

**ANALISIS SENTIMEN TERHADAP NATURALISASI PEMAIN
TIMNAS INDONESIA BERDASARKAN OPINI TWITTER
MENGUNAKAN NAIVE BAYES**

Mohammad Sofiyan¹, Sufajar Butsiyanto², Sophian Andika³
Universitas Pelita Bangsa

E-mail: m.sofian80820@gmail.com¹, sufajar@pelitabangsa.ac.id², sophian.andhika@gmail.com³

Abstract

Naturalization is a legal process in which a person who is not a citizen of a country obtains citizenship status in that country. This process is also carried out in the world of sports, one of which is football. In the context of sports, naturalization is carried out to allow foreign athletes to become citizens of a country and represent it in international competitions. In the case of the naturalization of foreign players to the Indonesian National Team, it became quite a sensational topic in Indonesia, especially on social media Twitter. Some people responded with positive opinions because it could improve the performance of team players by combining foreign players into the Indonesian National Team, and some people responded with negative opinions because it could hinder the performance of local players. Because of this, researchers are interested in analyzing public sentiment regarding the naturalization of foreign players to the Indonesian National Team. The goal is to measure and understand public perceptions and reactions to the policy. This analysis helps identify whether the public tends to support, reject, or be neutral towards naturalization, as well as the reasons behind these sentiments. With this understanding, policymakers and related parties can evaluate the social impact of naturalization policies, adjust communication strategies, and make more informed decisions to increase public support or address existing concerns. By using the Lexicon Based technique as a data labeling process and a classification method with the Naive Bayes algorithm as a measurement of sentiment data classification, a sentiment analysis model was obtained with evaluation results of accuracy 72.23%, precision 82.23% and recall 83.37%.

Keywords: Data Mining, Sentiment Analysis, Classification, Naive Bayes, National Football Team.

Abstrak

Naturalisasi adalah proses hukum di mana seseorang yang bukan warga negara suatu negara memperoleh status kewarganegaraan di negara tersebut. Pada proses ini juga dilakukan dalam dunia olahraga, salah satunya Sepak Bola. Dalam konteks olahraga, naturalisasi dilakukan untuk memungkinkan atlet asing menjadi warga negara suatu negara dan mewakilinya dalam kompetisi internasional. Pada kasus naturalisasi pemain asing ke Timnas Indonesia menjadi topik yang cukup menghebohkan di negara Indonesia, khususnya pada media sosial Twitter. Beberapa masyarakat menanggapi dengan pendapat positif karena bisa meningkatkan performa pemain tim dengan menggabungkan pemain asing ke Timnas Indonesia, dan juga sebagian masyarakat menanggapinya dengan pendapat yang negatif karena dapat menghalangi performa pemain lokal. Karena hal tersebut peneliti tertarik untuk menganalisis sentimen publik mengenai naturalisasi pemain asing ke Timnas Indonesia. Tujuannya adalah untuk mengukur dan memahami persepsi serta reaksi masyarakat terhadap kebijakan tersebut. Analisis ini membantu mengidentifikasi apakah publik cenderung mendukung, menolak, atau netral terhadap naturalisasi, serta alasan di balik sentimen tersebut. Dengan pemahaman ini, pembuat kebijakan dan pihak terkait dapat mengevaluasi dampak sosial dari kebijakan naturalisasi, menyesuaikan strategi komunikasi, dan mengambil keputusan yang lebih tepat untuk meningkatkan dukungan publik atau mengatasi kekhawatiran yang ada. Dengan menggunakan teknik Lexicon Based sebagai proses labelisasi data dan metode klasifikasi dengan algoritma Naive Bayes sebagai pengukuran klasifikasi data sentimen didapatkan sebuah model analisis sentimen dengan hasil evaluasi accuracy 72.23%, precision 82.23% dan recall 83.37%.

Kata Kunci: Data Mining, Analisis Sentimen, Klasifikasi, Naive Bayes, Timnas Sepakbola.

1. PENDAHULUAN

Teknologi informasi telah berkembang sangat cepat sejak awal dan sangat berbeda dari sebelumnya. Globalisasi saat ini memiliki dampak yang signifikan terhadap pesatnya kemajuan teknologi informasi dalam berbagai bidang, termasuk ekonomi, seni, budaya, pendidikan, dan bahkan politik. Teknologi informasi telah menjadi sangat penting dalam era globalisasi karena memiliki kemampuan untuk mewujudkan dunia yang tidak memiliki batas ruang, waktu, atau jarak, dan juga memiliki kemampuan untuk meningkatkan produktivitas dan efisiensi [1]. Teknologi telah membuat jarak tidak lagi menjadi masalah dalam berkomunikasi, dan sekarang sosial media menjadi kebutuhan wajib bagi sebagian kalangan masyarakat mengatakan komentar mereka di web media sosial mereka seperti Facebook, Twitter, Wechat, dan sejenisnya. Twitter adalah salah satu web media sosial yang paling sering digunakan untuk membicarakan masalah sosial dan politik [2].

Media sosial Twitter adalah salah satu media komunikasi yang diminati oleh masyarakat di dunia. Hal ini dapat dilihat dari peningkatan pengguna Twitter yang tercatat di seluruh dunia. Twitter memiliki jumlah pengguna aktif sebesar 313 juta per bulan pada tahun 2016 dan sebagian besar pengguna mengakses Twitter melalui perangkat mobile yaitu sebesar 82 persen. Jumlah pengguna yang banyak juga menimbulkan tweet yang banyak dari pengguna. Pengguna akan memberikan kabar terbaru atau komentar tentang hal yang sedang menjadi topik utama di dunia. Hal yang sedang menjadi topik utama dan banyak dikomentari oleh pengguna akan menimbulkan trending topik di Twitter [3]. Layanan yang disediakan Twitter untuk penggunaannya termasuk pembuatan pesan status (disebut "tweet"), yang dapat dibaca oleh pengguna Twitter lainnya dan biasanya pendapat pengguna tentang berbagai topik dengan batas 140 karakter, sehingga twitter menjadi salah satu situs yang menyediakan kumpulan data opini dari masyarakat di seluruh dunia [4]. Twitter telah berkembang menjadi platform penting untuk diskusi publik tentang berbagai topik, mulai dari topik sehari-hari hingga masalah global. Karena kemampuan untuk menyebarkan informasi dengan cepat dan luas, menjadikannya alat yang sangat baik untuk mengukur opini publik tentang berbagai topik, salah satu topik yang baru hangat dibahas dalam beberapa bulan di tahun ini yaitu tentang naturalisasi pemain asing untuk Timnas Indonesia.

