

**ANALISIS KINERJA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR*  
UNTUK KLASIFIKASI JUDUL ARTIKEL DALAM JURNAL  
ILMIAH**

**SKRIPSI SARJANA REKAYASA TEKNOLOGI  
INFORMATIKA**

Oleh

**MUHAMMAD IRSYAD MARWAN**

197064416095



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI KOMUNIKASI DAN  
INFORMATIKA**

**UNIVERSITAS NASIONAL**

**2022**

**ANALISIS KINERJA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI JUDUL ARTIKEL  
DALAM JURNAL ILMIAH**

**SKRIPSI SARJANA**

Diajukan sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Teknik Teknologi Informatika dari Fakultas Teknologi Komunikasi dan  
Informatika

Oleh

**MUHAMMAD IRSYAD MARWAN**

**197064416095**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI KOMUNIKASI DAN  
INFORMATIKA  
UNIVERSITAS NASIONAL**

**2022**

HALAMAN PENGESAHAN

TUGAS AKHIR

**ANALISIS KINERJA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI JUDUL ARTIKEL DALAM JURNAL ILMIAH**



Muhammad Irsyad Marwan

197064416095

Dosen Pembimbing 1

A handwritten signature in black ink, appearing to read "Hayati".

Nur Hayati, S.Si., MTI

Dosen Pembimbing 2

A handwritten signature in black ink, appearing to read "Dhieka".

Dhieka Avrilia Lantana, S.kom, M.kom

## LEMBAR PERSETUJUAN TUGAS AKHIR

Tugas Akhir dengan judul :

### **ANALISIS KINERJA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI JUDUL ARTIKEL DALAM JURNAL ILMIAH**

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Akhir Semester Ganjil 2022-2023 pada tanggal 22 Februari Tahun 2023



**Dosen Pembimbing 1**

(.....)

Nur Hayati, S.Si., MTI

NID. 0110140843

**Ketua Program Studi**

(.....)

Ratih Titi Komala Sari, ST,

MM, MMSI

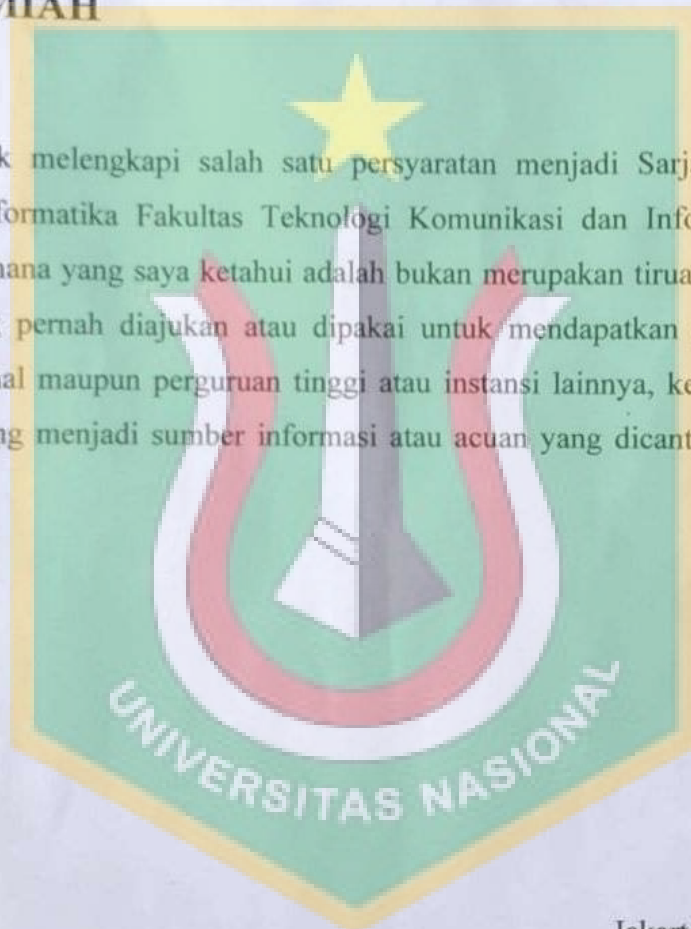
NID. 0103150850

## PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

### **ANALISIS KINERJA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI JUDUL ARTIKEL DALAM JURNAL ILMIAH**

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.



