

**ANALISIS SENTIMEN KOMENTAR AKUN *INSTAGAM*  
*PUBLIC FIGURE* POLITIK TERKAIT PEMBATALAN  
PIALA DUNIA U-20 DENGAN ALGORITMA K-  
*NEAREST NEIGHBORS* (KNN)**

**SKRIPSI**

**Oleh :**

**RIZKY AFFANDI**

**198160068**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MEDAN AREA  
2023**

**UNIVERSITAS MEDAN AREA**

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 24/6/24



Dipindai dengan CamScanner

Access From (repository.uma.ac.id)24/6/24

**ANALISIS SENTIMEN KOMENTAR AKUN *INSTAGAM*  
*PUBLIC FIGURE* POLITIK TERKAIT PEMBATALAN  
PIALA DUNIA U-20 DENGAN ALGORITMA *K-  
NEAREST NEIGHBORS* (KNN)**

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh  
Gelar Sarjana di Fakultas Teknik  
Universitas Medan Area**

**Oleh:**

**Rizky Affandi**

**198160068**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MEDAN AREA  
2023**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 24/6/24



Dipindai dengan CamScanner  
Access From (repository.uma.ac.id)24/6/24

## LEMBAR PENGESAHAN

Judul Skripsi : Analisis Sentimen Komentar Akun Instagram Public Figure Politik  
Terkait Pembatalan Piala Dunia U-20 Dengan Algoritma *K-Nearest Neighbors*  
(KNN)

Nama : Rizky Affandi

NPM : 198160068

Fakultas : Teknik Informatika

Disetujui Oleh  
Pembimbing

  
Andre Hasudungan Lubis, S. Ti, MSc



  
Dr. Eng. Supriatno, ST., MT  
Dekan Fakultas Teknik



  
Rizki Muliono, S. Kom, M. Kom  
Kepala Prodi Teknik Informatika

Tanggal Lulus: 27 Maret 2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 24/6/24

## HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun merupakan sebagai syarat untuk memperoleh gelar sarjana dari hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain yang telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah. Saya bersedia menerima sanksi akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.



Medan, 27 Maret 2024



Rizky Affandi  
198160068

## RIWAYAT HIDUP

Penulis dilahirkan di Kota Medan Provinsi Sumatera Utara pada tanggal 16 Oktober 2001, dari Ayah Anwar dan Ibu Wagini, dan penulis merupakan anak ke 2 dari 2 bersaudara. Penulis pertama sekali sekolah di SDS Al-Ulum Medan pada tahun 2007-2013 dan kemudian melanjutkan Sekolah Menengah Pertama di SMPN 4 Medan lulus pada tahun 2016.

Pada tahun 2019 Penulis lulus dari SMAN 7 Medan dan pada tahun 2019 terdaftar sebagai mahasiswa Fakultas Teknik Prodi Informatika Universitas Medan Area. Penulis mengikuti program Kampus Merdeka yaitu Pertukaran Mahasiswa Merdeka selama 4 bulan. Selama proses perkuliahan, penulis aktif berdiskusi belajar terhadap teman-teman dan mahasiswa selain jurusan penulis.

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI  
UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini :

Nama : Rizky Affandi  
NPM : 198160068  
Program Studi : Teknik Informatika  
Fakultas : Teknik  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti (*Non-Exclusive Royalti-Free Right*)** atas karya ilmiah saya yang berjudul : Analisis Sentimen Komentar Akun Instagram Public Figure Politik Terkait Pembatalan Piala Dunia U-20 Dengan Algoritma *K-Nearest Neighbors* (KNN)

Dengan Hak Bebas Royalti Non Eksklusif ini Universitas Medan Area berhak menyimpan, mengalih meida/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan  
Pada tanggal : 27 Maret 2024  
Yang menyatakan

  
/Rizky Affandi  
198160068

Piala dunia U-20 merupakan ajang pertandingan sepak bola tingkat internasional dikelola oleh PSSI, dimana Indonesia sebagai tuan rumah di tahun 2023. Namun sangat disayangkan, FIFA membatalkan Indonesia sebagai tuan rumah dan harus mengikuti kualifikasi agar dapat bermain di turnamen tersebut dikarenakan adanya penolakan terhadap kedatangan tim nasional negara Israel ke Indonesia oleh beberapa pihak termasuk *public figure* politik sehingga menimbulkan kericuhan dan amarah dari masyarakat. Dalam penelitian ini, algoritma *K-Nearest Neighbors* (KNN) digunakan sebagai metode untuk menganalisis sentimen terkait pembatalan piala dunia U-20 tahun 2023 pada akun *Instagram public figure* politik. *Instagram* menjadi media sosial yang sangat populer saat ini. Penelitian ini menggunakan metode KNN dimana sebelum perhitungan KNN, terlebih dahulu dilakukan proses *preprocessing* yaitu *cleaning, case folding, normalization, stemming, filtering, tokenizing, TF-IDF* yang kemudian dilakukan proses *labeling* data. Data yang digunakan sebanyak 2994 komentar. Dalam konteks klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN) dimana menggunakan berbagai nilai  $k$  yaitu 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. Proses ini melibatkan perbandingan pada tahap pengujian guna mencapai tingkat akurasi yang optimal, dan hasilnya dievaluasi menggunakan *confusion matrix*. Dari pelabelan, diketahui bahwa kelas sentimen negatif lebih unggul dengan jumlah 273 dibandingkan dengan sentimen positif yang hanya berjumlah 257 sehingga dapat diketahui jika sentimen masyarakat terhadap tokoh politik adalah negatif. Dari hasil pengujian diperoleh hasil dengan nilai  $k=3$  memiliki nilai akurasi yang optimal daripada nilai  $k$  yang lainnya yaitu 93 %.

**Kata kunci:** Komentar *Instagram*, Analisis Sentimen, KNN

## ABSTRACT

The U-20 world cup is an international football event managed by PSSI, where Indonesia is the host in 2023. Unfortunately, FIFA cancelled Indonesia as the host and had to follow the qualifications in order to play in the tournament due to the rejection of the arrival of the Israeli national team to Indonesia by several parties including political public figures, causing riots and anger from the community. In this research, the K-Nearest Neighbors (KNN) algorithm was used as a method to analyse sentiment related to the cancellation of the 2023 U-20 world cup on the Instagram accounts of political public figures. Instagram is a very popular social media today. This research used the KNN method where before the KNN calculation, a preprocessing process was first carried out, namely cleaning, case folding, normalization, stemming, filtering, tokenizing, TF-IDF which is then carried out the data labelling process. The data used were 2994 comments. In the context of classification using the K-Nearest Neighbor (K-NN) method which uses various  $k$  values namely 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. This process involved comparison at the testing stage in order to achieve an optimal level of accuracy, and the results are evaluated using a confusion matrix. From the labelling, it was known that the negative sentiment class was superior with a total of 273 compared to the positive sentiment which only amounts to 257 so that it could be known if the public sentiment towards political figures was negative. From the test results, it was obtained that the value of  $k = 3$  had an optimal accuracy value than other  $k$  values, which is 93%.

**Keywords:** Instagram Comments, Sentiment Analysis, KNN





## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Kuasa atas segala karuniaNya sehingga skripsi ini berhasil diselesaikan. Tema yang dipilih dalam penelitian ini ialah *Machine Learning* dengan judul “Analisis Sentimen Komentar Akun *Instagram Public Figure* Politik Terkait Pembatalan Piala Dunia U-20 Dengan Algoritma *K-Nearest Neighbors (KNN)*”.

Skripsi ini adalah salah satu syarat untuk menyelesaikan pendidikan untuk mencapai gelar sarjana di Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area. Pada kesempatan ini penulis mengucapkan terima kasih kepada :

1. Allah SWT yang telah memberikan kekuatan dan kesabaran yang sangat besar bagi penulis untuk menyelesaikan skripsi ini, tanpa bantuan-Nya, mungkin penulis tidak akan mempunyai kekuatan untuk terus berusaha menyelesaikan skripsi ini.
2. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc. selaku Rektor Universitas Medan Area.
3. Bapak Dr. Eng. Supriatno, ST., MT selaku Dekan Fakultas Teknik Universitas Medan Area.
4. Bapak Rizki Muliono, S.Kom., M.Kom selaku Kepala Program Studi Teknik Informatika Universitas Medan Area.
5. Bapak Andre Hasudungan Lubis, S. Ti, MSc selaku Dosen pembimbing yang telah membantu penulis dari segi materi dan moril sehingga penulis dapat menyelesaikan skripsi ini.

6. Orang tua penulis yaitu Bapak Anwar dan Ibu Wagini yang telah mendoakan tiada henti dan memberikan semangat serta membantu penulis dalam segi materi dan moril sehingga penulis dapat menyelesaikan skripsi ini dengan sebaik baiknya.
7. Seluruh Dosen dan Staf Program Studi Teknik Informatika Universitas Medan Area.
8. Seluruh teman-teman yang sudah memberikan dukungannya selama penulisan skripsi ini, khususnya teman-teman Teknik Informatika angkatan 2019.
9. The Guardian yang selalu memberikan dukungan serta semangat selama penulisan skripsi ini.

Penulis menyadari bahwa tugas akhir/skripsi ini masih memiliki kekurangan, oleh karena itu kritik dan saran yang bersifat membangun sangat penulis harapkan untuk kesempurnaan tugas akhir/skripsi ini. Penulis juga berharap pada tugas akhir/skripsi ini dapat bermanfaat bagi kalangan pendidikan maupun masyarakat. Penulis ucapkan terima kasih.

Medan, 27 Maret 2024



Rizky Affandi

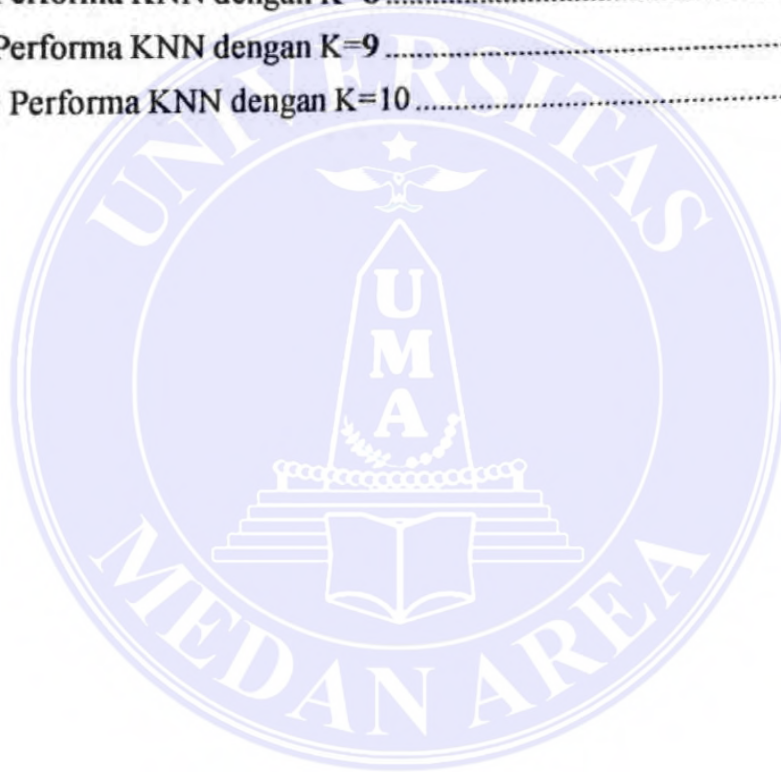
## DAFTAR ISI

LEMBAR PENGESAHAN .....	i
HALAMAN PERNYATAAN.....	ii
RIWAYAT HIDUP.....	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iv
ABSTRAK.....	v
<i>ABSTRACT</i> .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI .....	ix
DAFTAR GAMBAR .....	xi
DAFTAR TABEL .....	xii
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian. ....	4
1.5. Manfaat Penelitian. ....	5
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>6</b>
2.1. Analisis Sentimen.....	6
2.2. <i>K-Nearest Neighbors (KNN)</i> .....	7
2.3 Preprocessing Data.....	8
2.4 Klasifikasi.....	8
2.5 Pembobotan <i>TF-IDF</i> .....	10
2.6 <i>Confusion Matrix</i> .....	11
2.7 <i>Instagram</i> .....	12
2.8 <i>Public Figure</i> Politik .....	13
2.9 Penelitian Terdahulu.....	14
<b>BAB III METODE PENELITIAN.....</b>	<b>18</b>
3.1 Prosedur Penelitian .....	18
3.1.1 Identifikasi Masalah.....	20

