OPEN ACCESS

Volume 9 (1), January-March 2025, 149-158

E-ISSN:2580-1643

Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)

DOI: https://doi.org/10.35870/jtik.v9i1.3020

Analisis Sentimen Kepuasan Publik Terhadap Masa Kepemimpinan Shin Tae Yong Menggunakan Algoritma *Naïve Bayes*

Pramudya Nugraha [1*], Rasiban [2], Frencis Matheos Sarimole [3], Tundo [4]

[1,2,3,4] Program Studi Sistem Informasi, Fakultas Teknologi Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.*

a r t i c l e i n f o

Article history:
Received 31 July 2024
Received in revised form
15 September 2024
Accepted 25 October 2024
Available online January
2025.

Keywords:
Data Mining; Shin Tae Yong;
Naïve Bayes Algorithm;
Football.

a b s t r a c t

Shin Tae Yong is the coach of the Indonesian national team who has been a football player in South Korea and has coached the South Korean national team at the 2018 World Cup in Russia. Many people watch or pay attention to Shin Tae Yong's behavior and behavior when coaching the Indonesian national team. Shin Tae Yong has considerable worry with the Indonesian national team because of his strategy. However, there are several media that frame Shin Tae Yong's news differently so that differences in viewpoints and opinions on Shin Tae Yong are controversial, inviting many people to give their opinions. Therefore, people choose social media as a place to channel opinions. In this study, we will take tweets from X with search keywords for Shin Tae Yong and the Indonesian national team to process and classify the text using the sentiment analysis method. The text classification process is divided into two classes, namely positive sentiment classes and negative sentiment classes. The data used amounted to 2495 data that had been cleansed, which amounted to 2.348 Positive sentiment data and 147 data with negative sentiments so that they can be presented 98.94% positive and 60.00% negative, based on the classification of the Naïve Bayes algorithm model, using a split comparative data 0.8 : 0.2 With the value of k=3 for Shin Tae Yong's dataset, an accuracy value of 96.67%.

Kata Kunci:
Data Mining; Shin Tae Yong;
Algoritma Naïve Bayes;
Football.

a b s t r a k

Shin Tae Yong adalah pelatih timnas Indonesia Banyak orang menonton atau memperhatikan tingkah laku dan kelakuan Shin Tae Yong saat melatih timnas Indonesia. Shin Tae Yong memiliki kekhawatiran yang cukup besar dengan tim nasional Indonesia karena strateginya. Namun, ada beberapa media yang membingkai berita Shin Tae Yong secara berbeda sehingga perbedaan sudut pandang dan pendapat tentang Shin Tae Yong menjadi kontroversial, mengundang banyak orang untuk memberikan pendapatnya. Oleh karena itu, masyarakat memilih media sosial sebagai tempat untuk menyalurkan opini. Dalam studi ini, kami akan mengambil tweet dari X dengan kata kunci pencarian untuk Shin Tae Yong dan tim nasional Indonesia untuk memproses dan mengklasifikasikan teks menggunakan metode analisis sentimen. Proses klasifikasi teks dibagi menjadi dua kelas, yaitu kelas sentimen positif dan kelas sentimen negatif. Data yang digunakan berjumlah 2.495 data yang telah dibersihkan, yang berjumlah2.348 Data sentimen positif dan 147 data dengan sentimen negatif sehingga dapat disajikan 98.94% positif dan 60.00% negatif, berdasarkan klasifikasi model algoritma Naïve Bayes, menggunakan data perbandingan terpisah 0.8 : 0.2 Dengan nilai k=3 untuk himpunan data Shin Tae Yong, nilai akurasi 96.67%.

ACM Computing Classification System (CCS)
EBSCOhost
Communication and Mass Media Complete (CMMC)

Corresponding Author. Email: prmdnugraha@gmail.com [1].

1. Pendahuluan

Naïve Bayes Classifier memiliki kelebihan antara lain sederhana, cepat, dan berakurasi tinggi. Metode *Naïve Bayes Classifier* untuk klasifikasi teks menggunakan atribut kata yang muncul dalam suatu dokumen sebagai dasar klasifikasinya. Penelitian Rish (2001) menunjukkan bahwa meskipun asumsi independensi antar kata dalam dokumen tidak sepenuhnya dapat dipenuhi, kinerja *Naïve Bayes Classifier* dalam klasifikasi relatif sangat baik. Pada penelitian ini, terdapat beberapa identifikasi masalah, antara lain belum ada pengklasifikasian sentimen positif dan sentimen negatif pada Shin Tae Yong dan belum diketahui nilai akurasi klasifikasi algoritma *Naïve Bayes Classifier* terhadap data uji dalam pembentukan *sentiment analysis*.

Berdasarkan eksperimen dan pengujian dari banyaknya data *tweet* pengguna Tokopedia, diperoleh hasil bahwa nilai akurasi tertinggi didapatkan pada data ke-600 yang terdiri dari 300 data positif dan 300 data negatif. Berdasarkan eksperimen dan pengujian menggunakan parameter, diperoleh hasil bahwa nilai akurasi tertinggi ada pada pengujian dengan parameter 5, yaitu akurasi 81,50% serta nilai *AUC* sebesar 0,640. Implementasi *Naïve Bayes* mampu melakukan analisis sentimen opini masyarakat terhadap isu *New Normal* di Indonesia. Hasil pengujian terhadap rasio data *training* dan *testing* menunjukkan rasio terbaik sebesar 70% dan 30%, dengan nilai akurasi, presisi, dan *recall* secara berturut-turut sebesar 94,55%, 93,55%, dan 93,55%. Model analisis sentimen yang dirancang diharapkan dapat membantu pemerintah dalam memperoleh umpan balik dari opini masyarakat terhadap kebijakan terkait Covid-19 yang dikeluarkan.