Jakarta, 09 Maret 2023



Muhammad Irsyad Marwan

197064416095

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Muhammad Irsyad Marwan  
NPM : 197064416095  
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika  
Program Studi : Informatika  
Tanggal Sidang : 22 Februari 2023

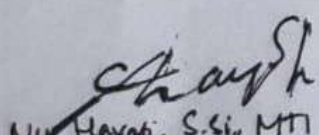

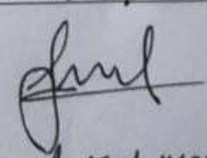
JUDUL DALAM BAHASA INDONESIA :

**ANALISIS KINERJA NAÏVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI JUDUL ARTIKEL DALAM JURNAL ILMIAH**

JUDUL DALAM BAHASA INGGRIS :

**PERFORMANCE ANALYSIS OF NAÏVE BAYES AND K-NEAREST NEIGHBOR FOR CLASSIFICATION OF ARTICLE TITLES IN SCIENTIFIC JOURNALS**

TANDA TANGAN DAN TANGGAL

| Pembimbing 1   | Ka. Prodi   | Mahasiswa   |
|--|---|---|
| TGL : 9/3/2023   | TGL : 9/3/2023  | TGL : 21/3/2023   |
| <br>Nur Hayati, S.Si, MTI | <br>Rakh Tihi Komala Sari, ST, M.Eng, M.Eng, M.Eng | <br>Muhammad Irsyad Marwan |

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Muhammad Irsyad Marwan  
NPM : 197064416095  
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika  
Program Studi : Informatika  
Tanggal Sidang : 22 Februari 2023

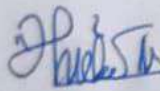

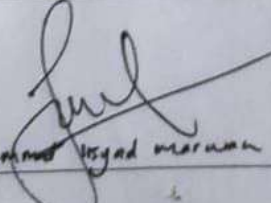
JUDUL DALAM BAHASA INDONESIA :

**ANALISIS KINERJA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI JUDUL ARTIKEL DALAM JURNAL ILMIAH**

JUDUL DALAM BAHASA INGGRIS :

**PERFORMANCE ANALYSIS OF *NAÏVE BAYES* AND *K-NEAREST NEIGHBOR* FOR CLASSIFICATION OF ARTICLE TITLES IN SCIENTIFIC JOURNALS**

TANDA TANGAN DAN TANGGAL

| Pembimbing 2  | Ka. Prodi   | Mahasiswa   |
|---|---|---|
| TGL : 9/3/2023  | TGL : 9/3/2023  | TGL : 9/3/2023  |
|  | <br>Rahm Feti Komala Sari, ST, Mde, MMSI | <br>Muhammad Irsyad Marwan |

## KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia sehingga penulis dapat menyelesaikan skripsi dengan judul **“ANALISIS KINERJA NAÏVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI JUDUL ARTIKEL DALAM JURNAL ILMIAH”** sebagai salah satu syarat kelulusan Program Studi Sarjana --- Fakultas Teknologi Komunikasi dan Informatika.

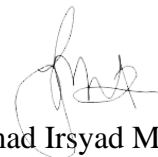
Penelitian dan penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak terima kasih terutama kepada dosen pembimbing Tugas Akhir, ibu Nur Hayati, S.Si., MTI dan ibu Dhieka Avrilia Lantana, S.Kom., M.Kom yang telah meluangkan banyak waktu, tenaga, pikiran, bimbingan, arahan, motivasi serta memaklumi segala kekurangan penulis selama penelitian tugas akhir dan penyusunan skripsi.

Penulis juga mengucapkan banyak terima kasih kepada :

1. Keluarga penulis yang telah banyak memberi dukungan salam segala bentuk yang tak terhitung.
2. Seluruh dosen pengajar di Program Studi Teknik Informatika FTKI maupun dosen di Program Studi lain yang memberikan banyak ilmu.
3. Teman-teman seangkatan dan sehimpuan berbagai angkatan yang telah membantu dan mendukung.

