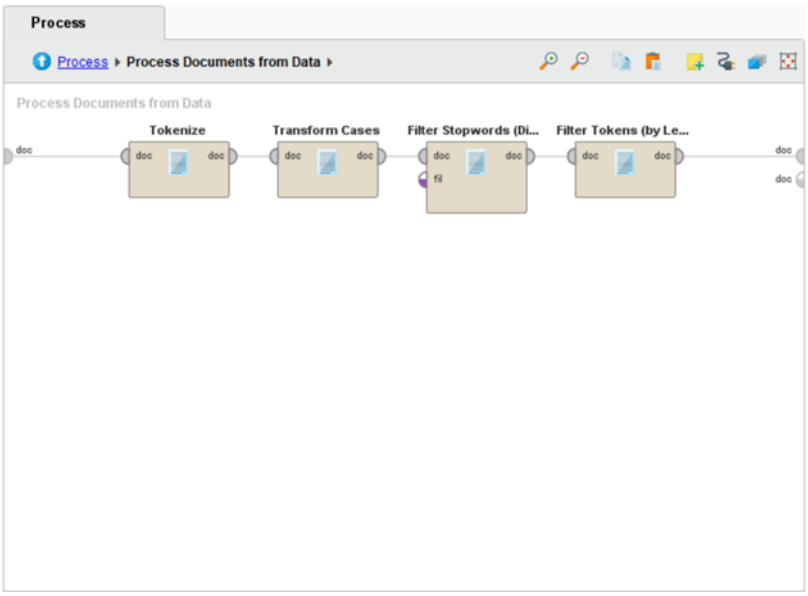


Gambar 11. Data Yang Sudah Dilabeli

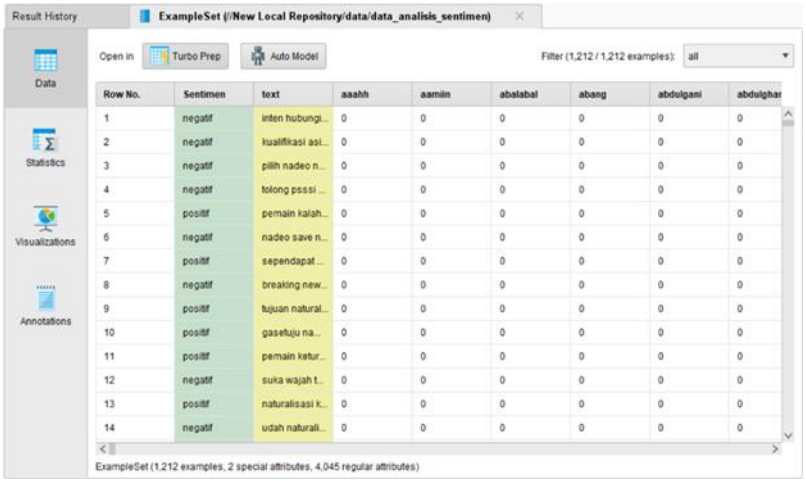
Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

4) *Tokenizing*

Pada proses ini adalah untuk mengubah teks dari kalimat-kalimat yang kompleks menjadi urutan token atau unit-unit yang lebih sederhana. Tujuannya untuk mempersiapkan data teks agar dapat diproses oleh algoritma analisis sentimen dengan lebih efektif dan akurat.



