

ANALISIS SENTIMEN TERHADAP SEPAK BOLA PADA MEDIA TWITTER MENGGUNAKAN METODE NAÏVE BAYES

Muhammad Iqbal Al Anbari
Universitas Ahmad Dahlan
Jl. Sanggrahan Giwangan Umbulharjo Kota Yogyakarta
Muhammad1900018124@webmail.uad.ac.id



kcc draft

Abstract

Football is a sport that is of interest to many groups of people, especially teenagers and children. Ball games, or usually known as soccer matches are leads to a competition between two teams consisting of 11 players. Even at this moment most people know football players better than government figures. Part of They even dream of becoming soccer players. This research aims to: Find out sentiment level of satisfaction with the Indonesian national football team using the Naïve Bayes method. And knowing the level of accuracy of the Naïve Bayes method. This research was carried out using the Naïve Bayes method to obtain positive, neutral and negative classification of Indonesian football comments or opinions on Twitter. There is There are eight stages in this research, namely collecting review data (crawling), data preprocessing, data labeling, word cloud labeling, dividing training data and test data, Naïve Bayes, Naïve labeling results Bayes and confusion matrices. Application of the Naïve Bayes algorithm to produce a classification between negative and positive classes. Then the accuracy calculation is carried out using confusion matrix to determine the accuracy of the labeling that has been done. The data used is data from crawling using Python for the period from 07 December 2021 to 01 January 2024 on the Twitter platform. Accuracy checking using Confusion Matrix with an accuracy of 85% by dividing training data and test data as many as 80:20, namely 883 training data and 221 test data from 1104 data respectively. Classification results produced 129 negative data, 43 neutral data, and 27 positive data.

Keywords: Sentiment Analysis, Twitter, Naive Bayes

Abstrak

Analisis sentimen adalah seperangkat metode, teknik dan alat yang bertujuan untuk mendeteksi atau mengekstraksi informasi yang subjektif seperti opini dari sebuah artikel. Analisis sentimen digunakan untuk mengekstrak informasi dari artikel dan mengkategorikannya menjadi sentimen tersebut ke bentuk sentimen positif, negatif, dan netral. Penelitian ini bertujuan untuk mengeksplorasi tingkat kepuasan terhadap timnas sepak bola Indonesia menggunakan metode Naïve Bayes. Tujuan lainnya adalah untuk menilai akurasi dari metode Naïve Bayes dalam menganalisis sentimen terhadap timnas sepak bola Indonesia. Penelitian ini menggunakan metode Naïve Bayes untuk mengklasifikasikan komentar atau pendapat mengenai sepak bola Indonesia di Twitter menjadi kategori positif, netral, atau negatif. Terdapat delapan tahap dalam penelitian ini, yaitu pengumpulan data ulasan (crawling), preprocessing data, pelabelan data, pelabelan, word cloud, pembagian data menjadi data latih dan data uji, penerapan Naïve Bayes, evaluasi hasil dari Naïve Bayes, dan penggunaan matriks konfusi. Algoritma Naïve Bayes diterapkan untuk menghasilkan klasifikasi antara kategori negatif dan positif. Selanjutnya, akurasi dari proses pelabelan dievaluasi menggunakan matriks konfusi untuk menilai ketepatan hasil klasifikasi tersebut.

Data yang digunakan dalam penelitian ini berasal dari hasil crawling menggunakan Python dari tanggal 07 Desember 2021 hingga 01 Januari 2024 di platform Twitter(X). Pada tahap pengujian labeling menghasilkan persentase sentiment netral 71%, positif 16% dan negatif 11%. Dengan pembagian data uji 221 dari 1104 data sedangkan data latih 883 dari 1104 data uji. Dihasilkan 793 sentimen netral, 186 positif dan 125 negatif. Tahap preprocessing menghasilkan akurasi 85% dengan hasil presisi negatif 100%, presisi netral 86%, presisi positif 69%, recall negatif 67%, recall netral 94% dan positif 62%. Analisis sentimen ini efektif dalam mengidentifikasi sentimen netral dan negatif, namun perlu peningkatan dalam mendeteksi sentimen positif.

