

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

# **KLASIFIKASI SENTIMEN MASYARAKAT DI TWITTER TERHADAP KENAIKAN HARGA BAHAN BAKAR MINYAK DENGAN METODE MODIFIED K-NEAREST NEIGHBOR**

## **TUGAS AKHIR**

Disusun Sebagai Salah Satu Syarat  
Untuk Memperoleh Gelar Sarjana Teknik  
Pada Jurusan Teknik Informatika

Oleh

**SOFIAH**

**NIM. 11950121742**



**UIN SUSKA RIAU**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU**

**PEKANBARU**

**2023**



**LEMBAR PERSETUJUAN**

**KLASIFIKASI SENTIMEN MASYARAKAT DI TWITTER  
TERHADAP KENAIKAN HARGA BAHAN BAKAR MINYAK  
DENGAN METODE MODIFIED K-NEAREST NEIGHBOR**

**TUGAS AKHIR**

Oleh

**SOFIAH**

**NIM. 11950121742**

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir  
di Pekanbaru, pada tanggal 12 Juli 2023

Pembimbing I,

**Yusra, S.T., M.T**

**NIP. 19840123 201503 2 001**

Pembimbing II,

**Muhammad Fikry, S.T., M.Sc**

**NIP. 19801018 200710 1 002**

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## LEMBAR PENGESAHAN

### KLASIFIKASI SENTIMEN MASYARAKAT DI TWITTER TERHADAP KENAIKAN HARGA BAHAN BAKAR MINYAK DENGAN METODE MODIFIED K-NEAREST NEIGHBOR

Oleh

**SOFIAH**

NIM. 11950121742

Telah dipertahankan di depan sidang dewan penguji  
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik  
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 12 Juli 2023

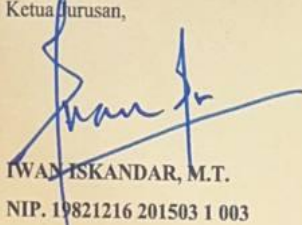
Mengesahkan,

Ketua Jurusan,



**DR. HARTONO, M.PD.**

NIP. 19640301 199203 1 003



**IWAN ISKANDAR, M.T.**

NIP. 19821216 201503 1 003

#### DEWAN PENGUJI

Ketua	: Iis Afrianty, ST., M.Sc.
Pembimbing I	: Yusra, S.T., M.T.
Pembimbing II	: Muhammad Fikry, S.T., M.Sc.
Penguji I	: Surya Agustian, S.T., M.Kom.
Penguji II	: Lola Oktavia, S.S.T., M.T.I.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**SURAT PERNYATAAN**

Saya yang bertandatangan di bawah ini:

Nama : SOFIAH  
NIM : 11950121742  
Tempat/Tgl. Lahir : S. Linau, 12 November 2002  
Fakultas : Sains dan Teknologi  
Prodi : Teknik Informatika  
Judul Skripsi : Klasifikasi Sentimen Masyarakat di Twitter terhadap Kenaikan Harga Bahan Bakar Minyak dengan Metode *Modified K-Nearest Neighbor*

Menyatakan dengan sebenar-benarnya bahwa:

1. Penulisan Skripsi dengan judul sebagaimana tersebut diatas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu, Skripsi saya ini, saya nyatakan bebas dari plagiat
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 17 Juli 2023

Yang membuat pernyataan



SOFIAH

NIM. 11950121742

## Klasifikasi Sentimen Masyarakat di Twitter Terhadap Kenaikan Harga Bahan Bakar Minyak dengan Metode *Modified K-Nearest Neighbor*

Sofiah<sup>1</sup>, Yusra<sup>2</sup>, Muhammad Fikry<sup>3</sup>, Lola Oktavia<sup>4</sup>

<sup>1</sup>Universitas Islam Negeri Sultan Syarif Kasim Riau, 11950121742@student.uin-suska.ac.id, Panam, Jl. HR. Soebrantas No.Km. 15, Simpang Baru, Riau 28293, Pekanbaru, Indonesia

<sup>2</sup>Universitas Islam Negeri Sultan Syarif Kasim Riau, yusra@uin.suska.ac.id, Panam, Jl. HR. Soebrantas No.Km. 15, Simpang Baru, Riau 28293, Pekanbaru, Indonesia

<sup>3</sup>Universitas Islam Negeri Sultan Syarif Kasim Riau, muhammad.fikry@uin-suska.ac.id, Panam, Jl. HR. Soebrantas No.Km. 15, Simpang Baru, Riau 28293, Pekanbaru, Indonesia

<sup>4</sup>Universitas Islam Negeri Sultan Syarif Kasim Riau, lola.oktavia@uin-suska.ac.id, Panam, Jl. HR. Soebrantas No.Km. 15, Simpang Baru, Riau 28293, Pekanbaru, Indonesia