Gambar 12. Model Tokenizing

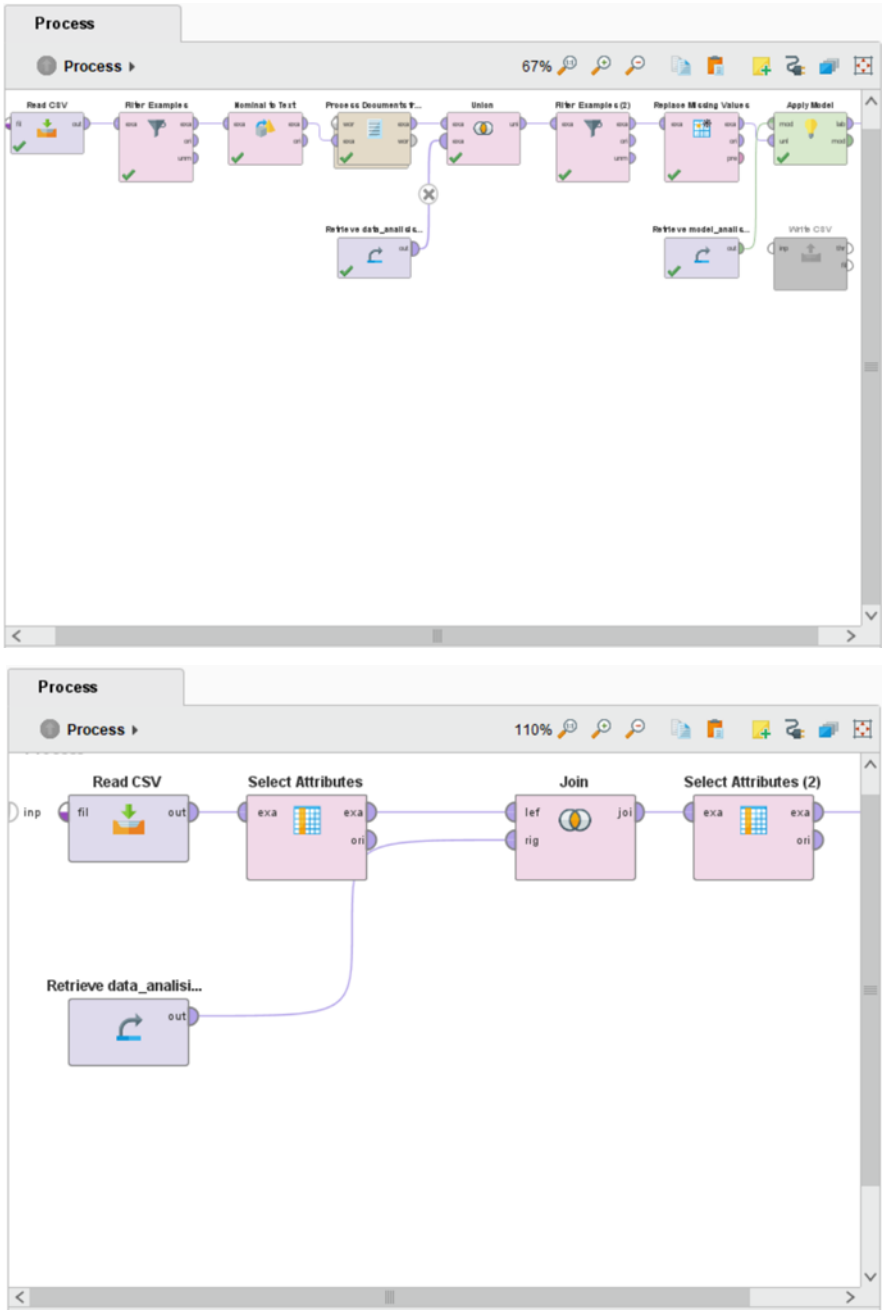


Gambar 13. Hasil Tokenizing Naïve Bayes

3.1.4 *Modeling*

Pada tahapan ini peneliti akan melakukan prepocessing pada dataset hal ini ditunjukkan untuk menyiapkan data yang bersih dan data yang bebas dari noise. Pada proses ini juga untuk menghitung pembobotan kata untuk keperluan pada proses modelling nanti. Setelah melewati tahapan text preparation, maka jumlah data tweet didapatkan menjadi 2344 data yang merupakan data bersih untuk dipakai pada tahap selanjutnya. Pada tahapan ini akan dilakukan pengukuran performa dua klasifikasi sekaligus dengan menggunakan algoritma Naïve bayes dan K-Nearest Neighbor

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E - ISSN: 2723-7079 | P - ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024