Naturalisasi adalah perolehan kewarganegaraan bagi penduduk asing (hal menjadikan warga negara) pewarganegaraan yang diperoleh setelah memenuhi syarat sebagaimana yang ditetapkan dalam peraturan perundang-undangan. Jadi secara mendasar, naturalisasi adalah proses perubahan status dari penduduk asing menjadi warga negara di suatu negara [5]. Sama halnya dengan Naturalisasi pemain asing yaitu proses di mana seorang pemain sepak bola yang bukan warga negaranya diberikan kewarganegaraannya sehingga dapat bermain untuk Tim Nasional di suatu negara tersebut, khususnya untuk negara Indonesia. Olahraga sepak bola adalah salah satu olahraga yang paling banyak diminati oleh masyarakat dunia. Indonesia sendiri termasuk negara penggemar sepak bola paling banyak di dunia. Di Indonesia sendiri merupakan salah satu negara dengan penggemar sepak bola terbanyak di dunia dengan 77% penduduk Indonesia memiliki ketertarikan dengan olahraga sepak bola [4]. Dalam pembahasan naturalisasi pemain asing ke Indonesia memicu banyak polemik mengenai Timnas Indonesia ke depannya, beberapa pendukung melihat naturalisasi sebagai cara untuk memperkuat tim nasional dengan bakat dari luar negeri, sementara yang lain merasa bahwa itu menghalangi perkembangan pemain lokal. Oleh karena itu, penting untuk memahami opini publik mengenai isu ini, dan Twitter menyediakan platform yang ideal untuk melakukan analisis sentimen terhadap pandangan publik.

Opinion Mining disebut juga sebagai Analisis Sentimen yaitu sistem yang dibangun untuk mengumpulkan dan mengevaluasi/menguji pendapat yang dituliskan di kiriman blog, komentar, ulasan atau media daring lainnya. Opinion Mining merupakan cabang penelitian dari Text Mining. Fokus dari Opinion Mining adalah melakukan analisis opini dari suatu

dokumen teks [6]. Analisis sentimen adalah teknik yang digunakan untuk mengidentifikasi dan mengkategorikan opini yang dinyatakan dalam sebuah teks, khususnya untuk menentukan sikap penulis terhadap suatu subjek, apakah positif atau negatif. Secara umum analisis sentimen dibagi menjadi 5 (lima) tahapan yaitu crawling data, preprocessing, feature selection, classification dan evaluation. Manfaat analisis sentimen yaitu sebagai sebuah ide atau evaluasi pada berbagai bidang [7].

Penelitian ini dilakukan dengan tujuan untuk mengetahui opini pengguna Twitter terhadap pendapat naturalisasi pemain asing ke Timnas Indonesia, apakah pendapat tersebut merupakan pendapat positif atau negatif. Data sentimen yang digunakan berasal dari proses crawling data pada media sosial Twitter dengan kueri “naturalisasi pemain Timnas Indonesia”. Berdasarkan hasil penjelasan yang di atas, penulis tertarik untuk melakukan analisis sentimen dengan menggunakan salah satu algoritma klasifikasi yaitu Naive Bayes. Algoritma Naive Bayes Classification dikenal sangat efektif untuk klasifikasi teks, seperti penyaringan spam dan analisis sentimen, karena asumsi independen kondisional antar fitur sering kali cukup valid dalam konteks ini, dan algoritma ini juga bekerja dengan baik pada dataset yang besar maupun yang kecil [8].

2. METODE

Untuk melakukan analisis sentimen ada beberapa algoritma klasifikasi yang dapat digunakan salah satunya adalah algoritma Naive Bayes. Naive Bayes Classifier merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yaitu memprediksi peluang berdasarkan pengalaman di masa sebelumnya (Teorema Bayes) dengan ciri utamanya adalah asumsi yang sangat kuat (naif) akan ketergantungan dari masing-masing kondisi/kejadian [9]. Naive Bayes Classifier memiliki beberapa keunggulan, antara lain yaitu proses komputasi yang cepat, mudah diterapkan dengan struktur yang sederhana, dan efektif. Naive Bayes telah digunakan dengan cukup sukses dalam konteks beragam aplikasi, dan sangat populer dalam konteks klasifikasi teks. Dalam melakukan klasifikasi teks, algoritme Naive Bayes mampu mendapatkan nilai akurasi yang tinggi dan kompleksitas runtime yang baik dengan jumlah data yang besar. Analisis Sentimen dapat diaplikasikan menggunakan sebuah metode klasifikasi untuk mempermudah dalam pengelompokan data berupa data positif atau data negatif yaitu dengan menggunakan metode Naive Bayes. Metode Naive Bayes digunakan pada proses klasifikasi dalam sebuah penelitian karena bekerja dengan cepat dan efisien sebagai metode pengklasifikasian teks [10]. Analisis sentimen atau opinion mining adalah proses mengolah data tekstual secara otomatis untuk mengetahui apakah opini yang terkandung dalam suatu kalimat adalah positif atau negatif dan Algoritma Naive Bayes Classifier dianggap cocok untuk digunakan dalam analisis sentimen karena bertujuan sebagai metode klasifikasi ke dalam kategori positif dan negatif. Adapun rumus dari Teroma Bayes sebagai berikut [7]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan:

X : Data yang belum diketahui.

H : Hipotesis data X merupakan suatu *class* spesifik.

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (*Posterior Probability*).

P(H) : Probabilitas hipotesis H (*Prior Probability*).

P(X|H) : Probabilitas X berdasarkan Hipotesis H.

P(X) : Probabilitas X.

3. HASIL DAN PEMBAHASAN

Hasil Penelitian

Penelitian ini bertujuan untuk mengukur kemampuan algoritma Naive Bayes dalam mengklasifikasikan sentimen positif dan negatif terkait pandangan masyarakat terhadap isu naturalisasi pemain asing ke Timnas Indonesia, berdasarkan data yang diperoleh dari media sosial Twitter. Setelah dilakukan analisis dan pengujian data, seperti yang telah dibahas pada bab sebelumnya, penulis akan memaparkan hasil pengolahan data yang dihasilkan dari proses pengujian dalam penelitian ini.

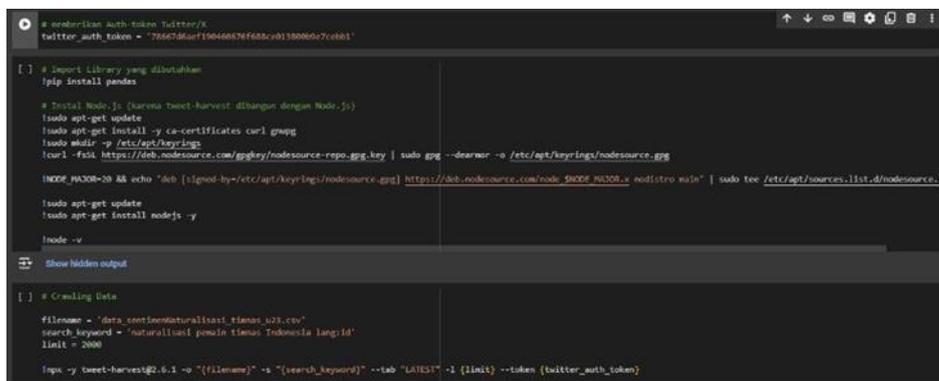
1. Dataset

Pada penelitian ini, dataset yang telah dikumpulkan dari hasil crawling data pada Twitter sebanyak 2,005 data tweet. Proses crawling data dengan query (kata kunci) “Naturalisasi Pemain Timnas Indonesia” sebagai objek utama dalam penelitian ini. Proses pengumpulan data dilakukan dengan proses crawling data yang menggunakan bahasa pemrograman Python dengan alat bantuan Google Colaboratory dan juga alat untuk penarikan datanya menggunakan pustaka Tweet Harvest. Setelah melalui banyak pemrosesan pada data didapat jumlah data yang telah dilakukan pemrosesan sebanyak 1,055 data. Data yang telah diperoleh dari hasil pemrosesan tersebut selanjutnya akan dilakukan pada tahapan-tahapan proses selanjutnya.