Hasil perbandingan dengan *SVM* menunjukkan bahwa *Naïve Bayes* menghasilkan nilai pengujian yang lebih baik untuk setiap perbandingan rasio. Hal ini disebabkan karena setiap kata merupakan variabel independen yang memiliki nilai probabilitas terhadap kelas tertentu. Berdasarkan uraian di atas dan beberapa penelitian sebelumnya, dapat disimpulkan bahwa metode yang sangat efektif dalam menghasilkan akurasi terbaik dalam klasifikasi sentimen adalah algoritma *Naïve Bayes*. Tujuan penelitian ini adalah untuk mengklasifikasi sentimen dan mengetahui tingkat nilai akurasi sentimen positif dan negatif terhadap komentar publik tentang Shin Tae Yong pada media sosial *X*, dengan menggunakan data sebanyak 2.000 data. Hasil penelitian ini dapat digunakan untuk melihat sentimen yang lebih dominan dari klasifikasi sentimen tentang kepuasan publik terhadap pelatih Shin Tae Yong pada media sosial *X*.

2. Metodologi Penelitian

Data penelitian merujuk pada informasi yang dikumpulkan dan digunakan dalam konteks penelitian ilmiah. Data penelitian dapat berupa fakta, angka, statistik, pernyataan, observasi, atau informasi lain yang relevan dengan topik yang diteliti. Data ini dapat diperoleh melalui berbagai metode, seperti survei, eksperimen, wawancara, observasi, analisis literatur, atau pengumpulan data sekunder. Data penelitian memainkan peran penting dalam proses penelitian ilmiah. Data yang dikumpulkan harus akurat, terpercaya, dan relevan dengan pertanyaan penelitian yang diajukan. Selain itu, data penelitian harus diorganisir, dianalisis, dan diinterpretasikan untuk menghasilkan informasi dan pengetahuan yang dapat menjawab pertanyaan penelitian, menguji hipotesis, serta mengungkap pola, tren, atau hubungan dalam fenomena yang diteliti. Pada penelitian ini, data yang digunakan adalah data publik. Data yang diperoleh berjumlah 5.000 *record* dengan dua atribut, yang dapat dilihat pada tabel 3.1 berikut:

Tabel 1. Nama Atribut

Atribut	Keterangan
Tweets	Komentar masyarakat melalui media X
Sentimen	Sentimen positif dan negatif

Metode Pengumpulan dan Pengolahan Data

Data penelitian ini diperoleh dari *Twitter* menggunakan tagar "Shin Tae Yong" dan "Timnas Indonesia." Setelah mengumpulkan *tweet* yang sesuai, digunakan *Google Colab* untuk mengambil data tersebut, diikuti oleh pemrosesan di *RapidMiner* dengan ekstensi *Google Colab Twitter* untuk memudahkan ekstraksi dan pengelolaan data *tweet*. Tahap berikutnya meliputi deskripsi data, pemilihan atribut, dan *cleansing* data. Data kemudian diekspor ke

format CSV agar lebih mudah dianalisis. Data yang dikumpulkan berasal dari *tweet* antara 16 Desember 2023 hingga 20 Mei 2024. Data yang digunakan dalam penelitian ini bersifat kualitatif, yaitu data yang mencirikan atau menggambarkan fenomena, dapat diamati, dan direkam. Data kualitatif dibagi menjadi dua jenis, yaitu:

Data Primer

Data primer adalah data yang diperoleh langsung dari sumber aslinya, memungkinkan peneliti untuk mengamati dan mencatat informasi langsung dari objek penelitian. Teknik pengumpulan data primer adalah melalui observasi langsung, yang memungkinkan peneliti memperoleh data dan informasi terkait objek tersebut secara menyeluruh.

Data Sekunder

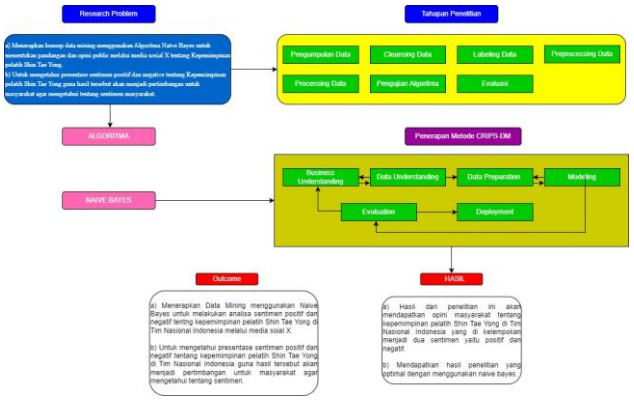
Data sekunder merupakan data yang sudah ada atau telah diolah sebelumnya dan diperoleh dari sumber lain sebagai tambahan informasi. Teknik pengumpulan data sekunder meliputi:

1) Studi pustaka
   Metode ini mencakup pengumpulan data dari literatur, membaca, mencatat, dan mengolah bahan penelitian.
2) Buku teks
   Pengumpulan informasi melalui buku-buku yang relevan dengan topik penelitian.