### Informasi Makalah

Submit : May 28, 2023  
Revisi : May 30, 2023  
Diterima : Juni 15, 2023

### Kata Kunci :

Kenaikan BBM  
MK-NN  
Twitter  
Klasifikasi Sentimen

### Abstrak

#### Abstract

Kenaikan harga Bahan Bakar Minyak menjadi salah satu *tranding topic* di kalangan masyarakat Indonesia, baik di dunia nyata maupun dunia maya khususnya di media sosial Twitter. Perkembangan teknologi informasi yang sangat pesat memudahkan masyarakat dalam menyebarkan informasi di sosial media. Naiknya harga BBM memunculkan opini masyarakat yang mengandung sentimen positif dan negatif. Penelitian ini dilakukan untuk mengetahui sentimen publik terkait kebijakan pemerintah yang menaikkan harga BBM serta menerapkan metode *Modified K-Nearest Neighbor* dalam pengklasifikasian sentimen pengguna Twitter terhadap kenaikan harga BBM. *Modified K-Nearest Neighbor* merupakan salah satu metode klasifikasi berdasarkan kemunculan kelas terbanyak pada data latih. Data yang digunakan adalah *tweet* dalam bahasa Indonesia berdasarkan kata kunci “kenaikan BBM” dengan jumlah dataset sebanyak 3.000 *tweet*. Pembobotan kata dengan menggunakan TF-IDF untuk melakukan klasifikasi sentimen ke dalam dua kelas positif dan negatif. Hasil dari penelitian ini adalah klasifikasi sentimen terhadap kenaikan harga BBM. Akurasi tertinggi didapat 83.33% pada data opini menggunakan perbandingan 90:10 dan K=3.

The increase in fuel prices has become one of the trending topics among the people of Indonesia, both in the real world and the virtual world, especially on social media Twitter. The rapid development of information technology makes it easier for people to spread information on social media. Rising fuel prices led to public opinion that contains positive and negative sentiments.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

This study was conducted to determine public sentiment related to government policies that

Sofiah  
Email: 11950121742@students.uin-suska.ac.id

increase fuel prices and apply the Modified K-Nearest Neighbor method in classifying Twitter user sentiment towards rising fuel prices. Modified K-Nearest Neighbor is one of the classification methods based on the most class occurrences in the training data. The Data used is Tweets in Indonesian based on the keyword "fuel increase" with a total dataset of 3,000 tweets. Weighting words using TF-IDF to classify sentiment into two classes of positive and negative. The result of this study is a classification of sentiment towards rising fuel prices. The highest accuracy was obtained 83.33% in opinion data using a ratio of 90:10 and K=3.

## 1. Pendahuluan

Perkembangan teknologi memberikan saat ini. Keputusan pemerintah yang kemudahan pada pertukaran informasi dan menaikkan harga BBM menimbulkan banyak komunikasi.

munculnya berbagai platform media sosial dari Twitter mendapat banyak tanggapan dari yang sudah banyak, seperti Facebook, masyarakat, yang berupa tanggapan positif dan negatif (TRISNO et al., 2020). Pendapat (Buntoro, 2017). Penyebaran informasi masyarakat mengenai kenaikan harga BBM menjadi sangat mudah bagi masyarakat, menunjukkan adanya sentimen positif dan dikarenakan banyaknya masyarakat yang negatif. Dengan demikian, penelitian ini akan sudah menggunakan media sosial. Media menguji tingkat akurasi pengklasifikasian sosial yang digunakan memiliki banyak sentimen masyarakat terhadap kenaikan manfaat seperti untuk berkomunikasi antar harga Bahan Bakar Minyak (BBM) di orang atau kelompok (Rozi & Sulistyawati,

2019).

*Microblogging* atau yang lebih dikenal

dengan Twitter adalah layanan jejaring untuk mengetahui apakah masyarakat pro internet yang di luncurkan pada tahun 2006 atau kontra di berlakukannya kenaikan harga oleh seorang pengembang *web* asal Amerika

Serikat yang bernama Jack Dorsey. Dapat di

lihat jumlah pengguna Twitter yang semakin pengamatan data dan atribut saat melakukan bertumbuh di setiap harinya, maka tidak pengelompokan terhadap data, sehingga data jarang Twitter kerap dimanfaatkan untuk yang belum memiliki kategori dapat menyampaikan informasi berupa komentar dikelompokkan menjadi data yang maupun saran mengenai sesuatu hal yang diklasifikasikan. Klasifikasi melibatkan sedang *trending* di kalangan masyarakat. pencarian pola atau fitur yang mewakili dan Twitter menjadi salah satu media sosial yang membedakan antara konsep atau kelas

harga BBM di Indonesia menjadi masalah besar sehingga menjadi topik utama hingga

Ini ditunjukkan dengan opini di masyarakat. Informasi yang didapat

Twitter dengan menggunakan metode *Modified K-Nearest Neighbor* (MK-NN).

masyarakat mengenai kenaikan harga BBM

Twitter dengan menggunakan metode

*Modified K-Nearest Neighbor* (MK-NN).

Hasil klasifikasi sentimen ini akan digunakan

BBM.