Gambar 14. Model Proses Data

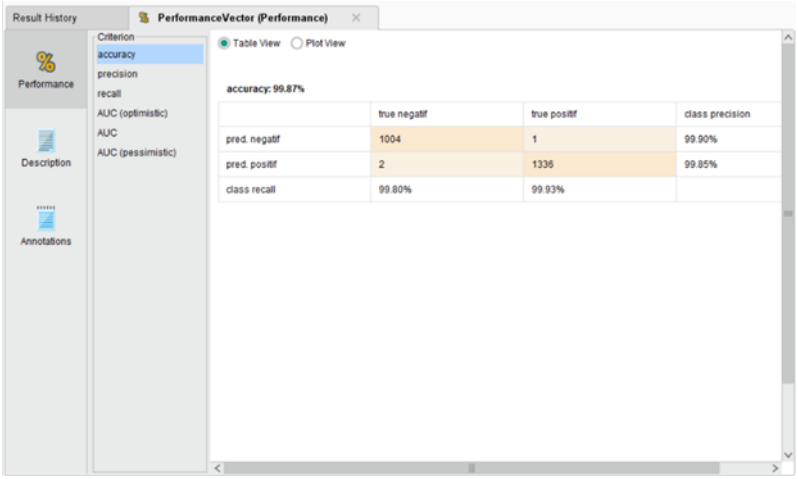
3.1.5 Evaluasi

Pada penelitian ini akan dilakukan pengujian menggunakan algoritma K-Nearest Neighbor dengan hasil sebagai berikut:

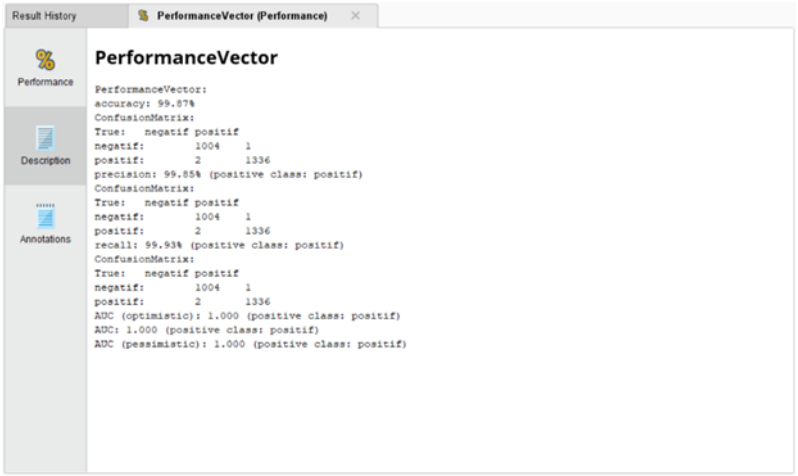
Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

1) Algoritma K-Nearest Neighbor

Total dataset yang dikumpulkan adalah 2344 data. Berikut ini adalah hasil dari tahapan modelling dengan menggunakan algoritma K-Nearest Neighbor yang dapat dilihat hasil dari perhitungan hasil RapidMiner dapat dilihat pada Gambar dibawah ini untuk menunjukkan dan membuktikan hasil prediksi.