Kata kunci: Analisis Sentimen, Twitter, Naïve Bayes

1. PENDAHULUAN

Sepak bola Indonesia adalah olahraga yang banyak diminati oleh sebagian besar masyarakat terutama remaja dan anak-anak. Bahkan pada saat ini, kebanyakan orang mengetahui lebih banyak mengenal tentang pemain sepak bola dibandingkan pejabat pemerintah. Beberapa orang bermimpi ingin menjadi pemain sepak bola. Banyak orang dewasa menghabiskan malamnya hanya dengan menonton pertandingan sepak bola di TV, bertemu teman di kafe, kedai kopi, atau bahkan pos ronda di sekitar rumah. Mayoritas penduduk Indonesia mengetahui olahraga Sepak bola Indonesia menduduki peringkat kedua dunia sebagai negara pecinta sepak bola, dengan 77% penduduk Indonesia tertarik dengan olahraga sepak bola [1].

Pada saat ini masyarakat Indonesia banyak memperbincangkan sepak bola di media sosial yaitu Twitter. Twitter adalah media sosial yang digunakan banyak orang. Salah satu tweet yang sedang ramai diperbincangkan adalah sepak bola, khususnya Timnas Indonesia, merupakan olahraga yang digemari masyarakat Indonesia. Pengguna Twitter akan memposting komentar dan opini, tentang sepak bola Timnas Indonesia melalui media sosial, termasuk Twitter [2]. Berbagai pendapat dan komentar telah diterima oleh para penggemar ketika tim nasional tampil baik, ada opini dan komentar gembira dan pujian, ketika tim nasional tampil buruk ada opini dan komentar kritis dan dalam beberapa kasus, bahkan pelecehan. Oleh karena itu, perlu adanya klasifikasi tweet agar pengguna dapat mengenali, melihat dan memberi nilai sebagai tweet yang positif atau negatif [1]. Analisis sentimen adalah seperangkat metode, teknik dan alat yang bertujuan untuk mendeteksi atau mengekstraksi informasi yang subjektif seperti opini dari sebuah artikel [3]. Dalam analisis sentimen minimal, terdapat tiga kategori untuk menentukan sentimen yaitu kategori sentimen positif, negatif, dan netral. Analisis sentimen digunakan untuk mengekstrak informasi dari artikel dan mengkategorikannya menjadi sentimen tersebut ke bentuk sentimen positif, negatif, dan netral [4]. Ada berbagai macam teknik untuk klasifikasi dokumen, seperti metode *Naïve Bayes Classifier*. Metode *Naïve Bayes* digunakan karena metode ini menawarkan teknik klasifikasi yang sederhana namun kuat [5].

Algoritma *Naïve Bayes* dapat memberikan manfaat dan sederhana, cepat, dan berakurasi tinggi, oleh karena itu dalam penelitian ini akan

dilakukan analisis sentimen sepak bola timnas Indonesia untuk menghasilkan klasifikasi dari tiga labeling yaitu positif, negatif, dan netral [6]. Untuk menguji akurasi dari klasifikasi tersebut menggunakan *confusion matrix* dari masing-masing label sehingga hasil dari analisis sentimen sepak bola Indonesia yang lebih akurat [7].

2. METODOLOGI PENELITIAN

2.1 Metode Pengumpulan data

Untuk memperoleh informasi yang diperlukan maka terdapat beberapa langkah untuk mengumpulkan data yaitu menggunakan metode pustaka (studi literatur). Metode pustaka atau studi literatur merupakan cara yang digunakan dalam melakukan penelitian dengan mencari sumber-sumber referensi terkait data mining dengan membaca buku, membaca jurnal, dan membaca paper mengenai penelitian-penelitian sebelumnya.

2.2 Software And Hardware

Dalam melakukan penelitian ini dibutuhkan alat dan bahan untuk membantu proses penelitian adalah sebagai berikut :

