

**KOMPARASI ALGORITMA NAIVE BAYES DAN K-NEAREST  
NEIGHBORS UNTUK KLASIFIKASI JUDUL BERITA**

Disusun sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Teknik Informatika



Disusun oleh:

**NAJMUNDA ZIA AKMAL**

**18106050051**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA YOGYAKARTA**

**2022**



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1961/Un.02/DST/PP.00.9/08/2022

Tugas Akhir dengan judul : Komparasi Algoritma Naive Bayes dan K-Nearest Neighbors untuk Klasifikasi Judul Berita

yang dipersiapkan dan disusun oleh:

Nama : NAJMUNDA ZIA AKMAL  
Nomor Induk Mahasiswa : 18106050051  
Telah diujikan pada : Senin, 29 Agustus 2022  
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.  
SIGNED

Valid ID: 630db9523a814



Penguji I

Nurochman, S.Kom., M.Kom.  
SIGNED

Valid ID: 630dkbc9a5f50



Penguji II

Agus Mulyanto, S.Si., M.Kom.  
SIGNED

Valid ID: 630db6f55e47f



Yogyakarta, 29 Agustus 2022  
UIN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi

Dr. Dra. Hj. Khurul Wardati, M.Si.  
SIGNED

Valid ID: 630d16c5aadd4e



**SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR**

Hal : Persetujuan Skripsi  
Lamp :

Kepada  
Yth. Dekan Fakultas Sains dan Teknologi  
UIN Sunan Kalijaga Yogyakarta  
di Yogyakarta

*Assalamu 'alaikum wr. wb.*

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:


Nama : Najmunda Zia Akmal  
NIM : 18106050051  
Judul Skripsi : Komparasi Algoritma Naive Bayes dan K-Nearest Neighbors untuk  
Klasifikasi Judul Berita

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

*Wassalamu 'alaikum wr. wb.*

Yogyakarta, 25 Agustus 2022  
Pembimbing,

  
Maria Ulfah Siregar, S.Kom. MIT., Ph.D.  
NIP. 19780106 200212 2 001

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : Najmunda Zia Akmal  
NIM : 18106050051  
Program Studi : Teknik Informatika  
Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul "**Komparasi Algoritma Naive Bayes dan K-Nearest Neighbors untuk Klasifikasi Judul Berita**" merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah diajukan untuk memperoleh gelar sarjana di UIN Sunan Kalijaga Yogyakarta, dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 25 Agustus 2022



Najmunda Zia Akmal  
NIM. 18106050051

STATE ISLAMIC UNIVERSITY  
**SUNAN KALIJAGA**  
YOGYAKARTA

## KATA PENGANTAR

Puja dan puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, nikmat dan petunjuk-Nya, sehingga penulis dapat menyelesaikan tugas akhir skripsi dengan judul “Komparasi Algoritma Naive Bayes dan K-Nearest Neighbors untuk Klasifikasi Judul Berita”. Shalawat serta salam penulis tujukan untuk Nabi Muhammad SAW.

Tugas akhir ini disusun sebagai salah satu syarat untuk menyelesaikan studi S-1 jurusan Teknik Informatika di UIN Sunan Kalijaga Yogyakarta. Dalam penyusunan laporan ini, penulis mendapat banyak bantuan yang diberikan dari berbagai pihak. Oleh karena itu pada bagian pengantar, izinkan penulis berterima kasih kepada:

1. Allah SWT atas limpahan rezeki-Nya, baik itu berupa kesehatan, kesempatan, ridho, dan segala hal yang tidak terkira macam dan jumlahnya.
2. Bapak dan Ibu saya atas segala kasih sayang, kesabaran, dan pengorbanan dalam merawat dan menemani setiap langkah saya.
3. Saudara saya, Nawwab Zia Ajnaden yang telah menemani saya selama masa studi dan mengantarkan saya pergi-pulang ke/dari kampus.
4. Ibu Maria Ulfah Siregar, S.Kom. MIT., Ph.D. selaku Dosen Pembimbing Skripsi atas waktu, nasihat, pikiran, dan bimbingannya selama pengerjaan skripsi.