2. Hasil Crawling Data

Dalam tahapan ini, akan menjelaskan bagaimana melakukan crawling data pada Twitter. Pada penelitian ini penulis melakukan crawling dengan menggunakan tools Google Colaboratory dengan bahasa pemrograman.

Python pustaka Tweet Harvest. Alasan penulis menggunakan alat-alat bantuan tersebut karena pada bulan Februari tahun 2023, Twitter mengumumkan bahwa mereka akan mengakhiri akses gratis ke API Twitter dan memperkenalkan model berlangganan berbayar. Perubahan ini memengaruhi banyak aplikasi dan layanan pihak ketiga yang menggunakan API Twitter, salah satunya tools RapidMiner. Pada akhirnya RapidMiner memutuskan untuk menghapus operator yang sangat penting untuk para peneliti sentimen yaitu operator “Search Twitter” pada RapidMiner. Maka dari itu penulis mencari jalan alternatif untuk tetap bisa menarik data API dari Twitter dengan cara menggunakan library Tweet Harvest pada bahasa pemrograman Python. Berikut adalah gambar dari proses dan hasil penarikan data yang didapat:



```
# memberikan auth token twitter/
twitter_auth_token = "756c78ac11944687cf68a9d013006e7c06d1"

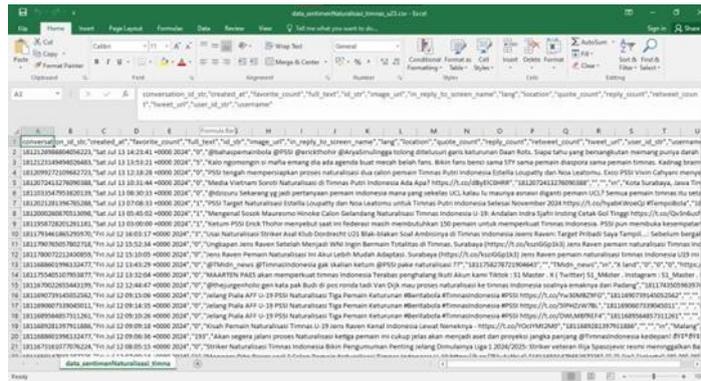
[ ] # Import library yang dibutuhkan
!pip install pandas

# Instal Node.js (karena tweet-harvest dibangun dengan Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg
!NODE_MAJOR=20 echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_20.x nodistro main" | sudo tee /etc/apt/sources.list.d/nodesource.
!sudo apt-get update
!sudo apt-get install nodejs -y
!node -v

[ ] # Crawling Data
filename = 'data_sentimenaturalisasi_tinmas_u23.csv'
search_keyword = 'naturalisasi pemain timnas Indonesia lang:Id'
limit = 2000

!python3 tweet-harvest@2.6.1 --o "{filename}" --s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Gambar 1. Proses Crawling Data Menggunakan Python Dengan Tweet Harvest



Gambar 2. Hasil Crawling Data

Pada gambar 1 memperlihatkan proses dari penarikan data pada platform media sosial Twitter menggunakan Google Colaboratory dengan bahasa pemrograman Python dan pustaka Tweet Harvest. Pada proses tersebut dibutuhkan beberapa masukan, seperti kode authorization token untuk dapat mengakses API Twitter dan juga beberapa pemanggilan pustaka untuk menjalankan library Tweet Harvest, karena Tweet Harvest bukan library langsung dari Python. Kemudian memberikan sebuah nama file untuk menyimpan hasil dari proses crawling data dari Twitter dengan format *.csv, setelah itu memberikan sebuah search_keyword untuk menentukan query (kata kunci) pada data yang ingin dicari, dan menentukan batas data yang ingin diperoleh. Setelah melalui proses-proses yang telah dilakukan didapat sebuah dataset dengan jumlah penarikan sebanyak 2,005 data seperti pada gambar 2 Pada hasil crawling data masih berupa data mentah dari Twitter yang di mana masih banyak ditemukan variabel-variabel data yang tidak berguna untuk analisis sentimen atau isi teks pada kolom “full_text” yang masih banyak menggunakan huruf lowercase atau uppercase, masih ada simbol-simbol yang tidak penting dan tautan URL yang tidak berguna dalam proses analisis sentimen nanti, sehingga masih perlu dinormalisasikan agar bisa dilakukan analisis sentimen untuk mendapatkan variabel teks yang berguna atau isi teks yang bersih dari data.

3. Hasil Preprocessing

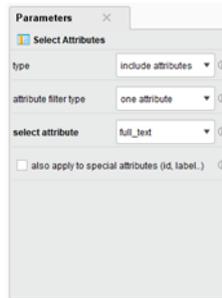
Preprocessing dilakukan sebelum melakukan pemberian label sentimen terhadap data. Pada tahap preprocessing data pada penelitian ini untuk memastikan data yang akan diolah adalah data yang baik atau bersih dengan cara melakukan seleksi data, pembersihan data dari isi-isi yang tidak berguna di dalamnya, terakhir menghapus data yang missing value dan menghapus data yang duplikat. Berikut adalah tahap preprocessing data yang dilakukan pada penelitian ini:

1) Seleksi Data

Pada tahap ini dilakukan untuk menyeleksi data yang akan digunakan, proses ini dilakukan karena hasil dari crawling data pada Twitter diperoleh beberapa banyak kolom seperti “conversation_id_str”, “created_at”, “favorite_count”, “image_url” dan masih banyak lagi kolom data yang tidak diperlukan dalam proses pengolahan data, hal ini perlu dilakukan agar proses pengolahan data menjadi terstruktur dan juga proses pengolahan data menjadi tidak terganggu karena banyaknya kolom data yang tidak diperlukan ini. Pada penelitian ini kolom data yang akan digunakan hanyalah kolom data yang berisikan data teks saja, yaitu kolom “full_text”. Berikut adalah gambar proses yang dilakukan untuk seleksi data:



Gambar 3. Proses Data Selection



Gambar 4. Parameter Pada Operator Select Attributes

Pada Gambar 4 dilihat pada proses seleksi data pada Rapid Miner. Pada proses ini dilakukan pemanggilan data hasil crawling dengan menggunakan operator “Read CSV” karena data yang kita simpan berformat file Comma Separated Values (CSV). Kemudian menggunakan operator “Select Attributes” untuk memilih kolom atribut apa saja yang ingin digunakan. Pada pengaturan operator “Select Attributes” untuk menyeleksi data bisa dilihat pada parameter operator “Select Attributes” pada Gambar 4.4, di sini penulis hanya memilih satu kolom atribut yaitu “full_text”. Berikut adalah hasil yang diperoleh dari proses seleksi data:

Row No.	conversatio...	created_at	favorite_cos...	full_text	id_str	image_url	in_reply_to_...	lang	location
1	1812125985...	Sat Jul 13 14...	0	@bahaspem...	1812130783...	?	bahaspemal...	in	?
2	1812123149...	Sat Jul 13 13...	5	Kalo ngomon...	1812123149...	?	?	in	written by a ti