Akhir kata, semoga Tuhan Yang Maha Esa membalas kebaikan dan bantuan yang telah diberikan dengan hal yang lebih baik. Penulis mengharapkan kritik dan saran yang bersifat membangun dan semoga skripsi ini dapat memberikan manfaat di bidang Teknologi Informatika.

Jakarta, 22 Februari 2023



Muhammad Irsyad Marwan



## ABSTRAK

Perkembangan teknologi Informatika membuat kebutuhan akan otomasi pada sistem kerja semakin meningkat. Termasuk kebutuhan akan klasifikasi artikel jurnal secara otomatis mulai dibutuhkan untuk menentukan kategori jurnal yang tepat. Metode text mining telah dapat digunakan untuk melakukan otomatisasi artikel jurnal berdasarkan kategori. Penelitian ini memakai beberapa metode text mining yakni *Naïve Bayes(NB)* dan *K-Nearest Neighbor(K-NN)*. Dataset yang digunakan berupa judul artikel jurnal sejumlah 531 data yang kemudian dipisahkan menjadi data uji dan data latih. Hasil akurasi yang didapat dengan menggunakan *K-NN* sebesar 75,70% dengan data uji 20% dan  $k=5$ . Hasil akurasi menggunakan *NB* adalah sebesar 77,77% dengan menggunakan data uji 10%. Akurasi tidak terlalu berbeda jauh namun *NB* menciptakan akurasi yang sedikit lebih besar dibandingkan *K-NN*.

**Kata Kunci:** Klasifikasi, *K-Nearest Neighbor*, *Naïve Bayes*.



## ABSTRACT

The development of Informatics technology has made the fear of automation in the work system increasing. Including the need for automatic classification of journal articles began to be needed to determine the right journal categories. The text mining method can be used to automate journal articles by category. This research used several text mining methods, namely *Naïve Bayes(NB)* and *K-Nearest Neighbor(K-NN)*. The dataset used is in the form of article titles and 531 data which are then separated into test data and training data. The accuracy obtained with *KNN* were 75.70% with test data of 20% and  $k = 5$ . The accuracy result using *NB* is 77.77% using 10% test data. Accuracy doesn't differ much but *NB* creates slightly greater accuracy than *K-NN*.

**Keywords:** Clasification, *K-Nearest Neighbor*, *Naïve Bayes*.



## DAFTAR ISI

|   |          |
|---|----------|
| HALAMAN JUDUL                             | i        |
| LEMBAR PERNYATAAN ORISINALITAS            | iii      |
| LEMBAR PENGESAHAN                         | iv       |
| KATA PENGANTAR                            | v        |
| LEMBAR PERSETUJUAN PUBLIKASI              | vi       |
| ABSTRAK                                   | vii      |
| ABSTRACT                                  | viii     |
| DAFTAR ISI                                | ix       |
| DAFTAR GAMBAR                             | xii      |
| DAFTAR TABEL                              | xiv      |
| <b>1. PENDAHULUAN</b>                     | <b>1</b> |
| 1.1 Latar Belakang                        | 1        |
| 1.2 Perumusan Masalah                     | 1        |
| 1.3 Tujuan Penelitian                     | 2        |
| 1.4 Batasan Masalah                       | 2        |
| 1.5 Kontribusi                            | 2        |
| <b>2. TINJAUAN PUSTAKA</b>                | <b>3</b> |
| 2.1 Tinjauan Penelitian Terdahulu         | 3        |
| 2.2 Teori Dasar yang Digunakan            | 4        |
| 2.2.1 Klasifikasi                         | 4        |
| 2.2.2 Algoritma <i>Naïve Bayes</i>        | 5        |
| 2.2.3 Algoritma <i>K-Nearest Neighbor</i> | 5        |
| 2.2.4 Confusion Matrix                    | 6        |
| 2.2.5 Python                              | 6        |
| <b>3. METODA PENELITIAN</b>               | <b>7</b> |