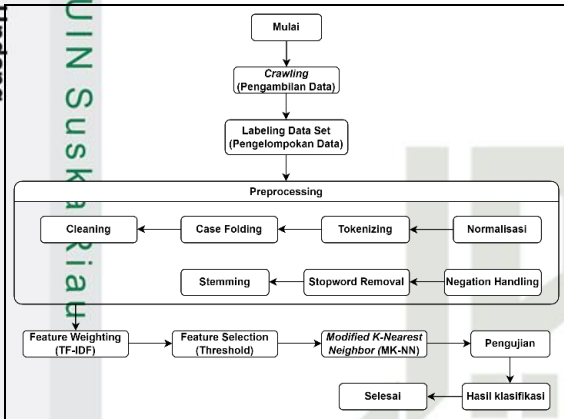
Klasifikasi didasarkan pada hasil



Twitter serta menguji tingkat akurasi metode MK-NN dalam mengklasifikasikan sentimen masyarakat terhadap kenaikan harga BBM.

## 2. Metode Penelitian

Tahapan dalam penelitian ini terdiri dari beberapa tahap. Berikut tahapan dapat dilihat pada Gambar 1.



Gambar 1. Desain Penelitian

### 2.1. Pengumpulan Data

Data dari *tweet* diambil melalui *crawling* media sosial Twitter. Semua data dikumpulkan dalam bahasa Indonesia.. Berikut pada Tabel 1 merupakan salah satu contoh *crawling* data.

Tabel 1. Contoh *Crawling* Data

Data Set
b'@teddgus krisis ekonomi dunia malah memaksakan bangun ikn, rakyat yg dibikin susah dgn kenaikan bbm, kenaikan pajak.. \xf0\x9f\x98\xaa'

### 2.2. Pelabelan Data

Pelabelan merupakan tahap memberikan label pada seluruh data *tweet* sesuai dengan kelas yang telah ditentukan (Anggraini et al., 2021). Untuk pelabelannya dilakukan oleh Ibu Elfina, S.Pd., M.Pd selaku dosen Bahasa Indonesia. Pada penelitian ini, klasifikasi sentimen *tweet* terhadap kenaikan harga BBM dibagi menjadi dua yaitu positif dan negatif. Dataset kemudian dibagi menjadi data uji dan data latihan. Contoh pelabelan data dapat dilihat pada Tabel 2.

Tabel 2. Contoh Pelabelan Data

Data Set	Pelabelan Data
b'@teddgus krisis ekonomi dunia malah memaksakan bangun ikn, rakyat yg dibikin susah dgn kenaikan bbm, kenaikan pajak.. \xf0\x9f\x98\xaa'	negatif

### 2.3. s

*Preprocessing* merupakan tahapan untuk mengolah data dengan mempersiapkan data mentah dari hasil *crawling* diubah kedalam bentuk yang terstruktur. Pada tahap *preprocessing* dilakukan 7 tahapan sebagai berikut:

#### 1. *Cleaning*

*Cleaning* merupakan proses menghapus *tweet* dari kata-kata yang tidak diperlukan, seperti *hashtag*, emoji, emoticon, *username*, *url*, dan e-mail. Dapat dilihat pada Tabel 3 di bawah.

Tabel 3. Contoh *Cleaning* Data

Data Set	Hasil Cleaning
b'@teddgus krisis ekonomi dunia malah memaksakan bangun ikn, rakyat yg dibikin susah dgn kenaikan bbm, kenaikan pajak.. \xf0\x9f\x98\xaa'	krisis ekonomi dunia malah memaksakan bangun ikn rakyat yg dibikin susah dgn kenaikan bbm kenaikan pajak

#### 2. *Case Folding*

*Case folding* berfungsi untuk mengubah semua huruf besar pada *tweet* menjadi huruf kecil, berikut adalah prosesnya yang dapat di lihat pada Tabel 4.

Tabel 4. Contoh *Case Folding*

Setelah Cleaning	Hasil Sase Folding
krisis ekonomi dunia malah memaksakan bangun ikn rakyat yg dibikin susah dgn kenaikan bbm kenaikan pajak	krisis ekonomi dunia malah memaksakan bangun ikn rakyat yg dibikin susah dgn kenaikan bbm kenaikan pajak

#### 3. *Tokenizing*

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



*Tokenizing* merupakan proses untuk memecah kalimat kedalam token unigram menggunakan pembatas spasi. Dapat dilihat pada Tabel 5 contoh *tokenizing*.

Tabel 5. Contoh *Tokenizing*

Setelah Cleaning	Hasil Hokenizing
skrisis ekonomi dunia malah memaksakan bangun ikn rakyat yg dibikin susah dgn kenaikan bbm kenaikan pajak	krisis ekonomi dunia malah memaksakan bangun ikn rakyat yg dibikin susah dgn kenaikan bbm kenaikan pajak

4. Normalisasi

Normalisasi merupakan tahapan proses untuk mengembalikan kata-kata gaul yang ada kedalam bahasa baku sehingga menjadi kata normal, sesuai kamus normalisasi. Dapat dilihat pada Tabel 6 di bawah.

Tabel 6. Contoh Normalisasi Data

Hasil Tokenizing	Hasil Normalisasi
krisis	krisis
ekonomi	ekonomi
dunia	dunia
malah	malah
memaksakan	memaksakan
bangun	bangun
ikn	ikn
rakyat	rakyat
yg	yang
dibikin	dibikin
susah	susah

dgn	dengan
kenaikan	kenaikan
bbm	bbm
kenaikan	kenaikan
pajak	pajak

5. *Negation Handling*

*Negation handling* merupakan pemrosesan teks awal menangani penambahan kata negasi, sehingga bisa mempertahankan sentimen aslinya. Berikut adalah prosesnya.