3	1812099272...	Sat Jul 13 12...	0	PSSI tengah ...	1812099272...	https://pos.twi...	?	in	?
4	1812072413...	Sat Jul 13 10...	0	Media Vietna...	1812072413...	?	?	in	Kota Suraba
5	1811033547...	Sat Jul 13 08...	0	@dzuru Sek...	1812041914...	?	dzuru	in	Pulau Rote
6	1812000260...	Sat Jul 13 05...	1	Mengenal So...	1812000260...	?	?	in	Indonesia
7	1811793461...	Fri Jul 12 16...	3	Usai Naturali...	1811793461...	https://pos.twi...	?	in	Pulo Gadung
8	1811688601...	Fri Jul 12 13...	0	@TMIdn_new...	1811758278...	?	TMIdn_news	in	X land
9	1811670022...	Fri Jul 12 12...	0	@thejurgenh...	1811743505...	?	thejurgenholic	in	?
10	1811689281...	Fri Jul 12 09...	0	Kisah Pemat...	1811689281...	?	?	in	Malang
11	1811688601...	Fri Jul 12 09...	193	Akan segera j...	1811688601...	https://pos.twi...	?	in	Indonesia
12	1811673161...	Fri Jul 12 08...	0	Striker Natura...	1811673161...	https://pos.twi...	?	in	Pulo Gadung
13	1811659147...	Fri Jul 12 07...	1	Mempora Dito...	1811659147...	?	?	in	Jakarta
14	1811658395...	Fri Jul 12 07...	0	Mempora Dito...	1811658395...	?	?	in	Bandung, Jal
15	1811643401...	Fri Jul 12 06...	0	Manajer Timn...	1811643401...	?	?	in	JAKARTA

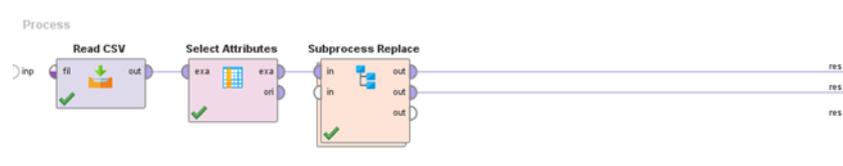
Gambar 5. Sebelum Proses Seleksi Data

Row No.	full_text
1	@bahaspemainbota @PSSI @ericthohir @AyaSimulingga telong ditelusuri garis keturunan Daan Rots. Siapa tahu yang bersangkutan memang punya da...
2	Kalo ngomongin si mafia emang dia ada agenda buat mecah belah fans. Bikin fans benci sama STY sama pemain diaspora sama pemain timnas. Kadha...
3	PSSI tengah mempersiapkan proses naturalisasi dua calon pemain Timnas Putri Indonesia Estella Loupaty dan Noa Leatomu. Exco PSSI Vvin Cahyani ...
4	Media Vietnam Soroti Naturalisasi di Timnas Putri Indonesia Ada Apa? https://t.co/IdyEIC0HRR
5	@dzuru Sekarang yg jadi pertanyaan pemain Indonesia mana yang sekelas UCL kalau lu maunya asnawi diganti pemain UCL? Semua pemain timnas it...
6	Mengenal Sosok Mauresmo Hinokle Calon Gelandang Naturalisasi Timnas Indonesia U-19 Andalan Indra Sjati Insting Cetak Gol Tinggi https://t.co/Qv3n6...
7	Usai Naturalisasi Striker Asai Klub Dordrecht U21 Blak-blakan Soal Ambisinya di Timnas Indonesia Jeens Raven. Target Pribadi Saya Tampai... Sebelum ...
8	@TMIdn_news @TimnasIndonesia gak skalian ketum @PSSI pake naturalisasi ??
9	@thejurgenholic gen kata pak Budi di pos ronda tadi Van Dijk mau proses naturalisasi ke Timnas Indonesia soalnya emaknya dari Padang
10	Kisah Pemain Naturalisasi Timnas U-19 Jens Raven Kenal Indonesia Lewat Neneknya - https://t.co/YOcdYMR2M0
11	Akan segera jalani proses Naturalisasi ketiga pemain ini cukup jelas akan menjadi aset dan proyeksi jangka panjang @TimnasIndonesia kedepan! 4YnB...
12	Striker Naturalisasi Timnas Indonesia Bikin Pengumuman Penting Jelang Dimulainya Liga 1 2024/2025. Striker veteran Ilya Spasojevic resmi meninggal...
13	Mempora Dito Bicara soal 3 Calon Pemain Naturalisasi Timnas Indonesia U-19 https://t.co/7Stu44Nc
14	Mempora Dito Bicara soal 3 Calon Pemain Naturalisasi Timnas Indonesia U-19 https://t.co/7Sha5u9y0d9
15	Manajer Timnas U-20 Indonesia Ahmad Zaki Iskandar angkat suara soal data pemain naturalisasi saat ini. Pihaknya mengaku sudah mengantongi banya...

Gambar 6. Setelah Proses Seleksi Data

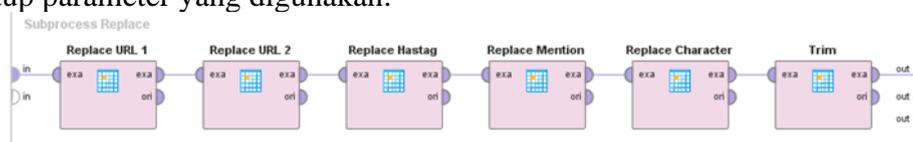
2) Cleaning

Pada tahapan ini merupakan proses pembersihan pada teks data, proses cleaning ini akan dilakukan pembersihan pada isi tweet yang mengandung isi-isi yang tidak berguna di dalamnya seperti simbol- simbol, karakter atau angka yang tidak berguna di dalamnya ketika melakukan proses pengolahan data nanti. Berikut adalah proses yang dilakukan pada Rapid Miner:



Gambar 7. Proses Cleaning Text Data

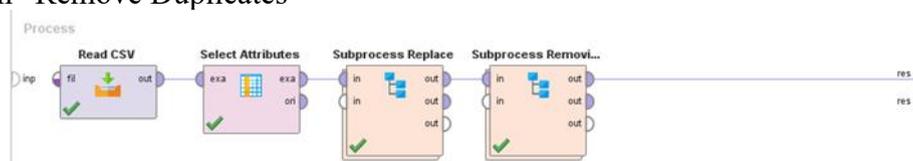
Pada Gambar 7 terdapat sebuah operator subprocess, operator tersebut digunakan untuk menampung proses-proses yang terlalu panjang atau terlalu banyak ke dalam satu subprocess menjadi satu bagian. Di dalam operator subprocess terdapat operator-operator untuk proses cleaning, di antaranya adalah “Replace URL 1” yang digunakan untuk menghapus link URL yang setelahnya terdapat teks lagi setelah link URL, kedua ada “Replace URL 2” yang digunakan untuk menghapus link URL yang setelahnya tidak terdapat sebuah teks lagi, yang ketiga ada “Replace Hastag” yang digunakan untuk menghapus hastag-hastag yang ada di dalam teks, yang keempat “Replace Mention” yaitu untuk menghapus tag mention username yang ada dalam teks, yang kelima ada “Replace Character” yang digunakan untuk menghapus karakter, simbol-simbol atau angka yang tidak penting digunakan dalam pemrosesan pengolahan data, dan yang terakhir ada operator “Trim” yang digunakan untuk menghapus spasi di awal atau di akhir pada string dalam data teks. Berikut adalah gambar proses-proses cleaning yang ada di dalam operator subprocess beserta setup parameter yang digunakan:



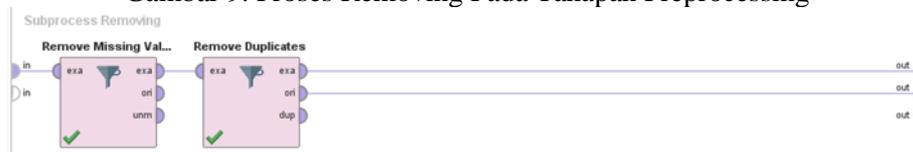
Gambar 8. Proses Cleaning di Dalam Operator Subprocess

3) Remove Duplicate & Missing Values

Tahapan ini merupakan tahapan untuk menghapus data yang duplikat dan missing values setelah dilakukannya proses cleaning. Pada tahapan ini digunakan dua operator untuk proses penghilangan data yang duplikat dan missing values, yaitu operator “Remove Missing Value” dan “Remove Duplicates”

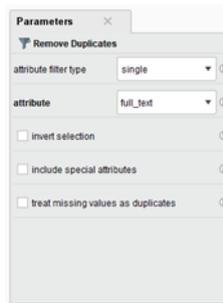


Gambar 9. Proses Removing Pada Tahapan Preprocessing



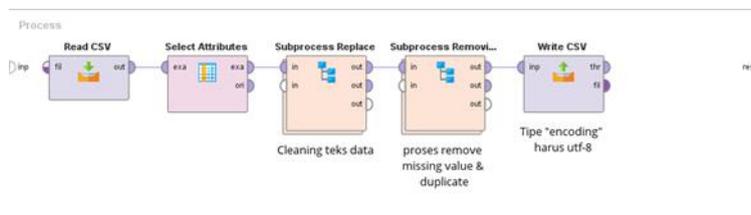
Gambar 10. Operator Remove Missing Values & Remove Duplicates pada subprocess