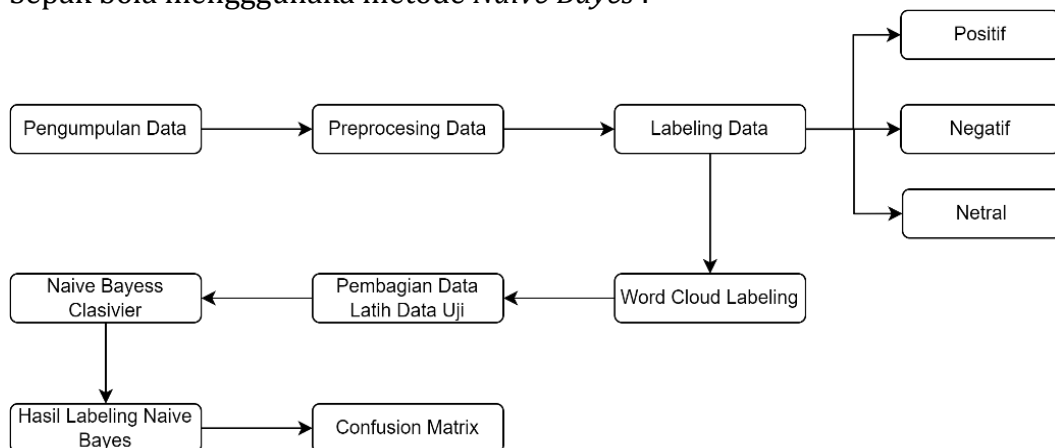
1. Microsoft windows version 22H2 (OS Build 19045.3324) Microsoft Comporation
2. Microsoft office 2010
3. Microsoft excel 2010
4. Web browser Google Chrome version 103.0.5060.66 Official Build (64-bit)
5. Web browser Microsoft Edge version 103.0.1264.44 Official build (64-bit)
6. Anaconda python
7. Jupyter notebook

Sedangkan perangkat kerasnya yang dibutuhkan yaitu :

1. Memory 4.00 GB (3.39 GB usable).
2. AMD A9-9425 REDEON R5, 5 COMPUTE CORES 2C+3G.

2.3 Tahap Penelitian

Berikut tahapan yang dilakukan dalam penelitian Analisis Sentimen Sepak bola menggunakan metode *Naïve Bayes* :



Gambar 2.1 Tahap Penelitian

Proses pengumpulan data ini dilakukan dengan *crawling* pada aplikasi tweet dengan menggunakan bahasa pemrograman *python*. Melakukan *pre-processing* pada data tweet, tahap *pre-processing* merupakan salah satu tahap proses yang harus dilakukan untuk memproses data pada prose mining diantaranya meliputi : *Cleaning, Case Folding, Tokenizing, Stemming, Stopword removal*. Melakukan *labeling* data agar untuk mengetahui hasil dari tahap sebelumnya yang akan dilakukan. Melakukan *Word Cloud* untuk mengetahui semakin besar ukuran kata-kata maka semakin sering kata tersebut muncul. *Naïve Bayes* untuk mendapatkan model atau memprediksi suatu kelas. *Confusion Matrix* alat yang digunakan untuk mempertimbangkan bentuk klasifikasi dalam memprediksi topic.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Data yang digunakan dalam penelitian ini “Analisis Sentimen sepak bola Indonesia Pada Media Twitter Menggunakan Metode *Naïve Bayes*” yaitu data yang dilakukan dengan menggunakan program python berupa cuitan-cuitan akun yang berhubungan sepak bola Indonesia.

3.2 Data Mining Tahapan dan Implementasi.

3.2.1 Pengumpulan Data

created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str	conversation_id_str	username
Thu Feb 01 21:34:34 +0000 2024	1753169996279386417	OFFICIAL toShayne Pattynama (25DF) resmi...	37	86	272	2347	in	811427632336146433	1753169996279386417	Indostransfer
Thu Feb 01 19:05:06 +0000 2024	1753132383430402078	RESMI: Pemain Timnas Indonesia, Elkan Bag...	8	30	128	1775	in	1316071146	1753132383430402078	FaktaSepakbola
Thu Feb 01 15:34:58 +0000 2024	1753079500081119319	Timnas Indonesia baru saja mengalahkan Vietnam...	55	152	325	3466	in	1316071146	1753079500081119319	FaktaSepakbola
Thu Feb 01 10:22:01 +0000 2024	1753000743395877047	SUDAH TAYANG!! "AMIN-KAN DARI JAUH" merupakan...	4	14	435	1776	in	1716995484961947649	1753000743395877047	UbahBareng
Thu Feb 01 07:22:19 +0000 2024	1752955522494747078	PP Muhammadiyah Puji Capaian Timnas	64	111	184	2014	in	1089010123	1752955522494747078	UpdateBolabola

Gambar 3. 1 Pengumpulan Data

Pada gambar 3.1 yaitu hasil crawling data menggunakan google Colaboraty yang didalamnya dimasukan atribut ‘created_at’, ‘id_str’, ‘full_text’, ‘quote_count’, ‘replay_count’, ‘retweet_count’, ‘favorite_count’, ‘lang’, ‘user_id-str’, ‘conversation_id_str’, ‘username’.