|   |           |
|---|-----------|
| 3.1 Waktu Penelitian  | 7         |
| 3.2 Fokus Penelitian  | 7         |
| 3.3 Sumber Data   | 7         |
| 3.4 Teknik Pengumpulan Data   | 7         |
| 3.5 Desain Penelitian   | 7         |
| 3.5.1 Studi Literatur   | 8         |
| 3.5.2 Identifikasi Masalah  | 9         |
| 3.5.3 Pengumpulan Data  | 9         |
| 3.5.4 Implementasi <i>Naïve Bayes</i>   | 9         |
| 3.5.5 Implementasi Klasifikasi <i>K-Nearest Neighbor</i>                              | 9         |
| 3.5.6 Pengujian Metode Klasifikasi  | 10        |
| <b>4. HASIL DAN PEMBAHASAN</b>  | <b>11</b> |
| 4.1 Pengumpulan Data  | 11        |
| 4.2 Klasifikasi <i>Naïve Bayes</i>  | 11        |
| 4.2.1 Klasifikasi <i>Naïve Bayes</i> dengan 50% data training dan 50% data latih      | 12        |
| 4.2.2 Klasifikasi <i>Naïve Bayes</i> dengan 70% data training dan 30% data latih      | 13        |
| 4.2.3 Klasifikasi <i>Naïve Bayes</i> dengan 80% data training dan 20% data latih      | 15        |
| 4.2.4 Klasifikasi <i>Naïve Bayes</i> dengan 90% data training dan 10% data latih      | 17        |
| 4.3 Klasifikasi <i>K-Nearest Neighbor</i>   | 20        |
| 4.3.1 Klasifikasi <i>K-Nearest Neighbor</i> dengan 50% data training dan 50% data uji | 20        |
| 4.3.2 Klasifikasi <i>K-Nearest Neighbor</i> dengan 70% data training dan 30% data uji | 28        |
| 4.3.3 Klasifikasi <i>K-Nearest Neighbor</i> dengan 80% data training dan 20% data uji | 35        |
| 4.3.4 Klasifikasi <i>K-Nearest Neighbor</i> dengan 90% data training dan 10% data uji | 42        |
| 4.4 Perbandingan Hasil <i>Naïve Bayes</i> dan <i>K-Nearest Neighbor</i>               | 50        |

|   |           |
|---|-----------|
| 4.5 Visualisasi Sistem  | 51        |
| 4.5.1 Tampilan Awal   | 51        |
| 4.5.2 Halaman List Data                                       | 52        |
| 4.5.3 Halaman klasifikasi <i>K-Nearest Neighbor</i>           | 52        |
| 4.5.4 Halaman Klasifikasi <i>Naïve Bayes</i>                  | 53        |
| 4.5.5 Halaman Perbandingan <i>K-NN</i> dan <i>Naïve Bayes</i> | 54        |
| <b>5. KESIMPULAN DAN SARAN</b>                                | <b>56</b> |
| 5.1 Kesimpulan  | 56        |
| 5.2 Saran   | 56        |
| <b>DAFTAR PUSTAKA</b>   | <b>57</b> |
| <b>LAMPIRAN</b>   | <b>58</b> |