Sesuai pada nama operator yang digunakan, operator “Missing Values” digunakan untuk menghapus data yang missing atau kosong. Pada RapidMiner data yang error sudah secara otomatis digantikan menjadi data yang missing value, maka dari itu penggunaan operator ini bukan hanya menghilangkan data yang kosong, tetapi menghilangkan data yang error juga. Operator terakhir yang digunakan untuk preprocessing data adalah “Remove Duplicates”. Operator ini digunakan untuk menghilangkan data yang memiliki isi yang sama pada data-data yang lain, sehingga hal ini dapat menyebabkan keseimbangan yang tidak merata pada bobot data. Berikut adalah setup parameter yang dilakukan pada operator “Remove Duplicates”:



Gambar 11. Setup Parameter Pada Operator “Remove Duplicates”

Setelah melakukan hal itu semua, terakhir data yang telah di preprocessing kemudian disimpan ke dalam file dengan format *csv. dengan tipe encoding “utf-8” untuk dilanjutkannya proses labelisasi sentimen dengan lexicon based berdasarkan kamus Bahasa Indonesia. Berikut adalah gambar proses lengkapnya dan hasil yang diperoleh pada tahap preprocessing:



Gambar 12. Data Preprocessing

Row No.	full_text
1	tolong ditelusuri ganis keturunan Daan Rots. Siapa tahu yang bersangkutan memang punya darah Indonesia dan mau di naturalisasi untuk memperfajam...
2	Kalo ngomongin si mafia emang dia ada agenda buat mecah belah fans. Bikin fans bendi sama STY sama pemain diaspora sama pemain timnas. Kadn...
3	PSSI tengah mempersiapkan proses naturalisasi dua calon pemain Timnas Putri Indonesia Estella Loupaty dan Noa Leatoma. Exco PSSI Wlin Cahyani ...
4	Media Vietnam Soroti Naturalisasi di Timnas Putri Indonesia Ada Apa
5	Sekarang yg jadi pertanyaan pemain indonesia mana yang sekelas UCL kalau tu maunya asnawi diganti pemain UCL. Semua pemain timnas itu setara s...
6	Mengenal Sosok Mauresmo Hinoka Calon Gelandang Naturalisasi Timnas Indonesia U19 Andalan Indra Sjathu Insang Cetak Gol Tinggi
7	Usai Naturalisasi Stiker Asal Klub Dordrecht U21 Blakblakan Soal Ambisinya di Timnas Indonesia Jeens Raven Target Pribadi Saya Tampi... Sebelum b...
8	gak skalian ketum pake naturalisasi
9	gen kata pak Budi di pos ronda tadi Van Dijk mau proses naturalisasi ke timnas Indonesia soalnya emaknya dari Padang
10	Kisah Pemain Naturalisasi Timnas U19 Jens Raven Kenal Indonesia Lewat Neneknya
11	Akan segera jalani proses Naturalisasi ketiga pemain ini cukup jelas akan menjadi aset dan proyeksi jangka panjang kedepan @Yl@6Yh@
12	Stiker Naturalisasi Timnas Indonesia Bikin Pengumuman Penting Jelang Dimulainya Liga 1 2024/2025 Stiker veteran Ijja Spasojevic resmi meninggalk...
13	Mengora Dito Bicara soal 3 Calon Pemain Naturalisasi Timnas Indonesia U19
14	Manajer Timnas U20 Indonesia Ahmad Zaki Iskandar angkat suara soal data pemain naturalisasi saat ini. Pihaknya mengaku sudah mengantongi banya...
15	Pujan Khusus Pelatih Legendaris Liga Indonesia pada Thom Haje Takjub dengan Nasionalisme The Professor Ada Rasa Bangsa... Pelatih legendaris L...

Gambar 13. Hasil Preprocessing Data

Labelisasi Sentimen

Setelah dataset mentah yang diperoleh dari hasil crawling pada Twitter telah dilakukan proses pembersihan data, maka selanjutnya akan dilakukan proses labelisasi sentimen atau penentuan attribute class pada data. Pada penelitian ini akan dilakukan labelisasi sentimen pada data menggunakan teknik Lexicon Based berdasarkan kamus Bahasa Indonesia. Lexicon-based sentiment labeling bekerja dengan menggunakan kamus atau daftar kata (lexicon) yang telah diberi label sentimen, seperti positif, negatif, atau netral. Setiap kata dalam teks yang akan dianalisis dan dicocokkan dengan kata-kata dalam lexicon ini. Skor sentimen dari teks dihitung berdasarkan akumulasi skor sentimen dari masing-masing kata yang ditemukan dalam teks. Misalnya, jika sebuah teks mengandung lebih banyak kata-kata dengan label sentimen positif dibandingkan negatif, maka teks tersebut akan diberi label sentimen positif. Pada penelitian ini tidak memasukan label netral karena tidak memberikan informasi apapun. Pada proses ini penulis akan menggunakan bantuan Google Colaboratory sebagai modul untuk menjalankan kode program lexicon base ini. Proses dilakukan kepada pemanggilan data terlebih dahulu yang di mana data yang dipanggil merupakan data yang sudah dilakukan proses pembersihan data pada proses tahapan preprocessing. Kemudian data akan ditentukan sentimennya berdasarkan kamus positif, negatif dan netral yang telah dibuat sebelumnya. Pada penelitian ini tidak ditentukan label sentimen netral, maka dari itu

data yang berlabel sentimen netral akan dianggap kosong sehingga dihasilkan dataset dengan sentimen labelnya sebagai berikut:

Row No.	full_text	sentiment
1	tolong dilelusu garis keturunan Daan Rots. Siapa tahu yang bersangkutan memang punya darah Indonesia dan mau di naturalisasi unt...	?
2	Kalo ngomongin si mafia emang dia ada agenda buat mecah belah fans. Bikin fans beno sama STY sama pemain diaspora sama pem...	negative
3	PSSI tengah mempersiapkan proses naturalisasi dua calon pemain Timnas Putri Indonesia Estella Loupaty dan Noa Leatomu. Exco PS...	positive
4	Media Vietnam Soroti Naturalisasi di Timnas Putri Indonesia Ada Apa	?
5	Sekarang yg jadi pertanyaan pemain Indonesia mana yang sekelas UCL, kalau lu maunya asnawi diganti pemain UCL, Semua pemain b...	negative
6	Mengenal Sosok Mauresmo Hinoke Calon Gelandang Naturalisasi Timnas Indonesia U19 Andalan Indra Sjafi Insting Cetak Gol Tinggi	?