Penelitian ini menggunakan algoritma *Naïve Bayes* dengan data dari *Twitter* yang dikumpulkan menggunakan tagar "Shin Tae Yong" dan "Timnas Indonesia." Tahapan penelitian mengacu pada metode *Cross Industry Standard for Data Mining (CRISP-DM)*, yang terdiri dari enam tahap berikut:

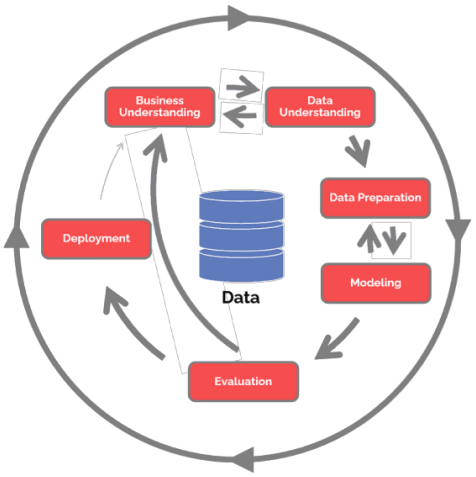
1) Pemahaman Bisnis
2) Pemahaman Data
3) Persiapan Data
4) Pemodelan
5) Evaluasi
6) Deployment

Tahapan ini ditampilkan dalam gambar 1 untuk lebih memperjelas metodologi penelitian yang diterapkan.



Gambar 1. Tahapan Penerapan Metodologi

Data tersebut diperoleh dari X dengan menggunakan GoogleColab ,cuitan yang diambil berupa cuitan berbahasa Indonesia dengan kata kunci Shin tae yong dan Timnas indonesia dengan limit 5.000 dan berhasil di dapat berjumlah 2.495 data cuitan yang telah selesai melalui tahap text preprocessing. Pada tahapan ini pendekatan yang digunakan menggunakan metode Cross Industry Standard for Data Mining (CRISP-DM). CRISP-DM merupakan metode yang menggunakan model proses pengembangan data yang banyak digunakan para ahli untuk memecahkan masalah. Proses penelitian ini mengacu pada enam tahap terdapat dalam CRISP-DM ini yakni dijelaskan sebagai berikut:



Gambar 2. Metodologi Crisp-DM

Pemahaman Bisnis (*Business Understanding*)

Pada tahap ini, penelitian menganalisis objek berupa sentimen publik terhadap Shin Tae Yong, pelatih tim nasional sepak bola Indonesia. Tujuan penelitian adalah menentukan sentimen publik terhadap

kepemimpinan Shin Tae Yong. Data diambil dari media sosial *X* menggunakan *Google Colab* dan aplikasi *RapidMiner*. Setelah data terkumpul, beberapa proses dilakukan untuk mengolah data, termasuk mengganti *RT*, tautan, tagar, *mention*, simbol, dan menghapus duplikasi agar menghasilkan *tweet* yang siap untuk analisis klasifikasi menggunakan algoritma *Naïve Bayes*.

Pemahaman Data (Data *Understanding*)

Berdasarkan data yang dikumpulkan, langkah berikutnya adalah memahami kebutuhan data untuk mencapai tujuan penelitian. Data yang dikumpulkan dari *X* menggunakan *Google Colab* ini akan dianalisis lebih lanjut untuk memastikan kecocokan dan kesesuaiannya dengan tujuan penelitian.

Persiapan Data (Data *Preparation*)

Persiapan data mencakup seluruh proses pengolahan untuk membangun *dataset* yang akan diterapkan pada alat pemodelan klasifikasi. Persiapan data ini padat dengan kegiatan pengolahan untuk memastikan bahwa data sudah siap digunakan. Tahapan persiapan data ini meliputi:

1) Pengumpulan Data
   Mengumpulkan data dari *X* dengan kata kunci "Shin Tae Yong", "STY", dan "Timnas".
2) Pembersihan Data (*Cleansing* Data)
   Membersihkan data dari karakter khusus, tautan, tanda baca, dan elemen lain yang tidak diperlukan, termasuk *stopwords* seperti "dan", "atau", dan "juga" yang tidak relevan terhadap analisis sentimen.
3) Tokenisasi
   Memecah teks menjadi unit-unit yang lebih kecil (kata atau frasa) untuk analisis lebih lanjut.
4) Stemming atau Lemmatisasi
   Mengembalikan kata-kata ke bentuk dasarnya, baik melalui stemming (memotong akhiran kata) atau lemmatisasi (pengembalian kata ke bentuk dasar sesuai konteks).
5) Menghilangkan Noise
   Menghapus angka, simbol, dan kata yang berulang secara berlebihan dalam teks agar tidak mengganggu analisis.
6) Normalisasi
   Melakukan normalisasi teks, seperti menghapus tanda baca atau mengganti kata-kata tertentu dengan representasi standar.
7) Pemberian Label Sentimen (Labeling Sentimen)
   Menandai data dengan sentimen yang sesuai, seperti "positif" atau "negatif". Pemberian label ini dapat dilakukan secara manual atau otomatis dengan menggunakan kamus sentimen atau model klasifikasi.

Setelah tahap persiapan data selesai, langkah berikutnya adalah melakukan analisis sentimen menggunakan *Naïve Bayes Algorithm Classification*.

Pemodelan (*Modeling*)

Pada tahap ini, metode klasifikasi menggunakan algoritma *Naïve Bayes* diterapkan. Data dibagi menjadi tiga atribut utama: *tweet*, *positive*, dan *negative*. Alat pemodelan yang digunakan adalah *RapidMiner* untuk mengelompokkan data sesuai klasifikasi yang diinginkan.