Tabel 7. Contoh *Negation Handling*

Hasil Normalisasi	Hasil Negation Handling
krisis	krisis
ekonomi	ekonomi
dunia	dunia
malah	malah
memaksakan	memaksakan
bangun	bangun
ikn	ikn
rakyat	rakyat
yang	yang
dibikin	dibikin
susah	susah
dengan	dengan
kenaikan	kenaikan
bbm	bbm
kenaikan	kenaikan
pajak	pajak
kenaikan	kenaikan

6. *Stopword Removal* *Stopword removal* adalah tahap mengambil kata-kata penting dari hasil token. Dapat dilihat pada Tabel 8 dibawah.

Tabel 8. Contoh *Stopword Removal*

Hasil Negation Handling	Hasil Stopword Removal
krisis	krisis
ekonomi	ekonomi

Hak Cipta Dilindungi Undang-Undang

© Hak Cipta milik UIN Suska Riau

State Islamic University of Sultanarif Kasim Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dunia	dunia
malah	
memaksakan	memaksakan
bangun	bangun
ikn	ikn
rakya	rakyat
yang	
dibikin	dibikin
susah	susah
dengan	
kenaikan	kenaikan
bbm	bbm
kenaikan	kenaikan
pajak	pajak

### 7. Stemming

Proses *stemming* menghilangkan kata imbuhan pada awalan dan akhiran untuk menjadi token kata dasar. Contoh *stemming* dapat dilihat pada Tabel 9.

Tabel 9. Contoh *Stemming*

Hasil Stopword Removal	Hasil Stemming
krisis	krisis
ekonomi	ekonomi
memaksakan	paksa
bangun	bangun
ikn	ikn
rakyat	rakyat
dibikin	bikin
susah	susah
kenaikan	naik

### 2.2. Feature Weighting

Setelah melakukan tahapan preprocessing, selanjutnya tahap feature weighting. Pembobotan yang digunakan adalah TF-IDF. TF-IDF adalah metode pembobotan yang mengaitkan antara term frequency (TF) dan invers dokument frequency (IDF). Proses pembobotan TF-IDF dimulai dengan pengalihan IDF untuk menemukan nomor kata yang dikenal

sebagai bobot atau TF di setiap dokumen. (Rozi & Sulistyawati, 2019).

### 2.3. Feature Selection

Setelah nilai DF dan IDF pada bobot kata ditentukan, tahap selanjutnya pemilihan fitur yaitu thrashing. Nilai yang digunakan sebagai threshold adalah nilai DF. Pada ambang batas DF diproses untuk TFIDF.

### 2.4. Modified K-Nearest Neighbor

Metode *Modified K-Nearest Neighbor* memiliki cara kerja yang hampir sama dengan metode K-NN dalam pengklasifikasian, pada metode MK-NN menambahkan proses validasi dan *weigh voting*. Berikut tahapan metode MK-NN.

#### a. Jarak Euclidian

Setelah mendapatkan bobot data latih dan data uji, selanjutnya menghitung jarak *Euclidean* untuk menghitung jarak antar data dengan rumus sebagai berikut:

$$d(x_i, y_i) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (1)$$

**Ket:**

- d = Jarak x dan y
- x = Data training yang digunakan
- y = Data testing yang digunakan
- i = Presentasi nilai atribut
- n = Dimensi atribut

#### b. Menghitung Nilai Validasi

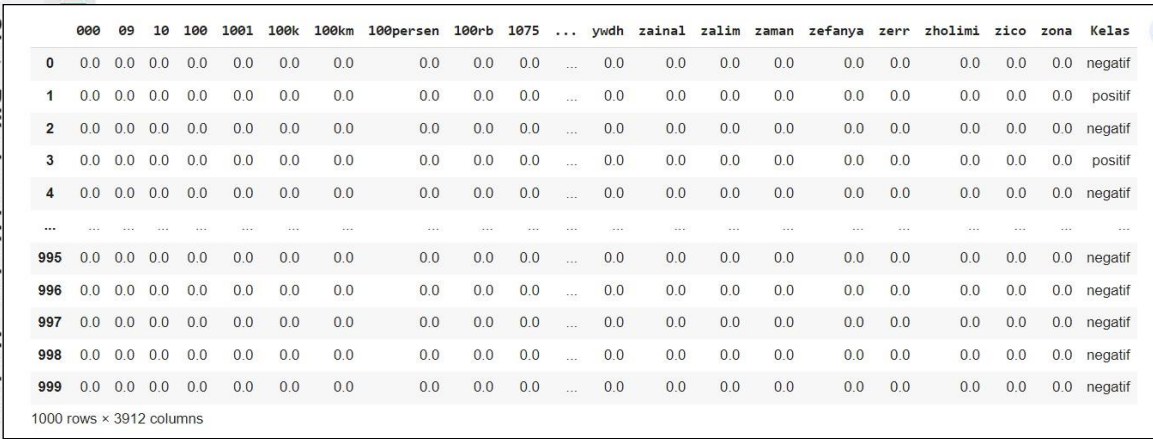
bbm	bbm
kenaikan	naik
pajak	pajak

Data yang memiliki nilai terdekat (jarak) dengan nilai data yang akan dicari nilai validasinya. Label kelas S, label yang sama bernilai = 1, label yang berbeda bernilai = 0. Tahap ini dirumuskan pada Persamaan (6).