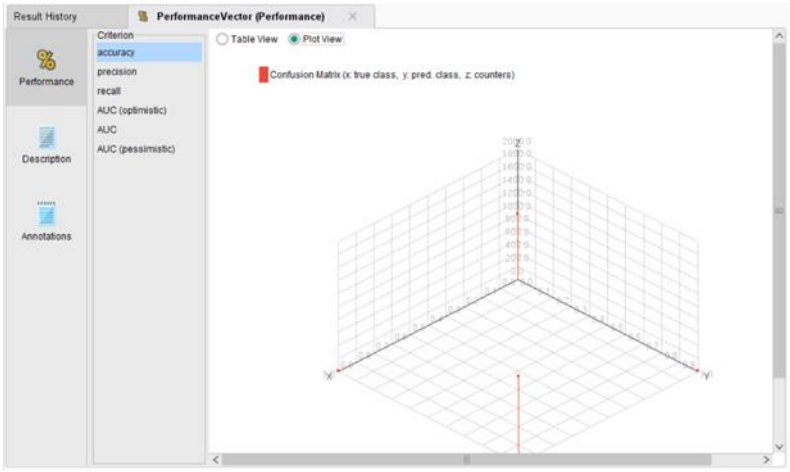


Gambar 15. Hasil Akurasi K-Nearest Neighbor

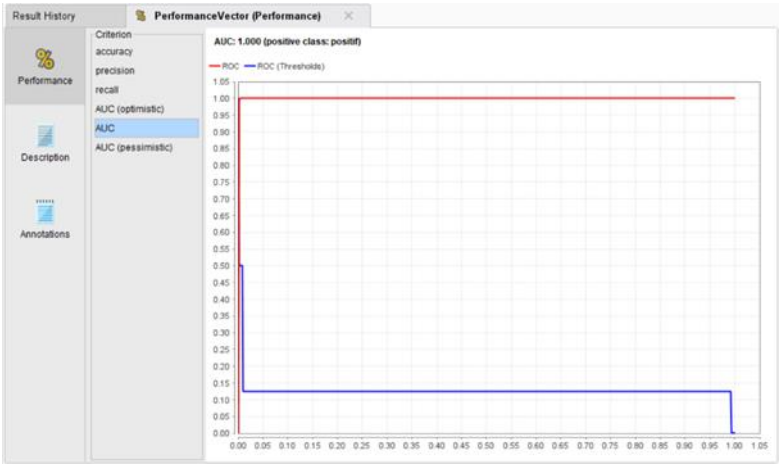


Gambar 16. Deskripsi Hasil Akurasi K-Nearest Neighbor

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024



Gambar 17. Plot View K-Nearest Neighbor



Gambar 18. Grafik AUC K-Nearest Neighbor

Berdasarkan hasil diatas dapat disimpulkan bahwa *accuracy* menunjukkan data yang benar diprediksi positif yang menghasilkan presentase ketepatannya adalah 99.85%, sedangkan untuk data bersentimen Negatif memiliki *precision* sebesar 99.90%. Untuk sentimen Positif memiliki recall (*Specificity*) adalah 99.93% , sedangkan pada sentimen negatif memiliki recall sebesar 99.80%. Nilai *accuracy* yang dihasilkan menggunakan model *K-Nearest Neighbor* adalah 99.87%, sehingga dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* dapat mengklasifikasi sentiment dengan baik menggunakan data *Artificial intelligence*.

3.1.6 Deployment

Dari hasil evaluasi dan pengujian diatas, untuk sentimen analisis dapat digunakan K-Nearest Neighbor sebagai algoritma pengkasifikasian.

1) Berdasarkan penelitian yang telah dilakukan mengenai klasifikasi opini masyarakat Indonesia terhadap *Artificial intelligence* dapat diambil kesimpulan yaitu hasil sentimen masyarakat terhadap *Artificial intelligence* dari 2489 data ditemukan 1202 atau 48% data bersentimen positif dan 1287 data atau 52% bersentimen negatif.
2) Dari hasil penelitian terdapa masyarakat Indonesia sebagian besar memiliki respon negatif terhadap *Artificial intelligence* dan berdasarkan klasifikasi model algoritma C4.5 terhadap dataset *Artificial intelligence*, didapatkan nilai accuracy sebesar 62.38%%. Sedangkan jika menggunakan algoritma Naïve Bayes mendapatkan nilai accuracy 97.47%.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

3) Sehingga dapat dikatakan bahwa algorima Naïve Bayes lebih baik dibandingkan algotitma C4.5 dalam mengklasifikasikan data secara baik dan benar.

3.2 Pembahasan

Penelitian ini menggunakan dua algoritma pembelajaran mesin, yaitu *Naive Bayes* dan *K-Nearest Neighbor* (KNN), untuk menganalisis sentimen masyarakat terhadap kebijakan naturalisasi pemain Tim Nasional Indonesia U-23. Hasilnya menunjukkan bahwa kedua algoritma tersebut efektif dalam mengklasifikasikan sentimen publik yang diambil dari platform Twitter, dengan tingkat akurasi yang tinggi untuk keduanya. Pada bagian ini, kita akan membahas lebih lanjut mengenai perbandingan kinerja kedua algoritma, persepsi publik terhadap kebijakan naturalisasi, efektivitas metodologi CRISP-DM, dan peluang untuk penelitian lebih lanjut, serta memperkaya pembahasan dengan referensi dari beberapa penelitian sebelumnya.

Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* dan *K-Nearest Neighbor* sama-sama mampu mengklasifikasikan sentimen dengan akurasi yang sangat tinggi. *Naive Bayes* mencapai akurasi sebesar 97,47%, sedangkan KNN mencapai akurasi sebesar 99,87%. Dalam penelitian ini, *Naive Bayes* berhasil menangani data teks yang bervariasi dan besar dengan menggunakan pendekatan probabilistik. Hal ini sejalan dengan penelitian yang dilakukan oleh Setiawan dan Zufria (2023), di mana *Naive Bayes* digunakan untuk mengklasifikasikan sentimen publik terhadap pembatalan Indonesia sebagai tuan rumah Piala Dunia U-20 dan memberikan hasil yang sangat akurat. Di sisi lain, KNN menggunakan pendekatan berbasis jarak yang mengklasifikasikan data berdasarkan kemiripan antara data baru dan data yang telah diklasifikasikan sebelumnya. KNN memberikan hasil yang sangat baik dalam penelitian ini, sebagaimana terlihat pada penelitian yang dilakukan oleh Pratama et al. (2022) yang juga menggunakan KNN untuk menganalisis sentimen publik terhadap Tim Nasional Indonesia selama Piala AFF 2020, dengan hasil akurasi yang tinggi. Selain itu, penelitian oleh Supriyanto et al. (2023) menunjukkan bahwa KNN juga efektif digunakan dalam analisis sentimen publik terkait pembelajaran daring, mendukung efektivitas algoritma ini dalam berbagai konteks penelitian sentimen.

Analisis sentimen terhadap kebijakan naturalisasi pemain dalam Tim Nasional Indonesia U-23 mengungkapkan bahwa sentimen publik terbagi menjadi dua bagian, dengan 48% menunjukkan sentimen positif dan 52% menunjukkan sentimen negatif. Hal ini mencerminkan adanya polarisasi pendapat di kalangan masyarakat terkait kebijakan tersebut. Publik yang mendukung kebijakan ini berpendapat bahwa naturalisasi dapat meningkatkan kualitas tim, terutama dengan adanya pemain yang memiliki pengalaman internasional. Pandangan ini didukung oleh penelitian Pratama et al. (2022) yang menunjukkan bahwa masyarakat cenderung mendukung langkah-langkah yang bertujuan untuk meningkatkan daya saing tim nasional di kancah internasional. Namun, sentimen negatif yang lebih dominan dalam penelitian ini mencerminkan kekhawatiran publik tentang dampak kebijakan tersebut terhadap pemain lokal. Publik merasa bahwa kebijakan naturalisasi dapat mengurangi kesempatan bagi pemain lokal untuk berkembang dan berpartisipasi dalam tim nasional. Selain itu, ada juga kekhawatiran mengenai identitas nasional yang mungkin terancam dengan adanya dominasi pemain asing dalam tim nasional. Penelitian oleh Sembada dan Prasetyo (2020) juga mencatat bahwa dalam konteks sepak bola Indonesia, ada ketegangan antara penerapan kebijakan internasional dan pentingnya menjaga nilai-nilai nasional, termasuk Pancasila sebagai landasan negara.

Metode *Cross Industry Standard for Data Mining* (CRISP-DM) yang digunakan dalam penelitian ini terbukti sangat efektif dalam memandu setiap tahapan pengolahan data, mulai dari pemahaman bisnis hingga penyebaran hasil. Salah satu keunggulan metode ini adalah kemampuannya dalam menangani berbagai tahap pengolahan data dengan struktur yang jelas, seperti pengumpulan data, pembersihan data, dan pemodelan. Dalam penelitian ini, proses pembersihan data berhasil mengurangi jumlah data dari 5.000 menjadi 2.344 tweet yang relevan untuk dianalisis, memastikan bahwa data yang digunakan benar-benar bersih dari elemen yang tidak relevan. Keberhasilan metode CRISP-DM dalam penelitian ini serupa dengan temuan Prabowo dan Wiguna (2021), di mana metode CRISP-DM membantu pengembangan sistem informasi berbasis web untuk UMKM bengkel. Metode ini memudahkan

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

dalam merancang sistem pengolahan data yang terstruktur dan sesuai dengan kebutuhan bisnis. Dalam konteks analisis sentimen, CRISP-DM memfasilitasi alur yang efisien dari awal hingga akhir, yang juga ditunjukkan oleh penelitian Salam et al. (2018) yang menggunakan metode ini untuk analisis sentimen data komentar media sosial pada layanan jasa ekspedisi.