3.1.2 Studi Literatur .....	20
3.1.3 Pengumpulan Data .....	20
3.1.4 TF-IDF .....	21
3.1.5 Labeling .....	22
3.1.6 KNN .....	23
3.1.7 Evaluasi .....	23
3.2 Perhitungan Manual .....	23
3.2.1 Cleaning .....	24
3.2.2 Case Folding .....	25
3.2.3 Normalization .....	26
3.2.4 Stemming .....	26
3.2.5 Filtering .....	27
3.2.6 Tokenizing .....	28
3.2.7 TF-IDF .....	28
3.2.9 Perhitungan KNN .....	33
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>34</b>
4.1 Hasil .....	34
4.1.1 Data yang Dikumpulkan .....	34
4.1.2 Proses <i>TF-IDF</i> .....	34
4.1.3 Labeling .....	40
4.1.4 Implementasi KNN .....	40
4.2 Pembahasan .....	50
<b>BAB V PENUTUP .....</b>	<b>54</b>
5.1 Kesimpulan .....	54
5.2 Saran .....	54
<b>DAFTAR PUSTAKA .....</b>	<b>56</b>
<b>LAMPIRAN .....</b>	<b>64</b>

## DAFTAR GAMBAR

Gambar 3.1 Diagram dalam tahapan penelitian .....	18
Gambar 4.1 Performa KNN dengan $K=1$ .....	41
Gambar 4.2 Performa KNN dengan $K=2$ .....	42
Gambar 4.3 Performa KNN dengan $K=3$ .....	43
Gambar 4.4 Performa KNN dengan $K=4$ .....	44
Gambar 4.5 Performa KNN dengan $K=5$ .....	45
Gambar 4.6 Performa KNN dengan $K=6$ .....	46
Gambar 4.7 Performa KNN dengan $K=7$ .....	47
Gambar 4.8 Performa KNN dengan $K=8$ .....	48
Gambar 4.9 Performa KNN dengan $K=9$ .....	49
Gambar 4.10 Performa KNN dengan $K=10$ .....	50



## DAFTAR TABEL

Tabel 2.1 Tabel <i>Confusion Matrix</i> .....	12
Tabel 3.1 Contoh komentar yang akan diolah .....	24
Tabel 3.2 Proses <i>cleaning</i> pada komentar .....	24
Tabel 3.3 Proses <i>case folding</i> pada komentar .....	25
Tabel 3.4 Proses <i>normalization</i> pada komentar .....	26
Tabel 3.5 Proses <i>stemming</i> pada komentar .....	26
Tabel 3.6 Proses <i>filtering</i> pada komentar .....	27
Tabel 3.7 Proses <i>tokenizing</i> pada komentar .....	28
Tabel 3.8 Tabel Perhitungan TF .....	28
Tabel 3.9 Tabel Perhitungan TF-IDF .....	30
Tabel 3.10 Hasil Perhitungan TF-IDF .....	31
Tabel 3.11 <i>Vector Length</i> .....	31
Tabel 3.13 Hasil Perhitungan KNN .....	33
Tabel 4.1 Data Komentar .....	34
Tabel 4.2 Hasil <i>Cleaning</i> .....	35
Tabel 4.3 Hasil <i>Case Folding</i> .....	36
Tabel 4.4 Hasil <i>Normalization</i> .....	36
Tabel 4.5 Hasil <i>Stemming</i> .....	37
Tabel 4.6 Hasil <i>Filtering</i> .....	38
Tabel 4.7 Hasil <i>Tokenizing</i> .....	38
Tabel 4.8 Hasil <i>TF-IDF</i> .....	39
Tabel 4.9 Hasil <i>Labeling</i> .....	40
Tabel 4.10 Akurasi Algoritma KNN .....	52

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Sepak bola adalah olahraga paling populer di dunia dan memiliki basis penggemar yang besar di berbagai negara, termasuk Indonesia (Lukmana & Tambunan, 2019). Olahraga sepak bola sudah ada di Indonesia sejak lama, yaitu dari tahun 1914 hingga saat ini menjadi bentuk olahraga yang sangat diminati oleh masyarakat luas. Indonesia juga memiliki federasi yang mengurus berbagai kegiatan olahraga ini yang disebut sebagai PSSI (Persatuan Sepakbola Seluruh Indonesia) (Emily & Wibisono, 2021). Dimana, PSSI juga berperan dalam berbagai pertandingan sepak bola yang melibatkan tim nasional Indonesia baik nasional maupun internasional (Harjo dkk., 2022).

Seperti salah satunya, penyelenggaraan piala dunia U-20 merupakan ajang pertandingan sepak bola tingkat internasional dikelola oleh PSSI, dimana Indonesia sebagai tuan rumah di tahun 2023. Namun sangat disayangkan, FIFA membatalkan Indonesia sebagai tuan rumah dan harus mengikuti kualifikasi agar dapat bermain di turnamen tersebut (Aryani dkk., 2023; Maulana, 2023). Hal ini dikarenakan adanya penolakan terhadap kedatangan tim nasional negara Israel ke Indonesia oleh beberapa pihak termasuk *public figure* politik (Ananda dkk., 2022).

Hal ini menimbulkan keributan dan keributan di masyarakat, dimana keributan terjadi antara pendukung dan penolak atas kejadian tersebut, khususnya di media sosial (Ananda dkk., 2022). Salah satu media sosial yang paling banyak digunakan sebagai sarana bagi masyarakat untuk mengemukakan pendapat adalah *Instagram*, khususnya di bidang sepak bola (Kosasi & Endri, 2022). Hal ini

dikarenakan, media sosial memiliki kemudahan untuk diakses dan memberikan kesempatan bagi pengguna untuk merespon suatu kejadian (Aliyudin, 2023).

Olahraga sepak bola di Indonesia, kerap dipengaruhi dan mempengaruhi oleh beberapa faktor salah satunya adalah politik. Berkaitan dengan media sosial, tokoh politik yang berperan dalam pembatalan piala dunia U-20 menjadi sarana masyarakat untuk menyampaikan aspirasi dan protes akan hal tersebut (Maulana, 2023). Sehingga, media sosial *Instagram* milik tokoh politik yang menolak kedatangan negara Israel menjadi sasaran bagi pendukung sepak bola untuk berkomentar negatif. Sebaliknya, terdapat pula pihak yang mendukung penolakan tersebut (Siallagan dkk., 2023).

Banyaknya beragam komentar yang memenuhi akun dari tokoh politik tersebut menimbulkan sulitnya untuk membedakan tanggapan positif, negatif, dan netral dari pengguna media sosial. Oleh karena itu, dibutuhkan suatu analisis untuk mengetahui perbedaan tanggapan yang dimiliki masyarakat, yaitu analisis sentimen (Aliyudin, 2023). Analisis sentimen adalah suatu metode dalam pemrosesan bahasa alami yang bertujuan untuk memahami dan menganalisis sentimen atau perasaan yang terkandung dalam teks atau dokumen tertentu. Tujuan utama dari analisis sentimen adalah untuk mengidentifikasi apakah pendapat yang dinyatakan dalam teks tersebut positif, negatif, atau netral (Darwis dkk., 2021).

Terdapat berbagai algoritma yang dapat digunakan untuk melakukan analisis sentimen di berbagai tujuan, seperti algoritma *Decision Tree*, *Naïve Bayes*, *SVM*, dan *K-Nearest Neighbors* (KNN) (Pamungkas & Kharisudin, 2021; Puspita & Widodo, 2021). KNN merupakan algoritma pembelajaran mesin (*machine learning*) yang banyak digunakan dalam klasifikasi dan regresi. KNN memiliki



prinsip dasar yaitu mencari k-nearest neighbors (k tetangga terdekat) dari sebuah data yang belum diklasifikasikan atau diprediksi berdasarkan data yang sudah diketahui. KNN telah banyak digunakan di berbagai penelitian untuk tujuan menganalisis sentimen.

Seperti yang dilakukan oleh Dede Sandi, Ema Utami dan Kusnawi (2023), melakukan penelitian yang bertujuan untuk menganalisis sentimen terhadap elektabilitas Ganjar Pranowo di tahun politik 2024 di *Twitter* dengan algoritma KNN dan *Naïve Bayes*. Adapun hasil dari penelitian tersebut menunjukkan bahwa perbandingan dari kedua pemodelan tersebut, pemodelan *K-Nearest Neighbor* yang paling baik akurasi dengan nilai akurasi sebesar 99% dari *K-Nearest Neighbor* dengan nilai *accuracy* sebesar 96%. Persentase sentimen dengan perbandingan 96,6% sentimen positif dan 3,4% sentimen negatif menyimpulkan bahwa sebagian besar masyarakat masih mendominasi sentimen positif. Penelitian yang dilakukan oleh Salim dan Mayary (2020) juga menggunakan KNN untuk analisis sentimen terhadap dompet elektronik pada media sosial *Twitter*. Hasil dari penelitian ini menunjukkan bahwa KNN memiliki tingkat akurasi yang cukup tinggi, yakni berkisar 94.05%. Sehingga, hal ini menunjukkan bahwa algoritma KNN memiliki performa yang cukup baik untuk digunakan dalam analisis sentimen. Algoritma KNN dipilih karena merupakan metode klasifikasi yang handal dalam memproses teks dan telah banyak digunakan.

Dalam penelitian ini, algoritma KNN digunakan sebagai metode untuk menganalisis sentimen terkait pembatalan piala dunia U-20 tahun 2023 pada akun *Instagram public figure* politik. *Instagram* menjadi media sosial yang sangat populer saat ini. Analisis sentimen dilakukan dengan memeriksa teks yang terdapat dalam

kolom komentar yang ditulis oleh masyarakat pada akun *Instagram public figure* politik.

## 1.2. Rumusan Masalah

Berdasarkan penjelasan dari masalah yang telah di paparkan pada latar belakang, berikut ini merupakan rumusan masalah dalam penelitian yaitu bagaimana menerapkan metode *K-Nearest Neighbors* (KNN) untuk melakukan analisis sentimen komentar pada akun *Instagram public figure* politik terkait pembatalan piala dunia U-20?

## 1.3. Batasan Masalah.

Adapun batasan masalah dalam penelitian ini yaitu sebagai berikut :

1. Data yang digunakan diambil langsung dari *Instagram* salah satu *public figure* politik terkait pembatalan piala dunia U-20 di Indonesia yaitu akun *Instagram @ganjar\_pranowo* sebanyak 2994 komentar.
2. Algoritma yang digunakan dalam menganalisis sentimen tersebut yaitu algoritma *K-Nearest Neighbors* (KNN).
3. Proses preprocessing data meliputi tahapan seperti *cleaning data*, melakukan *case folding*, melakukan *normalization*, melakukan *stemming*, melakukan *filtering*, melakukan *tokenizing*, *labeling* dan pembobotan *TF-IDF*.

