



Analisis Sentimen Transfer Pemain Klub La Liga Spanyol Pada Bursa Transfer Musim Dingin Eropa Di Twitter

Ahmad Adita Siddiq^{a,1,*}, Aris Wahyu Murdiyanto^{b,2}, Arif Himawan^{b,3}

^a Universitas Jenderal Achmad Yani Yogyakarta, Jalan Siliwangi - Gamping, Sleman 55294, Indonesia

^b Universitas Jenderal Achmad Yani Yogyakarta, Jalan Siliwangi - Gamping, Sleman 55294, Indonesia

^b Universitas Jenderal Achmad Yani Yogyakarta, Jalan Siliwangi - Gamping, Sleman 55294, Indonesia

¹ ahmad.adit123@gmail.com; ² ariswahyu@unjaya.ac.id*; ³ reef1881@gmail.com

* corresponding author

ABSTRACT

Dari beberapa kompetisi Sepak Bola yang ada, Liga Champions UEFA yang paling digemari oleh masyarakat. Pada tahun 2022 bursa transfer pemain Eropa dibuka, bursa transfer yang dilakukan merupakan cara jangka pendek untuk memperbaiki tim dalam mengejar prestasi sepak bola. Dengan media sosial sebagai wadah komunitas, para penggemar sepak bola dapat juga menyalurkan opini, informasi dan berita tentang klub kesayangan kepada masyarakat. Opini masyarakat terhadap transfer pemain Liga Spanyol memiliki peranan penting. Dengan dilakukannya analisis sentimen terhadap opini, dapat dijadikan suatu pola prediksi penilaian masyarakat terhadap transfer pemain serta dapat memberikan saran kepada tim sepak bola terkait bursa transfer pemain pada periode musim selanjutnya.

Membuat analisis sentiment penggemar sepak bola terhadap transfer pemain Liga Spanyol apakah bersifat positif dan negatif. Metode Naïve Bayes Classifier (NBC) dalam penelitian ini dipilih dikarenakan pada algoritma NBC dapat melakukan proses pengolahan data diskrit dan data kuantitatif dengan menggunakan sampel yang relative sedikit dan juga perhitungan pada algoritma NBC lebih cepat.

Pengambilan data berupa topik mengenai keyword "Transfer La Liga", "Transfer Real Madrid", "Transfer Barcelona", "Transfer Liga Spanyol" dan "Transfer Copa Del Ray". Data tweet di ambil dari periode 1 Januari 2020 sampai dengan 31 Mei 2022, dengan jumlah data total 11.282. Pada penelitian telah berhasil mendapatkan akurasi dengan nilai 81,67 % pada data training dan 85 % untuk data testing. Pada penelitian ini berhasil membuat model analisis sentimen berupa file.pickle yang dimana untuk melakukan klasifikasi dan prediksi pada data tweet untuk mendapatkan sebuah hasil sentimen positif dan negative. Penelitian ini telah berhasil mendapatkan akurasi dengan nilai 81,67 % pada data training dan 85 % untuk data testing. Hasil analisis sentimen akhir dalam klasifikasi penelitian ini bernilai "Sentimen Negatif"

ARTICLE INFO

Article history

Received: 17 Februari 2023

Revised: 2 Maret 2023

Accepted: 10 Mei 2023

Keywords

Analisis Sentimen

Naive Bayes Classifier

La Liga Spanyol

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



I. Pendahuluan

Sepak bola merupakan olahraga dengan jumlah tim sebanyak 4% dari populasi di dunia. Permainan sepak bola mengharuskan para pemain untuk ahli dalam menguasai bola, mencetak gol ke gawang lawan serta menjaga gawang tim. Untuk mencapai kemenangan diperlukan penguasaan bermain serta kerjasama antar pemain dengan baik. Beberapa klub sepakbola rela bersaing dalam ajang bursa transfer pemain, ditambah klub sepakbola tersebut rela menghabiskan waktu dan tenaga untuk mendapatkan pemain yang diinginkan (Fajrin et al., 2021).

Ada beberapa perlombaan kompetisi antar tim sepak bola yang resmi di dunia, salah satunya di Eropa. Di Eropa sendiri ada beberapa kompetisi antar tim sepak bola yang terdiri dari Liga Champions UEFA, Liga Eropa UEFA, Liga Konferensi Eropa dan Piala Super Eropa. Dari beberapa kompetisi yang ada, Liga Champions UEFA yang paling digemari oleh masyarakat. Adapun perlombaan dalam piala UEFA terbagi kedalam 20 Liga yaitu Liga Spanyol, Liga Inggris, Liga Italia, Liga Jerman, Liga Prancis, Liga Rusia, Liga Portugal, Liga Ukraina, Liga Belgia dan Liga Turki. Pada saat ini Liga Spanyol menduduki peringkat pertama dalam UEFA (Union of European Football Associations) (Utama & Widodo, 2021). Kelebihan Liga Spanyol dibanding dengan Liga yang lain berupa sejarah divisi sepakbola pertama, kualitas pemain-pemain sepak bola, kompetitif dan gaya hidup (Natadjaya et al., 2020).

Pada tahun 2022 bursa transfer pemain Eropa dibuka, bursa transfer pemain dilakukan pada musim dingin dan musim panas, tepat pada tanggal 1 Januari 2022 bursa transfer musim dingin Eropa telah dibuka. Beberapa klub sepakbola dalam Liga Spanyol seperti klub Atletico Madrid, FC Barcelona, Real Madrid serta Sevilla melakukan transfer pemain. Bursa transfer yang dilakukan pada bulan Januari merupakan cara jangka pendek untuk memperbaiki tim dalam mengejar prestasi sepak bola (Anshari & Qalbie Septizar Akbar, 2019). Penggemar sepak bola tentu sangat antusias dalam mengikuti perkembangan keputusan klub favorit terhadap bursa transfer pemain. Salah satu cara para penggemar sepak bola untuk ikut andil dalam acara bursa transfer, mereka antusias untuk mengeluarkan opini melalui sosial media. Dengan media sosial sebagai wadah komunitas, para penggemar sepak bola dapat juga menyalurkan opini, informasi dan berita tentang klub kesayangan kepada masyarakat (Jordan et al., 2021).

Twitter merupakan salah satu layanan media sosial yang digunakan untuk media komunikasi dan mendapatkan informasi. Pengguna Twitter dapat memposting pertanyaan, pendapat pribadi maupun komentar yang bersifat positif maupun negative (Ayuna Olenti, 2019). Laporan dari majalah republik tahun 2022 menyebutkan bahwa setiap hari sebanyak 217 juta pengguna Twitter melakukan aktifitas berupa mencari informasi, mengajukan pertanyaan, memberikan opini dan komentar (Septian et al., 2019). Opini masyarakat terhadap transfer pemain Liga Spanyol memiliki peranan penting. Dengan dilakukannya analisis sentimen terhadap opini, dapat dijadikan suatu pola prediksi penilaian masyarakat terhadap transfer pemain serta dapat memberikan saran kepada tim sepak bola terkait bursa transfer pemain pada periode musim selanjutnya. Hal ini dibuktikan dengan penelitian dari (Aryashila et al., 2020) penelitian ini menghasilkan bahwa adanya hubungan antara transfer pemain dengan komentar di media sosial.

