

## ANALISIS SENTIMEN KOMENTAR *NETIZEN* INSTAGRAM TERHADAP *RACISM* DI SEPAK BOLA INDONESIA DENGAN METODE NAIVE BAYES

Muhammad Ikhsan Yusuf<sup>1)</sup>

Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Halu Oleo,  
yusufmuhammadikhsan@gmail.com

Natalis Ransi<sup>2)</sup>, Arman<sup>3)</sup>, Andi Tenriawaru<sup>4)</sup>, La Ode Saidi<sup>5)</sup>, La Surimi<sup>6)</sup>

Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Halu Oleo, <sup>2</sup>natalis.ransi@uho.ac.id,  
<sup>3</sup>arman.mtmk@uho.ac.id, <sup>4</sup>atenriawaru36@gmail.com, <sup>5</sup>lm.saidi@yahoo.ac.id, <sup>6</sup>lasurimi@uho.ac.id,

### ABSTRAK

Penggunaan kalimat *racism* yang terjadi dalam dunia Sepak bola, dikarenakan sepak bola adalah sarana bentuk olahraga yang sangat digemari di berbagai belahan dunia dan hampir semua ras mengetahui jika sepak bola itu olah raga yang sangat disukai oleh berbagai kalangan dan golongan, hal tersebut membuat peneliti tertarik untuk melakukan penelitian dengan judul “Analisis Sentimen Komentar *Netizen* Instagram Terhadap *Racism* di Sepak Bola Indonesia Dengan Metode Naive Bayes” Penelitian ini dilakukan sebagai salah satu bentuk informasi agar melihat tingkat sentiman masyarakat di media sosial Instagram terhadap *Racism*.

Penelitian ini dengan menggunakan metode *Naive Bayes Classifier*. Percobaan dilakukan pada 650 data komentar. Enam ratus lima puluh komentar data tersebut dibagi menjadi 2 kategori. Untuk mengetahui pengaruh jumlah data latih terhadap efektifitas *Naive Bayes Classifier* maka diambil beberapa kombinasi banyaknya dokumen latih dan dokumen uji. Secara berurutan kombinasi dokumen latih dan uji tersebut adalah 90:10, 80:20, dan 70:30.

Dengan menggunakan metode *Naive Bayes Classifier* diperoleh *accuracy* hasil klasifikasi kombinasi dokumen latih dan uji tersebut adalah 90:10, 80:20, dan 70:30 secara berurut sebesar 82%, 77%, dan 80%. Berdasarkan hasil pengujian yang telah dilakukan, *accuracy* tertinggi diperoleh pada pembagian 90:10 dengan nilai *accuracy* 82%. Aplikasi ini dapat dikembangkan untuk platform *social media* lainnya seperti facebook, instagram, path, kaskus, dan lainnya.

**Kata Kunci** – Sepak Bola, *Racism*, Media Sosial, Komentar, *Naive Bayes Classifier*

### ABSTRACT

*The use of the phrase racism that occurs in the world of football, because football is a very popular form of sport in various parts of the world and almost all races know that soccer is a sport that is highly favored by various groups and groups, this makes researchers interested in conducting research. with the title "Instagram Netizen Comment Sentiment Analysis Against Racism in Indonesian Football Using the Naive Bayes Method" this research was conducted as a form of information in order to see the level of public sentiment on Instagram social media towards Racism.*

*In this study using the Naive Bayes Classifier. The experiment was carried out on 650 comment data. Six hundred and fifty data comments are divided into 2 categories. To determine the effect of the amount of training data on the effectiveness of the Naive Bayes Classifier, several combinations of the number of training documents and test documents were taken. Sequentially the combination of the training and test documents included 90:10, 80:20, 70:30.*

*By using the Naive Bayes Classifier method accuracy of the classification results is 82%, 77% and 80%. Based on the results of the tests that have been carried out, accuracy is obtained in the 90:10 division with an accuracy 82%. This application can be developed for other social media platforms such as Facebook, Instagram, Path, Kaskus, and others.*

**Keywords:** Football, *Racism*, Social Media, Comments, *Naive Bayes*

### 1. Pendahuluan

Pemakaian bahasa dalam kehidupan sehari-hari merupakan hal yang berguna untuk melakukan komunikasi sesama manusia dan dapat juga digunakan untuk mengukir beberapa besar intelektual yang dikuasainya. Penggunaan bahasa tidak sekedar berkomunikasi secara

langsung tetapi dapat juga menggunakan jejaring 187 sosial contohnya: instagram, twitter, facebook, whatsapp dan sebagainya. hal ini yang menyebabkan pengguna internet melonjak (Rahmawati Nur Arifah, 2020)