**Rumus:**

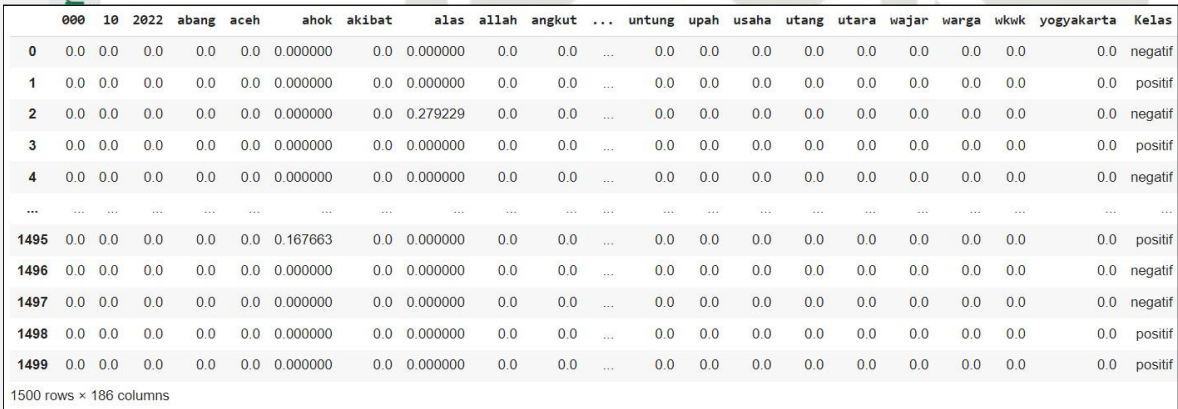
$$Validasi(x) = \frac{1}{k} \sum_{i=1}^n S(lb_l(x) \cdot lb_l(N_i(x))) \quad (2)$$

Hak Cipta Ditanggung Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2. Tampilan TF.IDF

3.2. *Feature Selection* menggunakan nilai *thresholding* 0,001. Tahapan *feature selection* untuk Gambar 3 dibawah ini tampilan *feature* mengurangi fitur yang tidak akurat. Dengan *selection*.



Gambar 3. Tampilan *Feature Selection*

**K:** = Jumlah titik terdekat  
**k** = Label  $l(x)$   
**L** = Label kelas x  
**Kelas x**  $l(N_i(x))$  = Label kelas titik terdekat x

**W(i)** : Perhitungan weigh voting Validasi  
**de(i)** : Jarak eucledian

### 3. Hasil dan Pembahasan

Setelah data set yang digunakan melalui tahap *preprocessing* seperti yang penulis jabarkan dibagian metodologi penelitian pada 2.3. Tahapan selanjutnya yaitu TF-IDF.

#### 3.1. TF-IDF

Pada tahapan TF-IDF untuk memberikan nilai bobot setiap kata pada dokumen. Berikut tampilan output TF-ID pada Gambar 2.

#### 3.3. Klasifikasi

c. Menghitung *Weigh Voting Weight Voting* menggunakan nilai K dengan menghitung tetangga terdekat, dilanjutkan dengan nilai validasi tiap data training yang dikalikan dengan *weigh* berdasarkan jarak setiap tetangganya. Perhitungan dilakukan menggunakan persamaan (3).

**Rumus:**

$$W(i) = Validasi(i) \cdot \frac{1}{de(i)+0.5} \quad (3)$$

**K:**

Klasifikasi tahapan untuk pengujian data yang telah melewati tahap *preprocessing* sampai *feature selection*. Dalam pengujian data uji yang digunakan yaitu 90:10, 80:20, dan 70:30 dengan menggunakan K=3, 5, 7, 9, 11, 13, dan 15. Output klasifikasi seperti pada Gambar 4 berikut ini.

menghasilkan akurasi tertinggi 83.33%, *precision* 83.33%, dan *recall* 83.33%. Tahapan pengujian nilai akurasi dengan menggunakan metode *modified k-nearest neighbor*. Berikut Gambar 5 tampilan *confusion matrix* pada pengujian.