## 1.4. Tujuan Penelitian.

Adapun menjadi tujuan dalam penelitian adalah sebagai berikut :

1. Menganalisis sentimen yang terkait dengan pembatalan Piala Dunia U-20 dalam akun *Instagram public figure* politik. Penelitian ini bertujuan untuk

memahami reaksi dan pandangan publik terkait dengan keputusan pembatalan tersebut.

2. Menerapkan algoritma *K-Nearest Neighbors* (KNN) dalam analisis sentimen. Penelitian ini bertujuan untuk menguji dan mengimplementasikan algoritma KNN dalam konteks analisis sentimen pada data akun *Instagram*.

### 1.5. Manfaat Penelitian.

Adapun manfaat penelitian ini yaitu sebagai berikut :

1. Penelitian ini akan memberikan wawasan tentang sentimen masyarakat terkait dengan pembatalan Piala Dunia U-20. Dengan menganalisis akun *Instagram public figure* politik, penelitian ini dapat memberikan gambaran tentang pandangan dan reaksi mereka terhadap keputusan tersebut.
2. Hasil penelitian ini dapat digunakan sebagai referensi atau landasan untuk penelitian lanjutan dalam bidang analisis sentimen.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Analisis Sentimen

Analisis sentimen adalah proses penggunaan teknik komputasional untuk mengidentifikasi, mengekstrak, dan memahami sentimen, opini, atau sikap dalam teks manusia (Sari & Wibowo, 2019). Tujuannya adalah untuk menentukan apakah teks tersebut memiliki sentimen positif, negatif, atau netral. Proses ini melibatkan beberapa langkah, seperti memecah teks menjadi unit-unit yang lebih kecil, menganalisis setiap unit secara individual, dan menggunakan pendekatan berbasis aturan atau berbasis mesin pembelajaran (Lestandy dkk., 2021).

Pendekatan berbasis aturan melibatkan penggunaan kamus kata-kata dengan nilai sentimen seperti "positif" atau "negatif" (Lestandy dkk., 2021). Setiap kata dalam teks diperiksa dengan kamus tersebut untuk menentukan sentimen yang terkait. Bobot sentimen setiap kata digabungkan untuk memberikan nilai sentimen keseluruhan teks. Di sisi lain, pendekatan berbasis mesin pembelajaran melibatkan pelatihan model komputer menggunakan data yang dikategorikan secara manual (Musfiroh dkk., 2021).

Selama analisis sentimen, beberapa aspek dapat diperhatikan, seperti intensitas sentimen dan kategorisasi sentimen menjadi positif, negatif, atau netral (Musfiroh dkk., 2021). Metode analisis sentimen juga dapat memperhatikan aspek subjek atau entitas dalam teks. Aplikasi analisis sentimen sangat luas, termasuk dalam bisnis untuk melacak respons pelanggan, memonitor tren opini publik, dan mengidentifikasi masalah yang mempengaruhi reputasi perusahaan (Sholeha dkk., 2022).

## 2.2. *K-Nearest Neighbors* (KNN)

*K-Nearest Neighbors* (KNN) adalah algoritma pembelajaran mesin yang digunakan dalam klasifikasi dan regresi (Laksono dkk., 2020). Prinsip dasar dari algoritma KNN adalah mencari *K-Nearest Neighbors* (k tetangga terdekat) dari sebuah data yang belum diklasifikasikan atau diprediksi berdasarkan data yang sudah diketahui. Algoritma ini beroperasi dengan mengukur jarak antara data yang belum diklasifikasikan dengan data yang sudah ada dalam ruang fitur menggunakan metrik *Euclidean*, *Manhattan*, atau *Minkowski* dan setelah itu, algoritma KNN mencari k data dengan jarak terdekat untuk digunakan sebagai prediksi kelas dalam klasifikasi atau prediksi nilai target dalam regresi (Mayasari dkk., 2022).

Salah satu keunggulan dari algoritma KNN adalah kemampuannya dalam bekerja dengan baik pada dataset yang kompleks atau tidak terstruktur (Qiudandra dkk., 2022). Namun, algoritma ini memiliki kelemahan dalam komputasi yang relatif mahal, terutama ketika jumlah data atau fitur yang besar (Supeli & Setiaji, 2023). Selain itu, algoritma KNN juga sensitif terhadap skala dan rentang data, sehingga *normalization* data diperlukan sebelum menerapkan algoritma ini.

Secara keseluruhan, algoritma *K-Nearest Neighbors* (KNN) adalah algoritma sederhana namun kuat yang digunakan untuk klasifikasi dan regresi (Ramadhani & Wahyudin, 2022). Dengan mencari k tetangga terdekat dari data yang belum diklasifikasikan, algoritma KNN dapat membuat prediksi berdasarkan mayoritas kelas atau nilai target dari tetangga-tetangga tersebut (Mayasari dkk., 2022). Meskipun memiliki beberapa kelemahan, algoritma ini tetap menjadi pilihan yang baik untuk banyak masalah pembelajaran mesin.

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_1 - x_2)^2} \quad (2.1)$$

### 2.3 Preprocessing Data

*Preprocessing* data adalah tahap pertama dalam *text mining* (Hermawati dkk., 2023). Proses ini bertujuan untuk mempersiapkan data agar dapat diproses selanjutnya dan meningkatkan kualitas data. *Preprocessing* penting dalam kegiatan *text mining* karena jika data yang digunakan memiliki kualitas rendah, maka pengetahuan yang dihasilkan juga akan rendah (Hidayah dkk., 2021).

*Preprocessing* melibatkan beberapa langkah seleksi dan pemeriksaan pada teks dimana langkah-langkah tersebut meliputi pembersihan data, perbaikan kesalahan, dan penyederhanaan teks agar dapat diproses dengan lebih baik (Zufria, 2022). Beberapa teknik *preprocessing* yang umum digunakan meliputi *cleaning*, *case folding*, *normalization*, *filtering*, *stemming*, dan *tokenizing* (Hidayah dkk., 2021). Dengan melakukan langkah-langkah ini, data teks dapat disiapkan dengan lebih baik sebelum dilakukan analisis lanjutan.

### 2.4 Klasifikasi

Klasifikasi adalah proses untuk menemukan fungsi atau model yang dapat membedakan kelas data (Sholeha dkk., 2022). Tujuannya adalah untuk memprediksi kelas objek yang tidak memiliki label. Teknik klasifikasi ini melibatkan evaluasi pada objek data dan mengkategorikannya ke dalam kelas-kelas yang ada dan model digunakan sebagai prototipe dan penyimpanan memori untuk melakukan prediksi terhadap objek data dan menentukan kelasnya (Ellif dkk., 2021).

Klasifikasi juga dapat diartikan sebagai pekerjaan yang menerapkan pembelajaran pada fungsi  $f$  (Sholeha dkk., 2022). Fungsi ini memetakan setiap atribut  $x$  ke salah satu label kelas  $y$  yang tersedia (Utomo & Mesran, 2020). Dalam proses ini, model dibuat dan disimpan sebagai memori untuk digunakan dalam prediksi.

Setelah tahap preprocessing selesai, data dilabeli sebagai sentimen positif atau negatif (Lestandy dkk., 2021). Pelabelan ini otomatis dilakukan menggunakan metode perhitungan skor sentimen dengan kamus lexicon. Kamus ini adalah metode klasifikasi yang menggunakan kamus yang berisi kata-kata opini untuk menentukan apakah suatu teks memiliki sentimen positif atau negatif (Pratiwi & Nudin, 2021).

Pada penelitian ini, kamus yang digunakan adalah terjemahan bahasa Indonesia dari kamus opini positif dan negatif. Skor pelabelan sentimen dihitung dengan mengurangi jumlah skor sentimen positif dengan skor sentimen negatif pada setiap kalimat komentar. Proses perhitungan skor menggunakan rumus berikut:

$$Skor = \left( \sum \text{kata positif} - \sum \text{kata negatif} \right) \quad (2.2)$$

Apabila hasil perhitungan skor suatu kalimat melebihi 0, maka kalimat tersebut diklasifikasikan sebagai kelas positif. Sebaliknya, jika skor kalimat kurang dari 0, maka kalimat dikategorikan sebagai kelas negatif. Sedangkan jika skor kalimat sama dengan 0, maka kalimat tersebut diklasifikasikan sebagai kelas netral (Septian dkk., 2019).

## 2.5 Pembobotan *TF-IDF*

Pembobotan *TF-IDF* adalah metode yang digunakan dalam pemrosesan bahasa alami untuk mengevaluasi pentingnya suatu kata dalam sebuah dokumen (Apriani dkk., 2021). Metode ini terdiri dari dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Term Frequency* menghitung seberapa sering sebuah kata muncul dalam dokumen tersebut namun semakin sering sebuah kata muncul, semakin penting kata tersebut dalam konteks dokumen tersebut.

Untuk mengatasi masalah ini, digunakan komponen IDF dimana IDF mengukur seberapa jarang sebuah kata muncul dalam semua dokumen yang ada dalam korpus (Apriani dkk., 2021). Jika sebuah kata hanya muncul dalam beberapa dokumen, maka kata tersebut dianggap memiliki bobot yang tinggi karena kemungkinan besar kata tersebut adalah kata yang spesifik dan memiliki nilai informatif yang tinggi. Dalam perhitungan *TF-IDF*, bobot TF dan bobot IDF dikalikan Bersama (Tanari dkk., 2019).

$$w_{ij} = tf_{ij} \times \log(D/df) \quad (2.3)$$

Hasil perkalian ini menghasilkan nilai bobot untuk setiap kata dalam dokumen. Kata-kata yang memiliki bobot tinggi berarti kata-kata tersebut penting dalam dokumen tersebut dan memberikan kontribusi signifikan terhadap pemahaman dan representasi dokumen (Widaningrum dkk., 2022). Pembobotan *TF-IDF* digunakan dalam berbagai aplikasi pemrosesan bahasa alami, seperti informasi retrieval, klasifikasi teks, dan penambangan teks sehingga dengan menggunakan metode ini, kita dapat mengidentifikasi kata-kata kunci yang



paling relevan dalam suatu dokumen, serta memahami tingkat kepentingan relatif dari kata-kata tersebut dalam korpus keseluruhan (Rofiqi dkk., 2019).

## 2.6 Confusion Matrix

*Confusion matrix* adalah alat evaluasi yang digunakan dalam pembelajaran mesin dan statistik untuk mengukur kinerja model klasifikasi (Normawati & Prayogi, 2021). *Confusion matrix* memberikan gambaran tentang sejauh mana model dapat mengklasifikasikan data dengan benar atau salah. *Confusion matrix* berisi empat nilai yang mendasarinya, yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) (Ali dkk., 2022).