Pengguna internet pada tahun 2019 tercatat sebanyak 171 juta pengguna internet di Indonesia setiap tahun pengguna internet tumbuh 10,2% atau 27 juta jiwa. 64,8%

jumlahituadalahpengguna internet aktif yang mengakseskontenhiburanseperti video dan film halitudisampaikan oleh DirjenPenyelenggara Pos dan InformatikaKementriankominfo Ahmad Muhammad Ramli, 3 juli 2019. Pemakaian media sosialsendiriuntukberbagi info tentangkeadaansaatininamunbanyak yang menggunakan media sosialuntukberkomentarburukterhadap orang lain. Penggunaankalimatsarkasmebanyakdijumpai pada jejaring sosial. Hal inilah yang memiciperselisihanantarsesama(Cahyanti & Sabardila, 2020).

Dalam media sosialbanyakmasyarakat yang menggunakan media sosialuntukmelakukan proses komunikasiterutama pada era digital sekarang. *Racism*sendirimerupakan gayabahasa yang mengandungcacia, hinaansertacelaan yang kurangenakdidengarsertamengandungkepahitan yang merendahkan status sosialmaupun yang lainyaitudapatmenyebabkankerugiankepada orang lain. Hal tersebutbisaterjadidengansengajaataukurangnyapema

hamanataubahkanketidaktahuan(Lestari, 2015).

*Racism*dapat terjadidisebuah elemenmasyarakat, khususnya dalam sepak bola, *racism*biasaterjadidaripenggemar yang melakukan kata-kata rasialterhadapklub dan pemain. *Racism*sudahmenjadimasalahumum di Eropa, dengannisminoritas dan pendatang yang paling banyakmendapatpelecehan dan diskriminasi(Tirahmawan dkk., 2021).

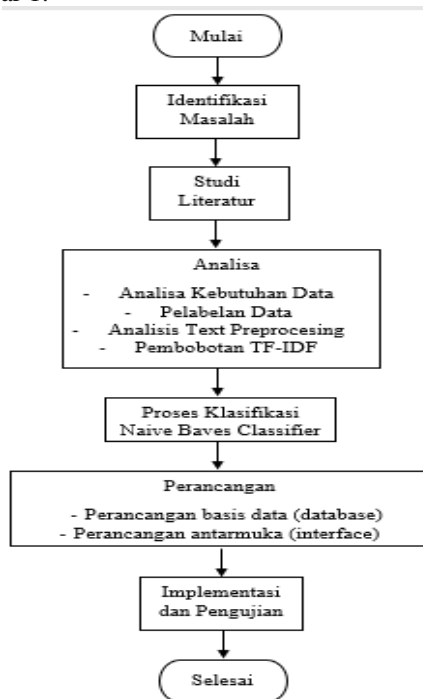
Penggunaankalimat*racism* yang terjadidalam dunia Sepak bola, dikarenakansepak bola adalahsarabentukolahraga yang sangat digemaridiberbagaibelahan dunia dan hampirsemuarasmengetahujiikasepak bola ituolahraga yang sangat disukai oleh berbagaikalangan dan golongan. Sepak bola sendiritelahmemilikiketerkaitandalamberbagaiaspek, mulaidari aspek Ekonomi, politik, budaya, dan sosial. Sehinggaaltersebutmembuatpenelititertarikuntukmelakukanpenelitiandenganjudul“*Analisis*

**SentimenKomentar Netizen Instagram Terhadap Racism Di Sepak Bola Indonesia Dengan Metode Naive Bayes**”penelitianinidilakukansebagai salah satubentukinformasi agar melihattingkatsentimanmasyarakat di media sosial Instagram terhadap*Racism*.

## 2. Metode

Metodologi pada penelitianiniterdiridari beberapatahapandiantaranyaid entifikasi data masukan, analisa, pelabelan manual, *textpreprocessing*, pembobotan TF-IDF, proses

klasifikasi*Naive Bayes Classifier* dan perancangan. Datamasukanberupa data latih dan data uji yang digunakanbersumberdari akun Instagram PSM Makassar danPersija. Pada data-data tersebutdilakukan*preprocessing*yaitucase folding, cleaning, tokenizing, filtering, dan stemming. Data yang telahdiproses padatextpreprocessingdiproses kembalimenggunakan metode*Naive Bayes Classifier*. Keluarandari proses tersebutyaituinformasisentimen. Penjelasantahapanpenelitiandapatdilihat pada Gambar 1.