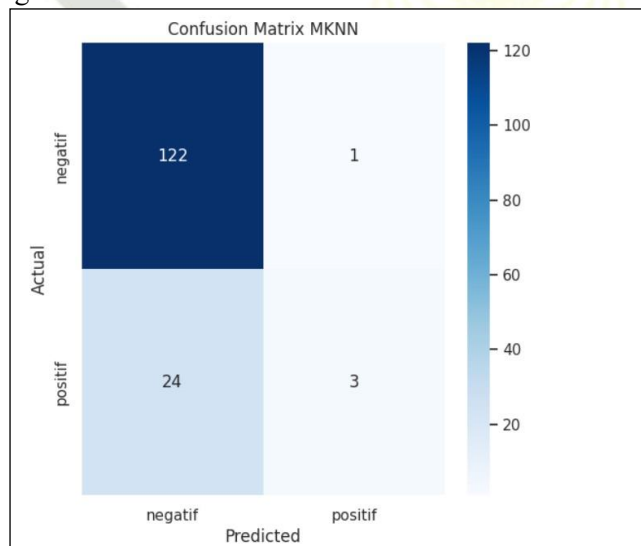
	000	10	2022	abang	aceh	ahok	akibat	alas	allah	angkut	...	upah	usaha	utang	utara	wajar	warga	wkwk	yogyakarta	Kelas Target	Hasil Klasifikasi		
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif		
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif	negatif	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.339803	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
145	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif	negatif
146	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
147	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
148	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
149	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif

50 rows x 187 columns

Gambar 4. Tampilan Hasil Klasifikasi

### 3.4. Nilai Akurasi

Hasil akurasi pengujian data opini menggunakan perbandingan 90:10 dan K=3



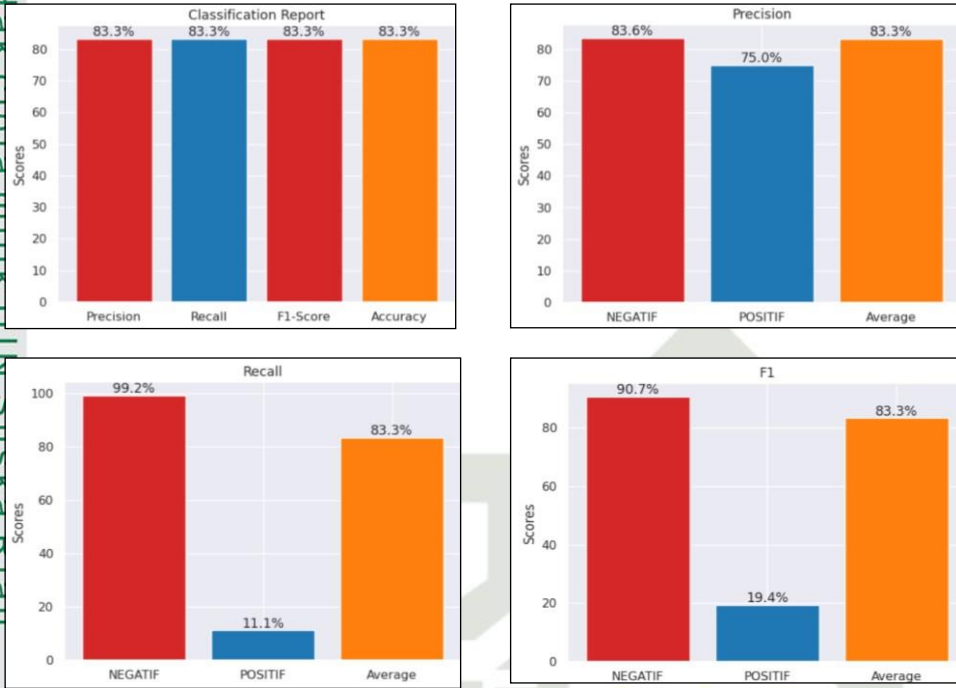
Gambar 5. Tampilan *Confusion Matrix* Opini

Dalam pengujian data opini memiliki grafik *classification report*, *Precision*, *recall*, dan *score* F1 seperti Gambar 6.

Hak Cipta Diindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

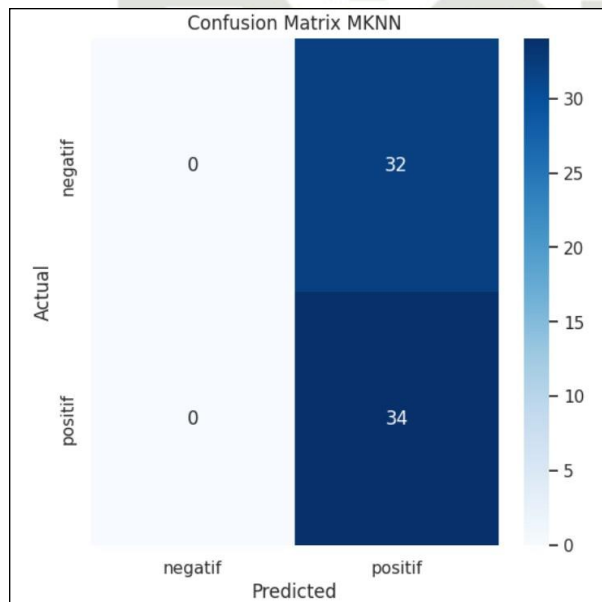
**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 6. Grafik *Classification Report* Opini