Evaluasi (*Evaluation*)

Tahap evaluasi bertujuan untuk menganalisis dan mengukur ketepatan pemodelan yang telah dilakukan. Evaluasi dilakukan untuk memastikan pemodelan yang dibuat sesuai dan tepat untuk kasus penelitian ini. Hasil evaluasi menentukan langkah berikutnya, apakah penelitian dapat dilanjutkan atau perlu diulang jika tidak sesuai dengan rencana awal.

Penyebaran Hasil (*Deployment*)

Tahap akhir adalah menyebarkan hasil penelitian, yang kemudian disusun dalam bentuk laporan atau presentasi, untuk menyampaikan temuan dan pengetahuan yang diperoleh dari pemodelan dan evaluasi proses *data mining* ini.

3. Hasil dan Pembahasan

Hasil

Dalam penelitian, membutuhkan alat untuk mendukung berjalannya penelitian. Alat penelitian yang digunakan yaitu berupa perangkat lunak (*software*) dan perangkat keras (*hardware*). Berikut alat perangkat lunak (*software*), versi , dan fungsinya dapat dilihat pada tabel 2.

Tabel 2. Spesifikasi *Software*

No	*Software*	Versi	Fungsi
1	Google Colabs	2024	Untuk Mengcrawling Data Dari X
2	Google Spreadsheets	2024	Untuk Memfilter data dan Menyimpan data
3	Rapid Miner	9.10	Untuk Melakukan Analisi Data Minng

Tabel 3. Spesifikasi *Hardware*

No	*Hardware*	Spesifikasi
1	Processor	Intel(R) Core(TM) i7-7600U CPU @ 2.80GHz 2.90 GHz
2	RAM	8 GB
3	System Type	64 BIT
4	Monitor	DELL LATITUDE 7480
5	Operating System	WINDOWS 11 PRO
6	Mouse	Inphic P-M6

Implementasi Dan Pengujian

Pada penelitian ini metode yang digunakan adalah *Naïve Bayes Algorithm Classification*. Data yang digunakan ialah tweet yang diambil dari aplikasi media sosial X menggunakan Google Colab dengan cara menggunakan ekstensi dari Rapidminer yaitu search X yang di ambil dari 1 januari 2023 sampai 30 mei 2024 setelah data tersebut di dapatkan lalu masuk ke proses penerapan CRISP-DM yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan penyebaran (*Deployment*). Data tersebut akan diproses dalam pengujian algoritma Naïve Bayes menggunakan tools Rapidminer supaya dapat nilai akurasi klasifikasi yang baik dan optimal.

CRISP-DM

Pengujian ini dapat dilakukan dengan menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang memiliki 6 tahapan.

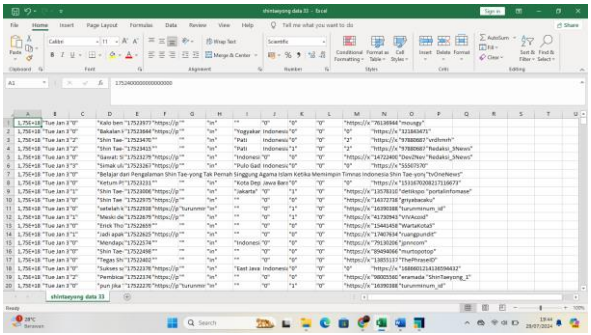
Pemahaman Bisnis (*Business Understanding*)

Shin Tae Yong atau yang biasa disebut Shin Tae Yong merupakan Timnas Indonesia yang pernah menjadi pemain sepak bola di Korea Selaran dan pernah melatih tim Nasional Korea Selatan di Piala Dunia 2018 Rusia. Banyak yang mengawasi atau memperhatikan tingkah dan perilaku Shin Tae Yong saat melatih Timnas Indonesia. Shin Tae Yong memiliki rasa khawatir yang cukup besar dengan Timnas Indonesia karena strateginya. Tujuan dasar adalah untuk menentukan mencari nilai akurasi, presisi dan recall dan menganalisa sentiment masyarakat tentang Shin tae yong.

Pemahaman Data (Data *Understanding*)

Memahami informasi data yang akan digunakan dalam penelitian, adapun beberapa tahap untuk melakukan penelitian ini yaitu Pengumpulan data awal yaitu melakukan pengumpulan data yang dibutuhkan untuk mendukung dalam melakukan pemahaman data. Adapun sumber data utama yang digunakan dalam penelitian ini adalah data tweet Shin tae yong pada 1 januari 2023sampai 30 mei 2024.

Gambar 3. Proses Get Data

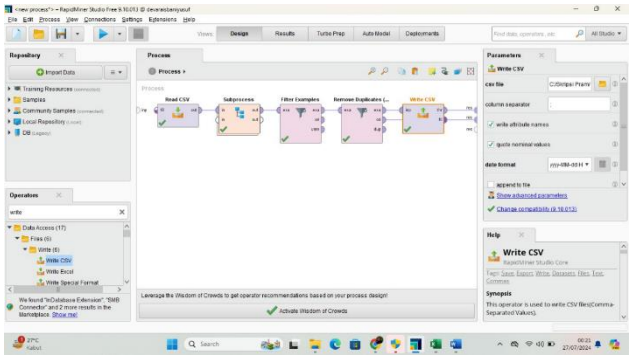


Gambar 4. Hasil Proses Data

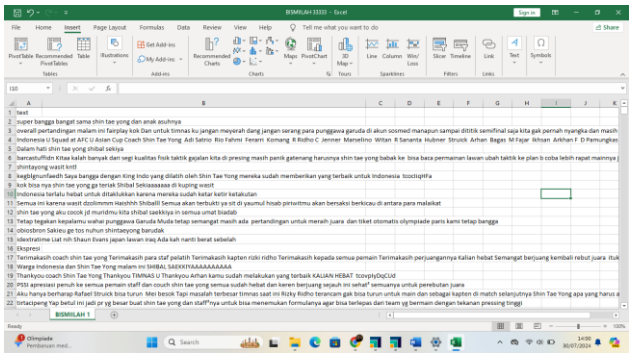
Data Preparition

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 5.000 data tweet pada file csv, kemudian setelah melewati proses *Cleansing* maka data tweet berkurang menjadi 2.495 data.



