

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP IBUKOTA  
BARU MENGGUNAKAN METODE K-NEAREST NEIGHBORS**

**SKRIPSI**



disusun oleh

**Adi Surya Basri**

**16.11.0808**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2022**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP IBUKOTA  
BARU MENGGUNAKAN METODE K-NEAREST NEIGHBORS**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Adi Surya Basri**

**16.11.0808**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2022**

# **PERSETUJUAN**

## **SKRIPSI**

### **ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP IBUKOTA BARU MENGGUNAKAN METODE K-NEAREST NEIGHBORS**

yang dipersiapkan dan disusun oleh

**Adi Surya Basri**

**16.11.0808**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 14 Mei 2016

**Dosen Pembimbing,**

**Erni Seniwati, S.Kom., M.Cs.**  
**NIK. 190302231**

**PENGESAHAN**

**SKRIPSI**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP IBUKOTA  
BARU MENGGUNAKAN METODE K-NEAREST NEIGHBORS**

yang dipersiapkan dan disusun oleh

**Adi Surya Basri**

**16.11.0808**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 20 April 2020

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Muhammad Rudyanto Arief, M.T.**  
**NIK. 190302098**

**Sharizata Dyah Anggita, M.Kom**  
**NIK. 190302285**

**Erni Seniwati, S.Kom, M.Cs**  
**NIK. 190302231**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 9 Mei 2020

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, M.Kom**  
**NIK. 190302096**

### PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 19 Mei 2022



Adi Surya Basri

NIM. 16.11.0808

## Motto

"Jangan menjelaskan dirimu kepada siapa pun, karena yang menyukaimu tidak butuh itu. Dan yang membencimu tidak percaya itu."

**(Ali bin Abi Thalib)**

"Memangnya kenapa kalau hidup kita tak sempurna? Toh ini bukanlah surga."

**(Nourman Ali Khan)**

"Selalu tinggalkan orang-orang yang kamu cintai dengan kata-kata indah. Mungkin inilah terakhir kalinya kamu melihatnya."

**(Mufti Ismail Menk)**



## PERSEMBAHAN

Puji syukur saya ucapkan kepada ALLAH SWT, atas izin-nya penulis dapat menyelesaikan Skripsi berjudul “**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP IBUKOTA BARU MENGGUNAKAN METODE K-NEAREST NEIGHBORSB**” sesuai yang di harapkan penulis. Dengan penuh rasa bangga dan haru penulis mempersembahkan skripsi ini kepada:

- 1.Allah SWT karena atas izin dan karunia-nya skripsi ini dapat dibuat dan diselesaikan.
- 2.Kedua orang tua yang selalu mensupport saya dalam kondisi apapun.
- 3.Kepada teman-teman yang mengenal saya yang di Mataram.
- 4.Kepada teman-teman penghuni kontrakan Tehanis.
- 5.Kepada akang burjo café.

## KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Puji dan syukur penulis persembahkan untuk Allah SWT yang telah memberikan rahmat, dan hidayah sehingga penulis dapat menyelesaikan skripsi ini. Tidak lupa sholawat dan salam penulis haturkan pada Nabi Muhammad SAW beserta para sahabatnya yang telah menuntun kita pada jalan yang terang dan penuh kebaikan.

Skripsi ini disusun dalam rangka memenuhi salah satu persyaratan kelulusan jenjang Program Sarjana Strata 1 pada Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Dengan selesainya skripsi ini, penulis tidak lupa mengucapkan terima kasih kepada:

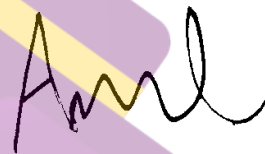
1. Ibu dan Bapak saya, yang selalu menyelipkan doa di setiap sujudnya agar penulis dapat menjadi pribadi yang lebih baik dan terus maju.
2. Bapak Prof. Dr. M. Suyanto, M.M. , selaku Rektor Universitas AMIKOM Yogyakarta.
3. Hanif Al Fatta, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
4. Ibu Erni Seniwati, S.Kom, M.Cs selaku Dosen Pembimbing yang telah memberikan bimbingan, saran, dan waktunya dengan sepenuh hati.
5. Bapak Muhammad Rudyanto Arief, M.Cs dan Ibu Sharazita Dyah Anggita, M.kom sebagai dosen penguji serta semua dosen Prodi Informatika Universitas Amikom Yogyakarta, terima kasih atas semua jasa Bapak dan Ibu Dosen.
6. Segenap Dosen dan Civitas Akademika Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu dan pengalaman kepada penulis selama menjalani perkuliahan.
7. Seluruh pihak yang telah banyak membantu.