7	Usai Naturalisasi Striker Asal Klub Dordrecht U21 Blakblakan Soal Ambisinya di Timnas Indonesia Jeens Raven Target Pribadi Saya Ta...	negative
8	gak skalian ketum pake naturalisasi	?
9	gen kata pak Budi di pos ronda tadi Van Dijk mau proses naturalisasi ke Timnas Indonesia soalnya emaknya dari Padang	negative
10	Kisah Pemain Naturalisasi Timnas U19 Jens Raven Kenal Indonesia Lewat Neneknya	?
11	Akan segera jatani proses Naturalisasi ketiga pemain ini cukup jelas akan menjadi aset dan proyeksi jangka panjang kedepan A'ā, Jé'ā...	positive
12	Striker Naturalisasi Timnas Indonesia Bikin Pengumuman Penting Jelang Dimulainya Liga 1 2024/2025 Striker veteran Raja Spasojevic re...	?
13	Mengora Dito Bicara soal 3 Calon Pemain Naturalisasi Timnas Indonesia U19	negative
14	Manajer Timnas U20 Indonesia Ahmad Zaki Iskandar angkat suara soal data pemain naturalisasi saat ini. Pihaknya mengaku sudah me...	negative
15	Pujian Khusus Pelatih Legendaris Liga Indonesia pada Thom Hays Takjub dengan Nasionalisme The Professor Ada Rasa Bangga... Pel...	negative

Gambar 14. Hasil Labelisasi Data

Pembobotan Kata

Pada tahapan ini akan dilakukan pembobotan kata pada data teks, pembobotan kata ini dilakukan untuk memberikan nilai atau bobot pada setiap kata dalam teks berdasarkan kontribusinya dalam menentukan sentimen keseluruhan. Metode ini membantu dalam mengidentifikasi kata-kata yang memiliki pengaruh besar terhadap sentimen positif atau negatif. Pada proses ini penerapannya menggunakan data yang telah dipreprocessing, akan tetapi sentimen labelnya tidak diikutsertakan, karena label sentimen tidak mempengaruhi hasil akhir pada pembobotan. Sebelum melakukan pembobotan kata dilakukan text preprocessing terlebih dahulu. Berikut adalah proses dari text preprocessing yang dilakukan pada penelitian ini:

a. Tokenizing

Proses ini merupakan pemrosesan bahasa alami atau NLP (Natural Language Processing) yang memecah teks menjadi unit-unit kecil yang disebut “token”. Proses ini penting dilakukan karena memungkinkan teks mentah diubah menjadi format yang lebih mudah diolah, seperti menghitung frekuensi kata atau membobotkan kata. Pada proses ini dilakukan menggunakan operator “Tokenize” pada tools RapidMiner, berikut adalah hasilnya:

b. Case Folding

Proses ini dilakukan untuk mengubah semua huruf dalam teks menjadi kecil, menghapus perbedaan kata antara huruf kapital dan huruf kecil. Tujuannya untuk menyederhanakan analisis teks dengan memastikan bahwa kata-kata yang sama diidentifikasi secara konsisten. Misalnya kita mempunyai kata yang salah satu hurufnya memiliki huruf kapital seperti “Naturalisasi” dan memiliki kata yang lainnya seperti “naturalisasi”, kedua kata itu akan dianggap berbeda pada proses tokenizing. Oleh karena itu dibutuhkan proses Case Folding untuk penyederhanaan teks untuk menghapus perbedaan teks antar huruf. Pada proses ini menggunakan operator “Transform Cases” pada tools RapidMiner.

c. Stopword Removal

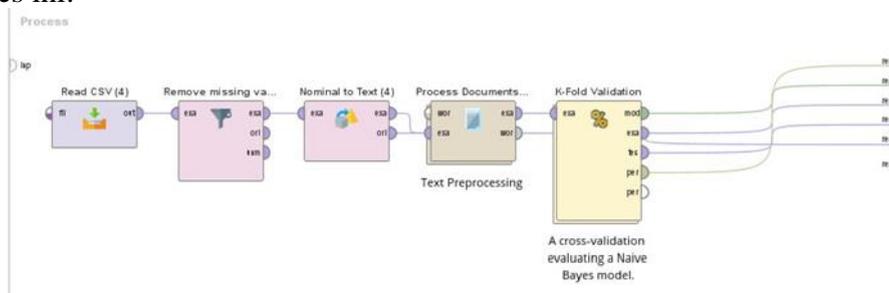
Tahapan ini merupakan proses penghapusan kata-kata umum yang sering muncul seperti kata “dan”, “atau”, “adalah” dan lain-lain yang di mana kata-kata tersebut tidak memberikan informasi apapun terkait analisis sentimen. Proses ini juga untuk mengurangi jumlah kata yang harus dianalisis dan mengurangi data noise yang ada, sehingga memfokuskan kepada kata-kata yang lebih signifikan dan relevan. Pada proses Stopword Removal ini kita tidak bisa menggunakan operator default dari RapidMiner, karena tools tersebut belum menyediakan operator Filter Stopword berbahasa Indonesia. Maka dari itu penulis menggunakan operator Filter Stopword (Dictionary) dengan file Indonesian Stoplist pada tautan <https://www.kaggle.com/datasets/oswinrh/indonesian-stoplist>.

d. Filter Tokens

Pada penelitian ini digunakan sebuah operator lainnya dalam proses text preprocessing ini yaitu Filter Tokens (by length). Pada penggunaan operator ini digunakan untuk menghapus kata-kata yang telah ditentukan jumlah minimum karakternya dan jumlah maksimum karakternya. Proses ini dilakukan pada penelitian ini tujuannya untuk menghilangkan kata-kata singkatan “yg”, “dri”, “sbg” dan kata-kata lainnya yang kemungkinan singkatan katanya memiliki karakter berlebih yang telah ditentukan pada operator tersebut. Pada kasus ini minimal karakter yang akan muncul yaitu 4 (empat) karakter dan maksimal 25 (dua puluh lima) karakter. Pada proses-proses di atas dikumpulkan dalam subprocess pada operator Process Documents from Data, karena pada operator inilah penulis akan menentukan metode pembobotan kata yang akan digunakan, yaitu TF-IDF. Berikut adalah susunan operator yang digunakan untuk pembobotan kata dengan TF-IDF.

Pengujian Dataset Dengan Algoritma Naive Bayes

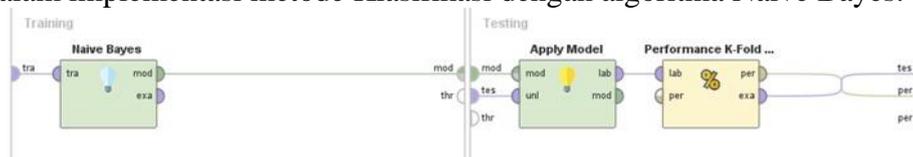
Pada tahapan ini merupakan tahapan inti dari penelitian ini, yaitu tahapan deployment atau proses implementasi metode Naive Bayes Classification, di mana proses tersebut dapat digunakan untuk mengambil keputusan atau memberikan wawasan dari data yang telah dianalisis. Tujuan dari proses ini yaitu membuat model Machine Learning untuk menentukan suatu sentimen apakah positif atau negatif. Pada proses ini juga akan menghitung tingkat accuracy, precision dan recall untuk penilaian pengujian pada model yang dibangun menggunakan K- Fold Cross Validation. Berikut adalah susunan operator yang digunakan dalam proses ini:



Gambar 15. Susunan Operator untuk analisis sentimen menggunakan Naive Bayes Classifier