Gambar 1 Tahapan Penelitian

### A. Text Preprocessing

Pada tahap analisis *text mining*, teks yang akan digunakan dipersiapkan lebih dulu, setelah itu baru digunakan pada proses utama. Proses mempersiapkan teks dokumen disebut juga dengan proses *text preprocessing*. *Text preprocessing* berguna untuk mengubah data teks yang tidak terstruktur menjadi terstruktur (C Sutami, 2015). Berikut ini proses tahap *text preprocessing*: *Case Folding*, *cleaning*, *tokenizing*, *filtering*, *stemming*.

### B. Pembobotan TF-IDF

Setelah proses *preprocessing*, perlu ekstraksi fitur untuk mendapatkan hasil klasifikasi yang lebih maksimal. Ekstraksi fitur berfungsi untuk mengurangi *noise* dengan menghapus *feature* yang tidak relevan, sehingga dapat meningkatkan akurasi klasifikasi. Pada penelitian (Pratama & Bambang, 2015) ekstraksi fitur dapat dilakukan dengan melakukan pengamatan terhadap distribusi frekuensi kemunculan kata

dan jumlah *feature*. Ekstraksi berfungsi untuk mendapatkan nilai *threshold* parameter dalam klasifikasi NBC nantinya. Hal tersebut dilakukan dengan melakukan pengamatan terhadap distribusi frekuensi kemunculan kata dan jumlah *feature* (Pratama, 2015). Pendekatan ekstraksi yang dilakukan dalam penelitian ini yaitu *Term Frequency Inverse Document Frequency* (TF-IDF). Perhitungan *Term Frequency* (TF).

### C. Naive Bayes Classifier

Algoritma *Naive Bayes Classifier* merupakan algoritma metode pengklasifikasi turunan dari teorema Bayes yang sederhana, berguna untuk mencari nilai probabilitas atau peluang tertentu untuk mengklasifikasi data *testing* (uji) pada kategori yang paling tepat (Feldman & Sanger, 2006). Menurut (Kusumadewi, 2009) metode *naive bayes classifier* adalah algoritma klasifikasi yang sangat efektif (mendapatkan hasil yang tepat) dan efisien (proses penalaran dilakukan memanfaatkan input yang ada dengan cara yang relatif cepat). Metode *Naive Bayes* memanfaatkan teori probabilitas yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi *probabilitas* di masa depan berdasarkan pengalaman di masa sebelumnya. Tujuan algoritma NBC (Kusumadewi, 2009) yaitu untuk melakukan klasifikasi data pada kelas tertentu. Kelebihan NBC adalah sederhana tetapi memiliki akurasi yang tinggi (Saputra dkk., 2019).

### D. Perancangan

Tahap ini merupakan tahap untuk memulai perancangan aplikasi berdasarkan analisa yang telah dilakukan sebelumnya. Adapun tahapannya sebagai berikut:

#### 1. Basis data (*database*)

Tahap ini dilakukan untuk perancangan basis data yang berisikan tabel, field, dan atribut untuk melengkapi kebutuhan aplikasi.

#### 2. Antarmuka (*interface*)

Tahap ini dilakukan untuk membuat pilihan menu yang akan ditampilkan pada aplikasi. User interface (antarmuka). Tahap ini dilakukan bertujuan menciptakan tampilan yang mudah digunakan oleh user (pengguna).

### E. Pengujian

Pada tahap ini aplikasi telah selesai dan dapat dilakukan pengujian.

Berikut merupakan penjelasan perangkat yang dibutuhkan dalam pembuatan aplikasi:

- Hardware
  - Processor : Intel Inside
  - Memory : 2 Gb

### 2. Software

- Sistem operasi : Windows 10
- Bahasa Pemrograman : PHP
- Tools : Notepad, Microsoft Visio
- Database : Mysql

Setelah dilakukan implementasi maka dilakukan tahap pengujian terhadap aplikasi yang telah dirancang. Tahap pengujian dilakukan untuk tujuan mengetahui bagaimana kinerja metode.

## 3. Hasil dan Pembahasan

### 3.1 Analisis Data

Akun Instagram yang akan digunakan untuk pengumpulan data pada penelitian ini: *Psm\_makassar* dan *persija*. Dari nama akun di atas akan dipilih dataset sebanyak 650 komentar menggunakan bahasa Indonesia. Klasifikasi terdiri dari 3 kelas dengan komentar 288 positif, 196 komentar negatif, dan 166 komentar netral (termasuk iklan dan promosi akudagang).