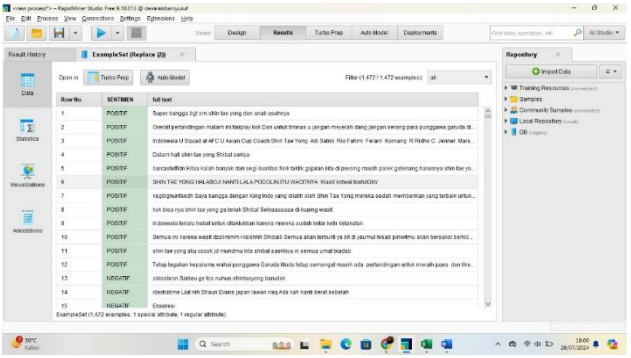
Gambar 5. ProsesModel cleansing Data



Gambar 6. Hasil Cleansing Data

Labeling Data

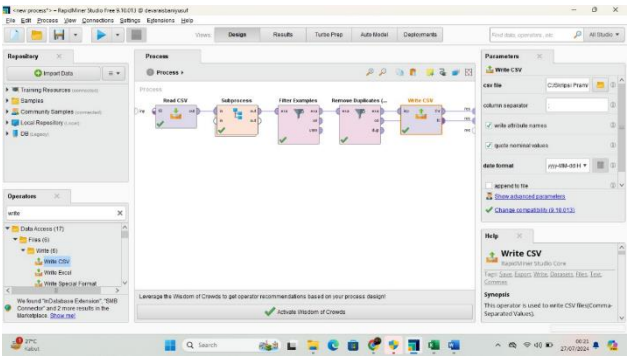
Kemudian tahap selanjutnya adalah melakukan klasifikasi atau penentuan sentimen berdasarkan masing-masing tweet. Selama proses penentuan sentimen berlangsung, disini peneliti melakukan pemberian sentimen secara manual. Berikut ini data yang sudah di beri label.



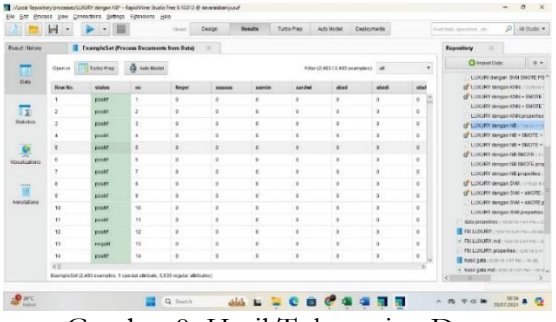
Gambar 7. Data Telah Dilabeli Sentimen

Modelling

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. Berikut proses *text preprocessing* dapat dilihat pada gambar 8. Setelah melewati tahapan text preparation, maka jumlah data tweet didapatkan menjadi 2.495 data yang merupakan data bersih untuk dipakai pada tahap selanjutnya.



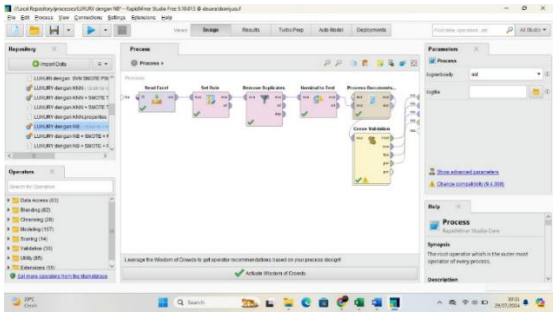
Gambar 8. *Text Propocessing*



Gambar 9. Hasil Tokenezing Data

Untuk tahapan modelling peneliti akan melakukan pengukuran performa klasifikasi dengan Pemodelling menggunakan split data. Pemodelan ini data akan

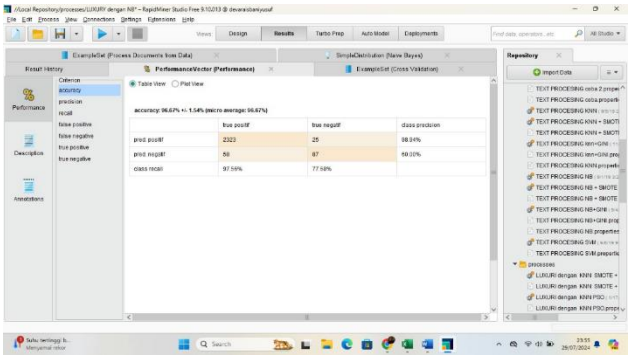
dibagi menggunakan split menjadi 2 model yaitu 0.8 dan 0.2. Pada tahapan ini akan dilakukan pengukuran performa klasifikasi dengan menggunakan split data. Berikut proses Modelling yang dilakukan di *software RapidMiner* yang ditunjukkan pada gambar 10.