5. Bapak Agus Mulyanto, S.Si., M.Kom. selaku Dosen Pembimbing Akademik yang senantiasa mengerti dan membantu jika saya mempunyai pertanyaan atau masalah terkait akademik dan perkuliahan.
6. Teman-teman Teknik Informatika Angkatan 2018 yang telah membuat lingkungan belajar yang nyaman, dan senantiasa menemani dan menerima saya selama masa studi.
7. Bapak dan Ibu Dosen Teknik Informatika UIN Sunan Kalijaga atas segala pembelajaran, kesabaran, dan nasehatnya.
8. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang terlibat dalam hal apapun terkait penyusunan tugas akhir skripsi maupun secara umum.

Akhir kata, penulis menyadari bahwa tugas akhir ini masih jauh dari kesempurnaan. Oleh karena itu, kritik dan saran dari segala aspek sangat diharapkan. Semoga tersusunnya tugas akhir ini dapat memberi bermanfaat untuk berbagai pihak tanpa terkecuali.

Yogyakarta, 23 Agustus 2022

Penulis,



Najmunda Zia Akmal

## **HALAMAN PERSEMBAHAN**

Untuk Bapak & Ibu atas kasih sayang, rasa sabar, pengorbanan tiada batas dan henti.

Untuk orang-orang yang pernah saya kenal dan temui atas segala kebaikan dan keikhlasan yang sudah diberi.



## HALAMAN MOTO

*“Sebelum kau sampaikan alasanmu  
Harus dicoba dulu dengan kekuatan yang ada”*

*“Di Sini Rhodes, Lompatlah!” oleh JKT48*





## DAFTAR ISI

<b>HALAMAN PENGESAHAN</b>	<b>II</b>
<b>SURAT PERSETUJUAN SKRIPSI</b>	<b>III</b>
<b>SURAT PERNYATAAN KEASLIAN</b>	<b>IV</b>
<b>KATA PENGANTAR</b>	<b>V</b>
<b>HALAMAN PERSEMBAHAN</b>	<b>VII</b>
<b>HALAMAN MOTO</b>	<b>VIII</b>
<b>DAFTAR ISI</b>	<b>IX</b>
<b>DAFTAR TABEL</b>	<b>XI</b>
<b>DAFTAR GAMBAR</b>	<b>XII</b>
<b>DAFTAR KODE</b>	<b>XIII</b>
<b>INTISARI</b>	<b>XIV</b>
<b>ABSTRACT</b>	<b>XV</b>
<b>BAB I PENDAHULUAN</b>	<b>1</b>
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian	2
1.4. Batasan Masalah	3
1.5. Manfaat Penelitian	3
1.6. Kebaruan Penelitian	3
<b>BAB II TINJAUAN PUSTAKA &amp; LANDASAN TEORI</b>	<b>5</b>
2.1. Tinjauan Pustaka	5
2.2. Landasan Teori	8
2.2.1. Preprocessing	8
2.2.2. Klasifikasi	9
2.2.3. Representasi Fitur	11
2.2.4. Evaluasi	13
<b>BAB III METODE PENELITIAN</b>	<b>16</b>

3.1. Kebutuhan Sistem dan Bahan Penelitian	16
3.1.1. Perangkat Keras	16
3.1.2. Perangkat Lunak	16
3.1.3. Bahan Penelitian	17
3.2. Metode Penelitian	18
3.2.1. Studi Pustaka	18
3.2.2. Pemilihan dan Pembagian Data	19
3.2.3. Preprocessing	19
3.2.4. Perhitungan TF-IDF	19
3.2.5. Training Model	20
3.2.6. Evaluasi	20
<b>BAB IV HASIL &amp; PEMBAHASAN</b>	<b>22</b>
4.1. Pemilihan dan Pembagian Data	22
4.1.1. Pemilihan Data	23
4.1.2. Pembagian Data	25
4.2. Preprocessing	26
4.2.1. Lower-casing	26
4.2.2. Penghilangan stopword, tanda baca, dan angka	26
4.2.3. Stemming	27
4.2.4. Tokenisasi	27
4.3. Perhitungan TF-IDF	28
4.4. Training Model	30
4.5. Evaluasi	34
4.6. Interface Klasifikasi Judul Berita	38
<b>BAB V PENUTUP</b>	<b>39</b>
5.1 Kesimpulan	39
5.2 Saran	40
<b>DAFTAR PUSTAKA</b>	<b>41</b>
<b>CURRICULUM VITAE</b>	<b>43</b>