#### A. Pengambilan Data

Pengambilan data dilakukan dengan menggunakan bantuan ekstensi *Google Chrome*, yaitu *Data Scraper*. Ekstensi *Data Scraper* merupakan salah satu ekstensi pada *Google Chrome* yang berfungsi sebagai *webscraper* untuk mengambil data atau konten yang terdapat pada situs online secara gratis. *Data Scraper* mengekstraksi data dari halaman web HTML dan mengimpornya ke dalam *spreadsheet* Microsoft Excel.

Sebelum menggunakan ekstensi *Data Scraper*, langkah pertama yang harus dilakukan yaitu menginstall dan mengaktifkannya ke dalam *Google Chrome*. Setelah *Data Scraper* terinstall, selanjutnya membuat daftar komentar dari Instagram yang telah ditentukan untuk diambil komentarnya. Sebagai contoh pada Gambar 2 komentar di akun Instagram.



Gambar 2 Komentar di Instagram

Selanjutnya, pengambilan data dimulai dari mengaktifkan *Data Scraper* dan

membuat sebuah query untuk menentukan bagian mana saja yang akan diambil datanya. Kemudian untuk melakukan scraping data dapat dimulai dengan mengklik tombol scrape in data miner.

Setelah semua komentar terkumpul, selanjutnya menyimpan hasil scraping data ke dalam file excel dengan format xlsx. Selanjutnya dilakukan pelabelan. Hasil scraping data dan labelling data dapat dilihat pada Gambar 3.

Gambar 3 Hasil Scraping Data dan Labelling Data

Pada proses pengambilan data yang dilakukan bersumber situs aplikasi Instagram untuk khasus racism di sepak bola Indonesia, dihasilkan data komentar dengan jumlah 650 komentar.

Data input yang digunakan adalah komentar netizen Instagram terhadap racism di sepak bola Indonesia. Perbandingan dataset yang digunakan adalah 90% data latih : 10% data uji, 80% data latih : 20% data uji dan 70% data latih : 30% data uji. Tabel 1, Tabel 2, dan Tabel 3 berisikan rincian dataset yang digunakan.

Tabel 1 Rincian Dataset 90% Data Latih : 10% Data Uji

Data Komentar	Data Latih	585(90%)		
	Data Uji	65(10%)		
		Latih	Uji	Jumlah
Kelas	Positif	254	26	280
	Negatif	174	29	203
	Netral	157	10	167
	Jumlah	585 Komentar	65 Komentar	650 Komentar

Tabel 2 Rincian Dataset 80% Data Latih: 20% Data Uji

Data Komentar	Data Latih	520(80%)		
	Data Uji	130(20%)		
		Latih	Uji	Jumlah
Kelas	Positif	228	52	280
	Negatif	254	49	203
	Netral	238	29	167

	Jumlah	520	130	650
--	--------	-----	-----	-----

Tabel 3 Rincian Dataset 70% Data Latih : 30% Data Uji

Data Komentar	Data Latih	455(70%)		
	Data Uji	195(30%)		
		Latih	Uji	Jumlah
Kelas	Positif	205	75	280
	Negatif	125	78	203
	Netral	125	42	167
	Jumlah	455	195	650

**B. Pelabelan Data**

Proses pelabelan komentar pada dataset dilakukan secara manual dengan cara memberikan label pada database Instagram dengan komposisi 90% : 10%, dengan 80% : 20%, dan 70% : 30% dari 650 komentar. Untuk setiap komentar diberi label oleh pakar Bahasa Indonesia.

**C. Analisis Text Preprocessing**

Preprocessing merupakan sebuah tahap dalam membersihkan data yang digunakan dalam penelitian. Adapun tahap yang terdapat pada Preprocessing yaitu, Case Folding, Cleaning, Tokenizing, Filtering, Stemming.

• Proses Case Folding

Merupakan tahapan untuk proses mengubah seluruh huruf pada data menjadi huruf kecil. Adapun hasil dari proses case folding ditunjukkan oleh Tabel 4.

Tabel 4 Proses Case Folding

Komentar sebelum case folding	Komentar sesudah case folding	Kelas	Komentar
Penjarakanaja yang rasisitu... Janganhanyamint ama after selesai... Biar jeradia... Biar lisannya di jaga.....	penjarakanaja yang rasisitu... janganhanyamint ama after selesai... i... biarjeradia... biarlisannya di jaga.....	Positif	Its.my_fajar
Suporterculunygr asisbiasangelem	suporterculunygr asisbiasangelem	Negatif	eky_lee
Btw siapa yang rasis min? Barukeluar dari goa maklum	btw siapa yang rasis min? barukeluar dari goa maklum	Netral	n.fhalxy

• Proses Cleansing

Merupakan proses untuk membersihkan data dari bagian yang tidak memiliki hubungan dengan informasi yang ada pada data dan penghilang tandabaca. Adapun Hasil dari proses case folding ditunjukkan oleh Tabel 5.