Berikut adalah elemen-elemen yang ada dalam confusion matrix ali:

1. *True Positive* (TP): Jumlah sampel positif yang secara benar diklasifikasikan oleh model sebagai positif. Misalnya, model mengidentifikasi dengan benar 100 pasien menderita penyakit dan mereka memang menderita penyakit.
2. *True Negative* (TN): Jumlah sampel negatif yang secara benar diklasifikasikan oleh model sebagai negatif. Misalnya, model mengidentifikasi dengan benar 200 pasien sehat dan mereka memang sehat.
3. *False Positive* (FP): Jumlah sampel negatif yang salah diklasifikasikan oleh model sebagai positif. Ini juga dikenal sebagai kesalahan Tipe I. Misalnya, model mengidentifikasi 50 pasien sehat sebagai pasien yang menderita penyakit, padahal sebenarnya mereka sehat.
4. *False Negative* (FN): Jumlah sampel positif yang salah diklasifikasikan oleh model sebagai negatif. Ini juga dikenal sebagai kesalahan Tipe II. Misalnya,

model mengidentifikasi 10 pasien yang menderita penyakit sebagai pasien sehat, padahal sebenarnya mereka menderita penyakit.

**Tabel 2.1** Tabel *Confusion Matrix*

Sumber: <https://ilmudatapy.com/apa-itu-confusion-matrix/>

		Nilai Sebenarnya	
		<i>True</i>	<i>False</i>
Nilai Prediksi	<i>True</i>	TP ( <i>True Positive</i> ) <i>Correct result</i>	FP ( <i>False Positive</i> ) <i>Unexpected result</i>
	<i>False</i>	FN ( <i>False Negative</i> ) <i>Missing result</i>	TN ( <i>True Negative</i> ) <i>Correct absence of result</i>

Dengan menggunakan TP, TN, FP, dan FN, kita dapat membangun confusion matrix yang terdiri dari dua baris dan dua kolom. *Confusion matrix* dapat digunakan untuk menghitung metrik evaluasi model yaitu akurasi (*accuracy*).

Rumus untuk masing-masing *confusion matrix* adalah sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{2.3}$$

### 2.7 Instagram

*Instagram* adalah platform media sosial yang populer di seluruh dunia (Supangat & Elda Prawinda, 2021). Diluncurkan pada tahun 2010, *Instagram* memungkinkan pengguna untuk berbagi foto dan video dengan pengikut mereka. Aplikasi ini memiliki antarmuka yang intuitif dan menarik, serta fitur-fitur yang

kaya untuk mengedit dan membagikan konten. Pengguna dapat mengambil foto atau video secara langsung melalui aplikasi, atau mengunggahnya dari galeri ponsel mereka (Restusari & Farida, 2020).

Setelah mengambil gambar, pengguna dapat menerapkan berbagai filter, mengatur kecerahan, kontras, dan saturasi, serta menambahkan teks atau stiker untuk memberikan sentuhan kreatif pada konten mereka (Anggraini dkk., 2021). Selain itu, *Instagram* memungkinkan pengguna untuk berbagi cerita sementara yang berupa foto atau video yang menghilang setelah 24 jam. *Instagram* juga memiliki fitur *Direct Message* yang memungkinkan pengguna untuk mengirim pesan pribadi satu-satu atau dalam grup (Supratman dkk., 2020).

*Instagram* telah menjadi tempat yang populer bagi individu, merek, dan selebritas untuk membangun kehadiran online mereka, mempromosikan produk atau jasa, dan terhubung dengan pengikut mereka (Martha, 2021). Dalam beberapa tahun terakhir, *Instagram* juga telah mengenalkan fitur-fitur baru seperti *Instagram Shopping*, yang memungkinkan pengguna untuk membeli produk langsung dari aplikasi, dan *Instagram Reels*, yang mirip dengan TikTok, di mana pengguna dapat membuat dan membagikan video pendek kreatif (Johana dkk., 2020). Secara keseluruhan, *Instagram* telah menjadi platform yang vital dalam budaya digital saat ini, memberikan kesempatan bagi pengguna untuk berbagi momen hidup mereka, mengekspresikan kreativitas, dan menjalin hubungan dengan komunitas online mereka.

## 2.8 Public Figure Politik

*Public figure* politik merujuk pada individu yang terkenal atau terkenal secara luas di kalangan masyarakat karena peran dan kontribusinya dalam politik

(Pureklolon, 2020). Mereka adalah tokoh-tokoh yang terlibat dalam kegiatan politik seperti menjadi pemimpin negara, anggota parlemen, menteri, duta besar, atau tokoh-tokoh politik terkemuka lainnya. *Public figure* politik memiliki pengaruh yang signifikan dalam membentuk kebijakan publik dan arah politik suatu negara (Purwanti & Alfirdaus, 2020).

Sebagai tokoh yang berada di bawah sorotan media dan perhatian publik, *public figure* politik sering kali menjadi simbol dan wajah dari partai politik atau gerakan politik tertentu (Purwanti & Alfirdaus, 2020). Mereka berbicara di depan umum, memberikan wawasan dan analisis politik, dan berusaha mempengaruhi pendapat publik tentang berbagai isu politik (M. Darwis, 2021). Selain itu, *public figure* politik juga sering berperan dalam negosiasi dengan pemimpin negara lainnya dan menjalin hubungan diplomatik .

Peran *public figure* politik tidak terbatas pada tingkat nasional, tetapi juga dapat meliputi tingkat regional atau internasional (Munzir, 2019). Beberapa *public figure* politik telah mencapai ketenaran global, menjadi pemimpin dalam forum internasional, atau berperan sebagai mediator dalam konflik antarnegara (Habibullah dkk., 2022). Namun, sebagai *public figure*, mereka juga rentan terhadap kritik dan penilaian publik yang tajam.

## 2.9 Penelitian Terdahulu

Penelitian terdahulu (atau sering disebut juga dengan "literatur terdahulu" atau "kajian literatur") merupakan tahapan awal dalam melakukan sebuah penelitian (Cahyono dkk., 2019). Pada tahap ini, peneliti melakukan studi tentang penelitian-penelitian yang sudah dilakukan sebelumnya yang relevan dengan topik yang akan diteliti. Penelitian terdahulu bertujuan untuk memperoleh informasi dan

pemahaman yang lebih baik tentang topik penelitian, serta mengetahui penemuan-penemuan sebelumnya yang telah dicapai oleh peneliti lain dalam bidang yang sama atau terkait (Ridwan dkk., 2021).

Penelitian ini menggunakan beberapa penelitian terdahulu sebagai pembandingan dan mendukung gambaran umum tentang tema penelitian. Beberapa penelitian terdahulu tersebut yaitu :

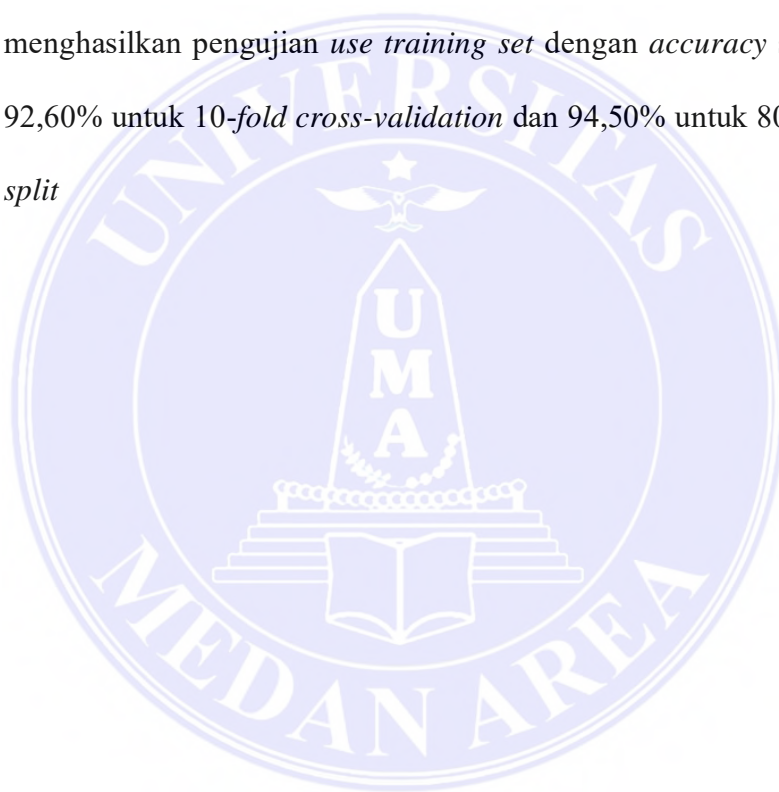
1. Penelitian yang dilakukan oleh Aluisius Dwiki Adhi Putra dan Safitri Juanita pada tahun 2021 dengan judul “Analisis Sentimen pada Ulasan Pengguna Aplikasi Bibit dan Bareksa dengan Algoritma KNN” dimana didapatkan hasil yaitu dalam tahapan pemodelan, dengan menggunakan algoritma *K-Nearest Neighbors* dan perbandingan 60:40 antara data *training* dan data *testing*. Hasil yang diperoleh menunjukkan nilai akurasi *precision* dan *recall* untuk setiap aplikasi. Untuk aplikasi bibit, kami mencapai akurasi sebesar 85,14% untuk *precision*, 91,91% untuk *recall*, dan 76,44% untuk kombinasi keduanya. Sementara itu, aplikasi bareksa mencapai akurasi sebesar 81,70% untuk *precision*, 87,15% untuk *recall*, dan 75,73% untuk kombinasi keduanya. Dengan menggunakan metode ini, maka berhasil mengukur dan membandingkan kinerja kedua aplikasi tersebut.
2. Penelitian yang dilakukan oleh Rani Puspita, Agus Widodo pada tahun 2020 dengan judul penelitian “Perbandingan Metode KNN, *Decision Tree*, dan *Naïve Bayes* Terhadap Analisis Sentimen Pengguna Layanan BPJS” dimana didapatkan hasil yaitu menunjukkan bahwa analisis sentimen terhadap data Twitter terhadap layanan BPJS dengan menggunakan metode KNN

mencapai tingkat akurasi 95.58% dengan *class precision* untuk *pred. negative* adalah 45.00%, *pred. positive* adalah 0.00%, dan *pred. neutral* adalah 96.83%. Lalu pada metode *Decision Tree* tingkat akurasinya mencapai 96.13% dengan *class precision* untuk *pred. negative* adalah 55.00%, *pred. positive* adalah 0.00%, dan *pred. neutral* adalah 97.28%. Dan yang terakhir adalah metode *Naïve Bayes* yang mencapai akurasi 89.14% dengan *class precision* untuk *pred. negative* adalah 16.67%, *pred. positive* adalah 1.64%, dan *pred. neutral* adalah 98.40%.

3. Penelitian yang dilakukan oleh Aditiya Yoga Pratama, Yuyun Umidah, Apriade Voutama pada tahun 2021 dengan judul penelitian yaitu “Analisis Sentimen Media Sosial Twitter Dengan Algoritma *K-Nearest Neighbor* Dan Seleksi Fitur *Chi-Square* (Kasus Omnibus Law Cipta Kerja)” dan didapatkan hasil yaitu Hasil pemodelan menggunakan *K-Fold Cross Validation* menunjukkan bahwa akurasi tertinggi diperoleh dengan skema penggunaan fitur sebesar 25% (seleksi fitur *Chi Square*), dan nilai  $k = 5$  pada metode KNN adalah 81,4%. Pengujian pada model dilakukan dengan menggunakan 100 data acak dan diperoleh akurasi sebesar 83%, presisi 100%, *recall* 15%, dan nilai *F-Measure* sebesar 26,08%. Dari 500 data yang diambil, kata "people" adalah kata yang paling dominan. Dari 500 data tersebut, 78,8% adalah sentimen negatif dan 21,2% adalah sentimen positif.
4. Penelitian yang dilakukan oleh Tobby Wiratama Putra, Agung Triayudi, Andrianingsih pada tahun 2022 dengan judul “Analisis Sentimen Pembelajaran Daring menggunakan Metode *Naïve Bayes*, KNN, dan *Decision Tree*” didapatkan hasil yaitu metode *Decision Tree* memiliki nilai

yang lebih tinggi diantara metode lainnya dengan akurasi sebesar 61.92%, presisi 73.63%, dan recall 11.42%.