Dalam melakukan analisis opini membutuhkan proses text mining yang akan memproses teks dari data yang diperoleh kemudian melakukan pola ekstraksi data sehingga informasi yang dihasilkan dapat berupa pengetahuan yang berguna (Fauziyyah, 2020). Analisis sentimen merupakan suatu kegiatan bidang studi yang berupa analisis dari suatu pendapat, penilaian, sikap, evaluasi serta emosi seseorang terhadap suatu isu, problem, layanan serta produk dari personal, kelompok atau even tertentu. Dengan dilakukannya analisis sentimen maka dapat diketahui apakah suatu pendapat terhadap topik tersebut memiliki penilaian yang positif atau negatif, sehingga bisa dijadikan referensi kebijakan untuk meningkatkan produk atau kualitas dari jasa tersebut (Nurjanah et al., 2017)

Salah satu metode klasifikasi yang dapat digunakan untuk text mining adalah Naïve Bayes Classifier. Metode Naïve Bayes Classifier merupakan klasifikasi yang memprediksi probabilitas suatu kelas, sehingga memiliki kelebihan tingkat akurasi yang tinggi dan waktu komputasi yang cepat (Gandhi et al., 2021). Kelebihan metode Naïve Bayes Classifier terbukti dapat digunakan secara efektif untuk mengklasifikasikan teks secara otomatis. Algoritma NBC memiliki karakteristik sederhana dan kecepatan tinggi dalam proses pelatihan dan klasifikasi. Algoritma ini menarik untuk digunakan sebagai salah satu metode klasifikasi (Zulfikar & Lukman, 2016)

Berdasarkan latar belakang pada uraian tersebut, maka dalam penelitian ini mengambil judul “Analisis Sentimen Transfer Pemain Klub La Liga Spanyol Pada Bursa Transfer Musim Dingin Eropa di Twitter”. Peneliti memilih untuk menggunakan metode Naive Bayes Classifier dikarenakan metode NBC memerlukan sample data sejumlah kecil untuk pelatihan dan klasifikasi, kemudian NBC juga dapat menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang dan metode NBC memiliki karakteristik cepat dan efisiensi terhadap memori.

Penelitian yang dilakukan oleh (Tuhuteru & Iriani, 2018) dengan judul “Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier”. Pada penelitian tersebut bertujuan untuk mengetahui sentiment masyarakat di Pulau Ambon terhadap kondisi kelistrikan yang terjadi dengan menggunakan metode analisis sentimen. Hasil dari penelitian ini yaitu pada metode NBC didapatkan sentiment positif sebesar 67%, sentiment netral sebesar 19% dan sentiment negative sebesar 14%. Sementara pada metode SVM didapatkan sentiment positif sebesar 24%, sentiment netral sebesar 29% dan sentiment negatif sebesar 47%.

Hasil penelitian dari (Sudiantoro et al., 2018) dengan judul “Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier”. Penelitian ini bertujuan untuk mendapatkan data dimedia social berupa opini masyarakat terhadap pilkada jawa barat kemudian dilakukan analisis sentiment untuk mengetahui klasifikasi positif atau negative. Hasil ini penelitian ini berupa 100 data uji yang klasifikasi menghasilkan 32 data sentiment positif dan 68 sentiment negatif, Adapun nilai akurasi algoritma NBC memberikan nilai sebesar 84%.

Hasil penelitian dari (Wibowo & Novianti, 2022) judul “Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi”. Penelitian ini bertujuan melakukan analisis sentiment dan memanfaatkan data untuk mendapatkan informasi yang berguna bagi pelanggan JD.ID serta dapat dijadikan bahan evaluasi bagi JD.ID. Hasil dari penelitian ini menunjukkan bahwa metode NBC tanpa penambahan TF-IDF mampu menghasilkan klasifikasi dengan nilai akurasi sebesar 96,44%, sementara jika ditambah dengan fitur TF-IDF disertai konversi ikon emosi mampu meningkatkan nilai akurasi sebesar 98%.

Hasil penelitian dari (Nurwahyuni, 2019) dengan judul “Analisis Sentimen Aplikasi Transportasi Online KRL Access Menggunakan Metode Naive Bayes Classifier”. Penelitian ini untuk analisis secara mendalam terkait opini public dalam perihal transportasi umum khususnya kereta KRL di jabodetabek sebagai solusi untuk mengurai kemacetan lalu lintas. Hasil dari penelitian ini berupa nilai akurasi metode NBC mendapatkan nilai sebesar 84%.

Hasil penelitian dari (Sari & Irhamah, 2020) dengan judul “Analisis Sentimen Nasabah Pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naive Bayes Classifier dan Support Vector Machine”. Penelitian ini bertujuan untuk melakukan analisis terhadap nasabah sehingga dapat diklasifikasikan menjadi sentiment positif atau sentiment negative, kemudian dapat dijadikan evaluasi bagi pihak Bank BRI dan Bank Mandiri untuk melakukan peningkatan pelayanan bagi para nasabah. Hasil dari penelitian ini menghasilkan metode terbaik dalam mengklasifikasikan sentiment pada layanan BRI menggunakan SMOTE-SVM dengan kernel RBF, sedangkan pada Bank Mandiri dengan menggunakan SMOTE-NBC.

Perbedaan penelitian peneliti dibandingkan dengan penelitian terdahulu dengan tema “Analisis Sentimen” berupa subjek penelitian kemudian pembuatan program pengambilan dan analisis data tweet. Pada penelitian terdahulu proses pengambilan dan melakukan analisis data menggunakan tools RStudio/ RapidMiner, sedangkan pada proses penelitian analisis bursa transfer pemain peneliti menggunakan bahasa pemrograman Python dengan tools dari Jupyter Notebook. Selanjutnya proses klasifikasi penelitian terdahulu secara umum menggunakan metode SVM, KNN, NBC dan Regresi Logistik Biner dengan bantuan tools RapidMiner. Pada penelitian bursa transfer pemain Liga Spanyol, peneliti menggunakan metode NBC dengan proses klasifikasi menggunakan bahasa Python.

2. Metode penelitian

2.1. Python

Python merupakan salah satu Bahasa pemrograman yang ditujukan untuk merancang dan membangun perangkat lunak dalam berbagai domain aplikasi. Python juga termasuk kedalam high level programming language yang artinya Python mudah untuk dipelajari dan dituliskan oleh pemula. Python menggabungkan kapabilitas, struktur, kemampuan dan sintaksis kode yang sangat jelas serta dilengkapi dengan fungsionalitas library yang lengkap dan komprehensif (Rahmadhika & Thantawi, 2021).

2.2. Twitter

Twitter merupakan sebuah platform berbasis website yang menawarkan jaringan sosial berbasis microblog sehingga pengguna dapat mengirim dan membaca pesan tweet. Tweet sendiri merupakan teks tulisan yang berisi maksimal 280 karakter yang akan ditampilkan pada halaman profil pengguna. Tweet bisa ditampilkan secara public dan juga secara privat, pengguna Twitter bisa saling melihat tweets pengguna lain yang dikenal dengan istilah pengikut (follower) (Giovani et al., 2020).

2.3. Analisis Sentimen

Analisis sentimen merupakan suatu Teknik untuk mengidentifikasi bagaimana sebuah sentiment di ekspresikan kedalam bentuk teks atau bagaimana sentiment di kategorikan kedalam sentiment positif atau sentiment negatif. Tujuan utama sentimen analisis ini adalah menganalisis secara mendalam tentang pendapat, mengidentifikasi sentiment yang masyarakat ungkapkan dan kemudian mengklasifikasikan polaritasnya. Dengan kata lain, analisis sentiment berfungsi untuk mengklasifikasikan teks kedalam kelas positif dan negatif (Rachman & Pramana, 2020). Penelitian terdahulu (Krisdiyanto, 2021) menyebutkan bahwa ada beberapa alasan media social Twitter digunakan sebagai analisis sentimen.

2.4. Naive Bayes Classifier

Algoritma Naïve Bayes Classifier merupakan algoritma yang diperuntukkan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang tepat, algoritma NBC merupakan salah satu metode dari machine learning yang menggunakan perhitungan probabilitas. Algoritma NBC memiliki keuntungan bahwa algoritma ini membutuhkan sampel kecil pelatihan untuk prediksi parameter yang digunakan untuk klasifikasi (Hayuningtyas, 2019).