Penelitian ini hanya berfokus pada data Twitter, padahal platform lain seperti Facebook dan Instagram mungkin memiliki dinamika yang berbeda dalam hal ekspresi sentimen publik. Menurut Salim dan Mayary (2020), penggunaan data dari berbagai platform sosial media dapat memperkaya hasil penelitian dan memberikan perspektif yang lebih luas mengenai opini publik. Selain itu, penelitian ini hanya menggunakan dua algoritma klasifikasi, yaitu *Naive Bayes* dan KNN. Algoritma lain seperti *Support Vector Machine* (SVM) dan *Random Forest* juga dapat dieksplorasi untuk melihat perbandingan akurasi dan efektivitas dalam analisis sentimen. Sari dan Suryono (2024) dalam penelitian mereka menunjukkan bahwa *Random Forest* dan SVM mampu memberikan hasil yang lebih baik dalam analisis sentimen terkait topik teknologi, seperti metaverse. Oleh karena itu, penelitian di masa depan dapat mempertimbangkan penggunaan algoritma tambahan untuk meningkatkan hasil.

4. Kesimpulan

Berdasarkan penelitian yang menggunakan algoritma C4.5 dan *Naive Bayes* dalam analisis sentimen terhadap *Artificial Intelligence* (AI), dapat disimpulkan bahwa dari 2.489 data yang dianalisis, 1.202 data (48%) menunjukkan sentimen positif terhadap AI, sementara 1.287 data (52%) menunjukkan sentimen negatif. Mayoritas masyarakat Indonesia tampaknya memiliki pandangan yang kurang mendukung terhadap perkembangan AI. Dari segi akurasi, model klasifikasi menggunakan algoritma C4.5 menghasilkan nilai akurasi sebesar 62,38%, sedangkan algoritma *Naive Bayes* mencapai akurasi yang jauh lebih tinggi, yaitu 97,47%. Hal ini menunjukkan bahwa *Naive Bayes* lebih efektif dalam mengklasifikasikan data sentimen dengan akurasi yang lebih baik dibandingkan dengan C4.5, terutama dalam menangani data teks dan mengidentifikasi sentimen publik terhadap AI.

5. Ucapan Terima Kasih

Terimakasih kepada Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika serta semua pihak yang memebantu dan mendukung semua penelitian ini. Penulis sadar bahwa penelitian ini masih banyak kesalahan dan kekurangan. Kekurangan tersebut tentunya dapat dijadikan peluang untuk peningkatan penelitian selanjutnya. Akhir kata penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dan penulis berharap semoga penelitian ini dapat bermanfaat bagi sesama.

6. Daftar Pustaka

Alfandi, S., & Hasan, F. N. (2023). Analisis sentimen masyarakat terhadap *Paylater* menggunakan metode *Naive Bayes* classifier. *ZONAsi: Jurnal Sistem Informasi, 5*(1), 59–70. https://doi.org/10.31849/zn.v5i1.12856

Asro'i, A., & Februariyanti, H. (2022). Analisis sentimen pengguna Twitter terhadap perpanjangan PPKM menggunakan metode *K-Nearest Neighbor*. *Jurnal Khatulistiwa Informatika, 10*(1), 17–24. https://doi.org/10.31294/jki.v10i1.12624

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

Deviyanto, A., & Wahyudi, M. D. R. (2018). Penerapan analisis sentimen pada pengguna Twitter menggunakan metode *K-Nearest Neighbor*. *JISKA (Jurnal Informatika Sunan Kalijaga), 3*(1), 1–10. https://doi.org/10.14421/jiska.2018.31-01

Ernawati, S., & Wati, R. (2018). Penerapan algoritma *K-Nearest Neighbors* pada analisis sentimen review agen travel. *Jurnal Khatulistiwa Informatika, 6*(1), 64–69.

Fauzi, M. A., Mentari, N. D., & Muflikhah, L. (2018). Analisis sentimen kurikulum 2013 pada sosial media Twitter menggunakan metode *K-Nearest Neighbor* dan *feature selection*. *Query Expansion Ranking Optimasi Sisa Bahan Baku Pada Industri Mebel Menggunakan Algoritma Genetika View Project Human Detection and Tra*, 2(8), 2739–2743. https://www.researchgate.net/publication/322959504

Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan metode *Naïve Bayes* dan *Support Vector Machine* pada analisis sentimen Twitter. *Smatika Jurnal, 10*(02), 71–76. https://doi.org/10.32664/smatika.v10i02.455

Mas Pintoko, B., & Muslim, K. (2018). Analisis sentimen jasa transportasi online pada Twitter menggunakan metode *Naïve Bayes* classifier. *E-Proceeding of Engineering, 5*(3), 8121–8130.