3.2.2 Pre-Processing

Preprocessing adalah proses dari menyiapkan data mentahan data yang terstruktur tahapan *preprocessing* dapat dilihat dibawah ini :

a. *Case Folding*


```

0  [ 'offici ', 'shayn ', 'pattynama ', 'df ' ...
1  [ 'resmi ', 'pemain ', 'timna ', 'indonesia...
2  [ 'timna ', 'indonesia ', 'baru ', 'saja ' ...
3  [ 'sudah ', 'tayang ', 'aminkan ', 'dari ' ...
4  [ 'pp ', 'muhammadiyah ', 'puji ', 'capaian...
5  [ 'timna ', 'indonesia ', 'main ', 'bagu ' ...
6  [ 'kita ', 'sudah ', 'pernah ', 'ada ', 'd...
7  [ 'pratama ', 'arhan ', 'df ', 'suwon ', '...
8  [ 'tebaktebakan ', 'masa ', 'kecil ', 'pema...
9  [ 'bambang ', 'pamungka ', 'legenda ', 'tim...
10 [ 'toni ', 'kroo ', 'mempertimbangkan ', 'b...
11 [ 'jersey ', 'away ', 'timna ', 'indonesia ...
12 [ 'petisi ', 'pertahankan ', 'shin ', 'taey...
13 [ 'timna ', 'amin ', 'pastikan ', 'ani ', ...
14 [ 'salah ', 'satu ', 'well ', 'decis ', 's...
15 [ 'apa ', 'bener ', 'lebih ', 'baik ', 'co...
16 [ 'break ', 'andr ', 'onana ', 'kini ', 'm...
17 [ 'ini ', 'bukan ', 'timna ', 'yang ', 'gu...
18 [ 'shin ', 'taeyong ', 'beserta ', 'jajaran...
19 [ 'yang ', 'jadi ', 'serba ', 'abuabu ', '...
Name: stemmed_text, dtype: object
    
```

Gambar 3. 4 Hasil *Stemming*

Pada gambar 3.4 yaitu hasil dari *Stemming* merupakan proses menghapus kata kata yang tidak diperlukan seperti imbuhan.

d. Stopword Removal

	full_text	username	created_at	text_tokens	text_steming	text_stopword
0	official shayne pattynama df resmi gabung klub...	IndoTransfer	Thu, Feb 1, 21:34:34, 2024	['official', 'shayne', 'pattynama', 'df', 'res...]	official shayne pattynama df resmi ...	official shayne pattynama df resmi ...
1	resmi pemain timnas indonesia elkan baggott ta...	FaktaSepakbola	Thu, Feb 1, 19:05:06, 2024	['resmi', 'pemain', 'timnas', 'indonesia', 'el...]	resmi main timnas indonesia elkan b...	resmi main timnas indonesia elkan b...
2	timnas indonesia baru saja mengalahkan vietnam...	FaktaSepakbola	Thu, Feb 1, 15:34:58, 2024	['timnas', 'indonesia', 'baru', 'saja', 'menga...]	timnas indonesia baru saja kalah vi...	timnas indonesia kalah vietnam ...
3	sudah tayang aminkan dari jauh merupakan progr...	UbahBareng	Thu, Feb 1, 10:22:01, 2024	['sudah', 'tayang', 'aminkan', 'dari', 'jauh', '...]	sudah tayang amin dari jauh rupa ...	tayang amin program inisiatif ...
4	pp muhammadiyah puji capaian timnas indonesia ...	UpdateBolabola	Thu, Feb 1, 7:22:19, 2024	['pp', 'muhammadiyah', 'puji', 'capaian', 'tim...]	pp muhammadiyah puji capai timnas i...	pp muhammadiyah puji capai timnas i...
5	timnas indonesia main bagus komentator luar vi...	gilabola_ina	Thu, Feb 1, 7:09:31, 2024	['timnas', 'indonesia', 'main', 'bagus', 'kome...]	timnas indonesia main bagus komentato...	timnas indonesia main bagus komentato...

Gambar 3. 5 Hasil *Stopword Removal*

Pada gambar 3.5 proses ini dilakukan dalam tahapan *stopword removal* menghapus kata kata imbuhan atau kata kata yang tidak penting yang terdapat pada twitter tersebut.