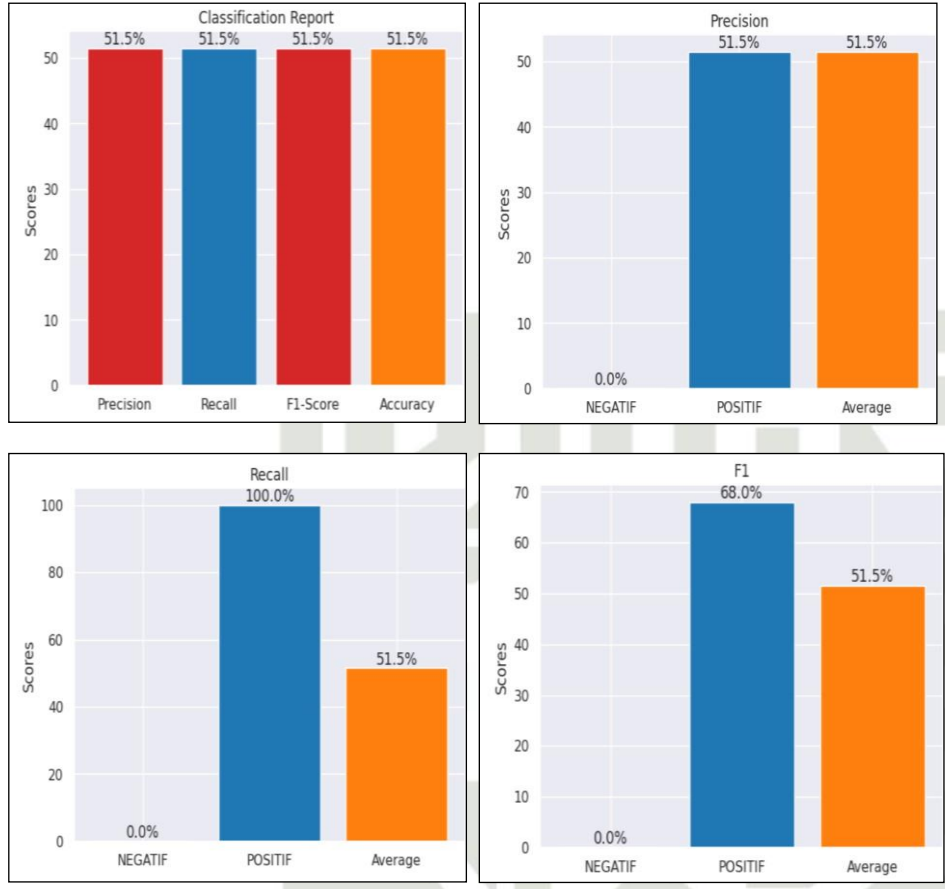
Sedangkan pengujian data balance *precision* 53.5%, dan *recall* 53.5%. dengan perbandingan 90:10 dan K=7 Tampilan grafik *confusion matrix* pada menghasilkan akurasi tertinggi 51.5%, Gambar 7.



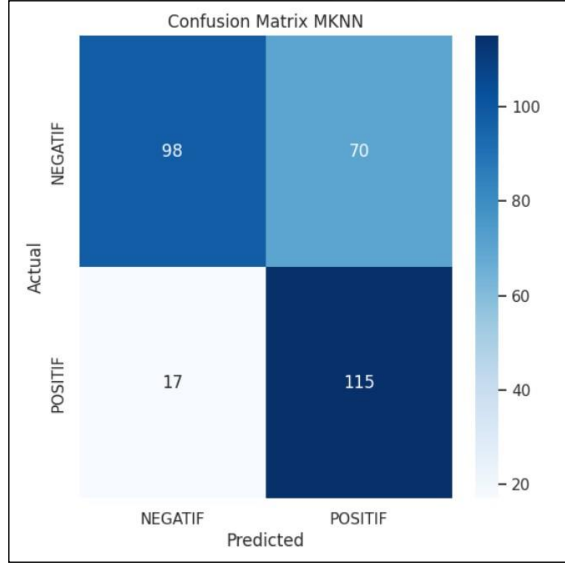
Gambar 7. Grafik *Confusion Matrix Balance*

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Gambar 8 tampilan grafik *classification* pengujian data balance dengan perbandingan *report*, *precision*, *recall*, dan *scor* F1 90:10 menggunakan K=3.



Gambar 8. Grafik *Classification Matrix Balance*



Pengujian data opini – non opini yang dilakukan dengan perbandingan 90:10 dan K=11 menghasilkan akurasi tertinggi 71.0%, *precision* 71.0%, dan *recall* 71.0%. Berikut tampilan *confusion matrix* pada Gambar 10.