## DAFTAR TABEL

Tabel 2.1 Tabel Tinjauan Pustaka	7
Tabel 4.1 Dataset	23
Tabel 4.2 Perubahan Dataset	25
Tabel 4.3 Perubahan Dataset pada Index 17995 setelah Dikenai Preprocessing dan Perhitungan TF-IDF	30
Tabel 4.4 Kemungkinan Parameter yang Digunakan pada Objek dari Kelas MultinomialNB dan KNeighborsClassifier	31
Tabel 4.5 Tabel Parameter Terbaik 6 Model	32
Tabel 4.6 Tabel Hasil Metrik Akurasi, Precision, Recall, F1, AUC, dan Waktu Train dan Prediksi	36

## DAFTAR GAMBAR

Gambar 2.1 Tahapan Training Multinomial Naive Bayes Classifier	10
Gambar 2.2 Tahapan Prediksi Multinomial Naive Bayes Classifier	10
Gambar 2.3 Tahapan Prediksi KNN	11
Gambar 3.1 Bagan Alur Tahapan Penelitian	18
Gambar 4.1 Flowchart Tahap Penelitian	22
Gambar 4.2 Tampilan Interface Sederhana	38



## DAFTAR KODE

Kode 4.1. Pemilihan Data dan Penghapusan Kolom	24
Kode 4.2. Pembagian Data Latih dan Data Uji	26
Kode 4.3. Fungsi Preprocessing	28
Kode 4.4. Preprocessing, Perhitungan TF-IDF, dan Normalisasi	30
Kode 4.5 Training Model Multinomial Naive Bayes	33
Kode 4.6 Training Model KNN	34
Kode 4.7 Fungsi Evaluasi	35



# KOMPARASI ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBORS UNTUK KLASIFIKASI JUDUL BERITA

Najmunda Zia Akmal

18106050051

## INTISARI

Keberadaan berita yang ada sejak dulu membuat jumlah berita yang dihasilkan sampai saat ini berjumlah sangat besar dan akan terus bertambah. Selain itu, tema berita juga sangat beragam, dimana hampir seluruh hal dapat dibahas di dalam berita. Sangat beragamnya tema membuat perusahaan pembuat berita harus mengelompokkan berita-berita tersebut, yang mana saat ini masih dilakukan secara manual. Penerapan algoritma pembelajaran mesin dapat menjadi solusi untuk mengelompokkan judul berita secara otomatis.

Pada penelitian ini, dibandingkan nilai akurasi, precision, recall, f1, AUC, dan waktu yang dibutuhkan dalam melatih dan melakukan prediksi oleh algoritma Naive Bayes dan KNN dalam kasus klasifikasi data berupa judul berita yang terdiri dari 4 kategori berita.

Hasil penelitian menunjukkan bahwa algoritma Naive Bayes bekerja lebih baik dibandingkan KNN dengan nilai akurasi, f1, AUC, dan waktu latih dan prediksi berurut 0,840; 0,840; 0,964; 0,090 sekon berbanding 0,796; 0,797; 0,944; 0,944 sekon untuk data dengan satu kata sebagai token; 0,851; 0,851; 0,968; 0,107 sekon berbanding 0,801; 0,802; 0,945; 0,927 sekon untuk data dengan 1 dan 2 kata sebagai token; dan 0,853; 0,853; 0,968; 0,115 sekon berbanding 0,795; 0,796; 0,944; 1,095 sekon untuk data dengan 1, 2, dan 3 kata sebagai token.