Penulis menyadari jika pembuatan skripsi ini masih terdapat banyak kekurangan dan kelemahannya. Maka dari itu penulis sangat berharap kepada semua pihak agar dapat memberikan saran dan kritik yang membangun untuk menambah kesempurnaan skripsi ini. Tentu saja penulis memiliki harapan besar agar skripsi ini dapat memberikan manfaat bagi semua pihak yang membacanya.

Wassalamualaikum Wr. Wb.

Yogyakarta, 9 Mei 2022



Adi Surya Basri

NIM : 16.11.0808

## DAFTAR ISI

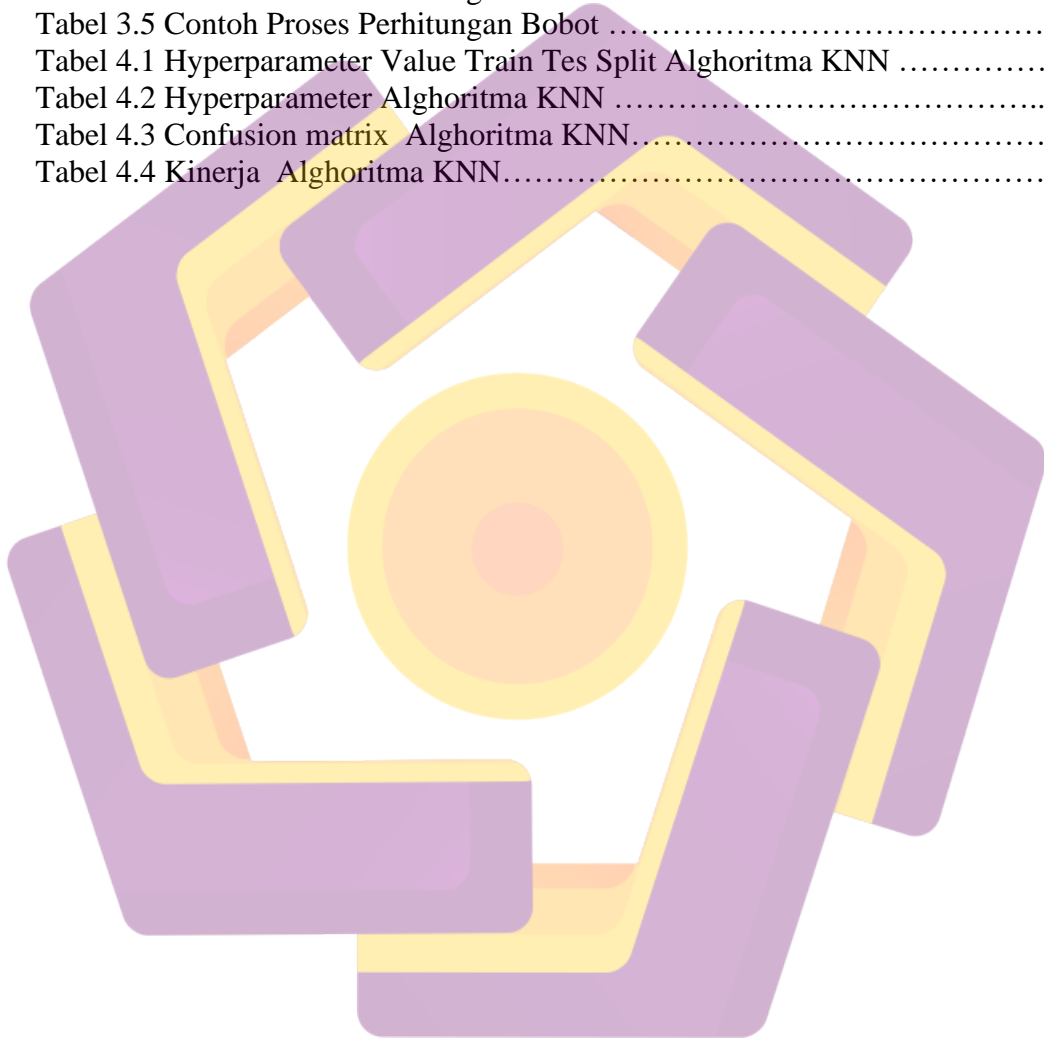
JUDUL.....	i
PERSETUJUAN .....	ii
PENGESAHAN .....	iii
PERNYATAAN.....	iv
MOTTO .....	v
PERSEMBAHAN.....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	1
1.1    LATAR BELAKANG.....	1
1.2    RUMUSAN MASALAH .....	2
1.3    BATASAN MASALAH.....	3
1.4    TUJUAN PENELITIAN .....	3
1.5    METODE PENELITIAN .....	3
1.6    SISTEMATIKA PENULISAN .....	4
BAB II LANDASAN TEORI.....	6
2.1    KAJIAN PUSTAKA.....	6
2.2    TWITTER .....	8
2.3    TEXT MINING.....	8
2.4    PREPROCESSING .....	9
2.5    PEMBOBOTAN TF-IDF .....	10
2.6    KLASIFIKASI.....	10
2.7    K-NEAREST NEIGHBORS (KNN).....	11
2.8    PENGUJIAN DENGAN CONFUSION MATRIX.....	11
2.9    BAHASA PEMROGRAMAN PHYTON.....	12