5. Penelitian yang dilakukan oleh Mhd Furqan, Sriani Sriani, Susan Mayang Sari pada tahun 2022 dengan judul “Analisis Sentimen Menggunakan *K-Nearest Neighbor* Terhadap New Normal Masa Covid-19 Di Indonesia” sehingga didapatkan hasil yaitu pelabelan untuk sentimen positif berjumlah 811 dan 189 untuk sentimen negatif. Klasifikasi K-NN dengan nilai  $k = 1$  menghasilkan pengujian *use training set* dengan *accuracy* sebesar 100%, 92,60% untuk *10-fold cross-validation* dan 94,50% untuk *80% percentage split*

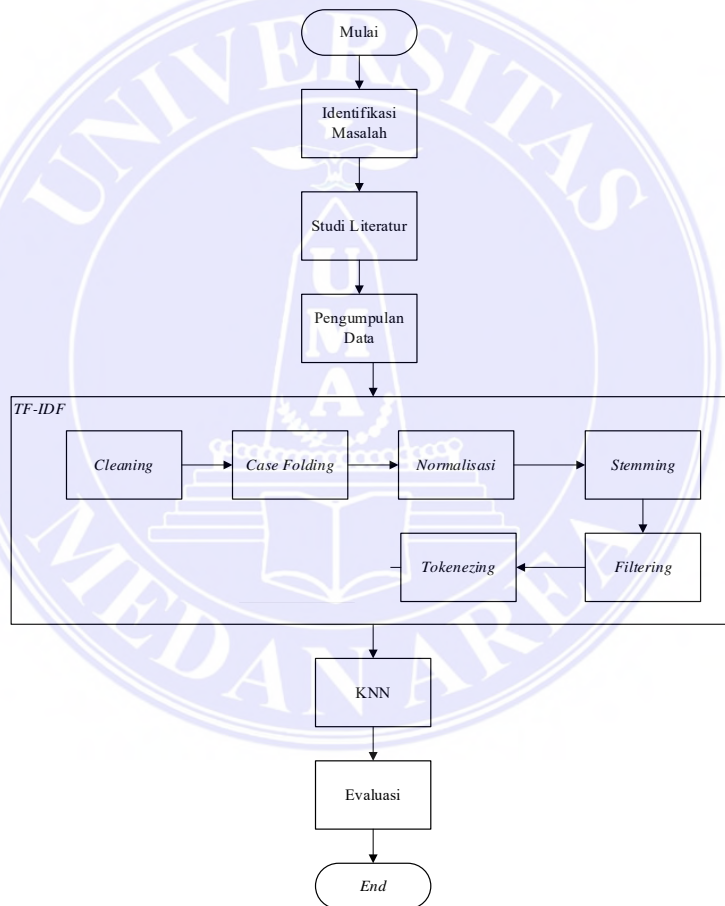


## BAB III

### METODE PENELITIAN

#### 3.1 Prosedur Penelitian

Dalam penelitian ini, dilakukan serangkaian prosedur yang mencakup persiapan penelitian, pengumpulan data, pengolahan data, dan pengambilan keputusan data (analisis atau hasil). Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 3.1 untuk memperjelasnya.



**Gambar 3.1** Diagram dalam tahapan penelitian



### 3.1.1 Identifikasi Masalah

Pada tahapan penelitian ini, penulis mengidentifikasi permasalahan yang terdapat di kalangan masyarakat, khususnya permasalahan pada penggunaan media sosial. Permasalahan yang terjadi yaitu adanya komentar-komentar terkait pembatalan piala dunia U-20 di Indonesia di akun *Instagram public figure* politik. Sehingga penulis menggunakan analisis metode klasifikasi sebagai salah satu upaya kritis dalam menyelesaikan permasalahan dengan memanfaatkan perkembangan teknologi. Analisis yang digunakan menggunakan algoritma KNN pengklasifikasian komentar di *Instagram*. Penelitian ini dilakukan untuk memahami bagaimana masyarakat merespons pembatalan Piala Dunia U-20 dan apakah mereka memiliki pandangan positif, negatif, atau netral terhadap keputusan tersebut. Hal ini dapat memberikan wawasan tentang pandangan umum masyarakat terhadap pemerintah dan kebijakan yang diambil.

### 3.1.2 Studi Literatur

Penelitian ini melakukan studi literatur dengan tujuan mencari referensi dan rujukan yang relevan untuk pengembangan penelitian ini. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam pengembangan objek pembahasannya. Dalam hal ini, literatur yang diambil berhubungan dengan platform *Instagram*, analisis sentimen, *data mining*, *text-mining*, metode klasifikasi, algoritma KNN, dan penelitian terdahulu.

### 3.1.3 Pengumpulan Data

Penulis melakukan observasi langsung dengan menggunakan API *Instagram* untuk mengumpulkan data opini masyarakat tentang kebijakan pemerintah terkait

pembatalan piala dunia U-20 di Indonesia di media sosial seorang *public figure* politik. Data yang digunakan berupa komentar berbahasa Indonesia yang diperoleh melalui *crawling* menggunakan *tools Instagram comment picker and exporter*. Sebanyak 2994 komentar diambil dari *instagram*, hanya mengambil data berupa teks tanpa gambar. Data yang telah dikumpulkan akan menjadi input sistem untuk proses lebih lanjut.

### 3.1.4 TF-IDF

Dalam tahap ini, penulis akan menggunakan Google Colab untuk melakukan pengolahan data yang telah diperoleh dengan menggunakan algoritma KNN. Proses pengolahan data akan melibatkan langkah-langkah sebagai berikut: ganti dengan *cleaning* sampai *tokenizing*

1. Proses *cleaning* dilakukan untuk menghilangkan karakter-karakter yang tidak perlu pada komentar, seperti tanda baca, angka, URL, nama pengguna, *mention*, *hashtag*, simbol seperti (, " ~ & ? ! > < # % { } [ ( [ 0 - 9 ] + ; : ; " ) [ 1 1 2 2 ] (Syah & Witanti, 2022).
2. Selanjutnya, data akan melalui proses *case folding*, dimana dokumen akan mengalami perubahan dengan mengubah setiap karakter menjadi huruf kecil (Albab & Fawaiq, 2023). Selain itu, dalam setiap tahap *preprocessing*, dilakukan penghilangan spasi berlebihan. Dalam tahap ini, semua perubahan yang terjadi pada dokumen adalah pengubahan huruf besar menjadi huruf kecil seperti huruf "P" pada kata "Pengendalian", "Perbandingan", dan "Peramalan" akan diubah menjadi "pengendalian", "perbandingan", dan "peramalan".

3. Kemudian, dilanjutkan dengan proses *normalization* yang bertujuan mengubah dan memperbaiki kata-kata yang disingkat ke dalam bentuk yang memiliki arti yang sama, sesuai dengan KBBI (Rifaldi dkk., 2023). Hal ini dilakukan agar informasi tersebut dapat diproses dengan mudah. Sebagai contoh, kata "utk" akan diubah menjadi "untuk" dan "yg" akan diubah menjadi "yang".
4. Setelah data *normalization*, data kemudian masuk ke proses *stemming* yang berfungsi untuk menghilangkan semua kata imbuhan dalam komentar, termasuk awalan, akhiran, dan tengahan kata (Rifaldi dkk., 2023).
5. Lalu, proses berikutnya adalah *filtering* dimana pada tahap ini, dilakukan penghapusan kata-kata yang tidak memiliki makna atau kurang penting dalam data (Septian dkk., 2019). Algoritma yang digunakan adalah stoplist atau stopword. Stopword merupakan daftar kata-kata umum yang sering digunakan namun kurang jelas maknanya, seperti "dari", "yang", "untuk", "dan", "di", dan sebagainya.
6. Dan berikutnya yaitu *tokenizing*. Pada tahap tokenisasi, dilakukan penghapusan angka, tanda baca, dan karakter lainnya kecuali huruf abjad (Hidayah dkk., 2021). Hal ini dilakukan karena karakter non-alphabet pada huruf abjad tidak berpengaruh pada pemrosesan teks.
7. Langkah berikutnya yaitu mencari nilai TF-IDF dari setiap kata yang telah melalui proses *tokenizing*.

### 3.15 Labeling

Setelah melalui proses *TF-IDF*, tahap selanjutnya yaitu melabel data untuk menentukan sentiment dari setiap kata. Sentimen kata yang digunakan yaitu positif

dan negatif. Tahap *labeling* pada penelitian ini menggunakan algoritma *clustering* yaitu *agglomerative clustering*. Menurut (Briggs dkk., 2020) *agglomerative clustering* adalah sebuah metode dalam analisis data dan machine learning yang digunakan untuk mengelompokkan data dengan pendekatan hirarkis. Pendekatan ini dimulai dengan menganggap setiap titik data sebagai satu kluster dan kemudian secara bertahap menggabungkan kluster yang mirip satu sama lain hingga membentuk satu kluster utuh.

### 3.1.6 KNN

Metode *k-Nearest Neighbor* mengaplikasikan klasifikasi berbasis lingkungan sebagai estimasi nilai pada contoh data yang baru. Dalam konteks ini, jarak antar tetangga dihitung dengan memanfaatkan kesamaan kosinus. Kesamaan kosinus digunakan untuk membandingkan tingkat kemiripan antara dokumen-dokumen. Penggunaannya melibatkan perbandingan antara kueri dan dokumen yang ada dalam set data pelatihan. Setelah mendapatkan hasil dari perhitungan *cosine similarity*, langkah melibatkan penggunaan algoritma *k-Nearest Neighbor* untuk melakukan proses klasifikasi data (Adlinnas dkk., 2020).

### 3.1.7 Evaluasi

Setelah dilakukan perhitungan KNN, maka selanjutnya digunakan *confusion matrix* untuk mengevaluasi seberapa besar tingkat akurasi, presisi

## 3.2 Perhitungan Manual

Pada penelitian ini, algoritma KNN digunakan untuk melakukan analisis sentimen terhadap akun *Instagram* terkait pembatalan piala dunia U-20. Berikut ini,

akan dilakukan contoh perhitungan manual dengan menggunakan algoritma KNN dengan data yang ada pada Tabel 3.1 sampai dengan Tabel 3.6.