3.2.3 Labeling

	sentiment	percentage
neutral	793	71.83%
positive	186	16.85%
negative	125	11.32%

gambar 3. 6 Hasil *Labeling*

Gambar 3. 9 Hasil *Word Cloud* Sentimen *Positive*

3.2.5 Pembagian Data Latih dan Data Uji

Jumlah data latih: 883
 Jumlah data uji: 221

Gambar 3. 10 Hasil Data Latih dan Data Uji

Hasil dari gambar 3.10 adalah jumlah data testing (data uji) berjumlah 221 data dari keseluruhan data yaitu 1104, data *training* (data latih) berjumlah 883 data dari keseluruhan data yaitu 1104.

3.2.6 Klasifikasi *Naïve Bayes Classifier*

Akurasi Test: 0.8552036199095022

Gambar 3. 11 Hasil Klasifikasi *Naïve Bayes Classifier*

Hasil gambar 3.11 merupakan hasil klasifikasi *naïve bayes classifier* dan menghasilkan score sebesar 0.855.

3.2.7 Klasifikasi Dari Hasil *Labeling Naive Bayes*

	Tweet	Sentimen	Prediksi
309	medali emas tahun timnas indonesia ...		neutral
1041	tembus eropa main timnas indonesia ...		neutral
381	bicara counter attack tim ...		neutral
497	owner corp gilang widya pramana b...		neutral
551	anggap dosa politik kait mimpi ...		neutral
...			...
1057	thanks timnas indonesia you did gre...		positive
321	tanyarl timnas valorant indonesia pil...		neutral
982	resmi jersey away timnas indonesia ...		neutral
924	bayang suporter indonesia bikin k...		neutral
771	buah atur aneh timnas kriket putr...		negative

Gambar 3. 12 klasifikasi Dari Hasil *Labeling Naïve Bayes*

Hasil dari gambar 3.12 akan diprediksi hasil dari tweet sentiment setelah menghitung akurasi dari metode *Naïve Bayes*, maka hasil labeling menggunakan *Naïve Bayes*

3.2.8 Confusion Matrix

	precision	recall	f1-score	support
negative	1.000	0.677	0.808	31
neutral	0.868	0.948	0.906	153
positive	0.697	0.622	0.657	37
accuracy			0.855	221
macro avg	0.855	0.749	0.790	221
weighted avg	0.858	0.855	0.851	221

Gambar 3. 13 Hasil *Confusion Matrix*

Hasil dari gambar 3.13 menunjukkan bahasa hasil evaluasi dari *confusion matrix* yang sering kali diukur menggunakan matrix seperti *Persion, Recall, Fi-score*, dan Support.

4. SIMPULAN

Dari uraian penerapan *Naïve Bayes* pada klasifikasi tweet tentang sentiment terhadap sepak bola Indonesia yang telah dilakukan dengan 1104 dataset yang diperoleh dari media twitter. Pada tahap pengujian labeling menghasil persentase sentiment netral 71%, positif 16% dan negative 11%. Dengan pembagian data uji 221 dari 1104 data sedangkan data latih 883 dari 1104 data uji. Dihasilkan 793 sentimen netral,186 positif dan 125 negatif. Tahap preprocessing menghasilkan akurasi 85% dengan hasil presisi negatif 100%, presisi netral 86%, presisi positif 69%, recal negatif 67%, recal netral 94% dan positif 62%.

DAFTAR PUSTAKA

- [1] M. Astiningrum, M. Haniah, and Y. rahmat yoga Pradana, "Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter," *Semin. Inform. Apl. Polinema*, p. 35— 39, 2020.
- [2] N. Cahyono and D. Setiyawan, "Analisa Sentimen Pengguna Sosial MediaTwitter Terhadap Perokok di Indonesia," vol. 12, no. 1, pp. 262–272, 2023.
- [3] I. Indriati, M. Marji, and S. Pakpahan, "Analisis Sentimen Tentang Opini Performa Klub Sepak Bola Pada Dokumen Twitter Menggunakan Support Vector Machine Dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, pp. 7259–7267, 2019.
- [4] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [5] N. Riyanah and F. Fatmawati, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 2, no. 4, pp. 206–213, 2021, doi: 10.35746/jtim.v2i4.117.
- [6] V. No, Y. K. Putra, M. Adrian, J. Hidayat, and V. No, "Penerapan Algoritma Naïve Bayes Untuk Analisis Pengaruh Faktor Pendidikan Terhadap Peningkatan Kesehatan Masyarakat," vol. 7, no. 2, 2024.
- [7] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 697–711, 2021.