<b>BAB III METODOLOGI PENELITIAN.....</b>	<b>14</b>
3.1 <b>MENGIDENTIFIKASI MASALAH.....</b>	<b>15</b>
3.2 <b>MERUMUSKAN MASALAH.....</b>	<b>15</b>
3.3 <b>PENENTUAN TUJUAN.....</b>	<b>15</b>
3.4 <b>STUDI LITERATUR.....</b>	<b>16</b>
3.5 <b>ANALISIS DAN PERANCANGAN.....</b>	<b>16</b>
3.5.1 <b>Alat dan Bahan Penelitian.....</b>	<b>16</b>
3.5.2 <b>Pengumpulan Data.....</b>	<b>17</b>
3.5.3 <b>Penentuan Query.....</b>	<b>17</b>
3.5.4 <b>Preprocessing.....</b>	<b>18</b>
3.5.5 <b>Pembobotan Dengan TF-IDF.....</b>	<b>20</b>
3.5.6 <b>Klasifikasi Menggunakan KNN.....</b>	<b>22</b>
3.5.7 <b>Perancangan Sistem.....</b>	<b>23</b>
3.5.8 <b>Pengukuran Akurasi.....</b>	<b>24</b>
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>	<b>25</b>
4.1 <b>IMPLEMENTASI.....</b>	<b>25</b>
4.1.1 <b>Pengumpulan Data/ Crawling.....</b>	<b>25</b>
4.2 <b>PREPROCESSING DATA.....</b>	<b>26</b>
4.2.1 <b>Normalisasi Data.....</b>	<b>26</b>
4.2.2 <b>Pelabelan Data.....</b>	<b>28</b>
4.2.3 <b>Tokenizing Data.....</b>	<b>29</b>
4.2.4 <b>Stopword Removal/Filtering.....</b>	<b>30</b>
4.2.5 <b>Stemming.....</b>	<b>31</b>
4.3 <b>SPLIT DATASET.....</b>	<b>32</b>
4.4 <b>PEMBOBOTAN TF-IDF.....</b>	<b>32</b>
4.5 <b>KLASIFIKASI MENGGUNAKAN KNN.....</b>	<b>33</b>
4.6 <b>HASIL PENGUJIAN K-NEAREST NEIGHBORS (KNN).....</b>	<b>35</b>
<b>BAB V PENUTUP.....</b>	<b>40</b>
5.1 <b>KESIMPULAN.....</b>	<b>40</b>
5.2 <b>SARAN.....</b>	<b>40</b>
<b>DAFTAR PUSTAKA.....</b>	<b>41</b>



## DAFTAR TABEL

Tabel 2.1 Persamaan dan Perbedaan dengan Penelitian Sebelumnya.....	7
Tabel 2.2 Confusion matrix.....	22
Tabel 3.1 Spesifikasi Perangkat Keras.....	16
Tabel 3.2 Contoh Proses Perhitungan TF.....	20
Tabel 3.3 Contoh Proses Perhitungan DF .....	21
Tabel 3.4 Contoh Proses Perhitungan IDF.....	21
Tabel 3.5 Contoh Proses Perhitungan Bobot .....	22
Tabel 4.1 Hyperparameter Value Train Tes Split Algoritma KNN .....	35
Tabel 4.2 Hyperparameter Algoritma KNN .....	36
Tabel 4.3 Confusion matrix Algoritma KNN.....	36
Tabel 4.4 Kinerja Algoritma KNN.....	37