TF-IDF (Term Frequency Inverse Document Frequency TF-IDF) merupakan sebuah metode yang digunakan untuk menentukan nilai frekuensi sebuah kata dari dokumen atau artikel. Pada Algoritma TF-IDF melakukan pemberian bobot pada setiap kata kunci di setiap kategori untuk mencari kemiripan kata dengan kategori yang tersedia. Sebelum melakukan pembobotan maka akan dilakukan tahapan pemecah kalimat, case folding, tokenizing, filtering dan steaming, lalu pada tahap selanjutnya dilakukan proses menghitung bobot TF- IDF, bobot query relevance dan bobot similarity (Naf'an et al., 2019)

3. Method

Metode Naïve Bayes Classifier dipilih dalam proses penelitian ini. Penelitian ini merupakan penelitian analisis sentimen positif dan sentimen negative. Penggunaan metode Naïve Bayes Classifier (NBC) dalam penelitian ini dipilih dikarenakan pada algoritma NBC dapat melakukan proses pengolahan data diskrit dan data kuantitatif dengan menggunakan sampel yang relative sedikit dan juga perhitungan pada algoritma NBC lebih cepat (Mustafa et al., 2018). Penelitian ini menggunakan data tweet, kemudian dilakukan proses pengolahan data berupa preprocessing. Hasil dari tahap pengolahan data digunakan untuk memetakan informasi opini masyarakat terhadap proses transfer pemain Liga Spanyol.

Penelitian ini bermula dari identifikasi permasalahan, kemudian pengolahan data yang sudah didapat dan menentukan sentimen analisis, sehingga hasil penelitian ini berupa informasi yang tepat dan sesuai.

Pada kasus term frequency $tf(t, d)$, cara yang sederhana menggunakan raw frequency di dalam dokumen. Jika menyatakan raw frequency t sebagai $f(t, d)$, maka skema tf yang sederhana adalah $tf(t, d) = f(t, d)$. Nilai idf pada sebuah term (kata) dapat dihitung menggunakan rumus sebagai berikut :

$$IDF = \log_{10} D / D_{fi} \quad (1)$$

D adalah jumlah dokumen yang berisi term (t) dan D_{fi} adalah jumlah kemunculan (frekuensi) kata terhadap D. Sedangkan algoritma yang digunakan untuk perhitungan bobot (W) masing-masing dokumen terhadap kata kunci (query) yaitu :

$$W_{d,t} = t f_{d,t} * IDF_t \quad (2)$$

Keterangan :

d = dokumen ke -d

t = kata ke -t dari kata kunci

W = bobot dokumen ke-d terhadap kata ke-t

Tf = term frekuensi

Setelah bobot (W) pada masing masing dokumen diketahui, maka dilakukan proses pengurutan dimana semakin besar nilai W, maka semakin besar juga tingkat kesamaan dokumen tersebut terhadap kata yang dicari (Krisdiyanto, 2021).

3.1. Bahan Penelitian

Pada penelitian ini bahan utama yaitu berupa data tweet dan retweet di media sosial Twitter yang berkaitan tentang tema Transfer Pemain Liga Spanyol. Penelitian ini menggunakan data tweet yang didapatkan dari Twitter dengan keyword/hastag “Bursa pemain La Liga Spanyol”, “Transfer La Liga”, “Transfer Real Madrid”, “Transfer Barcelona”, “Transfer Liga Spanyol” dan “Transfer Copa Del Ray” serta akun-akun sosial media Twitter yang meliput berita tentang dunia sepak bola seperti @idextratime, @PanditFootball.com, dan @SuperSoccerTV.

3.2. Jalan Penelitian

Pada penelitian ini tahap pertama adalah menggunakan Bahasa pemrograman Python untuk melakukan pengambilan data yang akan di tampilkan pada Microsoft Office Excel dan divisualisasikan dengan bantuan library Python. Pada gambar 1 merupakan alur penelitian :

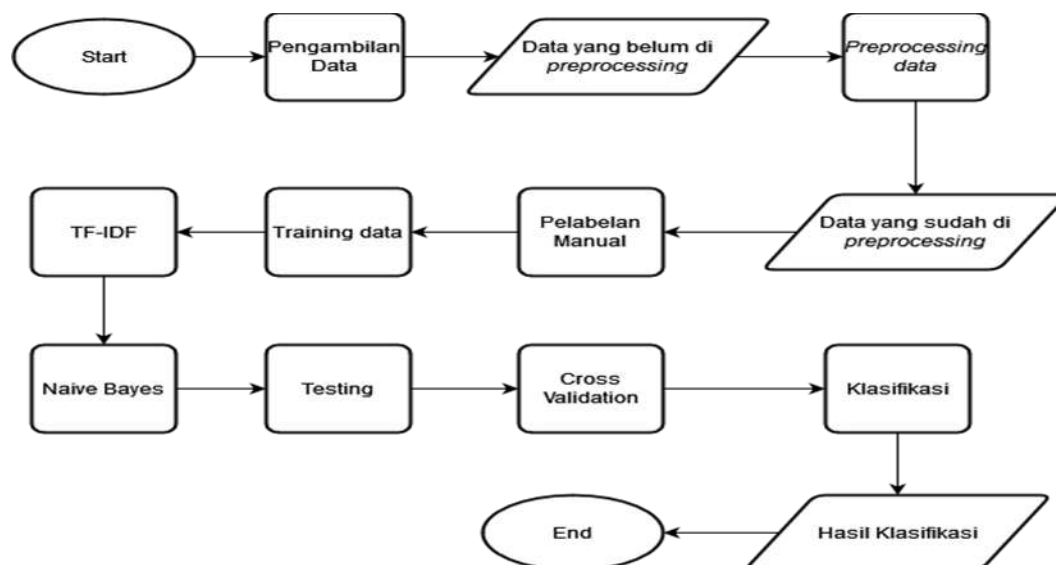


Fig. 1. Flowchart Jalan Penelitian

Pada tahap preprocessing terdapat subproses meliputi cleaning, casefolding, tokenization, stopword dan stemming. Pada gambar 2 merupakan alur dari subproses pada preprocessing data :

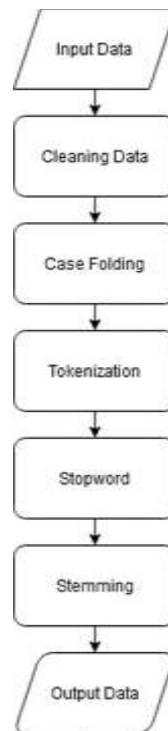


Fig. 2. Flowchart Preprocessing

Setelah melalui tahap data training terdapat subproses yaitu, TF-IDF untuk melakukan perhitungan term pada setiap dokumen yang dapat dilihat pada gambar 3 :



Fig. 3. Alur perhitungan TF-IDF

Fig. 4.

3.3. Pengambilan Data

Pengambilan data merupakan tahap pertama untuk pengambilan data berupa topik mengenai “Bursa pemain La Liga Spanyol” dengan menggunakan Jupyter Notebook. Data yang di ambil berupa keyword “Transfer La Liga”, “Transfer Real Madrid”, “Transfer Barcelona”, “Transfer Liga Spanyol” dan “Transfer Copa Del Ray”. Data tweet di ambil dari periode 1 Januari 2020 sampai dengan 31 Mei 2022, dengan jumlah data total 11.282.

Proses pengambilan data pada penelitian ini menggunakan library snsrape.module.twitter, peneliti menggunakan library snsrape dikarenakan tanpa API dan tidak terbatas pada periode waktu tertentu. Kemudian library pandas untuk manipulasi dan menampilkan data. Pada gambar 4 merupakan source code import library snsrape dan pandas.

```

import snsrape.modules.twitter as sntwitter
import pandas as pd
  
```

Fig. 5. Import library crawling data

Setelah melakukan import library, kemudian melakukan kode untuk pengambilan data. Dimulai dengan berapa jumlah data yang ingin di ambil, kata kunci/topik yang ingin di ambil serta periode waktu pengambilan data. Pada gambar 5 merupakan source code crawling data.

```
# Setting variables to be used below
maxTweets = 5000

# Creating list to append tweet data to
tweets_list2 = []

# Using TwitterSearchScrapper to scrape data and append tweets to list
for i,tweet in enumerate(sntwitter.TwitterSearchScrapper('transfer Copa Del Rey since:2020-01-01 until:2022-05-31 lang:id').get_items());
    if i>maxTweets:
        break
    tweets_list2.append([tweet.id, tweet.date, tweet.username, tweet.content])
```

Fig. 6. Crawling data

Data yang sudah berhasil diambil kemudian disimpan kedalam format Microsoft Office Excel atau dalam bentuk format CSV. Pada tabel 1 merupakan contoh data tweet yang diambil sebagai berikut:

Table 1. Contoh data tweet dan re-tweet

No	Tweet
1	Deal! Perisic Ke Tottenham Selamat Tinggal Abramovich Real Madrid Juara Berita Transfer Pemain https://t.co/Zk98NPhSFa https://t.co/HZxLdGmMQb
2	@Marcelo akan meninggalkan Real Madrid setelah 15 tahun membela klub, dengan status bebas transfer. Berpamitan pada momen yang tepat setelah memenangkan LaLiga dan Champions League sebagai kapten tim!!. https://t.co/NjZ2ZDAXnm
3	Real Madrid!!! Sapa mau senggol haa?? Terlepas dari si masalah transfer gajelas yang kemarin, kita sudah menang dan buktikan siapa rajanya
4	Dapat Dorongan Transfer Untuk Bintang Leeds Target Utama Barcelona https://t.co/0hOmBZWSaY https://t.co/GkdGZcZpN7
5	"Terdepan! Chelsea Segera Segel Transfer Dembele dari Barcelona - https://t.co/kpmQtIcweG
6	Mbappe minta Paris Saint-Germain untuk memboyong gelandang Barcelona Frenkie de Jong saat jendela transfer musim panas dibuka pada Juni nanti. https://t.co/QkCqwaTdmz
7	.Xavi Hernandez membenarkan bahwa Robert Lewandowski masuk radar Barcelona. Barcelona sedang mengupayakan transfer Lewandowski ke Camp Nou. https://t.co/CASqpbm52
8	.Kata Barcelona Soal Transfer Frenkie De Jong ke MU https://t.co/0ve6ptR6qL
9	@bolasportdotcom,"Ingin Berkhianat ke Barcelona, Gelandang Real Madrid Siap Ulangi Transfer Haram 14 Tahun Lalu https://t.co/D4naf3kqA2 "
10	@KhbrkNews, Barcelona plan bargain transfer deal for Man Utd star Nemanja Matic https://t.co/rWwjIqonHs https://t.co/iz5529v4GE

3.4. Preprocessing

Pada tahap preprocessing, data teks disiapkan agar dapat dipergunakan dalam proses selanjutnya. Pada proses preprocessing dibutuhkan beberapa library untuk membantu jalannya tahap preprocessing. Beberapa library yang dibutuhkan seperti library pandas yang berguna untuk manipulasi data, library numpy untuk komputasi serta library nltk untuk membantu pengolahan natural language. Pada gambar 6 merupakan import library yang perlu dilakukan sebelum melakukan preprocessing, kemudian perintah untuk menampilkan data tweet yang berupa file CSV akan di panggil ke dalam dataframe. Gambar 6 merupakan source code untuk melakukan import library pandas dan pemanggilan data.

```

import pandas as pd, numpy as np, nltk, string, emoji, re
from pandas import DataFrame

# membaca data set .csv
def load_data():
    data = pd.read_csv('laligaspanyol.csv', nrows=None, header=0, names=['Tanggal', 'Username', 'Text'])
    return data

data = load_data()

data.head()

```

Fig. 7. Import library dan pemanggilan data

Pada proses cleaning data, proses ini berfungsi untuk menghilangkan link url, tanda baca, angka, simbol dan username yang berada pada tweet dan re-tweet. Pada gambar 7 merupakan source code cleaning data :

```

def cleaning_text(text):
    # hapus tab, newline, dan backslash
    tab = text.replace('\t', ' ').replace('\n', ' ').replace('\r', ' ')
    # hapus underscore
    score = tab.replace('_', '')
    # hapus user mention
    user = re.sub('@[A-Za-z0-9]+', '', score)
    # hapus link
    link = re.sub(
        '((https?):(//)|(\w\w\w))+([\w\d:#@%/;$()~_?+\-=\|\.\&]{1,6})+', '', user)
    # menghapus url
    url = re.sub(r'http\S+', '', link)
    # menghapus punctuation
    punc = re.sub(r'[\w\s]', '', url)
    # menghapus retweet (rt)
    rt = re.sub(r'RT[\s]', '', punc)
    # menghapus angka
    no = re.sub('[0-9]+', '', rt)
    # menghapus slang
    slang = re.sub(r'\n', " ", no)
    # menghapus regex
    reg = re.sub("b'", " ", slang)
    # hapus hashtag
    hashtag = re.sub('/#[\w_]+[\s]*', '', reg)
    # menghapus emoticon
    emot = emoji.get_emoji_regexp().sub("", hashtag).strip()

```

Fig. 8. Source Code Cleaning Data

Setelah dilakukan proses cleaning data, maka hasil akhir dari data yang setelah di bersihkan seperti pada tabel 2 sebagai berikut.

Table 2. Data tweet bersih

No	Tweet
1	Deal Perisic Ke Tottenham Selamat Tinggal Abramovich Real Madrid Juara Berita Transfer Pemain
2	Marcelo akan meninggalkan Real Madrid setela tahun membela klub, dengan status bebas transfer. Berpamitan pada momen yang tepat setelah memenangkan LaLiga dan Champions League sebagai kapten tim
3	Real Madrid Sapa mau senggol haa Terlepas dari si masalah transfer gajelas yang kemarin kita sudah menang dan buktikan siapa rajanya
4	Dapat Dorongan Transfer Untuk Bintang Leeds Target Utama Barcelona
5	Terdepan Chelsea Segera Segel Transfer Dembele dari Barcelona
6	Mbappe minta Paris Saint Germain untuk memboyong gelandang Barcelona Frenkie de Jong saat jendela transfer musim panas dibuka pada Juni nanti
7	Xavi Hernandez membenarkan bahwa Robert Lewandowski masuk radar Barcelona. Barcelona sedang mengupayakan transfer Lewandowski ke Camp Nou
8	Kata Barcelona Soal Transfer Frenkie De Jong ke MU
9	Ingin Berkhianat ke Barcelona, Gelandang Real Madrid Siap Ulangi Transfer Haram Tahun Lalu
10	KhbrkNews Barcelona plan bargain transfer deal for Man Utd star Nemanja Matic

Pada proses *Tokenization* bertujuan untuk melakukan pemecahan dari sebuah kalimat menjadi sebuah potongan kata. Dari sebuah potongan kata dapat disimpak kedalam sebuah dokumen. Pada gambar 8 merupakan source code *Tokenization*.

```

def clean_text(tweet):
    for i in tweet:
        cleaned.append(cleaning_text(
            re.sub("[\n\r\t\xa0]", " ", i).strip()))
    clean_text(data["Text"])

```

Fig. 9. Source code Tokenization

Dalam proses penarikan data *tweet* bursa transfer pemain LaLiga Spanyol, data masih disimpan dalam keadaan *raw data*/data mentah. Sehingga apabila dilakukan analisis tanpa memiliki standard, maka hasil analisis tidak akurat. Sehingga, diperlukan proses *Case Folding* untuk melakukan konversi dari bentuk awal menjadi bentuk standard. *Case Folding* dalam penelitian bursa transfer pemain La Liga Spanyol berguna untuk merubah kalimat ke bentuk *uppercase* atau *lowercase* secara standard. Pada gambar 9 merupakan source code *Case Folding*.

```
def lowercase():
    lower_word = data['Clean_Text'].str.lower()
    return lower_word

lower_tweet = lowercase()

print(lower_tweet)
```

Fig. 10. Source code Case Folding

Stopword Removal merupakan proses penghapusan kata-kata yang tidak memiliki informasi. Pada proses ini mengambil kata-kata penting dan membuang kata-kata yang kurang penting seperti “yang”, “di”, “dan”, “dari”, dll. Peneliti menggunakan proses *Stopword Removal* untuk menghapus kata-kata yang memiliki informasi rendah dari teks *tweet* bursa transfer pemain La Liga Spanyol. Pada gambar 3.10 merupakan source code untuk *stopword removal*.

```
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

factory = StopWordRemoverFactory()
stopword = factory.create_stop_word_remover()
stopwords = factory.get_stop_words()
print(stopwords)
```

Fig. 11. source code stopword removal

Pada *stopwordremoval* diperlukan penghapusan kata-kata yang terdapat pada *library sastrawi*. Pada tabel 3 merupakan kata-kata yang terdapat pada *library sastrawi*.