Gambar 5 Proses Cleansing

KomentarsebelumCleansing	KomentarSesudahCleansing	Kelas	Komentator
penjarakanaja yang rasisitu... janganhanyamintama afterusselesai... biarjeradia... biarlisannya di jaga.....	penjarakanaja yang rasisitujanganhany amintamaafterusse lesaibiarjeradiabia rlisannya di jaga	Positif	Its.my_fajar
suporterculunygrasis biasangelem	suporterculunygrasis biasangelem	Negatif	eky_lee
btw siapa yang rasis min? barukeluar dari goa maklum	btw siapa yang rasis min barukeluar dari goa maklum	Netral	n.fhalxy

• Proses *Tekonizing*

Adalah proses memisahkan setiap kata didalam data yang awalnya berpakalimat menjadi kata. Adapun hasil dari proses *Tokenizing* dapat dilihat pada Tabel 6.

Tabel 6 Proses *Tokenizing*

Data 1	Data 2	Data 3
penjarakan	suporter	btw
aja	culun	siapa
yang	Yg	yang
rasis	rasis	rasis
itu	biasa	min
jangannya	ngelem	baru
hanya		keluar
minta		dari
maaf		goa
terus		maklum
selesai		
biar		
jera		
dia		
biar		
lisannya		
di		
jaga		

• Proses *Filtering*

Adalah tahapan membuang kata yang tidak penting dengan menggunakan *stopwords*. Adapun hasil dari proses *remove stopwords* dapat dilihat pada Tabel 7.

Tabel 7 Proses *Filtering*

Data sebelum di filtering			Data setelah di filtering		
Data 1	Data 2	Data 3	Data 1	Data 2	Data 3
penjarakan	suporter	btw	penjarakan	suporter	btw
aja	culun	siapa	rasis	culun	rasis
yang	yg	yang	maaf	rasis	goa
rasis	rasis	rasis	selesai	ngelem	maklum
itu	biasa	min	jera		
jangannya	ngelem	baru	lisannya		
hanya		keluar	jaga		
minta		dari			
maaf		goa			
terus		maklum			
selesai					
biar					
jera					
dia					
biar					

• Proses *Stemming*

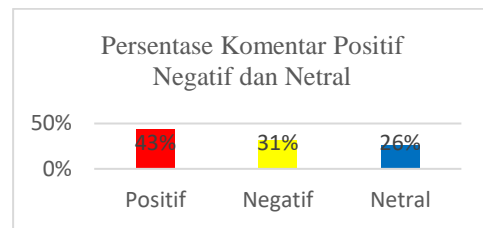
Adalah tahapan pengubahan kata menjadi kata dasarnya. Algoritma yang digunakan adalah ECS (*Enhanced Confix Stripping*). Adapun hasil dari proses *stemming* dapat dilihat pada Tabel 8.

Tabel 8 Proses *Stemming*

Data 1	Data 2	Data 3
penjara	suporter	btw
ras	culun	ras
maaf	ras	goa
selesai	lem	maklum
jera		
lisan		
jaga		

D. Visual Data

Tahap visualisasi data telah dilakukan dengan menggunakan *Worldcloud* dan Diagram Batang. Visualisasi data dengan *Worldcloud* menghasilkan kata yang sering muncul dalam teks yang telah dianalisis sebelumnya, penggunaan *font-size* huruf yang berbeda menggambarkan frekuensi kemunculan kata pada data yang dianalisis. Presentase data komentar positif, negatif, dan netral ditunjukkan pada Gambar 4.



Gambar 4 Presentase Data Komentar Positif, Negatif, dan Netral

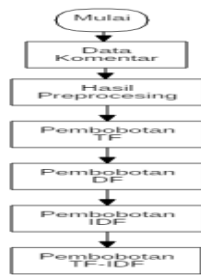
E. *Wordcloud* Komentar Data

*Wordcloud* merupakan representasi grafis dari sebuah dokumen yang dilakukan dengan *plotting* kata-kata yang sering muncul pada sebuah dokumen pada ruang dua dimensi.

F. Pembobotan TF-IDF

Pembobotan adalah proses mengubah kata menjadi bentuk *kaata* atau *vector*. TF-IDF berfungsi untuk menentukan nilai frekuensi sebuah kata di dalam sebuah dokumen dan pemberian bobot pada

setiap kata pada setiap dokumen untuk mencari dan menghitung berapa kali kata itu muncul.