## DAFTAR GAMBAR

Gambar 3. 1 Alur Metode Penelitian.....	14
Gambar 3. 2 Flowchart Proses Query.....	17
Gambar 3. 3 Contoh Normalisasi Data.....	18
Gambar 3. 4 Contoh Casefolding.....	18
Gambar 3. 5 Contoh Tokenizing.....	19
Gambar 3. 6 Contoh Stopword.....	20
Gambar 3. 7 Contoh Stemming.....	20
Gambar 3. 8 Diagram Alir Proses Perancangan Sistem.....	23
Gambar 3. 9 Proses Pengukuran Akurasi Algoritma KNN.....	24
Gambar 4. 1 Source Code dan Output Pengambilan Data Tweet.....	26
Gambar 4. 2 Source Code dan Ouput Normalisasi Data.....	27
Gambar 4. 3 Source Code dan Output Pelabelan Data.....	28
Gambar 4. 4 Source Code dan Output Proses Tokenizing.....	29
Gambar 4. 5 Source Code dan Output Proses Stopword Removal / Filtering.....	30
Gambar 4. 6 Source Code dan Output Proses Stemming.....	31
Gambar 4. 7 Source Code Proses Split Dataset.....	32
Gambar 4. 8 Source Code dan Ouput Proses Pembobotan dengan TF-IDF.....	33
Gambar 4. 9 Source Code Proses Klasifikasi Menggunakan KNN.....	34
Gambar 4. 10 Grafik Pengujian Variasi Nilai k Tetangga.....	35

## INTISARI

Angka pengguna internet yang mengakses media sosial mencapai 85,7%. Sosial media Twitter memiliki pengguna 7.05% atau sekitar 12,06 juta pengguna di Indonesia. Terdapat banyak informasi dan tweet dari para penggunanya tentang berbagai hal, salah satunya kebijakan pemerintah. Kebijakan pemerintah yang trending topic di Twitter adalah tentang rencana pembangunan Ibu Kota Baru.

Metode K-NN digunakan untuk classification atau clustering tweet berdasarkan tetangga terdekat pada data latih. Dengan begitu penelitian ini dapat digunakan untuk menentukan apakah tweet tentang rencana pembangunan Ibu Kota Baru tersebut berorientasi sentimen positif atau negatif serta mengetahui presentase akurasi dari Metode K-NN dalam classification tweet Bahasa Indonesia.

Studi ini menganalisis sentimen pengguna Twitter dalam bentuk data tweet dalam bahasa Indonesia dengan Metode K-NN (K-Nearest Neighbors). Tes tersebut menggunakan data tweet Maximal dari 1000 tweet dengan kata kunci ibu kota baru, dilakukan dengan berbagi 80% data pelatihan dan 20% data pengujian untuk melihat keakuratan dan kinerja hasil klasifikasi. Berdasarkan hasil tes diperoleh akurasi 80%.

Hasilnya, 83% pengguna twitter merespons positif ibu kota baru, sementara 17% merespons negatif.

**Kata Kunci:** Data mining, Klasifikasi, Sentimen Analisis, Algoritma K-NN.

## **ABSTRACT**

*The number of internet users accessing social media reached 85.7%. Twitter has 7.05% of users or about 12.06 million users in Indonesia. There is a lot of information and tweets from its users about various things, one of which is government policy. The government's policy that is trending topic on Twitter is about the development plan of the New Capital City.*

*The K-NN method is used for classification or clustering of tweets based on the nearest neighbor on the training data. That way this research can be used to determine whether tweets about the New Capital Development Plan are oriented towards positive or negative sentiments and find out the percentage of accuracy of the K-NN Method in the classification of Indonesian tweets.*

*The study analyzed Twitter user sentiment in the form of tweet data in Indonesian with the K-NN (K-Nearest Neighbors) Method. The test used Maximal tweet data of 1000 tweets with new capital city keywords, conducted with the sharing of 80% of training data and 20% of test data to see the accuracy and performance of classification results. Based on the test results obtained 80% accuracy.*

*As a result, 83% of twitter users responded positively to the new capital city while 17% responded negatively.*

**Keyword:** *Data mining, classification, sentiment analysis, K-NN Algorith*