Gambar 10. Proses Modeling

Evaluasi

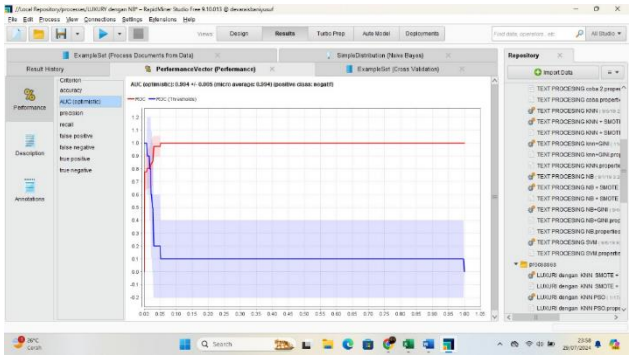
Pada penelitian ini akan dilakukan pengujian menggunakan confusion matrix untuk melihat hasil pengujian data yang diperoleh dari tahapan modelling dengan menggunakan algoritma *Naïve Bayes*. Total dataset yang dikumpulkan adalah 2.495 data. Berikut ini adalah confusion matrix hasil dari tahapan modelling yang dapat dilihat hasil dari perhitungan hasil RapidMiner dapat dilihat pada gambar 11 untuk menunjukkan dan membuktikan hasil prediksi dari model *Naïve Bayes*.



Gambar 11. Hasil Proses Akurasi

Tabel 4. Hasil Proses Akurasi

	True Positif	True Negatif	Class Precision
Pred.Positif	2323	25	98.94%
Pred.Negatif	58	87	60.00%
Class recall	97.56%	77.68%	



Gambar 12. Grafik AUC

Berdasarkan pada Gambar 10 dapat dibuat kesimpulan bahwa precision menunjukkan tingkat ketepatan data yang diprediksi positif terhadap banyaknya data yang benar diprediksi positif yang menghasilkan presentase ketepatannya adalah 61.16% sedangkan untuk data bersentimen negatif memiliki precision sebesar 13.96% Untuk sentimen positif memiliki recall (*Specificity*) adalah 77.58% sehingga dapat disimpulkan bahwa model dapat menemukan kembali informasi atau data yang benar-benar negatif dengan sangat baik, sedangkan pada sentimen negatif memiliki recall sebesar 15.02% Nilai *accuracy* yang dihasilkan menggunakan model *Naïve Bayes* adalah 96.67% sehingga dapat disimpulkan bahwa algoritma *Naïve bayes* dapat mengklasifikasi sentimen dengan baik menggunakan data Shin tae yong.

Deployment

Deployment merupakan tahap akhir dalam pembuatan laporan hasil kegiatan data mining. Laporan akhir yang berisi tentang pengetahuan yang diperoleh atau pengenalan pola dalam proses data mining

Pembahasan

Penelitian ini menggunakan algoritma *Naïve Bayes* untuk menganalisis sentimen publik terhadap pelatih tim nasional sepak bola Indonesia, Shin Tae Yong, melalui platform media sosial. Penerapan *Naïve Bayes*

pada analisis sentimen telah terbukti efektif dalam berbagai penelitian sebelumnya, terutama pada media sosial seperti *Twitter*, yang menjadi sumber opini publik tentang isu-isu sosial dan tokoh publik (Lisangan *et al*., 2022). Algoritma *Naïve Bayes* dipilih karena kemampuannya yang terbukti dalam klasifikasi teks, meskipun menghadapi keterbatasan asumsi independensi antar variabel. Penelitian Lisangan *et al*. (2022) menunjukkan bahwa *Naïve Bayes* berhasil mengklasifikasikan sentimen masyarakat terkait kebijakan *New Normal* di Indonesia dengan akurasi yang tinggi. Begitu pula dalam studi Pratama *et al*. (2022), *Naïve Bayes* diterapkan pada opini masyarakat terhadap tim nasional pada Piala AFF 2020. Kedua studi ini mengonfirmasi akurasi *Naïve Bayes* dalam memproses data media sosial untuk menghasilkan klasifikasi yang mendekati hasil sesungguhnya. Penelitian ini, seperti halnya yang dilakukan Apriani *et al.* (2019) pada platform Tokopedia, menunjukkan bahwa *Naïve Bayes* efektif dalam mengklasifikasikan opini publik meski dengan data yang sangat variatif. Penerapan algoritma ini untuk mengidentifikasi sentimen publik terhadap Shin Tae Yong berfungsi sebagai bentuk aplikasi yang penting dalam memahami opini masyarakat terhadap tokoh publik.

Data penelitian dikumpulkan dari media sosial *X* dengan kata kunci "Shin Tae Yong" dan "Timnas Indonesia." Pengumpulan data dilakukan menggunakan *Google Colab* dan *RapidMiner*, di mana data diolah melalui tahap pembersihan, tokenisasi, normalisasi, dan pemberian label sentimen. Pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang diterapkan dalam penelitian ini memudahkan proses data mining, dengan fokus pada akurasi klasifikasi. Penelitian oleh Susana dan Suarna (2022) mendukung pentingnya persiapan data yang komprehensif untuk hasil klasifikasi yang akurat. Selain itu, penelitian yang mengaplikasikan CRISP-DM dalam analisis sentimen mengakui pentingnya pembersihan dan tokenisasi, seperti yang dilakukan oleh Pratama *et al.* (2022), yang menghasilkan data yang terstruktur dan siap untuk analisis lanjut. Hal ini menegaskan bahwa algoritma *Naïve Bayes* dalam penelitian ini dapat memberikan hasil yang representatif tentang sentimen publik setelah melalui proses persiapan yang teliti.