Table 3. Daftar Kata Library Sastrawi

Daftar Kata library Sastrawi
['yang', 'untuk', 'pada', 'ke', 'para', 'namun', 'menurut', 'antara', 'dia', 'dua', 'ia', 'seperti', 'jika', 'jika', 'sehingga', 'kembali', 'dan', 'tidak', 'ini', 'karena', 'kepada', 'oleh', 'saat', 'harus', 'sementara', 'setelah', 'belum', 'kami', 'sekitar', 'bagi', 'serta', 'di', 'dari', 'telah', 'sebagai', 'masih', 'hal', 'ketika', 'adalah', 'itu', 'dalam', 'bisa', 'bahwa', 'atau', 'hanya', 'kita', 'dengan', 'akan', 'juga', 'ada', 'mereka', 'sudah', 'saya', 'terhadap', 'secara', 'agar', 'lain', 'anda', 'begitu', 'mengapa', 'kenapa', 'yaitu', 'yakni', 'daripada', 'itulah', 'lagi', 'maka', 'tentang', 'demi', 'dimana', 'kemana', 'pula', 'sambil', 'sebelum', 'sesudah', 'supaya', 'guna', 'kah', 'pun', 'sampai', 'sedangkan', 'selagi', 'sementara', 'tetapi', 'apakah', 'kecuali', 'sebab', 'selain', 'seolah', 'seraya', 'seterusnya', 'tanpa', 'agak', 'boleh', 'dapat', 'dsb', 'dst', 'dll', 'dahulu', 'dulunya', 'anu', 'demikian', 'tapi', 'ingin', 'juga', 'nggak', 'mari', 'nant', 'melainkan', 'oh', 'ok', 'seharusnya', 'sebetulnya', 'setiap', 'setidaknya', 'sesuatu', 'pasti', 'saja', 'toh', 'ya', 'walau', 'tolong', 'tentu', 'amat', 'apalagi', 'bagaimanapun']

Stemming merupakan proses penghilangan infleksi kata menjadi bentuk dasar. Peneliti menggunakan *Python Sastrawi* dalam proses *Stemming*, hal ini dikarenakan data *tweet* yang berbahasa Indonesia semua kata imbuhan sufiks dan prefix harus dihilangkan. Pada gambar 11 merupakan source code *stemming*.

```

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stemmed_wrapper(term):
    return stemmer.stem(term)

term_dict = {}

for document in stopwords_tweet:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ""

print(len(term_dict))
print("-----")

for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term,":", term_dict[term])

print(term_dict)
print("-----")

def get_stemmed_term(document):
    return [term_dict[term] for term in document]

stem_tweet = stopwords_tweet.apply(get_stemmed_term)

print(stem_tweet)

```

Fig. 12. Source Code Stemming

Normalization merupakan penyeragaman pada term yang mengalami kesalahan penulisan atau menggunakan bahasa yang tidak baku. Peneliti membuat *dataset* tentang *term* baku sehingga *dataset* berguna untuk menyeragamkan kata yang tidak sesuai. Hasil dari proses *Normalization* berupa data *tweet* yang lebih terstruktur serta dapat dilakukan perhitungan pada proses selanjutnya. Pada gambar 12 merupakan source code untuk proses normalisasi.

```

normalizad_word = pd.read_excel("normalisasi.xlsx")

normalizad_word_dict = {}

for index, row in normalizad_word.iterrows():
    if row[0] not in normalizad_word_dict:
        normalizad_word_dict[row[0]] = row[1]

def normalized_term(document):
    return [normalizad_word_dict[term] if term in normalizad_word_dict else term for term in document]

normal_tweet = stem_tweet.apply(normalized_term).str.join(" ")

print(normal_tweet)

```

Fig. 13. Source Code Normalisasi

Berikut contoh data pada file *normalisasi.xlsx* yang telah dibuat berdasarkan topik pembahasan sebagai dataset memperbaiki kata yang salah dapat dilihat pada Tabel 4.

Table 4. Dataset Normalisasi

No	Kata Sebelum	Kata Sesudah
1	Yg	Yang
2	Dl	Dahulu
3	Stlh	Setelah
4	Akn	Akan
5	Stju	Setuju
6	Lnjtkn	Lanjutkan
7	Dg	Dengan
8	Lg	Lagi
9	Gt	Gitu
10	Pny	Punya

Pada tahap pelabelan manual, tahap ini merupakan proses memberikan level terhadap kata pada dokumen sehingga dapat dianalisis lebih lanjut mengenai sifat yang positif atau negatif. Pada

proses pelabelan manual peneliti membagi data tweet yang berisi tentang bursa transfer pemain dengan data training label positif dan label negatif. Data *tweet* yang sudah dilabeli dengan jumlah data 600 tweet dari 600 *tweet* dengan masing-masing 300 *tweet* positif dan 300 *tweet* negatif dari data *training*. Pada gambar 13 merupakan hasil pelabelan manual

no	tanggal	username	tweet	label	kelas	
0	1	2012-03-07 17:32:43+00:00	agentaruhanbola	account transfer main bola situs taruh bola li...	Positif	1
1	2	2010-06-20 09:15:32+00:00	baskaramp	aduh aduh kangen liga inggris spanyol itali ni...	Positif	1
2	3	2021-04-15 16:30:00+00:00	Bolanet	akhir spekulasi barcelona segera umum transfer...	Negatif	0
3	4	2021-04-15 10:28:14+00:00	M88Indo	akhir spekulasi barcelona segera umum transfer...	Negatif	0
4	5	2020-08-30 15:51:35+00:00	_fresshare	aku masuk opini kalau macam insentif bisnis sp...	Negatif	0

Fig. 14. Hasil Pelabelan Manual

Ditunjukkan bahwa label positif diberi kelas 1 dan nilai kelas 0 untuk label negatif. Pelabelan manual dilakukan untuk perhitungan akurasi yang telah diberi sentimen positif dan negatif.

Pada proses training data menggunakan metode *Naïve Bayes Classifier*. Pada tahap ini diawali dengan fitur ekstraksi pada teks menggunakan TF-IDF, kemudian dilakukan proses training data untuk membuat model klasifikasi sentiment. Berikut contoh perhitungan TF-IDF secara manual dengan Microsoft Office Excel dapat dilihat pada Tabel 5 :

Table 5. Data Training

No	Kata Sebelum
d1	real madrid lineup ganti raphael varane tengah minat transfer manchester united.
d2	real madrid tumbal vinicius junior transfer kylian mbappe
d3	update transfer main sergio aguero sepakat gabung barcelona kontrak tahun depan
d4	rumor transfer sergio aguero sepakat gabung barcelona musim depan

Dalam melakukan perhitungan *Term Frequency* (TF) ini menggunakan beberapa komponen yaitu *term* atau kata, dan *d* merupakan jumlah data yang akan digunakan terdiri dari d1, d2, d3 dan d4 dan *df* untuk menghitung jumlah term atau kata yang muncul pada setiap dokumen. Contoh perhitungan TF dapat dilihat pada Tabel 6.

Table 6. Perhitungan TF

Term (kata)	d1	d2	d3	d4	df
real	1	1			2
madrid	1	1			2
lineup	1				1
ganti	1				1
raphael	1				1
varane	1				1
tengah	1				1
minat	1				1
transfer	1	1	1	1	4
manchester	1				1
united	1				1
tumbal		1			1
vinicius		1			1
junior		1			1
kylian		1			1
mbappe		1			1
update			1		1
main			1		1
sergio			1	1	2
aguero			1	1	2
sepakat			1	1	2
gabung			1	1	2

barcelona			1	1	2
kontrak			1		1
tahun			1		1
depan			1	1	2
rumor				1	1
musim				1	1

Perhitungan *Invers Document Frequency* (IDF) menggunakan beberapa komponen seperti *term* atau kata, *df* dan *idf* yang berhubungan antara ketersediaan suatu *term* di semua dokumen, dihitung dengan *N* atau jumlah dokumen. Berikut contoh perhitungan IDF dapat dilihat pada Tabel 7.