**Gambar 5** Flowchart Pembobotan TF-IDF

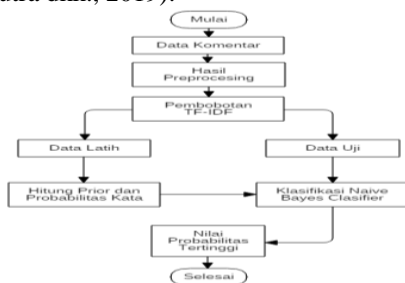
Flowchart Pembobotan TF-IDF ditunjukkan pada Gambar 5. Sedangkan Hasil pembobotan kata ditunjukkan pada Gambar 6.

Kata	D1 TF	D2 TF	D3 TF	TF Normalisasi			IDF <sup>log(Cd/Td)</sup>	TF-IDF			
				D1	D2	D3		Data1	Data2	Data3	
penjara	1	0	0	0,142857143	0	0	1	0,477121255	0,068160179	0	0
ras	1	1	1	0,142857143	0,25	0,25	3	0	0	0	0
maaf	1	0	0	0,142857143	0	0	1	0,477121255	0,068160179	0	0
selesai	1	0	0	0,142857143	0	0	1	0,477121255	0,068160179	0	0
jera	1	0	0	0,142857143	0	0	1	0,477121255	0,068160179	0	0
lisan	1	0	0	0,142857143	0	0	1	0,477121255	0,068160179	0	0
laga	1	0	0	0,142857143	0	0	1	0,477121255	0,068160179	0	0
superte	0	1	0	0	0,25	0	1	0,477121255	0	0,11928	0
culum	0	1	0	0	0,25	0	1	0,477121255	0	0,11928	0
jen	0	1	0	0	0,25	0	1	0,477121255	0	0,11928	0
bhv	0	0	1	0	0	0,25	1	0,477121255	0	0	0,11928
gea	0	0	1	0	0	0,25	1	0,477121255	0	0	0,11928
maklum	0	0	1	0	0	0,25	1	0,477121255	0	0	0,11928
Jumlah	7	3	3								

**Gambar 6** Pembobotan Kata

**G. Proses Klasifikasi Naive Bayes Classifier**

Pada proses ini akan menjelaskan proses-proses yang akan dilakukan dalam klasifikasi data. Gambar 7 merupakan Flowchart dari Naive Bayes Classifier (Saputra dkk., 2019).



**Gambar 7** Flowchart Klasifikasi Naive Bayes Classifier

**3.2 Implementasi**

Implementasi adalah tahapan yang dilakukan untuk menerapkan perancangan yang telah dilakukan pada tahapan sebelumnya.

**A. Lingkup Implementasi**

Dasarnya setiap desain aplikasi yang dirancang memerlukan sarana pendukung untuk memudahkan dalam menyelesaikan aplikasi yang akan dibuat. Sarana yang dimaksud seperti hardware yaitu kebutuhan perangkat eraskomputer dan software yang

digunakan untuk mengoperasikan aplikasi yang dibuat. Berikut ini beberapa sarana pendukung yang digunakan untuk mengimplementasikan aplikasi yang akan didesain yaitu:

1. Perangkat Keras
  - Prosesor : Intel Core i5-7200U
  - Memory : 4 GB DDR4
2. Perangkat Lunak
  - Sistem operasi : Windows 10
  - Bahasa Pemrograman : PHP
  - Tools : Notepad ++, Microsoft Visio
  - Web Server : Apache

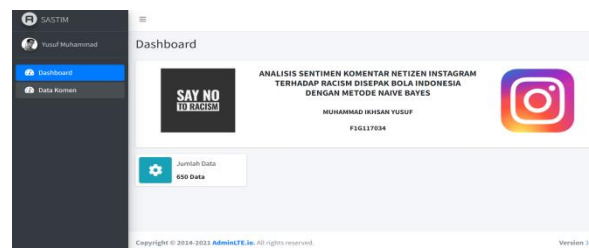
**B. Implementasi Aplikasi**

Tahapan implementasi Aplikasi ini merupakan tahap implementasi hasil perancangan antarmuka yang telah dibuat.

Berikut ini merupakan implementasi aplikasi klasifikasi sentimen berdasarkan komentar netizen terhadap racism di sepak bola Indonesia.