Pada gambar 15 di atas diperlihatkan sebuah penggunaan operator “Filter Examples”, operator tersebut digunakan karena pada proses labelisasi sentimen menggunakan Lexicon Based berdasarkan kamus Bahasa Indonesia secara otomatis didapatkan sebuah label sentimen “netral”, akan tetapi penelitian ini tidak menginginkan sebuah sentimen berlabel “netral”, maka dari itu sentimen yang berlabel “netral” dikosongkan dan pada RapidMiner dianggap sebagai data yang missing value. Penggunaan operator tersebut dilakukan untuk menghilangkan data yang missing value dengan cara memfilter data yang tidak kosong. Dengan menggunakan operator “Nominal to Text”, setiap nilai kategori dalam atribut nominal akan dikonversi menjadi string atau teks. Ini memungkinkan pengguna untuk mengaplikasikan teknik-teknik pemrosesan teks pada data yang awalnya berbentuk kategori. Kemudian operator “Process Document from Data”, operator tersebut digunakan untuk melakukan pembobotan kata seperti proses yang telah dijelaskan pada sub-bab Pembobotan Kata. Selanjutnya terdapat sebuah operator baru yang muncul yaitu “Cross Validation”, operator tersebut adalah operator khusus untuk menentukan nilai akurasi hasil uji dengan teknik K-Fold Cross Validation. Untuk menentukan nilai “K”-nya (atau kelipatan) dalam penilaian evaluasi terdapat pada parameter operator tersebut, pada penelitian ini nilai “K” yang ditentukan yaitu 10 lipatan. Pada parameter operator tersebut kita dapat memilih tipe sampling yang ingin digunakan dalam pengujian evaluasi K-Fold Cross Validation, pada penelitian ini tipe sampling yang digunakan yaitu “automatic” yang telah disediakan oleh

tools RapidMiner. Operator Cross Validation juga layaknya seperti operator Subprocess, yaitu operator yang dapat menyimpan operator-operator yang lain. Akan tetapi operator ini hanya dapat menyimpan operator modeling, yaitu operator yang dapat menjalankan Data Training dan Data Testing. Operator ini khusus untuk digunakan dalam melakukan modeling sekaligus sebagai modul untuk penilaian hasil pengujian model. Berikut adalah susunan operator dalam implementasi metode Klasifikasi dengan algoritma Naive Bayes:



Gambar 16. Susunan modeling Naive Bayes Classification di dalam operator Cross Validation

Pada proses-proses yang telah dilakukan pada proses implementasi metode Naive Bayes Classification di atas, telah dibuat sebuah model untuk menganalisis sentimen. Pada hasil pembuatan model tersebut akan dilakukan pengujian untuk mengetahui seberapa efektif model tersebut melakukan prediksi analisis sentimen. Maka dari itu pengujian akan dilakukan menggunakan K-Fold Cross Validation dengan lipatan “K”-nya sebanyak 10 kali untuk mengetahui tingkat accuracy, precision dan recall yang didapatkan.

a. Accuracy

Accuracy merupakan tingkat kedekatan antara hasil prediksi dengan hasil fakta, prediksi adalah tingkat ketepatan antara informasi yang diminta oleh user dengan jawaban yang diberikan oleh sistem. Setelah data sudah terbagi kemudian ditentukan nilai prediksinya. Berikut adalah hasil pengujian prediksi pada data:

accuracy: 72.23% +/- 4.29% (micro average: 72.23%)

	true negative	true positive	class precision
pred. negative	707	152	82.31%
pred. positive	141	55	28.06%
class recall	83.37%	26.57%	

Gambar 17. Hasil Evaluasi Model menggunakan K-Fold Cross Validation

Prediksi dari data sentimen negatif yang terdeteksi negatif adalah 707 dan yang salah adalah 152, sedangkan data sentimen positif yang terdeteksi positif adalah 55 dan yang salah adalah 141. Hasil rata-rata accuracy yang diperoleh dari pengujian model Naive Bayes Classification adalah 72.23%. Adapun hasil accuracy diperoleh dengan cara menjumlah data True Positive (TP) ditambah dengan True Negative (TN) dibagi dengan total keseluruhan data. Berikut cara perhitungan manual dari accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Accuracy = \frac{707 + 55}{707 + 55 + 152 + 141} \times 100\%$$

$$Accuracy = \frac{1055}{1055} \times 100\%$$

$$Accuracy = 72.23\%$$

b. Precision

Precision merupakan pengukuran seberapa ketepatan informasi yang diberikan oleh sistem. Pada pengujian ini akan dihitung presisi terhadap data testing. Adapun hasil precision diperoleh dengan cara data True Positive (TP) dibagi dengan total jumlah data True Positive (TP) ditambah dengan False Positive (FP). Berikut cara perhitungan manual dari precision:

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Precision = \frac{707}{707 + 152} \times 100\%$$

$$Precision = \frac{TP}{859} \times 100\%$$

$$Precision = 82.31\%$$

c. Recall

Recall merupakan pengukuran tingkat keberhasilan penagmbilan data yang relevan. Pada pengujian ini akan dihitung nilai recall terhadap data testing. Adapun hasil recall diperoleh dengan cara data True Positive (TP) dibagi dengan total jumlah data True Positive (TP) ditambah dengan False Negative (FN). Berikut cara perhitungan manual dari recall:

$$Precision = \frac{TP}{TP + FN} \times 100\%$$

$$Precision = \frac{707}{707 + 152} \times 100\%$$

$$Precision = \frac{707}{859} \times 100\%$$

$$Precision = 82.31\%$$

4. KESIMPULAN