Table 7. Perhitungan IDF

Term (kata)	df	Idf	Idf(N=4)	Idf(N=600)
real	2	0.5	0.30103	2,477121
madrid	2	0.5	0.30103	2,477121
lineup	1	1	0.60206	2,778151
ganti	1	1	0.60206	2,778151
raphael	1	1	0.60206	2,778151
varane	1	1	0.60206	2,778151
tengah	1	1	0.60206	2,778151
minat	1	1	0.60206	2,778151
transfer	4	0.25	0	2,176091
manchester	1	1	0.60206	2,778151
united	1	1	0.60206	2,778151
tumbal	1	1	0.60206	2,778151
vinicius	1	1	0.60206	2,778151
junior	1	1	0.60206	2,778151
kylian	1	1	0.60206	2,778151
mbappe	1	1	0.60206	2,778151
update	1	1	0.60206	2,778151
main	1	1	0.60206	2,778151
sergio	2	0.5	0.30103	2,477121
aguero	2	0.5	0.30103	2,477121
sepakat	2	0.5	0.30103	2,477121
gabung	2	0.5	0.30103	2,477121
barcelona	2	0.5	0.30103	2,477121
kontrak	1	1	0.60206	2,778151
tahun	1	1	0.60206	2,778151
depan	2	0.5	0.30103	2,477121
rumor	1	1	0.60206	2,778151
musim	1	1	0.60206	2,778151

Tabel 8 menjelaskan perhitungan TF-IDF secara manual dengan menggunakan Microsoft Office Excel dari hasil perkalian *tf* dan *idf*.

Table 8. Perhitungan TF-IDF

Term (kata)	d1	d2	d3	d4
real	0.30103	0.30103		
madrid	0.30103	0.30103		
lineup	0.60206			
ganti	0.60206			
raphael	0.60206			
varane	0.60206			
tengah	0.60206			

minat	0.60206			
transfer	0.60206	0.60206	0.60206	0.60206
manchester	0.60206			
united	0.60206			
tumbal		0.60206		
vinicius		0.60206		
junior		0.60206		
kylian		0.60206		
mbappe		0.60206		
update			0.60206	
main			0.60206	
sergio			0.30103	0.30103
aguero			0.30103	0.30103
sepakat			0.30103	0.30103
gabung			0.30103	0.30103
barcelona			0.30103	0.30103
kontrak			0.60206	
tahun			0.60206	
depan			0.30103	0.30103
rumor				0.60206
musim				0.60206

Pada proses perhitungan TF-IDF ini menggunakan *library sklearn.feature_extraction.text* dan *TfidfVectorizer* untuk menjalankan proses perhitungan secara otomatis. Proses perhitungan TF-IDF dibantu menggunakan *library Multinomial Naïve Bayes* yang dimana membantu dalam mengklasifikasi teks pada data *Data Training*. Pada gambar 14 merupakan source code perhitungan TF-IDF pada jupyter notebook.

```
from sklearn.feature_extraction.text import TfidfVectorizer

d1 = "tim buruk terlalu lebih utk line up yang isi mantan top skor liga prancis jerman main mahal afrika main muda potensial"
d2 = "kurang puas barcelona siap belanja banyak main bursa transfer musin panas"

vect = TfidfVectorizer()
X = vect.fit_transform([d1, d2])

X.toarray()
```

Fig. 15. Kode perhitungan TF-IDF

Contoh hasil perhitungan TF-IDF dari sistem dapat dilihat pada Gambar 15 dan 16.

```
[(0.15904861514904767, 'afrika'),
 (0.15904861514904767, 'aku'),
 (0.15904861514904767, 'apa'),
 (0.15904861514904767, 'arti'),
 (0.0, 'banyak'),
 (0.0, 'barcelona'),
 (0.0, 'belanja'),
 (0.0, 'bursa'),
 (0.31809723029809533, 'buruk'),
 (0.15904861514904767, 'defend'),
 (0.15904861514904767, 'gagal'),
 (0.15904861514904767, 'gak'),
 (0.15904861514904767, 'isi'),
 (0.15904861514904767, 'jerman'),
 ... ..]
```

Fig. 16. Hasil Perhitungan TF-IDF Pertama

```

[(0.0, 'afrika'),
 (0.0, 'aku'),
 (0.0, 'apa'),
 (0.0, 'arti'),
 (0.3160304990863645, 'banyak'),
 (0.3160304990863645, 'barcelona'),
 (0.3160304990863645, 'belanja'),
 (0.3160304990863645, 'bursa'),
 (0.0, 'buruk'),
 (0.0, 'defend'),
 (0.0, 'gagal'),
 (0.0, 'gak'),
 (0.0, 'isi'),
 (0.0, 'jerman'),

```

Fig. 17. Hasil Perhitungan TF-IDF Pertama

Selanjutnya melakukan pencarian nilai akurasi data *training* yang telah dilakukan pelabelan secara manual. Pada gambar 17 merupakan source code untuk mencari nilai akurasi data training.

```

from sklearn.metrics import accuracy_score, f1_score, confusion_matrix

print("Accuracy: {:.2f}%".format(accuracy_score(y_test, y_pred) * 100))
print("\nF1 Score: {:.2f}".format(f1_score(y_test, y_pred, average='weighted') * 100))
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))

```

Fig. 18. Kode Akurasi Data Training

Cross validation merupakan sebuah metode untuk memperoleh hasil akurasi dengan melakukan percobaan sebanyak K kali agar nilai parameter mempunyai hasil yang sama. Prinsip *cross-validation* membagi data menjadi dua bagian, yaitu data latih dan data uji. Dengan menggunakan *library from sklearn.model_selection* dan import *ShuffleSplit* untuk menghitung rata-rata dalam 10 kali. Pada gambar 18 merupakan source code untuk menghitung *cross validation*.

```

fig, (ax1, ax2) = plt.subplots(2, 1, sharex=True, figsize=(16,9))

acc_scores = [round(a * 100, 1) for a in accs]
f1_scores = [round(f * 100, 2) for f in f1s]

x1 = np.arange(len(acc_scores))
x2 = np.arange(len(f1_scores))

ax1.bar(x1, acc_scores)
ax2.bar(x2, f1_scores, color='#559ebf')

# Place values on top of bars
for i, v in enumerate(list(zip(acc_scores, f1_scores))):
    ax1.text(i - 0.25, v[0] + 2, str(v[0]) + '%')
    ax2.text(i - 0.25, v[1] + 2, str(v[1]))

ax1.set_ylabel('Accuracy (%)')
ax1.set_title('Naive Bayes')
ax1.set_ylim([0, 100])

ax2.set_ylabel('F1 Score')
ax2.set_xlabel('Runs')
ax2.set_ylim([0, 100])

sns.despine(bottom=True, left=True) # Remove the ticks on axes for cleaner presentation
plt.show()

```

Fig. 19. Kode Cross-validation

Maka hasil grafik dari cross validation seperti pada gambar 19 berikut :

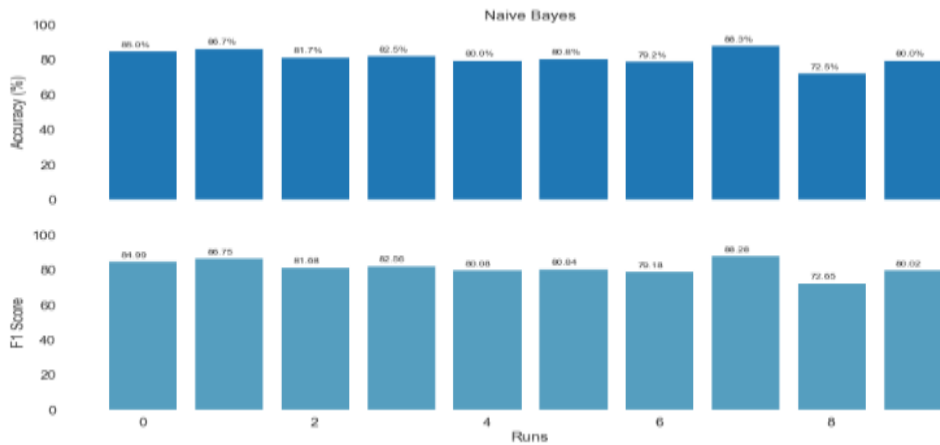


Fig. 20. Grafik Cross Validation

Selanjutnya dilakukan pembuatan model klasifikasi dengan variabel X dan Y dengan data *training* yang sudah dilakukan. Model dibuat pada sebuah fungsi agar lebih mudah dalam pemanggilannya dan dijalankan untuk tahap berikutnya, sehingga membuatnya lebih efektif dan efisien. model klasifikasi tersebut menggunakan *library sklearn.pipeline* dengan *import pipeline*. Pada gambar 20 merupakan source code import library dan pembuatan model klasifikasi

```
import os
import pickle
from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import TfidfTransformer

X = df.tweet
y = df.kelas

txt_classifier = Pipeline([('vect', TfidfVectorizer()),
                           ('tfidf', TfidfTransformer()),
                           ('classifier', MultinomialNB(alpha=1.0)),
                           ])
X_train = np.asarray(X)
txt_classifier = txt_classifier.fit(X_train, np.asarray(y))
```