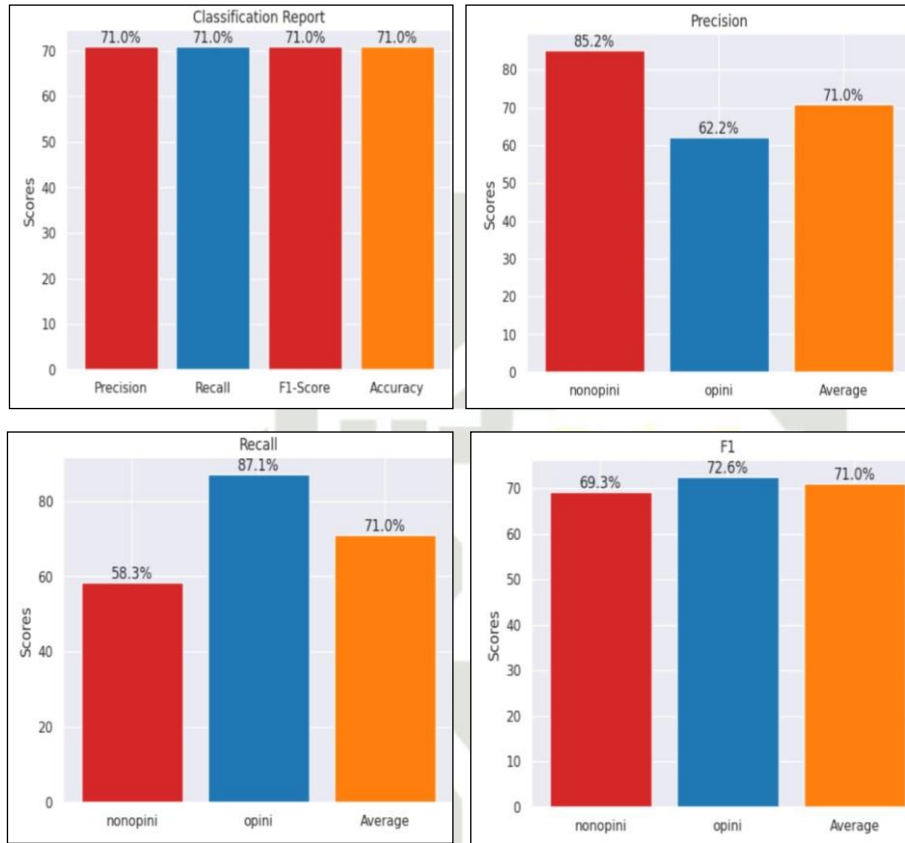
Gambar 9. Grafik *Confusion Matrix*

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Nonopini

Kredit Barang Elektronik dan Furniture.

Berikut tampilan grafik *classification report*, *precision*, *recall*, dan F1 pada pengujian data opini-nonopini menggunakan perbandingan 90:10 dan K=11.



Gambar 10. Grafik *Classification Report*

#### 4. Kesimpulan

Berdasarkan hasil penelitian dan pengujian metode *Modified K-Nearest Neighbor* untuk pengklasifikasian kenaikan harga Bahan Bakar Minyak dengan *tweet* berbahasa Indonesia didapat kesimpulan bahwa penggunaan metode MK-NN mendapatkan tingkat akurasi tertinggi pada data opini sebesar 83.33% dengan perbandingan 90:10 dan K=3, Yang artinya baik dalam penelitian mengenai kenaikan BBM.

#### 5. Referensi

Amelia, S., Nor Hayati, M., & Prangga, S. (2022). Penerapan Metode Modified K-

Nearest Neighbor Pada Pengklasifikasian Status Pembayaran

*Statistika*, 22(1), 95–104.

Anggraini, W. P., Utami, M. S., Berlianty, J. M., & Sellya, E. (2021). *Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Prakerja di Indonesia*. 13(4).

Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, 2(1), 32–41.

Firmansah, R. A., Indrajir, R. E., & Dazki, E. (2022). Perancangan Digital Enterprise Architect Smart course Pada Industri Pendidikan. *SATIN - Sains Dan Teknologi Informasi*, 8(2).  
<https://doi.org/10.33372/stn.v8i2.876>

Nofiar, A., & Muhammad Ridwan. (2022). Alat Pendeteksi Ketepatan Penggunaan Masker Berbasis Arduino Menggunakan Bahasa Pemrograman Python. *SATIN - Sains Dan Teknologi Informasi*, 8(1), 69–81.  
<https://doi.org/10.33372/stn.v8i1.835>

Nuriza, A., & Hidayat, N. (2020). *Klasifikasi Review Produk Kecantikan Pada Aplikasi Sociolla Menggunakan Algoritme Modified K-Nearest Neighbor (MK-NN) dengan Pembobotan BM25*. 4(10), 3426–3431.

Paramitha, A. A., Indriati, & Sari, Y. A. (2020). Analisis Sentimen Terhadap Ulasan Pengguna MRT Jakarta Menggunakan Information Gain dan Modified K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(4), 1125–1132.

Rozi, F. N., & Sulistyawati, D. H. (2019). Klasifikasi Berita Hoax Pilpres Menggunakan Metode Modified K-Nearest Neighbor Dan Pembobotan Menggunakan Tf-Idf. *Konvergensi*, 5(1).

Saputra, A. I., Oktavianto, H., Azizah, H., & Faruq, A. (2021). Penerapan Algoritma

Modified K-Nearest Neighbour (MKNN) pada Klasifikasi Masa Studi Mahasiswa Teknik Informatika. *Jurnal Smart Teknologi*, 3(1), 2774–1702.

Setiawan, R. (n.d.). *Sentiment Analysis on River Normalization for Flood Prevention with Modified K-Nearest Neighbor (MKNN)*.

TRISNO, M. O. H., Primartha, R., & Utami, A. S. (2020). *Perbandingan Kinerja Algoritma K-Nearest Neighbor Dan Modified K-Nearest Neighbor Pada Klasifikasi Website Phishing*.

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
  2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.