Pada penelitian ini, algoritma *Naïve Bayes* mencapai akurasi sebesar 96,67% dalam mengklasifikasikan sentimen publik terhadap Shin Tae Yong. Tingkat akurasi ini menunjukkan bahwa metode ini efektif untuk mengidentifikasi opini positif atau negatif dari publik, terutama di media sosial. Dalam penelitian Apriani *et al.* (2019), *Naïve Bayes* menghasilkan performa yang baik dalam klasifikasi sentimen pengguna Tokopedia, dan ini sejalan dengan penelitian Susanto dan Suarna (2022), di mana *Naïve Bayes* menunjukkan performa kuat dalam berbagai situasi klasifikasi yang melibatkan data besar. Evaluasi model menggunakan matriks konfusi untuk menghitung akurasi, presisi, dan *recall* menunjukkan bahwa *Naïve Bayes* tidak hanya akurat tetapi juga konsisten dalam menangani berbagai kategori data. Hasil ini diperkuat oleh penelitian Damayanti dan Maulidiyah (2023) yang menyoroti kemampuan *Naïve Bayes* untuk memberikan hasil klasifikasi yang dapat diandalkan ketika diterapkan pada data dengan distribusi yang berbeda. Implementasi matriks konfusi juga menunjukkan kesesuaian model dengan data *tweet* yang bervariasi, serupa dengan yang ditemukan dalam penelitian Pratama *et al.* (2022) pada data sentimen masyarakat terhadap tim nasional.

Penerapan *Naïve Bayes* dalam penelitian ini berperan penting dalam memberikan wawasan kepada pihak-pihak terkait di bidang olahraga Indonesia, seperti manajemen tim nasional dan pengambil kebijakan olahraga. Selain itu, algoritma ini dapat diaplikasikan untuk isu-isu lain yang berhubungan dengan opini publik mengenai tokoh atau kebijakan tertentu. Penelitian Putra dan Syafira (2023) pada topik politik dengan algoritma *Naïve Bayes* menyoroti pentingnya metode ini dalam analisis cepat opini masyarakat terhadap topik sensitif. Kemampuan *Naïve Bayes* dalam meninjau sentimen publik juga dapat membantu pengambilan keputusan yang lebih responsif berdasarkan opini masyarakat. Meski *Naïve Bayes* menunjukkan performa yang baik, studi lain seperti yang dilakukan Damayanti dan Maulidiyah (2023) menyarankan kombinasi *Naïve Bayes* dengan algoritma lain, seperti *Support Vector Machine* atau *Random Forest*, untuk menghasilkan akurasi yang lebih tinggi dalam klasifikasi sentimen. Ini membuka peluang bagi penelitian di masa depan untuk menggabungkan algoritma *Naïve Bayes* dengan model lain dalam pendekatan *ensemble*, sehingga

meningkatkan akurasi dan ketelitian prediksi. Penelitian Susanto dan Dzulkarnain (2023) pada *Random Forest* dan *Naïve Bayes* menunjukkan bahwa kombinasi ini dapat memperkuat hasil klasifikasi dalam topik ekonomi dan industri. Selain itu, pengembangan model otomatis untuk penandaan sentimen (labeling) akan mempercepat proses analisis dan mengurangi kesalahan manual. Dengan demikian, model yang dihasilkan lebih kuat dalam memberikan prediksi dan hasil yang lebih optimal dalam klasifikasi data sentimen.

Secara keseluruhan, algoritma *Naïve Bayes* terbukti sebagai metode yang efektif untuk analisis sentimen publik. Dengan akurasi 96,67% dalam penelitian ini, algoritma ini mampu mengidentifikasi sentimen publik terhadap pelatih Shin Tae Yong dengan tepat, memperlihatkan bahwa publik Indonesia memiliki respons positif terhadap kepemimpinannya. Hasil ini relevan dengan temuan lain yang menunjukkan bahwa *Naïve Bayes* dapat menangani variasi sentimen publik pada tokoh dan isu penting. Untuk penelitian selanjutnya, penggunaan algoritma tambahan seperti *Support Vector Machine* atau *Random Forest* dapat mempertajam hasil klasifikasi sentimen, terutama pada data yang beragam dan besar.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan mengenai klasifikasi opini masyarakat Indonesia terhadap Shin Tae Yong, dapat disimpulkan bahwa mayoritas sentimen masyarakat terhadap Shin Tae Yong bersifat positif. Hasil penelitian menunjukkan bahwa 99,85% dari total data, yaitu sebanyak 2.125 data, bersentimen positif, sedangkan hanya 145 data yang memiliki sentimen negatif. Dari hasil ini, dapat disimpulkan bahwa sebagian besar masyarakat Indonesia memberikan respons positif terhadap kepemimpinan Shin Tae Yong.