Berikut ini adalah beberapa contoh kalimat yang akan digunakan sebagai contoh untuk perhitungan manual pada dataset komentar dengan algoritma KNN:

**Tabel 3.1 Contoh komentar yang akan diolah**

No	Komentar
1.	"Respect Gubernurku, rela menolak Israel demi menyakiti hati masyarakat sendiri,"
2.	"Terimakasih sudah menghancurkan mimpi anak bangsa,"
3.	"Semangat terus pak Ganjar dalam menyatukan masyarakat Indonesia"

### 3.2.1 *Cleaning*

*Clening* merupakan proses pembersihan dilakukan untuk menghilangkan karakter-karakter yang tidak perlu pada komentar. Pada penelitian ini, proses *cleaning* dapat dilihat pada Tabel 3.2 sebagai berikut:

**Tabel 3.2 Proses *cleaning* pada komentar**

No	Teks Komentar	Hasil <i>Cleaning</i>
1.	"Respect Gubernurku, rela menolak Israel demi menyakiti hati masyarakat sendiri,"	Respect Gubernurku rela menolak Israel demi menyakiti hati masyarakat sendiri

2.	"Terimakasih sudah menghancurkan mimpi anak bangsa,"	Terimakasih sudah menghancurkan mimpi anak bangsa
3.	"Semangat terus pak Ganjar dalam menyatukan masyarakat Indonesia"	Semangat terus Pak Ganjar dalam menyatukan masyarakat Indonesia

### 3.2.2 Case Folding

Pada tahap *case folding*, dokumen akan mengalami perubahan dengan mengubah setiap karakter menjadi huruf kecil. Pada penelitian ini, proses *case folding* dapat dilihat pada Tabel 3.3 sebagai berikut:

**Tabel 3.3 Proses *case folding* pada komentar**

No	Teks Komentar	Hasil Case Folding
1.	"Respect Gubernurku, rela menolak Israel demi menyakiti hati masyarakat sendiri,"	respect gubernurku rela menolak israel demi menyakiti hati masyarakat sendiri,
2.	"Terimakasih sudah menghancurkan mimpi anak bangsa,"	terimakasih sudah menghancurkan mimpi anak bangsa,
3.	"Semangat terus pak Ganjar dalam menyatukan masyarakat Indonesia"	semangat terus pak ganjar dalam menyatukan masyarakat indonesia

### 3.2.3 Normalization

Tahap *normalization* bertujuan mengubah dan memperbaiki kata-kata yang disingkat ke dalam bentuk yang memiliki arti yang sama, sesuai dengan KBBI. Pada penelitian ini, tahapan *normalization* dapat kita lihat pada Tabel 3.4 sebagai berikut:

**Tabel 3.4 Proses *normalization* pada komentar**

No	Teks Komentar	Hasil <i>Normalization</i>
3.	respect gubernurku rela menolak israel demi menyakiti hati masyarakat sendiri	hormat gubernurku rela menolak israel demi menyakiti hati masyarakat sendiri
4.	terimakasih sudah menghancurkan mimpi anak bangsa	terima kasih sudah menghancurkan mimpi anak bangsa
6.	semangat terus pak ganjar dalam menyatukan masyarakat indonesia	semangat terus pak ganjar dalam menyatukan masyarakat indonesia

### 3.2.4 Stemming

Fungsi tahapan ini adalah untuk menghilangkan semua kata imbuhan dalam komentar, termasuk awalan, akhiran, dan tengahan kata. Pada penelitian ini, proses *stemming* dapat dilihat pada Tabel 3.5 sebagai berikut:

**Tabel 3.5 Proses *stemming* pada komentar**

No	Teks Komentar	Hasil <i>Stemming</i>
----	---------------	-----------------------

1.	hormat gubernurku, rela menolak israel demi menyakiti hati masyarakat sendiri	hormat gubernurku rela tolak israel demi sakit hati masyarakat sendiri
2.	terima kasih sudah menghancurkan mimpi anak bangsa	terimakasih sudah hancur mimpi anak bangsa
3.	semangat terus pak ganjar dalam menyatukan masyarakat indonesia	semangat terus pak ganjar dalam satukan masyarakat indonesia

### 3.2.5 Filtering

Pada tahap ini, dilakukan penghapusan kata-kata yang tidak memiliki makna atau kurang penting dalam data seperti “yang”, dan “di”. Pada penelitian ini, proses *filtering* dapat dilihat pada Tabel 3.6 sebagai berikut:

**Tabel 3.6 Proses *filtering* pada komentar**

No	Teks Komentar	Hasil <i>Filtering</i>
1.	hormat gubernurku rela tolak israel demi sakit hati masyarakat sendiri	hormat gubernurku rela tolak israel demi sakit hati masyarakat sendiri
2.	terima kasih sudah hancur mimpi anak bangsa	terima kasih sudah hancur mimpi anak bangsa
3.	semangat terus pak ganjar satukan masyarakat indonesia	semangat terus pak ganjar satukan masyarakat indonesia



### 3.2.6 Tokenezing

Pada tahap tokenisasi, dilakukan penghapusan angka, tanda baca, dan karakter lainnya kecuali huruf abjad. Pada penelitian ini, proses *tokenezing* dapat dilihat pada Tabel 3.7

**Tabel 3.7** Proses *tokenezing* pada komentar

No	Teks Komentar	Hasil <i>Tokenezing</i>
1.	hormat gubernurku rela tolak israel demi sakit hati masyarakat sendiri	'hormat', 'gubernurku', 'rela', 'tolak', 'israel', 'demi', 'sakit', 'hati', 'masyarakat', 'sendiri'
2.	Terima kasih sudah hancur mimpi anak bangsa	'terima kasih', 'sudah', 'hancur', 'mimpi', 'anak', 'bangsa'
3.	semangat terus pak ganjar satukan masyarakat indonesia	'semangat', 'terus', 'pak', 'ganjar', 'satukan', 'masyarakat', 'indonesia'

### 3.2.7 TF-IDF

Berikut merupakan proses pembobotan nilai pada setiap dokumen. Langkah pertama yaitu menentukan jumlah kemunculan kata dari seluruh data. Berikut adalah hasil dari kemunculan kata, sebagai berikut:

**Tabel 3.8** Tabel Perhitungan TF

Text Processing	TF			df	d/df
	D1	D2	D3		

hormat	1	0	0	1	3
gubernurku	1	0	0	1	3
rela	1	0	0	1	3
tolak	1	0	0	1	3
israel	1	0	0	1	3
demi	1	0	0	1	3
sakit	1	0	0	1	3
hati	1	0	0	1	3
masyarakat	2	0	0	2	1,5
sendiri	0	1	0	1	3
terima	0	1	0	1	3
kasih	0	1	0	1	3
sudah	0	1	0	1	3
hancur	0	1	0	1	3
mimpi	0	1	0	1	3
anak	0	1	0	1	3
bangsa	0	1	0	1	3
semangat	0	0	1	1	3
terus	0	0	1	1	3
pak	0	0	1	1	3
ganjar	0	0	1	1	3
satukan	0	0	1	1	3
indonesia	0	0	1	1	3

Setelah diketahui jumlah kemunculan kata disetiap data, proses berikutnya yaitu dilakukan proses pembobotan nilai pada setiap dokumen dengan menggunakan perhitungan pada persamaan (2.3):

**Tabel 3.9 Tabel Perhitungan TF-IDF**

Text Processing	IDF log 10 (d/df)	Dokumen		
		D1	D2	D3
hormat	0,477121255	0,477121255	0,698	0
gubernurku	0,477121255	0,477121255	0	0
rela	0,477121255	0,477121255	0	0
tolak	0,477121255	0,477121255	0	0
israel	0,477121255	0,477121255	0	0
demi	0,477121255	0,477121255	0	0
sakit	0,477121255	0,477121255	0	0
hati	0,477121255	0,477121255	0	0
masyarakat	0,352182518	0,352182518	0,352182518	0
sendiri	0,477121255	0,477121255	0	0
terima	0,477121255	0	0,477121255	0
kasih	0,477121255	0	0,477121255	0
sudah	0,477121255	0	0,477121255	0
hancur	0,477121255	0	0,477121255	0
mimpi	0,477121255	0	0,477121255	0
anak	0,477121255	0	0,477121255	0
bangsa	0,477121255	0	0,477121255	0

semangat	0,477121255	0	0	0,477121255
terus	0,477121255	0	0	0,477121255
pak	0,477121255	0	0	0,477121255
ganjar	0,477121255	0	0	0,477121255
satukan	0,477121255	0	0	0,477121255
indonesia	0,477121255	0	0	0,477121255

Langkah berikut yaitu dengan perkalian antara D3 dengan D1, dan D2.

**Tabel 3.10 Hasil Perhitungan TF-IDF**

<b>Wd3*wdi</b>	
D1	D2
0,031008132	0

Langkah selanjutnya adalah menghitung panjang setiap dokumen termasuk D10 dengan mengkuadratkan bobot setiap istilah dalam setiap dokumen, menjumlahkan nilai kuadrat, dan kemudian mengakarkan hasil dari dokumen kuadrat tersebut.

**Tabel 3.11 Vector Length**

<b>Vector Length</b>		
D1	D2	D3
0,227645	0	0
0,227645	0	0
0,227645	0	0
0,227645	0	0



### 3.2.9 Perhitungan KNN

Dari hasil perhitungan *vector lengt*, lalu dilanjutkan dengan perhitungan KNN sehingga mendapatkan hasil:

**Tabel 3.13 Hasil Perhitungan KNN**

Dokumen	Hasil	Sentimen
(D3,D1)	0,03855	Positif
(D3,D2)	0	Negatif

Dari hasil perhitungan KNN tersebut, diketahui jika sentiment positif terdapat pada dokumen satu dan sentimen negatif terdapat pada dokumen dua. Hal ini ditentukan dengan nilai pada dokumen satu yang bernilai 0,03855 sedangkan pada dokumen dua bernilai 0.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Penelitian ini melakukan analisis sentimen untuk mengetahui sentimen komentar pada akun *Instagram public figure* politik dikarenakan pembatalan piala dunia U-20 menggunakan algoritma KNN. Pengumpulan data pada penelitian ini dilakukan dengan menggunakan *web scrapiing* pada komentar dari *Instagram @ganjar\_pranowo* sebanyak 2994 komentar dimana paa proses *labeling* didapatkan jumlah sentimen negatif sebanyak 2737 dan sentimen positif sebanyak 257. Dalam proses pengujian algoritma KNN dilakukan dengan berbagai nilai  $k$  yaitu 1, 2, 3, 4, 5, 6, 7, 8, 9, dan 10. Dari hasil pengujian di dapatkan bahwa nilai  $k = 3$  memiliki akurasi paling optimal yaitu 93%.

Hasil pada penelitian menunjukkan bahwa sentimen masyarakat terhadap tokoh politik Ganjar Pranowo tergolong negatif berdasarkan jumlah sentiment yang negatif jauh lebih banyak dibanding dengan sentimen positif. Penelitian ini juga menunjukkan bahwa algoritma KNN berhasil dalam melakukan analisis sentimen dengan nilai akurasi yang cukup baik.

#### 5.2 Saran

Berdasarkan kesimpulan di atas, penulis menyarankan untuk:

1. Menambah jumlah data yang digunakan untuk mengetahui hasil klasifikasi dengan jumlah data yang lebih besar atau dengan menggunakan dua atau tiga akun *Instagram* untuk mendapatkan perbandingan antar akun *Instagram*.

2. Menggunakan metode lain seperti *Naïve Bayes*, *Support Vector Machine* untuk mengetahui perbandingan nilai hasil klasifikasi.
3. Menggunakan metode pembobotan lain selain *TF-IDF*.





## DAFTAR PUSTAKA

- Adlinnas, J., Lhaksana, K. M., & Richasdy, D. (2020). Implementasi Metode Tf-idf Dan K-nearest Neighbor Untuk Seleksi Pelamar Kerja. *EProceedings of Engineering*, 7(3).
- Albab, M. U., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*, 20(2), 1–12.
- Ali, M. M., Faishol, M., & Anwar, K. (2022). DETEKSI JALAN BERLUBANG MENGGUNAKAN METODE GREY LEVEL CO-OCCURRENCE MATRIX DAN NEURAL NETWORK. *Jurnal Kecerdasan Buatan*, 3(1).
- Aliyudin, A. (2023). ANALISIS SENTIMEN TERHADAP PIALA DUNIA U-20 YANG BATAL DISELENGGARAKAN DI INDONESIA PADA MEDIA YOUTUBE DENGAN MENGGUNAKAN METODE NAÏVE BAYES. *Informasi Interaktif: Jurnal Informatika Dan Teknologi Informasi*, 8(2).
- Ananda, I. M. J. K., Hafiih, M., & Akmal, M. H. (2022). Pro Dan Kontra Pembatalan Sanksi FIFA Terhadap PSSI. *Prosiding Seminar Nasional Ilmu Ilmu Sosial (SNIIS)*, 1, 791–796.
- Anggraini, S. Y., Supriatna, A. R., & Soleh, D. A. (2021). Pengembangan Filter Game Edukasi Berbasis Instagram Pada Muatan IPA Kelas V Sekolah Dasar. *OPTIKA: Jurnal Pendidikan Fisika*, 5(2), 145–151.
- Apriani, A., Zakiyudin, H., & Marzuki, K. (2021). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta. *Jurnal Bumigora Information Technology (BITe)*, 3(1), 19–27.