Fig. 21. Import library dan pembuatan model klasifikasi

Selanjutnya file *pickle* yang sudah dibuat model klasifikasi akan digunakan untuk eksekusi data *testing* dari data yang digunakan adalah 200 *tweet* yang sudah dilakukan pelabelan secara manual dari data *training* yang berjumlah 600 *tweet* dan 200 *tweet* yang digunakan data *testing* mengambil dari total data 11.283 *tweet*. Pada gambar 21 merupakan source code untuk variable pemanggilan hasil dari prediksi Naive Bayes.

```
result_tweet=[]
for i in range(len(predicted)):
    if(predicted[i]==1):
        sentiment_result='Positif'
    elif(predicted[i]==0):
        sentiment_result='Negatif'
# result_tweet.append({'class':prediction_Linear[i], 'result_nbc':sentiment_result})
result_tweet.append({'Cleaned_Text':data_tweet[i], 'class':predicted[i] })
```

Fig. 22. Kode Pemanggilan Prediksi

4. Hasil dan Pembahasan

Pada tahap testing merupakan tahapan untuk mengetahui tingkat keakuratan pemodelan yang telah dibangun serta untuk memprediksi label atau kelas dari data uji yang tersedia. Pada gambar 22 merupakan hasil pelabelan manual.

Unnamed: 0	Cleaned_Text	actual	predicted
0	afk transfer deal done everton resmi dapat jam...	0	1
1	agen lautaro martinez datang italia barcelona ...	1	0
2	akhir saga transfer lionel messi bintang asal ...	0	0
3	aku manajemen chelsea jdi lebih bagus skrng so...	0	0
4	anchester united pasti dapat untung real madri...	1	1
...
195	video bursa transfer chelsea pinjam bintang ba...	1	0
196	wujud ingin ronald koeman belanja main petingg...	1	1
197	wujud ingin ronald koeman belanja main petingg...	1	1
198	yang takut kalo saing macam real madridjuventu...	0	0
199	yaudah lah ikhlasin aja banyak ujung tombak gu...	0	0

200 rows × 4 columns

Fig. 23. Hasil Pelabelan Manual dan Naive Bayes

Setelah mendapatkan nilai akurasi yang baik dari proses *training* dan *testing* dari pemodelan klasifikasi maka dari itu dilakukannya tahap klasifikasi untuk data keseluruhan agar mendapat hasil sentimen positif dan negatif yang telah diuji dari tahapan *training* dan *testing*. Pada gambar 23 merupakan hasil klasifikasi seluruh tweet.

tweet	class
0 mgoalcomidid spesial transfer main lima main p...	1
1 kaka liverpool mgoalcomidid spesial transfer m...	1
2 dlvr spesial transfer main bintang la liga spa...	0
3 infobarca spesial transfer main bintang la lig...	0
4 fobinews spesial transfer main bintang la liga...	0
...	...
11278 real madrid transfer buat tahun w aurelien toh...	0
11279 isco umum diri resmi tinggal real madrid statu...	1
11280 pilih transfer real madrid ganti mbappe bintan...	1
11281 mungkin serius mo salah tinggal liverpool stat...	1
11282 depan chelsea segera segel transfer dembele ba...	0

11283 rows × 2 columns

Fig. 24. Hasil Klasiikasi

Pengambilan data berupa topik mengenai keyword “Transfer La Liga”, “Transfer Real Madrid”, “Transfer Barcelona”, “Transfer Liga Spanyol” dan “Transfer Copa Del Ray”. Data tweet di ambil dari periode 1 Januari 2020 sampai dengan 31 Mei 2022, dengan jumlah data total 11.282.

Pada tahap training data yang digunakan sejumlah 600 *tweet* dengan masing-masing 300 positif dan 300 negatif yang sudah diberi label secara manual. Sedangkan untuk data testing jumlah data yang digunakan adalah 200 data yang sudah diberi label positif dan negatif. Hasil dari penelitian ini pada data testing mendapatkan nilai akurasi sebesar 85%. Sedangkan, pada hasil klasifikasi total dari “Bursa Transfer Pemain La Liga Spanyol” mendapatkan nilai 3631 ulasan Positif dan 7652 Ulasan negatif. Hasil prediksi pada data tweet keseluruhan yang sudah diketahui hasil positif dan negatif dapat dilihat pada Gambar 24.

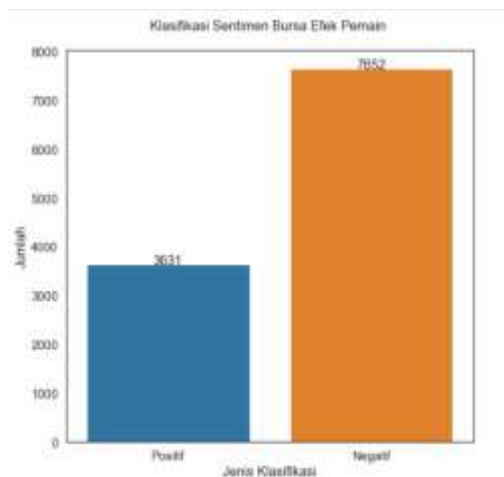


Fig. 25. Histogram Data Positif dan Negatif

Pada gambar 24 telah didapatkan hasil histogram dengan jumlah 3631 Label Positif dan 7652 Label Negatif. Pada sentimen negatif pengguna Twitter dikarenakan banyak klub sepak bola sering mengeluarkan banyak uang tetapi tidak ada prestasi demi klub sepak bola itu sendiri. Kemudian klub sepak bola juga sering salah membeli pemain, sehingga tidak menghasilkan prestasi demi klub sepak bola di tim tersebut. Berikut merupakan contoh data tweet negatif pada Tabel 9.

Table 9. Data Tweet Negatif

No	Data Tweet
1	jorjoran belanja bursa transfer musim panas juara laliga spanyol real madrid jual empat main tak guna tutup rugi finansial
2	luka modric transfer buruk liga spanyol main gelandang barca
3	gaji main transfer liga spanyol musim depan kurang akibat pandemi
4	iya bener barcelona rugi kalau kontrak messi habis tahun free transfer
5	kisruh soal lionel messi dampak transfer lautaro martinez lautaro kabar tolak barcelona

Sedangkan sentimen Positif pada data tweet banyak yang membahas tentang peluang transfer pemain, tentang penempatan posisi pemain di dalam klub sepak bola dan nilai gaji dari pemain. Berikut Contoh beberapa data *tweet* positif terdapat pada Tabel 10.

Table 10. Data Tweet Positif

No	Data Tweet
1	peluang arsenal dapat gelandang real madrid bursa transfer musim dingin buka lebar bursatransfer
2	roberto chen gabung malaga malaga prlm klub anggota primera liga spanyol malaga rampung transfer main bertahan
3	rejeki nompok real madrid proses transfer achraf hakim homealone
4	mantan asisten pelatih real madrid jose morais nilai erling haaland cocok jadi terus melanjutkan kontrak
5	real madrid mulai tabung transfer kylian mbappe

Berdasarkan hasil klasifikasi sentimen pada data *tweet*, banyak yang diklasifikasikan ke negatif mengenai pandangan pengguna Twitter terhadap transfer pemain klub sepak bola yang tidak sesuai serta tidak adanya prestasi yang membanggakan terhadap klub tersebut.