**Kata kunci:** klasifikasi, Naive Bayes, KNN, berita

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## **COMPARISON OF NAIVE BAYES AND K-NEAREST NEIGHBORS ALGORITHM FOR NEWS TITLES CLASSIFICATION**

**Najmunda Zia Akmal**

**18106050051**

### **ABSTRACT**

The existence of news that has existed for a long time makes the amount of produced news very large and will continue to grow. In addition, news themes are also very diverse, where almost everything can be discussed in the news. The variety of themes makes news-making companies have to group the news, which is still done manually. The application of machine learning algorithms can be a solution to group news automatically.

This study aims to compare the value of accuracy, precision, recall, f1, AUC, and Train-Prediction Time of the Naive Bayes and KNN algorithms in the classification of news titles that consist of 4 news categories.

The results showed that the Naive Bayes algorithm worked better than KNN with accuracy value, f1, AUC, and Train-Prediction Time are 0.840, 0.840, 0.964, 0.090 second versus 0.796, 0.797, 0.944, 0.944 second for data with one word as token; 0.851, 0.851, 0.968, 0.107 second versus 0.801, 0.802, 0.945, 0.927 second for data with 1 and 2 words as tokens; and 0.853, 0.853, 0.968, 0.115 second versus 0.795, 0.796, 0.944, 1.095 second for data with 1, 2, and 3 words as tokens.

**Keywords:** Classification, Naive Bayes, KNN, News

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Dari zaman dahulu hingga saat ini, berita senantiasa menjadi salah satu sumber fakta untuk semua orang. Cara berita disampaikan juga berkembang mengikuti zaman, yaitu dimulai dari mulut ke mulut, dicetak pada kertas, hingga saat ini dapat mudah ditemukan di telepon pintar. Keberadaan berita yang ada sejak dulu membuat jumlah berita yang dihasilkan sampai saat ini berjumlah sangat besar dan akan terus bertambah. Selain itu, tema berita juga sangat beragam, dimana hampir seluruh hal dapat dibahas di dalam berita. Sangat beragamnya tema membuat perusahaan pembuat berita harus mengelompokkan berita-berita tersebut, yang mana saat ini masih dilakukan secara manual.

Mengingat berita yang baru ditulis dan sudah lama ditulis berjumlah sangat besar, serta beragamnya tema berita, otomatisasi pengelompokan tema-tema berita dengan metode-metode pembelajaran mesin dapat menjadi solusi. Dengan pemilihan algoritma yang tepat, akurasi pengelompokan berita secara otomatis dapat meningkatkan akurasi pengelompokan secara manual.

Pada penelitian ini, penulis mencoba untuk membandingkan algoritma klasifikasi Naive Bayes dan K-Nearest Neighbor (KNN) untuk melakukan klasifikasi berita menggunakan judul berita yang merupakan data teks. Kedua algoritma dipilih karena keduanya merupakan algoritma yang umum digunakan, dan memiliki tahapan kerja yang sederhana sehingga membutuhkan sumber daya



yang kecil. Selain itu, keduanya dipilih untuk mewakili 2 jenis learning, dimana Naive Bayes mewakili probability-based learning dan KNN mewakili similarity-based learning (Ansari dkk, 2018).

Selain membandingkan kedua algoritma, pada penelitian ini juga dilakukan pembobotan dengan perhitungan Term Frequency - Inverse Document (TF-IDF) dan tokenisasi dengan perbedaan jumlah kata sebagai satu token (1 kata; 1 dan 2 kata; dan 1, 2, 3 kata).

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang di atas, maka rumusan masalah yang ada pada penelitian ini adalah:

- a. Bagaimana cara menerapkan algoritma Naive Bayes dan K-Nearest Neighbor untuk klasifikasi judul berita?
- b. Bagaimana perbandingan algoritma Naive Bayes dan K-Nearest Neighbor dalam klasifikasi judul berita?