## DAFTAR GAMBAR

|  |    |
|--|----|
| Gambar 3.1 Desain penelitian                             | 8  |
| Gambar 4.1 Data Dri Internet                             | 11 |
| Gambar 4.2 Data Traning 50% <i>Naïve Bayes</i>           | 12 |
| Gambar 4.3 Data Uji 50% <i>Naïve Bayes</i>               | 12 |
| Gambar 4.4 <i>Confusion matrix NB Data Uji 50%</i>       | 13 |
| Gambar 4.5 Data Traning 70% <i>Naïve Bayes</i>           | 13 |
| Gambar 4.6 Data Uji 30% <i>Naïve Bayes</i>               | 14 |
| Gambar 4.7 <i>Confusion matrix NB Data Uji 30%</i>       | 14 |
| Gambar 4.8 Data Training 80% <i>Naïve Bayes</i>          | 15 |
| Gambar 4.9 Data Uji 20% <i>Naïve Bayes</i>               | 15 |
| Gambar 4.10 <i>Confusion matrix NB Data Uji 20%</i>      | 16 |
| Gambar 4.11 Data Training 90% <i>Naïve Bayes</i>         | 17 |
| Gambar 4.12 Data Uji 10% <i>Naïve Bayes</i>              | 17 |
| Gambar 4.13 <i>Confusion matrix NB Data Uji 10%</i>      | 18 |
| Gambar 4.14 Grafik Akurasi <i>Naïve Bayes</i>            | 19 |
| Gambar 4.15 K=1 Data Uji 50% K-Nearest Neighbor          | 20 |
| Gambar 4.16 Akurasi K-Nearest Neighbor K=1 Data uji 50%  | 21 |
| Gambar 4.17 <i>Confusion matrix KNN Data Uji 50% K=1</i> | 21 |
| Gambar 4.18 K=3 Data Uji 50% K-Nearest Neighbor          | 22 |
| Gambar 4.19 Akurasi K-Nearest Neighbor K=3 Data uji 50%  | 22 |
| Gambar 4.20 <i>Confusion matrix KNN Data Uji 50% K=3</i> | 23 |
| Gambar 4.21 K=5 Data Uji 50% K-Nearest Neighbor          | 23 |
| Gambar 4.22 Akurasi K-Nearest Neighbor K=5 Data uji 50%  | 24 |
| Gambar 4.23 <i>Confusion matrix KNN Data Uji 50% K=5</i> | 24 |
| Gambar 4.24 K=7 Data Uji 50% K-Nearest Neighbor          | 25 |
| Gambar 4.25 Akurasi K-Nearest Neighbor K=7 Data uji 50%  | 26 |
| Gambar 4.26 <i>Confusion matrix KNN Data Uji 50% K=7</i> | 26 |
| Gambar 4.27 Grafik Akurasi <i>K-NN</i> Data Uji 50%      | 27 |
| Gambar 4.28 K=1 Data Uji 30% K-Nearest Neighbor          | 28 |
| Gambar 4.29 Akurasi K-Nearest Neighbor K=1 Data uji 30%  | 28 |
| Gambar 4.30 <i>Confusion matrix KNN Data Uji 30% K=1</i> | 29 |
| Gambar 4.31 K=3 Data Uji 30% K-Nearest Neighbor          | 29 |
| Gambar 4.32 Akurasi K-Nearest Neighbor K=3 Data uji 30%  | 30 |
| Gambar 4.33 <i>Confusion matrix KNN Data Uji 30% K=3</i> | 30 |
| Gambar 4.34 K=5 Data Uji 30% K-Nearest Neighbor          | 31 |
| Gambar 4.35 Akurasi K-Nearest Neighbor K=5 Data uji 30%  | 32 |
| Gambar 4.36 <i>Confusion matrix KNN Data Uji 30% K=5</i> | 32 |
| Gambar 4.37 K=7 Data Uji 30% K-Nearest Neighbor          | 33 |
| Gambar 4.38 Akurasi K-Nearest Neighbor K=7 Data uji 30%  | 33 |
| Gambar 4.39 <i>Confusion matrix KNN Data Uji 30% K=7</i> | 34 |
| Gambar 4.40 Grafik Akurasi <i>K-NN</i> Data Uji 30%      | 35 |