- Aryani, D. K., Putri, E. W., Aulia, N., Rafly, M., Fadhilah, P. M., & Siregar, T. M. (2023). The Impact of The Cancellation of Indonesia as Host of The 2023 U-20 World Cup on The Indonesian Economy. *JURNAL KREASI RAKYAT*, 1(2), 122–128.
- Briggs, C., Fan, Z., & Andras, P. (2020). Federated learning with hierarchical clustering of local updates to improve training on non-IID data. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–9.
- Cahyono, E. A., Sutomo, N., & Hartono, A. (2019). Literatur review; panduan penulisan dan penyusunan. *Jurnal Keperawatan*, 12(2), 12.
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131–145.
- Darwis, M. (2021). Peran Sosial Media dalam Budaya Politik. *Jurnal Sipatokkong Bpsdm Sulsel*, 2(1), 93–103.
- Ellif, E., Sitorus, S. H., & Hidayati, R. (2021). Klasifikasi Kematangan Pepaya Menggunakan Ruang Warna HSV dan Metode Naive Bayes Classifier. *Coding Jurnal Komputer Dan Aplikasi*, 9(01), 66–75.
- Emily, N. F., & Wibisono, H. (2021). Literature Review: Hubungan Dynamic Balance Terhadap Resiko Terjadinya Cedera Pada Pemain Sepak Bola. *Indonesian Journal of Physiotherapy*, 1(1), 21–26.
- Habibullah, A. Z., Purnama, C., & Rezasyah, T. (2022). PUBLIC DIPLOMACY OF INDONESIAN FILM AGENCY THROUGH FILM INDUSTRY AS INTERNATIONAL POLITICAL INSTRUMENTS. *Jurnal Wacana Politik*, 7(2).

- Harjo, I. W. W., Rifaj, M., & Fathoni, S. R. (2022). *Gerakan Sosial Berbasis Identitas: Kajian Pendukung Sepak Bola di Indonesia*. Universitas Brawijaya Press.
- Hermawati, L., Berland, V., Rahmadiyah, A., Hutabarat, E., & Saputra, D. D. (2023). Komparasi Metode Text Mining Terhadap Masalah Pengklasifikasian Narasi Informatif & Non Informatif Pada twitter@ PLN\_123. *Jurnal Sistem Informasi Dan Teknologi*, 109–120.
- Hidayah, N., Sahibu, S., & others. (2021). Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 820–826.
- Johana, K., Lestari, F. D., & Fauziah, D. N. (2020). Penggunaan fitur Instagram Story sebagai media self disclosure dan perilaku keseharian mahasiswi Public Relations Universitas Mercu Buana. *Jurnal Ilmu Manajemen Terapan*, 1(3), 280–289.
- Kosasi, S. B., & Endri, E. P. (2022). New Media: Instagram@ bolalobfootball Sebagai Media Informasi Sepakbola Masyarakat Kota Tangerang. *AL MUNIR: Jurnal Komunikasi Dan Penyiaran Islam*, 13(01), 108–117.
- Laksono, E., Basuki, A., Bachtiar, F., & others. (2020). Optimization of k value in knn algorithm for spam and ham email classification. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(2), 377–383.
- Lestandy, M., Abdurrahim, A., Syafa'ah, L., & others. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*,

5(4), 802–808.

Lukmana, D. S. A., & Tambunan, S. M. G. (2019). Media Sosial, Budaya Penggemar Sepak Bola dan (RE) Artikulasi Diskursus Dominan Mengenai Keislaman. *Jurnal Muara Ilmu Sosial, Humaniora, Dan Seni*, 3(2), 568–574.

Martha, Z. (2021). Penggunaan fitur media sosial instagram stories sebagai media komunikasi. *Jurnal Komunikasi Nusantara*, 3(1), 26–32.

Maulana, M. A. (2023). Politik, Olahraga, dan Islam Studi Kasus Pembatalan RI Menjadi Tuan Rumah Piala Dunia U-20 2023. *Islamic Education*, 1(3), 16–24.

Mayasari, C. N., Soeleman, M. A., & Pujiono, P. (2022). Deteksi Pornografi pada Karakter Animasi 2D dengan KNN (K-Nearest Neighbors) Menggunakan Fitur HSV. *COMSERVA: Jurnal Penelitian Dan Pengabdian Masyarakat*, 2(08), 1236–1250.

Munzir, A. A. (2019). Beragam peran media sosial dalam dunia politik di Indonesia. *JPPUMA: Jurnal Ilmu Pemerintahan Dan Sosial Politik UMA (Journal of Governance and Political Social UMA)*, 7(2), 173–182.

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

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 697–711.

- Pamungkas, F. S., & Kharisudin, I. (2021). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 628–634.
- Pertiwi, M. W. (2019). Analisis sentimen opini publik mengenai sarana dan transportasi mudik tahun 2019 pada twitter menggunakan algoritma Naive Bayes, neural network, KNN dan SVM. *Inti Nusa Mandiri*, 14(1), 27–32.
- Pratiwi, S. Y. A., & Nudin, S. R. (2021). Analisis Sentimen terhadap Facebook Marketplace Menggunakan Metode Lexicon Based dan Support Vector Machine. *Jurnal Ilmiah Teknologi Informasi Dan Robotika*, 3(2), 9–15.
- Pureklolon, T. (2020). Komunikasi Politik: Kajian Substansial dalam Pendekatan Politik. *Jurnal Visi Komunikasi, Universitas Mercu Buana*, 19(2), 205–222.
- Purwanti, U. D., & Alfirdaus, L. K. (2020). Dinasti Politik dalam Pemerintahan Lokal di Kota Bontang, Kalimantan Timur. *Journal of Politic and Government Studies*, 9(04), 309–323.
- Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naive Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *J. Inform. Univ. Pamulang*, 5(4), 646.
- Qiudandra, E., Akram, R., & others. (2022). SISTEM PAKAR DIAGNOSA PENYAKIT OSTEOARTHRITIS DENGAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR. *Jurnal Ilmiah Teknik Informatika METHOTIKA*, 2(2), 37–48.
- Ramadhani, S. H., & Wahyudin, M. I. (2022). Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naive Bayes

dan K-NN. *Jurnal JTIC (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(4), 526–534.

Restusari, F. N., & Farida, N. (2020). Instagram Sebagai Alat Personal Branding Dalam Membentuk Citra Diri (Studi Pada Akun Bara Pattiradjawane). *Mediakom: Jurnal Ilmu Komunikasi*, 3(2), 176–186.

Ridwan, M., Suhar, A. M., Ulum, B., & Muhammad, F. (2021). Pentingnya penerapan literature review pada penelitian ilmiah. *Jurnal Masohi*, 2(1), 42–51.

Rifaldi, D., Fadlil, A., & others. (2023). Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet “Mental Health.” *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 161–171.

Rofiqi, M. A., Fauzan, A. C., Agustin, A. P., & Saputra, A. A. (2019). Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 58–64.

Salim, S. S., & Mayary, J. (2020). Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K--Nearest Neighbor. *Jurnal Ilmiah Informatika Komputer*, 25(1), 1–17.

Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 10(2), 681–686.

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. *INSYST: Journal of Intelligent System and Computation*, 1(1), 43–49.
- Sholeha, E. W., Yunita, S., Hammad, R., Hardita, V. C., & Kaharuddin, K. (2022). Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 3(4), 203–208.
- Siallagan, L., Sitorus, W., Saragih, K. T., Purba, L. A., Ginting, D. A., & Putrimina, M. (2023). Analisis Praanggapan dalam Video “Eksklusif: Ganjar Pranowo dan Piala Dunia” pada Wawancara Mata Najwa. *Innovative: Journal Of Social Science Research*, 3(2), 11868–11879.
- Supangat, S., & Elda Prawinda, E. (2021). Digital Konten Media Strategi Marketing Dalam Instagram Makwinfood. *Digital Konten Media Strategi Marketing Dalam Instagram Makwinfood*.
- Supeli, M. F. F., & Setiaji, S. (2023). Klasifikasi Sentimen Positif Dan Negatif Pada Aplikasi Vidio Dengan Algoritma K-Nearest Neighbor. *Indonesian Journal Computer Science*, 2(1), 7–15.
- Supratman, L. P., Rachmansyah, M., & others. (2020). Peran Media Instagram dalam Memasarkan Produk Fashion Dollies. *Jurnal Studi Komunikasi Dan Media*, 24(1), 73–90.
- Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1), 59–67.
- Tanari, A., Handojo, A., & Andjarwirawan, J. (2019). Aplikasi Pencarian Jurnal

Ilmiah Dengan Term Frequency-Inverse Document Frequency. *Jurnal Infra*, 7(1), 46–50.

Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437–444.

Widaningrum, I., Mustikasari, D., Arifin, R., Tsaqila, S. L., & Fatmawati, D. (2022). Algoritma Term Frequency--Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen. *Prosiding SISFOTEK*, 6(1), 145–149.

Zufria, I. (2022). Penerapan Text Mining Pada Sistem Penyeleksian Judul Skripsi Menggunakan Algoritma Latent Dirichlet Allocation (LDA). *Indonesian Journal of Computer Science*, 11(3).




## LAMPIRAN

### Lampiran 1. Hasil Turnitin



## Lampiran 2. SK Pembimbing

 **UNIVERSITAS MEDAN AREA**  
**FAKULTAS TEKNIK**

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎(061) 7366876, 7360168, 7364348, 7366781, Fax.(061) 7366998 Medan 20223  
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122  
Website: www.teknik.uma.ac.id E-mail: univ\_medanarea@uma.ac.id

---

Nomor : 779/FT.6/01.10/XI/2023 20 November 2023  
Lamp : -  
Hal : **Perubahan Judul Tugas Akhir**

Yth, Pembimbing Tugas Akhir  
**Andre Hasudungan Lubis, S. Ti, MSc**  
di  
Tempat

Dengan hormat, Sehubungan dengan adanya perubahan judul tugas akhir maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa tersebut :

Nama : Rizky Affandi  
N P M : 198160068  
Jurusan : Teknik Informatika

Maka dengan hormat kami mengharapkan kesediaan saudara :


**Andre Hasudungan Lubis, S. Ti, MSc** (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :


**“Analisis Sentimen Komentar Akun Instagram Public Figure Politik Terkait Pembatalan Piala Dunia U-20 dengan Algoritma K-Nearest Neighbors (KNN)”**.

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.