Berdasarkan model klasifikasi menggunakan algoritma *Naïve Bayes* dengan rasio pembagian data 0,8 : 0,2 dan nilai k=3k = 3k=3 pada dataset Shin Tae Yong, diperoleh akurasi sebesar 96,67%. Hal ini menunjukkan bahwa algoritma *Naïve Bayes* efektif dalam mengklasifikasikan data secara akurat. Adapun saran yang disampaikan berdasarkan hasil pengamatan dan analisis dalam penelitian ini adalah sebagai berikut. Pertama, penerapan algoritma *Naïve Bayes* terbukti efektif dalam analisis sentimen, namun akan lebih optimal jika dikombinasikan dengan algoritma lain untuk meningkatkan akurasi serta presisi dalam proses perhitungan. Kedua, meskipun penelitian ini sudah memberikan hasil yang signifikan, namun penelitian ini masih memiliki beberapa keterbatasan, oleh karena itu, diharapkan pengembangan lebih lanjut dilakukan dengan menambahkan metode lain agar proses identifikasi objek lebih cepat dan akurat. Terakhir, untuk memaksimalkan efisiensi analisis, diperlukan metode otomatis untuk penentuan label sentimen agar proses penentuan sentimen terhadap objek dapat dilakukan dengan lebih cepat dan meminimalkan kesalahan manual.

5. Daftar Pustaka

Aponno, J. C. (2022). Penerapan Algoritma Sentimen Analysis dan Naïve Bayes terhadap opini pengunjung di tempat wisata pantai Pintu Kota, Kota Ambon. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, *9*(4), 3180-3188. DOI: https://doi.org/10.35957/jatisi.v9i4.2697.

Apriani, R., & Gustian, D. (2019). Analisis Sentimen dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia. *Jurnal Rekayasa Teknologi Nusa Putra*, *6*(1), 54-62.

Bara, E. B., Nasution, K. A., Ginting, R. Z., & Kartini, K. (2022). Penelitian tentang Twitter. *Jurnal Edukasi Nonformal*, *3*(2), 167-172.

Damayanti, A., & Maulidiyah, A. K. (2023). Analisis Sentimen Masyarakat Terhadap Pembatalan Piala Dunia U20 di Indonesia Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM). *Jurnal Ilmiah Terapan, Sains dan Teknologi (JITSI)*, *1*(2), 97-103. DOI: https://doi.org/10.25139/jitsi.v1i2.6607.

Hidayat, A. M., & Syafrullah, M. (2018). Algoritma Naïve Bayes Dalam Analisis Sentimen Untuk Klasifikasi Pada Layanan Internet PT. XYZ. *Telematika MKOM*, *9*(2), 91-95.

Lisangan, E. A., Gormantara, A., & Carolus, R. Y. (2022). Implementasi Naive Bayes pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi New Normal di Indonesia. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, *2*(1). DOI: https://doi.org/10.24002/konstelasi.v2i1.560.

Mahbubah, L. D., & Zuliarso, E. (2019). Analisa Sentimen Twitter Pada Pilpres 2019 Menggunakan Algoritma Naive Bayes.

Munawaroh, A., Ridhoi, R., & Rudiman, R. (2024). SENTIMENT ANALYSIS DENGAN NAÏVE BAYES BERBASIS ORANGE TERHADAP RESIKO PEMBANGUNAN IKN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 587-592. DOI: https://doi.org/10.36040/jati.v8i1.8454.

Nadhifa, F. H. (2023). ANALISIS FRAMING TERHADAP PEMBERITAAN PELATIH TIMNAS INDONESIA. *CommLine*, *7*(2), 106-111. DOI: http://dx.doi.org/10.36722/cl.v7i2.1313.

Prajamukti, R., Jayanta, J., & Santoni, M. M. (2021). Klasifikasi Dan Analisis Sentimen Pada Data Twitter Menggunakan Algoritma Naïve Bayes (Studi Kasus: Timnas Indonesia Senior, U-23, Dan U-19). *PROSIDING SEINASI-KESI*, *4*(1), 102-109.

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. OI: https://doi.org/10.70309/ticom.v10i3.33.

Purbayanto, B., & Suharsono, T. N. (2023). Analisis Sentimen Pengguna X terhadap Chatgpt dengan Algoritme Naive Bayes. *Jurnal Telematika*, *18*(2), 63-71. DOI: https://doi.org/10.61769/telematika.v18i2.614.

Putra, A. P., & Syafira, A. F. (2023). Analisis Sentimen Data Twitter Topik Politik Dengan Metode Naive Bayes Dan Convolutional Neural Networks (Cnn). *Jurnal Ilmiah Wahana Pendidikan*, *9*(20), 36-41. DOI: https://doi.org/10.5281/zenodo.8396579.

Ratnaningtyas, R. P., & Muhammad, Y. A. (2023). Analisis Pemberitaan Timnas Indonesia pada Media Daring. *MUKASI: Jurnal Ilmu Komunikasi*, *2*(1), 45-52. DOI: https://doi.org/10.54259/mukasi.v2i1.1492.

Sidik, M. H., Widiyanesti, S., & Ramadhani, D. P. (2022). Analisis Sentimen dan Topic Modelling Terhadap Tim Nasional Indonesia di Kejuaraan AFF Suzuki Cup 2020 Berdasarkan Opini Pengguna Twitter. *eProceedings of Management*, *9*(5).

Susana, H. (2022). Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, *4*(1), 1-8.

Susanto, A., & Dzulkarnain, I. A. (2023). Analisis Sentimen Data Twitter Topik Ekonomi Dan Industri Dengan Metode Naive Bayes Dan Random Forest. *Jurnal Ilmiah Wahana Pendidikan*, *9*(20), 59-65. DOI: https://doi.org/10.5281/zenodo.8398895.

Widowati, T. T., & Sadikin, M. (2020). Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, *11*(2), 626-636. DOI: https://doi.org/10.24176/simet.v11i2.4568.