### **1.3. Tujuan Penelitian**

Tujuan penelitian ini adalah:

- a. Mengetahui cara menerapkan algoritma Naive Bayes dan K-Nearest Neighbor untuk klasifikasi judul berita.
- b. Mengetahui perbandingan algoritma Naive Bayes dan K-Nearest Neighbor dalam klasifikasi judul berita.

#### 1.4. Batasan Masalah

Agar penelitian ini lebih fokus, maka perlu adanya batasan masalah penelitian. Adapun batasan masalah penelitian ini adalah:

- a. Judul berita yang diklasifikasi merupakan judul berita berbahasa Inggris.
- b. Judul berita yang diklasifikasi merupakan berita dengan kategori *business*, *sports*, *politics* dan *entertainment*.
- c. Judul berita yang diklasifikasi berasal dari situs <https://www.huffingtonpost.com>, dan dimuat pada rentang tahun 2012-2018.
- d. Bahasa Pemrograman yang digunakan adalah Python.

#### 1.5. Manfaat Penelitian

Hasil penelitian ini diharapkan memiliki manfaat penelitian sebagai berikut:

- a. Sebagai referensi untuk menyelesaikan masalah yang sejenis.
- b. Menambah pengetahuan penulis dan pembaca dalam pemahaman algoritma Naive Bayes dan KNN, serta algoritma pembelajaran mesin secara umum.

#### 1.6. Kebaruan Penelitian

Penelitian tentang klasifikasi (implementasi/komparasi) menggunakan Naive Bayes dan KNN untuk data dengan jenis text sudah banyak dilakukan. Namun, Penelitian yang membandingkan kedua metode dengan menggunakan data berupa hanya judul berita saja belum pernah dilakukan sebelumnya. Selain

itu, penelitian ini juga menyertakan *interface* sederhana terkait pelabelan judul berita.



## **BAB V**

### **PENUTUP**

#### **5.1. Kesimpulan**

Setelah melakukan rangkaian proses penelitian, dapat disimpulkan bahwa:

1. Dalam kasus klasifikasi 4 kategori (*Sports, Business, Politics, dan Entertainment*) judul berita menggunakan dataset “News Category Dataset” dengan data latih sebesar 70% (12600 judul berita dengan 3150 tiap kategori) dari dataset dan data uji sebesar 30% (5400 judul berita dengan 1350 tiap kategori) dan diukur dengan metrik akurasi, recall, precision, F1, dan AUC, serta mengukur waktu yang dibutuhkan untuk melatih model dan melakukan prediksi, algoritma Naive Bayes menghasilkan metrik lebih baik (untuk 3 macam data dengan jumlah kata berbeda setiap satu token) daripada algoritma K-Nearest Neighbor.
2. Berdasarkan nilai akurasi terhadap data uji untuk Naive Bayes (0.840, 0.851, 0.853) menandakan bahwa algoritma Naive Bayes sudah baik untuk otomasi pemberian label kategori sebuah judul berita. Sementara itu berdasarkan nilai AUC, kedua algoritma mencapai nilai di atas 0.95. Sehingga berdasarkan nilai yang didapat dari metrik AUC, menunjukkan bahwa kedua algoritma memiliki performa yang baik untuk otomasi pemberian label kategori sebuah judul berita.
3. Menggunakan 1 dan 2 kata setiap satu token dapat meningkatkan akurasi, F1, dan nilai AUC Naive Bayes dan K-Nearest Neighbor. Namun,

penggunaan 1, 2, dan 3 kata setiap satu token hanya meningkatkan akurasi, F1, dan nilai AUC Naive Bayes saja.