 Dekan,  
**Dr. Rahmad Syah, S. Kom, M. Kom**

### Lampiran 3. Surat Pengantar Riset

 **UNIVERSITAS MEDAN AREA**  
**FAKULTAS TEKNIK**

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax. (061) 7366998 Medan 20223  
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122  
Website: [www.teknik.uma.ac.id](http://www.teknik.uma.ac.id) E-mail: [univ\\_medanarea@uma.ac.id](mailto:univ_medanarea@uma.ac.id)

Nomor : 752 /FT.6/01.10/X/2023 23 Oktober 2023  
Lamp : -  
Hal : **Penelitian Dan Pengambilan Data Tugas Akhir**

Yth. Wakil Rektor Bid. Pengembangan SDM & Adm. Keuangan  
Jln. Kolam No.1  
Di  
Medan

Dengan hormat, kami mohon kesediaan ibu kiranya berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	N A M A	N P M	PRODI
1	Rizky Affandi	198160068	Teknik Informatika


Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir di **Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area**.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan Ilmiah dan Skripsi, yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul :

**Analisis Sentimen Akun Instagram Public Figure Politik terkait Pembatalan Piala Dunia U-20 dengan Algoritma K-Nearest Neighbor (KNN).**


Mohon kiranya tanggal Surat Izin Pengambilan Data Tugas Akhir agar disesuaikan dengan tanggal Terbitnya SK ini.

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.

  
Dr. Rahmad Syah, S. Kom, M. Kom

Tembusan :  
1. Ka. BAMA I  
2. Mahasiswa  
3. File

## Lampiran 4. Surat Selesai Riset



**UNIVERSITAS MEDAN AREA**  
Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, 7366878, 7364348 📠 (061) 7368012 Medan 20223  
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 8225602 📠 (061) 8226331 Medan 20122  
Website: www.uma.ac.id E-Mail: univ\_medanarea@uma.ac.id

---

**SURAT KETERANGAN**  
Nomor : /UMA/B/01.7/1/2024

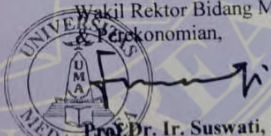
Rektor Universitas Medan Area dengan ini menerangkan bahwa :

Nama	: Rizky Affandi
No.Pokok Mahasiswa	: 198160068
Fakultas	: Teknik
Program Studi	: Teknik Informatika

Benar telah selesai Pengambilan Data Tugas Akhir di Laboratorium Komputer Universitas Medan Area dengan Judul Skripsi “Analisis Sentimen Akun Instagram Public Figure Politik terkait Pembatalan Piala Dunia U-20 dengan Algoritma *K-Nearest Neighbor (KNN)*”.

Dan kami harapkan Data tersebut kiranya dapat membantu yang bersangkutan dalam penyusunan skripsi dan dapat bermanfaat bagi mahasiswa khususnya Fakultas Teknik Universitas Medan Area.

Demikian surat ini diterbitkan untuk dapat digunakan seperlunya

Medan, 12 Januari 2024.  
Wakil Rektor  
Bidang Mutu SDM  
dan Ekonomian,  
  
Prof. Dr. Ir. Suswati, MP

CC:  
- Arsip

## Lampiran 5. Source Code

```
import pandas as pd
from google.colab import files
import io

uploaded = files.upload()

df = pd.read_csv('Komentar_KNN.csv', on_bad_lines='skip')
print(df)

import string
import nltk
import nest_asyncio
import itertools
import gensim
import urllib.request
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from textblob import TextBlob as tb
from collections import Counter
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from sklearn import svm
from sklearn.metrics import classification_report,
confusion_matrix
from sklearn.metrics import precision_score, recall_score,
f1_score, accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from gensim import corpora, models
from gensim.models import Phrases
from gensim.models.coherencemodel import CoherenceModel
from gensim.models.ldamodel import LdaModel
from gensim.corpora.dictionary import Dictionary
from wordcloud import WordCloud
from itertools import chain

nest_asyncio.apply()
nltk.download('punkt')
nltk.download('stopwords')

from nltk.corpus import stopwords
#preprocessing
```

```

nltk.download('punkt')
nltk.download('stopword')
nltk.download('wordnet') #stemming

# ----- Case Folding -----
# gunakan fungsi Series.str.lower() pada Pandas
df['clean_text'] = df['text'].str.lower()

print('Case Folding Result : \n')
df.head()
import re, string, unicodedata #modul regular expression
import nltk
from nltk import word_tokenize, sent_tokenize #Paket ini
membagi teks input menjadi kata-kata.,
from nltk.corpus import stopwords
nltk.download('punkt')
from nltk import word_tokenize, sent_tokenize
#remove stopwords and remove slang word
New_word_stop_list = [ 'tak','yg', 'ya', 'dg', 'dgn', 'nich',
'nih', 'sih', 'si', 'nya',
                    'lg', 'dulu', 'jgn', 'kl', 'klu', 'klo',
'kalo', 'nge',
                    'sip', 'spt', 'hallo', 'halo', 'ny', 'd',
'biar', 'skrg',
                    'bikin', 'bilang', 'tau', 'utk', 'jd',
'yah', 'loh', 'lho', 'aj', 'aja',
                    'cm', 'banget', 'deh', 'dimana', 'for',
'i', 'coba', 'it', 'sy', 'lumayan',
                    'kasih', 'tingkat',
                    "min", "tidak", "saya", "kenapa", "wib",
"dah", "oke", "gaada",
                    "tidak_ada", 'tdk', 'gak', 'enggak', 'ga', '-
' ]
def removeStopword(str):
    stop_words = set(stopwords.words('indonesian'))
    word_tokens = word_tokenize(str)
    filtered_sentence = [w for w in word_tokens if not w in
stop_words and w not in New_word_stop_list ]
    return ' '.join(filtered_sentence)
def cleaning(str):
    #remove non-ascii
    str = unicodedata.normalize('NFKD', str).encode('ascii',
'ignore').decode('utf-8', 'ignore')
    #remove URLs

```

```

    str = re.sub(r'(?i)\b(?:https?://|www\d{0,3}[.]|[a-z0-9.\-]+[.][a-z]{2,4}/)(?:[^\s()<>+|\\((([^\s()<>+|\\((([^\s()<>+|\\)))*\\)))+(?:|[^\s()<>+|\\((([^\s()<>+|\\)))*\\))|^[^\s!()\[\]{};:\'".,<>?«»"'''])', '', str)
    #remove punctuations
    str = re.sub(r'^\w|_', ' ', str)
    #remove digit from string
    str = re.sub("\S*\d\S*", "", str).strip()
    #remove digit or numbers
    str = re.sub(r"\b\d+\b", " ", str)
    #to lowercase
    str = str.lower()
    #Remove additional white spaces
    str = re.sub('[\s]+', ' ', str)
    return str
def preprocessing( title):

    title = str( title )
    title = removeStopword( title )
    title=cleaning(title)
    #print( str )

    return title
#MENGHITUNG TF-IDF dengan modul sklearn.feature extraction
text
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
#melihat kemunculan kata
vectorizer = CountVectorizer()
k_kata=
vectorizer.fit_transform(df_data['clean_text'].astype('U'))
print(k_kata)
#menghitung bobot kata atau tf-idf
tfidf= TfidfVectorizer()
h_tfidf=
tfidf.fit_transform(df_data['clean_text'].astype('U'))
print('-----BOBOT-----')
print(h_tfidf)
df = pd.DataFrame(h_tfidf.todense().T,
                  index = vectorizer.get_feature_names_out(),
                  columns=[f'D{i+1}' for i in range
(len(df_data))])
df
# Splitting data
import collections, numpy

```

```

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(h_tfidf,
df_data['Sentimen'], test_size=0.3, random_state=250)
print("Jumlah Data Uji:", X_test.shape)
print("Jumlah Data Latih:",X_train.shape)

pos = (y_test == 'positif').sum()
neg = (y_test == 'negatif').sum()
postrain = (y_train == 'positif').sum()
negtrain = (y_train == 'negatif').sum()
total = pos + neg
print("Jumlah data uji dengan sentimen positif:", pos)
print("Jumlah data uji dengan sentimen negatif:",neg)
print("Jumlah data latih dengan sentimen positif:", postrain)
print("Jumlah data latih dengan sentimen negatif:",negtrain)
df_data['Sentimen'].value_counts()
# perform algoritma KNN
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.neighbors import KNeighborsClassifier










clf = KNeighborsClassifier(n_neighbors=10).fit(X_train,
y_train)
predicted = clf.predict(X_test)
print(f'confusion matrix:\n {confusion_matrix(y_test,
predicted)}')
print('===== \n')
tn, fp, fn, tp = confusion_matrix(y_test, predicted).ravel()
print("TN:", tn)
print("FP:", fp)
print("FN:", fn)
print("TP:", tp)
print(classification_report(y_test, predicted,
zero_division=0))
print('===== \n')
print("Hasil Klasifikasi Sentimen Analisis :")
print("Accuracy:" , accuracy_score(y_test,predicted))
print("Precision:" , precision_score(y_test,predicted,
average="binary", pos_label="positif"))
print("Recall:" , recall_score(y_test,predicted,
average="binary", pos_label="positif"))
print("f1_score:" , f1_score(y_test,predicted,
average="binary", pos_label="positif"))
print("error_rate:", 1-accuracy_score(y_test,predicted))
saveKnn = round(accuracy_score(y_test,predicted), 3)

```



## Lampiran 6. Komentar Instagram

Komentar
sangat di sayangkan padahal saya sangat berharap bpk bisa menjadi penerus presiden berikutnya

Sudah terlanjur kecewa pak sulit buat move on
s
Semangat pak perjuangan untuk negara nkri kita #hidupprabowopuan
Semangat terus pakk, semoga menjadi lebih baik lagi 
Mengakui kesalahan sangat tak mudah, semangat dijalan yg benar
Semangat terus pak,semoga menjadi lebih baik lagi.... 
Rada kecewa sih , dulu sayang senang dengan nya . Sekarang saya kecewa
@avnxx 15 aku suka gayamu 
Melawan intoleransi,terorisme,radikalisme sangat jelas dalam undang undang negara,keadilan masih dijunjung tinggi di negara ini 
Mending bu puan deh kayanya ganjar udah kurang siip, di liat liat di lirik lirik kayanya puan sma prabowo sama sama jpss #HidupPrabowoPuan
Anda berani safari politik ke luar jawa tengah, bakal dimampusin
#HidupPrabowoPuan mungkin bisa menjadi pilihan yang cocok.
Maju terus pantang mundur never give up 
@indahmurniasari nah ke gep juga 
Pengen pemimpin yang anti kritik seperti Prabowo dan bu puan. #HidupPrabowoPuan
@anjaz_pulamba24 
Semoga kedepannya kita bisa berdamai dengan diri kita dan bisa selalu berbuat baik pada semua orang.
#Hidupprabowopuan butuh ilmu dan guru yang benar benar bisa memberi petunjuk pada kebaikan agar kita tidak salah dalam bertindak
dari sini kita tahu bahwa pak ganjar blunder, kalau gini mah mending ganti aja ama mba puan untuk maju capres apalagi diduetkan dengan pak Prabowo #HidupPrabowoPuan
Bagus sih kalo banteng jadi nyalon. Apalagi Prabowo Puan karakternya tegas dan gak bikin kisruh. Om ganjar udah gak se oke dulu  #HidupPrabowoPuan
Nasi Goreng jengkol enak kayanya ya
semog bapak menjadi bakal calon wakil presiden dan bu puna menjadi bakal calon presiden #HidupPrabowoPuan
@vilaakbar05 Siapa saja, yang penting bukan dari kalangan merah    #weskapok
@realfajarizky sok bijak lu njing
@fajaraguspradana 
@ferrysbachdim blog
@muhammadariwibowo2391 terus bola gunanya apa goblok ?? Mau bangga lolos kualifikasi bukan tuan rumah sok paling bola