**1. Implementasi Antarmuka Halaman Dashboard**

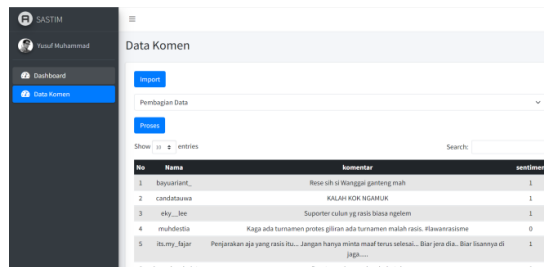
Dashboard memuat informasi tentang judul penelitian dan nama penelitian. Menu-menu tersebut yaitu dashboard, data komentar, data testing dan data training. Tampilan antarmuka menu dashboard dapat dilihat seperti Gambar 8 berikut ini.



**Gambar 8** Implementasi Antarmuka Halaman Dashboard

**2. Implementasi Antarmuka Halaman Data Komentar**

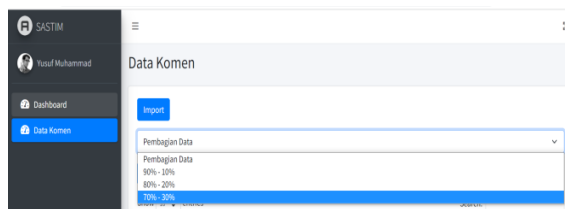
Halaman ini berisi data training informasi seluruh komentar yang datanya sudah diinput ke dalam sistem. Tampilan antarmuka data komentar dapat dilihat seperti pada Gambar 9.



**Gambar 9** Implementasi Antar Muka Data Komentor

3. ImplementasiAntarmuka Halaman Pembagian Data

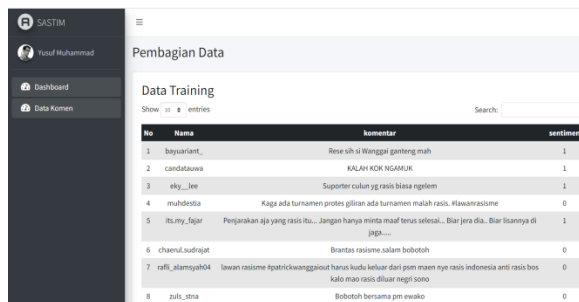
Pada halamaninitedakadabatasanuntukmembagi data *testing* yang akandigunakan. Proses data *training* 90 yang berartiapabilapenggunamemasukkanangka 90 makapenggunaanmelakukan proses *training* untuk 90% data, begitu juga dengan proses data *training* 80 dan proses data *training* 70. Dapat dilihat pada Gambar 10.



**Gambar 10** ImplementasiAntarmuka Halaman Pembagian Data

4. ImpementasiAntarmuka Pada Halaman Data Training

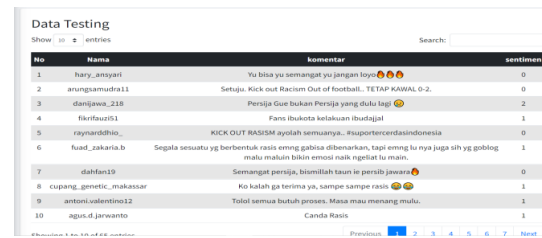
Pada halamaniniberisi data *training* yang nantiakandijadikanebagaibahanuntuk data *testing*. ImplementasiAntarmuka Halaman Data *Training* ditunjukkan pada Gambar 11.



**Gambar 11** ImplementasiAntarmuka Halaman Data Training

5. ImplementasiAntarmukahalaman Data Testing

Halaman inisistemakanmelakukantesting pada data apakahsesuaihasilnyadengan yang ada di data *Training*. ImplementasiAntarmuka Data *Testing* ditunjukkan pada Gambar 12.



**Gambar 12** ImplementasiAntarmuka Data Testing

C. Pengujian *Confusion Matrix*

Pada tahap ini,

dilakukanbeberapaskenariopengujian, yaitudenganpembagian data latih dan data uji dari total 650 komentar yang tersediadenggunakan 90% data latih dan 10% data uji, 80% data latih dan 20% data uji, 70% data latih dan 30% data uji (Puspitasari dkk, 2018). Pembagian data uji dan data latih pada setiap skenario pengujian ditunjukkan pada Tabel 8.

Pengujian dilakukandengan metode *Confusion Matrix*. Adapun hasil pengujian untuk setiap skenario ditunjukkan pada Tabel 9.