Pada penelitian ini berhasil melakukan analisis sentimen terhadap opini masyarakat Indonesia di Twitter mengenai naturalisasi pemain asing ke Timnas Indonesia menggunakan algoritma Naive Bayes. Dari hasil analisis, ditemukan bahwa sentimen negatif lebih dominan dibandingkan dengan sentimen positif. Hal ini menunjukkan bahwa mayoritas masyarakat Indonesia kemungkinan memiliki pandangan yang kurang mendukung terhadap kebijakan naturalisasi pemain asing untuk tim nasional Indonesia. Dengan menggunakan algoritma Naive Bayes, penelitian ini dapat mengidentifikasi dan mengklasifikasikan sentimen dengan akurasi yang cukup baik dengan hasil evaluasi model pada nilai accuracy sebesar 72.23%, precision sebesar 82.31% dan recall sebesar 83.37%. Hasil dari penelitian ini diharapkan dapat memberikan wawasan penting bagi para pembuat kebijakan dan peneliti selanjutnya. Bagi pembuat kebijakan, hasil analisis ini dapat digunakan untuk mengevaluasi dan menyesuaikan kebijakan naturalisasi agar lebih sesuai dengan harapan dan pandangan masyarakat, sehingga dapat meningkatkan dukungan publik. Bagi peneliti, analisis ini membuka peluang untuk mengeksplorasi lebih lanjut faktor-faktor yang mempengaruhi sentimen publik serta mengembangkan metode analisis yang lebih canggih untuk memprediksi dan memahami opini masyarakat di masa mendatang.

5. DAFTAR PUSTAKA

- A. E. Pratama, A. Ariesta, and G. Gata, "Analisis Sentimen Masyarakat Terhadap Tim Nasional Indonesia Pada Piala AFF 2020 Menggunakan Algoritma K-Nearest Neighbors," *J. TICOM Technol. Inf. Commun.*, vol. 10, no. 3, pp. 187–196, 2022, [Online]. Available: <https://jurnal-ticom.jakarta.aptikom.org/index.php/Ticom/article/view/33/47>.
- A. F. Watratan, A. P. B., and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 di Indonesia," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- A. Turmudi Zy, A. Nugroho, A. Rivaldi, and I. Afriantoro, "Analisis Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes," *J. Teknol. Inform. dan Komput.*, vol. 8, no. 2, pp. 202–213, 2022, doi: 10.37012/jtik.v8i2.1240.
- A. Y. Luntungan, "Naturalisasi Warganegara Asing Menjadi Warganegara Indonesia Menurut Undang-undang Nomor 12 Tahun 2006 Tentang Kewarganegaraan," *Lex Soc.*, vol. I, no. 5, p. 41, 2013, [Online]. Available: file:///C:/Users/user/Downloads/jurnal_naturalisasi_kewarganegaraan.pdf.
- B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online Dengan Menggunakan Pembobotan TF-IDF dan Cosine Similarity," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 306–312, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- D. Novianti, "Implementasi Algoritma Naive Bayes Pada Data Set Hepatitis Menggunakan Rapid

- Miner,” *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, 2019, doi: 10.31294/p.v21i1.4979.
- E. Widodo and P. Sa’adah, “Klasifikasi Data Penjualan Alat Tulis Kantor (ATK) Terlaris Untuk Optimasi Strategi Pemasaran Di Toko Citramedia Menggunakan Metode Naive Bayes,” vol. 10, pp. 1–19, 2019.
- F. Amaliah and I. K. Dwi Nuryana, “Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter,” *J. Informatics Comput. Sci.*, vol. 3, no. 03, pp. 384–393, 2022, doi: 10.26740/jinacs.v3n03.p384-393.
- Fuad Amirullah, Syariful Alam, and M. Imam Sulisty S, “Analisis Sentimen Terhadap Kinerja KPU Menjelang Pemilu 2024 Berdasarkan Opini Twitter Menggunakan Naive Bayes,” *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 69–76, 2023, doi: 10.55123/storage.v2i3.2293.
- I. P. W. K. Gumi, Hartatik, and A. Syafrianto, “Perbandingan Algoritma Naive Bayes dan Decision Tree Pada Sentimen Analisis,”
- I. Wickramasinghe and H. Kalutarage, “Naive Bayes: Applications, Variations And Vulnerabilities: A Review Of Literature With Code Snippets For Implementation,” *Soft Comput.*, vol. 25, no. 3, pp. 2277–2293, 2021, doi: 10.1007/s00500-020-05297-6.
- Indones. J. Comput. Sci. Res.*, vol. 1, no. 2, pp. 1–15, 2022, doi: 10.59095/ijcsr.v1i2.11.
- M. A. Kurniawan, A. S. Yuniarti, and S. Makin, “Analisis Sentimen AFC U23 Asian Cup 2024 menggunakan Algoritma Naive Bayes,” *Insa. Pembang. Sist. Inf. dan Komput.*, vol. 12, no. 1, pp. 25–28, 2024, doi: 10.58217/ipsikom.v12i1.275.
- M. K. Khoirul Insan, U. Hayati, and O. Nurdiawan, “Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 478–483, 2023, doi: 10.36040/jati.v7i1.6373.
- N. Habibah, E. Budianita, M. Fikry, and I. Iskandar, “Analisis Sentimen Mengenai Penggunaan E-Wallet Pada Google Play Menggunakan Lexicon Based dan K-Nearest Neighbor,” *J. Ris. Komputer*, vol. 10, no. 1, pp. 2407–389, 2023, doi: 10.30865/jurikom.v10i1.5429.
- Optimization (PSO),” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 1806, 2022, doi: 10.30865/mib.v6i4.4119.
- R. Sulastiyono, A. Setiawan, and S. Nugroho, “Sentimen Analisis Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 Menggunakan Metode Naive Bayes,” *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1387–1394, 2023, doi: 10.47065/josh.v4i4.3737.
- R. Tuntun, K. Kusriani, and K. Kusnawi, “Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K- Fold Cross Validation,” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2111, 2022, doi: 10.30865/mib.v6i4.4681.
- S. Andayani, “Pembentukan Cluster Dalam Knowledge Discovery in Database dengan Algoritma K-Means,” 2007.
- S. H. W. Putra and D. Febriawan, “Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naive Bayes pada Google Play Store,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, pp. 1962–1971, 2024, doi: 10.30865/klik.v4i4.1600.
- S. Hendrian, “Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan,” *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- T. Astuti and Y. Astuti, “Analisis Sentimen Review Produk Skincare Dengan Naive Bayes Classifier Berbasis Particle Swarm
- V. Vibrianti, E. Wahyudin, K. Kaslani, D. Pratama, and G. Dwilestari, “Klasifikasi Barang Produksi Pada TNT. Gitar Workshop Dengan Metode Naive Bayes Menggunakan Rapid Miner,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1432–1438, 2024, doi: 10.36040/jati.v8i2.8966.