## 5.2. Saran

Dalam melakukan penelitian selanjutnya, peneliti menyarankan bahwa:

1. Penerapan dan penggunaan dataset untuk masalah multilabel, mengingat dalam sebuah judul berita bisa saja memuat lebih dari satu tema.
2. Penggunaan nilai  $k$  dan metode perhitungan jarak yang berbeda dalam implementasi algoritma K-Nearest Neighbour karena penelitian ini hanya menyertakan opsi nilai  $k$  yaitu 3, 5, 7, 9, 11, 13, 15, dan opsi metode perhitungan jarak yaitu euclidean, manhattan, cosine.
3. Pencarian rentang waktu antar *update* model, mengingat judul berita baru berisi kata-kata baru.
4. Perbandingan algoritma klasifikasi lain selain KNN dan Naive Bayes.
5. Penggunaan lebih dari satu dataset untuk mendapatkan hasil perbandingan yang lebih valid.

## DAFTAR PUSTAKA

- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B. and Kochut, K., 2017. A brief survey of text mining: Classification, clustering and extraction techniques. arXiv preprint arXiv:1707.02919.
- Ansari, F., Erol, S. and Sihn, W., 2018. Rethinking human-machine learning in industry 4.0: how does the paradigm shift treat the role of human learning?. *Procedia manufacturing*, 23, pp.117-122.
- Fanny, F., Muliono, Y. and Tanzil, F., 2018. A comparison of text classification methods k-NN, Naïve Bayes, and support vector machine for news classification. *Jurnal Informatika: Jurnal Pengembangan IT*, 3(2), pp.157-160.
- Ghosh, S. and Desarkar, M.S., 2018, April. Class specific TF-IDF boosting for short-text classification: Application to short-texts generated during disasters. In *Companion Proceedings of the The Web Conference 2018* (pp. 1629-1637).
- Huang, X., Li, Z., Wang, C. and Ning, H., 2019. Identifying disaster related social media for rapid response: a visual-textual fused CNN architecture. *International Journal of Digital Earth*.
- Kusrini, E.T.L. and Taufiq, E., 2009. *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Limbong, J.J.A. and Sembiring, I., 2022. Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(2), pp.347-356.
- Linoff, G.S. and Berry, M.J., 2011. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.
- Loosemore, S., Stallman, R.M., McGrath, R., Oram, A. and Drepper, U., 2012. *The GNU C library reference manual: for version 2.17*. Free Software Foundation.
- Misra, R. (06 2018) 'News Category Dataset'. doi: 10.13140/RG.2.2.20331.18729.
- Misra, R. and Grover, J. (01 2021) *Sculpting Data for ML: The first act of Machine Learning*.
- MUHAMAD, K., Fachrurrozi, M. and Arsalan, O., 2019. PERBANDINGAN METODE NAÏVE BAYES DENGAN METODE K-NN UNTUK KLASIFIKASI BERITA.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, pp.2825-2830.
- Prasetyo, E., 2013. *Data Mining - Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

- Prasetyo, E., 2014. Data Mining - Mengolah Data menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi Offset.
- Pratama, B.Y. and Sarno, R., 2015, November. Personality classification based on Twitter text using Naive Bayes, KNN and SVM. In 2015 International Conference on Data and Software Engineering (ICoDSE) (pp. 170-174). IEEE.
- Powers, D.M., 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.
- Rajaraman, A. and Ullman, J.D., 2011. Mining of massive datasets. Cambridge University Press.
- Sasaki, Y., 2007. The Truth Oh the F--Measure. Manchester: School of Computer Science, University of Manchester.
- Schütze, H., Manning, C.D. and Raghavan, P., 2008. Introduction to information retrieval.
- Syarif, I., Prugel-Bennett, A. and Wills, G., 2016. SVM parameter optimization using grid search and genetic algorithm to improve classification performance. TELKOMNIKA (Telecommunication Computing Electronics and Control), 14(4), pp.1502-1509.
- Wahyono, T., 2018. Fundamental of Python for Machine Learning: Dasar-Dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan. Yogyakarta: Penerbit Gava Media.
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S. and Zhou, Z.H., 2008. Top 10 algorithms in data mining. Knowledge and information systems, 14(1), pp.1-37.
- Zhang, S., Zhang, C. and Yang, Q., 2003. Data preparation for data mining. Applied artificial intelligence, 17(5-6), pp.375-381.