5. Kesimpulan

Kesimpulan yang dapat ditarik setelah melakukan pengolahan dan pengujian data tweet pada media sosial Twitter dengan keyword “Transfer La Liga”, “Transfer Real Madrid”, “Transfer Barcelona”, “Transfer Liga Spanyol” dan “Transfer Copa Del Ray” yang diperoleh dari periode 1 Januari 2020 sampai dengan 31 Mei 2022, dengan jumlah data total 11.282 menggunakan metode Naïve Bayes Classifier, maka dapat diambil beberapa kesimpulan seperti :

1. Pada penelitian telah berhasil mendapatkan akurasi dengan nilai 81,67 % pada data training dan 85 % untuk data testing.
2. Pada penelitian ini berhasil membuat model analisis sentimen berupa file .pickle yang dimana untuk melakukan klasifikasi dan prediksi pada data tweet untuk mendapatkan sebuah hasil sentimen positif dan negatif.

3. Pada penelitian dengan studi kasus “Bursa Transfer Pemain La Liga Spanyol” telah mendapatkan hasil sentimen negatif dari hasil klasifikasi yang telah didapatkan.

Beberapa saran yang dapat disampaikan mengenai hasil penelitian yang telah dilakukan adalah sebagai berikut:

1. Diharapkan pada proses preprocessing bisa mendapatkan data yang sesuai yang diinginkan seperti terdapat kata yang masih belum terbersihkan dengan benar.
2. Diharapkan pada penelitian selanjutnya terdapat sebuah pengembangan aplikasi yang dimana dapat mencari sebuah kata kunci yang ingin dilakukan analisis sentimen dengan jumlah data dan tanggal yang bisa di custom oleh user itu sendiri.

Acknowledgment (HEADING 5)

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g.” Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

Daftar Pustaka

- [1] Anshari, F., & Qalbie Septizar Akbar, F. (2019). Analisis Model Pengelolaan Platform New Media Oleh Klub Sepakbola Di Indonesia. *Coverage: Journal Of Strategic Communication*, 10(1), 8–18. <https://doi.org/10.35814/Coverage.V10i1.1228>
- [2] Aryashila, N. M. D. L., Haliza, S. N., Maulana, N. F., Heniawan, D. A., & Yudianto, Y. (2020). ANALISIS JARINGAN TRANSFER PEMAIN BOLA MENGGUNAKAN METODE SOCIAL NETWORK ANALYSIS (STUDI KASUS : LIGA SPANYOL (LA LIGA) TAHUN 2015 – 2020). *JBMI (Jurnal Bisnis, Manajemen, Dan Informatika)*, 17(1), 88–95. <https://doi.org/10.26487/Jbmi.V17i1.9862>
- [3] Ayuna Olienti, N. (2019). Pendidikan Dan Pengajaran Bahasa Tindak Tutur Ekspresif Dalam Twitter. 1(2), 148–155. <https://ejournal.unri.ac.id/index.php/JTUAH/>
- [4] Fajrin, S. N., Agustiyawan, A., Purnamadyawati, P., & Mahayati, D. S. (2021). Literature Review : Hubungan Koordinasi Terhadap Keterampilan Menggiring Bola Pada Pemain Sepak Bola. *Indonesian Journal Of Physiotherapy*, 1(1), 6. <https://doi.org/10.52019/Ijpt.V1i1.2605>
- [5] Fauziyyah, A. K. (2020). Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python. *Jurnal Ilmiah SINUS*, 18(2), 31. <https://doi.org/10.30646/Sinus.V18i2.491>
- [6] Gandhi, B. S., Megawaty, D. A., & Alita, D. (2021). Aplikasi Monitoring Dan Penentuan Peringkat Kelas Menggunakan Naive Bayes Classifier. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 2(1), 54–63. <https://doi.org/10.33365/Jatika.V2i1.722>
- [7] Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/Jti.V14i2.679>
- [8] Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naive Bayes Untuk Rekomendasi Pakaian Wanita. *Jurnal Informatika*, 6(1), 18–22. <https://doi.org/10.31311/Ji.V6i1.4685>
- [9] Jordan, J., Listia, N., & Daniel, K. (2021). Perancangan Desain Konten Sosial Media Tentang Sepak Bola Melalui Instagram Dengan Berbasis Microblog. *Jurnal DKV Adiwarna*, 1(18), 10.
- [10] Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Naive Bayes Classifiers. *Jurnal Coreit: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32. <https://doi.org/10.24014/Coreit.V7i1.12945>
- [11] Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). Implementasi Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Creative Information Technology Journal*, 4(2), 151. <https://doi.org/10.24076/Citec.2017v4i2.106>
- [12] Naf'an, M. Z., Burhanuddin, A., & Riyani, A. (2019). Penerapan Cosine Similarity Dan Pembobotan TF-IDF Untuk Mendeteksi Kemiripan Dokumen. *Jurnal Linguistik Komputasional*, 2(1), 23–27.

- [13] Natadjaya, P., Raihanto, M. F., & Tiara Kania Ladzuardini, H. (2020). Tinjauan Implementasi Prinsip Koperasi Pada Klub Sepak Bola Indonesia. *Jurnal Ilmiah Akutansi Dam Keuangan*, 2(2), 163–182.
- [14] Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat Pada Media Sosial Twitter Menggunakan Metode K-Nearest Neighbor Dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 1(12), 1750–1757.
- [15] Nurwahyuni, S. (2019). Analisis Sentimen Aplikasi Transportasi Online Krl Access Menggunakan Metode Naive Bayes. *Swabumi*, 7(1), 31–36. <https://doi.org/10.31294/Swabumi.V7i1.5575>
- [16] Rachman, F. F., & Pramana, S. (2020). Analisis Sentimen Pro Dan Kontra Masyarakat Indonesia Tentang Vaksin COVID-19 Pada Media Sosial Twitter. *Health Information Management Journal*, 8(2), 100–109. <https://inohim.esaunggul.ac.id/index.php/INO/Article/View/223/175>
- [17] Rahmadhika, M. K., & Thantawi, A. M. (2021). Rancang Bangun Aplikasi Face Recognition Pada Pendekatan CRM Menggunakan Opencv Dan Algoritma Haarcascade. *IKRA-ITH INFORMATIKA: Jurnal Komputer Dan Informatika*, 5(1), 109–118.
- [18] Sari, E. D. N., & Irhamah, I. (2020). Analisis Sentimen Nasabah Pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naive Bayes Classifier (NBC), Dan Support Vector Machine (SVM). *Jurnal Sains Dan Seni ITS*, 8(2). <https://doi.org/10.12962/J23373520.V8i2.44565>
- [19] Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF Dan K-Nearest Neighbor. *Journal Of Intelligent System And Computation*, 1(1), 43–49. <https://doi.org/10.52985/Insyst.V1i1.36>
- [20] Sudiantoro, A. V., Zuliarso, E., Studi, P., Informatika, T., Informasi, F. T., Stikubank, U., & Mining, T. (2018). Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier. *Dinamika Informatika*, 10(2), 398–401.
- [21] Tuhuteru, H., & Iriani, A. (2018). Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine Dan Naive Bayes Classifier. *Jurnal Informatika: Jurnal Pengembangan IT*, 3(3), 394–401. <https://doi.org/10.30591/Jpit.V3i3.977>
- [22] Utama, C. A., & Widodo, A. (2021). Analisis Penyebab Kekalahan Dari Faktor Kegagalan Shooting On Target Manchester United Vs Paris Saint Germain Di Uefa Champions League 2020. *Jurnal Kesehatan Olahraga*, 09(04), 129–138.
- [23] Wibowo, Wahyu, & Novianti, Wahyu Ela. (2022). Analisis Sentimen Pengguna Twitter Terhadap Program Kartu Prakerja Di Tengah. 11(1).
- [24] Zulfikar, W. B., & Lukman, N. (2016). Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata. *Jurnal Online Informatika*, 1(2), 82–86. <https://doi.org/10.15575/Join.V1i2.33>