Muttaqin, M. N., & Kharisudin, I. (2021). Analisis sentimen pada ulasan aplikasi Gojek menggunakan metode *Support Vector Machine* dan *K-Nearest Neighbor*. *UNNES Journal of Mathematics, 10*(2), 22–27. http://journal.unnes.ac.id/sju/index.php/ujm

Nugroho, D. G., Chrisnanto, Y. H., & Wahana, A. (2015). Analisis sentimen pada jasa ojek online. *Jurnal*, 156–161.

Nurhuda, F., Widya Sihwi, S., & Doewes, A. (2016). Analisis sentimen masyarakat terhadap calon presiden Indonesia 2014 berdasarkan opini dari Twitter menggunakan metode *Naive Bayes* classifier. *Jurnal Teknologi & Informasi ITSmart, 2*(2), 35. https://doi.org/10.20961/its.v2i2.630

Prabowo, W. A., & Wiguna, C. (2021). Sistem informasi UMKM bengkel berbasis web menggunakan metode SCRUM. *Jurnal Media Informatika Budidarma, 5*(1), 149. https://doi.org/10.30865/mib.v5i1.2604

Pratama, A. E., Ariesta, A., & Gata, G. (2022). Analisis sentimen masyarakat terhadap Tim Nasional Indonesia pada Piala AFF 2020 menggunakan algoritma *K-Nearest Neighbors*. *Jurnal TICOM: Technology of Information and Communication, 10*(3), 187–196. https://jurnal-ticom.jakarta.aptikom.org/index.php/Ticom/article/view/33/47

Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Analisis sentimen data komentar sosial media Facebook dengan *K-Nearest Neighbor* (Studi kasus pada akun jasa ekspedisi barang J&T Express Indonesia). *Prosiding SINTAK*, 480–486.

Salim, S. S., & Mayary, J. (2020). Analisis sentimen pengguna Twitter terhadap dompet elektronik dengan metode *Lexicon Based* dan *K-Nearest Neighbor*. *Jurnal Ilmiah Informatika Komputer, 25*(1), 1–17. https://doi.org/10.35760/ik.2020.v25i1.2411

Sari, P. K., & Suryono, R. R. (2024). Komparasi algoritma *Support Vector Machine* dan *Random Forest* untuk analisis sentimen metaverse. *Jurnal Mnemonic, 7*(1), 31–39. https://doi.org/10.36040/mnemonic.v7i1.8977

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1024

Sari, R. (2020). Analisis sentimen pada review objek wisata Dunia Fantasi menggunakan algoritma *K-Nearest Neighbor* (*K-NN*). *EVOLUSI: Jurnal Sains dan Manajemen, 8*(1), 10–17. https://doi.org/10.31294/evolusi.v8i1.7371

Sembada, A. D., & Prasetyo, D. (2020). Aktualisasi Pancasila dalam sepak bola Indonesia. *CIVICUS: Pendidikan-Penelitian-Pengabdian Pendidikan Pancasila dan Kewarganegaraan, 8*(2), 1. https://doi.org/10.31764/civicus.v8i2.2410

Setiawan, H., & Zufria, I. (2023). Analisis sentimen pembatalan Indonesia sebagai tuan rumah Piala Dunia FIFA U-20 menggunakan *Naïve Bayes*. *Jurnal Media Informatika Budidarma, 7*(3), 1003–1012. https://doi.org/10.30865/mib.v7i3.6144

Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan sistem analisis sentimen komentar pelanggan menggunakan metode *Naive Bayes* classifier. *Jurnal Sistem Informasi (JSI), 8*(1), 2355–4614. http://ejournal.unsri.ac.id/index.php/jsi/index

Supriyanto, J., Alita, D., & Isnain, A. R. (2023). Penerapan algoritma *K-Nearest Neighbor* (*K-NN*) untuk analisis sentimen publik terhadap pembelajaran daring. *Jurnal Informatika dan Rekayasa Perangkat Lunak, 4*(1), 74–80. https://doi.org/10.33365/jatika.v4i1.2468.