**Tabel 8** Pembagian Data untuk Setiap Skenario Pengujian

No	Latih %	Uji %	Latih Pos	Uji Pos	Latih Neg	Uji Pos	Latih Net	Uji Net
1	90	10	244	33	181	23	160	9
2	80	20	221	56	150	54	149	20
3	70	30	194	83	129	75	132	37

**Tabel 9** Hasil Pengujian *Confusion Matrix* untuk Setiap Skenario Pengujian

1.	<b>PENGUJIAN 90% DAN 10%</b>	
	<i>Confusion Matrix Data Training</i>	
	<i>Accuracy</i>	72.8%
	<i>Recall</i>	73.3%
	<i>Precision</i>	70.0%
	<i>Confusion Matrix Data Testing</i>	
2.	<b>PENGUJIAN 80% DAN 20%</b>	
	<i>Confusion Matrix Data Training</i>	
	<i>Accuracy</i>	72.9%
	<i>Recall</i>	73.3%
	<i>Precision</i>	70.0%
	<i>Confusion Matrix Data Testing</i>	
3.	<b>PENGUJIAN 70% DAN 30%</b>	
	<i>Confusion Matrix Data Training</i>	
	<i>Accuracy</i>	71.2%
	<i>Recall</i>	70.0%
	<i>Precision</i>	70.0%
	<i>Confusion Matrix Data Testing</i>	
<i>Accuracy</i>		80.0%

	Recall	76.7%
	Precision	76.7%

Pada pengujian yang menggunakan 585 data *training* dan 65 data *testing* diperoleh *accuracy* pada data *training* sebesar 72% dan data *testing* sebesar 82%. Pada pengujian yang menggunakan 530 data *training* dan 130 data *testing* diperoleh *accuracy* pada data *training* sebesar 72% dan data *testing* sebesar 77%. Sedangkan pada pengujian yang menggunakan 455 data *training* dan 195 data *testing* diperoleh *accuracy* pada data *training* sebesar 71% dan data *testing* sebesar 80%.

#### 4. Penutup

##### 4.1 Kesimpulan

Berdasarkan pembahasan pada bagian 3, hal-hal yang dapat disimpulkan adalah sebagai berikut:

1. Tingkat sentimen netizen terhadap *practicism* di sepak bola Indonesia dengan menggunakan metode *Naive Bayes Classifier* diperoleh 650 data komentar yang terdiri dari, 43% komentar positif, 31% komentar negatif dan 26% komentar netral.
2. Metode *Naive Bayes Classifier* dapat diterapkan dengan baik untuk melakukan klasifikasi sentimen pada analisis komentar netizen Instagram terhadap *practicism* di sepak bola Indonesia menggunakan metode *Naive Bayes Classifier*.
3. Metode *Naive Bayes Classifier* mudah untuk di terapkan dan memiliki hasil yang bagus.

##### 4.2 Saran

Adapun saran yang dapat diberikan untuk penelitian lanjutan yang terkait dengan penelitian ini adalah Aplikasi klasifikasi komentar ini bisa dikembangkan dengan menggunakan metode-metode klasifikasi lain.

#### REFERENSI

[1] Aaputra, S. A. (2019). Sentiment Analysis Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 377–382.

[2] Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>

[3] Cahyanti, A. S., & Sabardila, A. (2020). Analisis

Penggunaan Kalimat Sarkasme oleh Netizen di Media Sosial Instagram. *Lingua Franca: Jurnal Bahasa, Sastra, Dan Pengajarannya*, 186–195.

[4] Kurniasari, I., & Fatta, H. Al. (2021). Analisis Sentimen Opini Publik pada Instagram mengenai Covid-19 dengan SVM. *JTECS: Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem & Komputer*, 1(1), 67–74.

[5] Lestari, D. P. (2015). Analisis Strategi Internet Marketing Butik Online Di Surabaya Melalui Instagram. *Jurnal Commonline Departemen Komunikasi*, 4(2), 412–424.

[6] Musfiroh, D., Khaira, U., Eko, P., Utomo, P., & Suratno, T. (2021). Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(April), 24–33.

[7] Puspitasari, A. A., Santoso, E., & Indriati. (2018). Klasifikasi Dokumen Tumbuhan Obat Menggunakan Metode Improved k-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(2), 486–492.

[8] Rahayu, S., Kusriani, & Sismoro, H. (2018). Sentimen Analisis Review Pengguna Marketplace Online Menggunakan Naive Bayes Classifier. *Jurnal Informatika Dan Teknologi Informasi*, 3(3), 181–186.

[9] Sholeh, M. (2014). Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 Yogyakarta, 15 November 2014 ISSN: 1979-911X. *Snast, November*, 211–216

[10] Tirahmawan, J., Melody, B. A. L., & Ahly, M. N. N. (2021). Rasisme Terhadap Kulit Hitam dalam Iklan H&M. *Jurnal Audiens*, 2(1). <https://doi.org/10.18196/jas.v2i1.8708>

Diterima pada tanggal 1 September 2022.  
Terbit online pada tanggal 30 Desember 2022