|  |    |
|--|----|
| Gambar 4.41 K=1 Data Uji 20% K-Nearest Neighbor                            | 35 |
| Gambar 4.42 Akurasi K-Nearest Neighbor K=1 Data uji 20%                    | 36 |
| Gambar 4.43 <i>Confusion matrix KNN Data Uji 20% K=1</i>                   | 36 |
| Gambar 4.44 K=3 Data Uji 20% K-Nearest Neighbor                            | 37 |
| Gambar 4.45 Akurasi K-Nearest Neighbor K=3 Data uji 20%                    | 37 |
| Gambar 4.46 <i>Confusion matrix KNN Data Uji 20% K=3</i>                   | 38 |
| Gambar 4.47 K=5 Data Uji 20% K-Nearest Neighbor                            | 38 |
| Gambar 4.48 Akurasi K-Nearest Neighbor K=5 Data uji 20%                    | 39 |
| Gambar 4.49 <i>Confusion matrix KNN Data Uji 20% K=5</i>                   | 39 |
| Gambar 4.50 K=7 Data Uji 20% K-Nearest Neighbor                            | 40 |
| Gambar 4.51 Akurasi K-Nearest Neighbor K=7 Data uji 20%                    | 40 |
| Gambar 4.52 <i>Confusion matrix KNN Data Uji 20% K=7</i>                   | 41 |
| Gambar 4.53 Grafik Akurasi <i>K-NN</i> Data Uji 20%                        | 42 |
| Gambar 4.54 K=1 Data Uji 10% K-Nearest Neighbor                            | 42 |
| Gambar 4.55 Akurasi K-Nearest Neighbor K=1 Data uji 10%                    | 43 |
| Gambar 4.56 <i>Confusion matrix KNN Data Uji 10% K=1</i>                   | 43 |
| Gambar 4.57 K=3 Data Uji 10% K-Nearest Neighbor                            | 44 |
| Gambar 4.58 Akurasi K-Nearest Neighbor K=3 Data uji 10%                    | 44 |
| Gambar 4.59 <i>Confusion matrix KNN Data Uji 10% K=3</i>                   | 45 |
| Gambar 4.60 K=5 Data Uji 10% K-Nearest Neighbor                            | 45 |
| Gambar 4.61 Akurasi K-Nearest Neighbor K=5 Data uji 10%                    | 46 |
| Gambar 4.62 <i>Confusion matrix KNN Data Uji 10% K=5</i>                   | 46 |
| Gambar 4.63 K=7 Data Uji 10% K-Nearest Neighbor                            | 47 |
| Gambar 4.64 Akurasi K-Nearest Neighbor K=7 Data uji 10%                    | 47 |
| Gambar 4.65 <i>Confusion matrix KNN Data Uji 10% K=7</i>                   | 48 |
| Gambar 4.66 Grafik Akurasi <i>K-NN</i> Data Uji 10%                        | 49 |
| Gambar 4.67 Grafik Semua Akurasi <i>K-NN</i> Data Uji 20%                  | 50 |
| Gambar 4.68 Grafik Perbandingan Akurasi <i>K-NN</i> dan <i>Naïve Bayes</i> | 51 |
| Gambar 4.69 Tampilan Awal  | 51 |
| Gambar 4.70 Halaman list data  | 52 |
| Gambar 4.71 Halaman klasifikasi <i>K-Nearest Neighbor</i>                  | 52 |
| Gambar 4.72 Grafik Akurasi Halaman <i>K-Nearest Neighbor</i>               | 53 |
| Gambar 4.73 Halaman klasifikasi <i>Naïve Bayes</i>                         | 53 |
| Gambar 4.74 Grafik Akurasi Halaman <i>Naïve Bayes</i>                      | 54 |
| Gambar 4.75 Halaman Perbandingan <i>K-NN</i> dan <i>Naïve Bayes</i>        | 54 |
| Gambar 4.76 Grafik Perbandingan Akurasi <i>K-NN</i> dan <i>Naïve Bayes</i> | 55 |

## DAFTAR TABEL

|   |    |
|---|----|
| Tabel 2.1 Tinjauan penelitian terdahulu                                 | 3  |
| Tabel 4.1 Hasil Akurasi Naïve Bayes                                     | 18 |
| Tabel 4.2 Hasil Akurasi K-Nearest Neighbor Data Uji 50%                 | 27 |
| Tabel 4.3 Hasil Akurasi K-Nearest Neighbor Data Uji 30%                 | 34 |
| Tabel 4.4 Hasil Akurasi K-Nearest Neighbor Data Uji 20%                 | 41 |
| Tabel 4.5 Hasil Akurasi K-Nearest Neighbor Data Uji 10%                 | 48 |
| Tabel 4.6 Hasil Akurasi K-Nearest Neighbor                              | 49 |
| Tabel 4.7 Hasil Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor | 50 |





## DAFTAR LAMPIRAN

Lampiran 1 : Lembar Publikasi Jurnal

58

