

**TUGAS AKHIR**

**PERBANDINGAN KINERJA ALGORITMA GAUSSIAN NAIVE BAYES  
DAN *K*-NEAREST NEIGHBOR (KNN) UNTUK MENGKLASIFIKASI  
PENYAKIT HEPATITIS C VIRUS (HCV)**

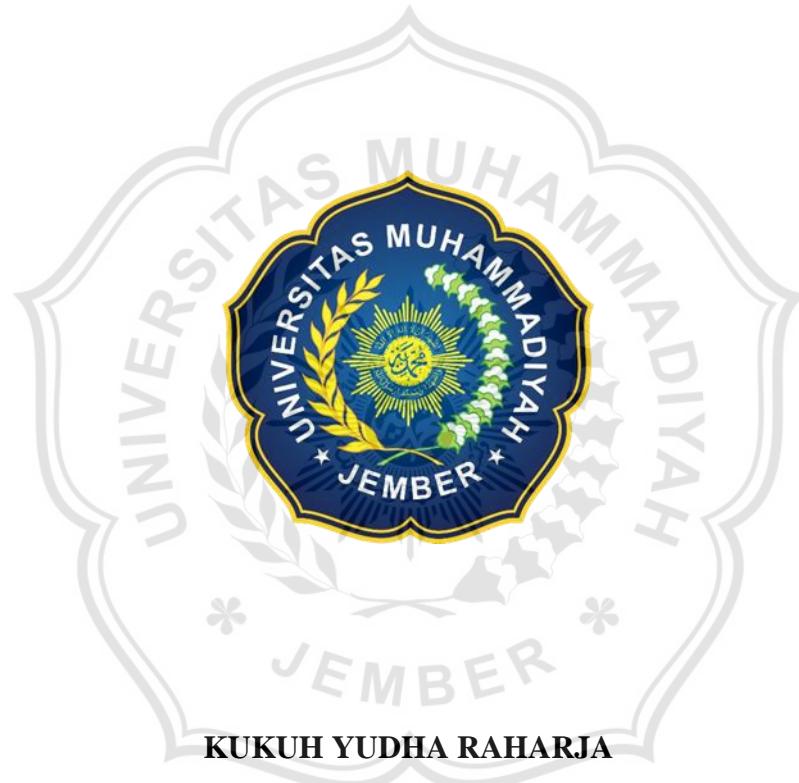


**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH JEMBER  
2021**

## **HALAMAN JUDUL**

# **PERBANDINGAN KINERJA ALGORITMA GAUSSIAN NAIVE BAYES DAN K-NEAREST NEIGHBOR (KNN) UNTUK MENGIKLASIFIKASI PENYAKIT HEPATITIS C VIRUS (HCV)**

Disusun untuk Melengkapi dan Memenuhi Syarat Kelulusan Guna Meraih Gelar Sarjana Komputer Teknik Informatika Universitas Muhammadiyah Jember



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH JEMBER  
2021**

**HALAMAN PERSETUJUAN**

**PERBANDINGAN KINERJA ALGORITMA GAUSSIAN NAIVE BAYES  
DAN *K*-NEAREST NEIGHBOR (KNN) UNTUK MENGKLASIFIKASI  
PENYAKIT HEPATITIS C VIRUS (HCV)**

**Oleh:**

**Kukuh Yudha Raharja**

**1610651047**

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada Sidang  
Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar  
**Sarjana Komputer (S.Kom)**

**di**

**Universitas Muhammadiyah Jember**

**Disetujui oleh:**

**Pembimbing I**

Hardian Oktavianto, S.Si., M.Kom.

NIDN. 0722108105

**Pembimbing II**

Reni Umilasari, S.Pd, M.Si.

NIDN.0728079101

## HALAMAN PENGESAHAN

### PERBANDINGAN KINERJA ALGORITMA GAUSSIAN NAIVE BAYES DAN K-NEAREST NEIGHBOR (KNN) UNTUK MENKLASIFIKASI PENYAKIT HEPATITIS C VIRUS (HCV)

Oleh:

Kukuh Yudha Raharja

1610651047

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada Sidang Tugas Akhir tanggal 15 Januari 2021 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui Oleh,

Penguji I

Ilham Saifudin, S.Pd., M.Si.

NIDN. 0731108903

Pembimbing I

Hardian Oktavianto, S.Si., M.Kom.

NIDN. 0722108105

Penguji II

Rosita Yanuarti, S.Kom., M.Cs.

NIDN. 0629018601

Pembimbing II

Reni Umilasari, S.Pd., M.Si.

NIDN. 0728079101

Mengesahkan,

Dekan Fakultas Teknik



Dr. Nanang Saiful Rizal, S.T., M.T.  
NIDN. 0705047806

Mengetahui,

Ketua Program Studi Teknik Informatika

Triawan Adi Cahyanto, M.Kom.  
NIDN. 0702098804

## **LEMBAR PERNYATAAN KEASLIAN**

Yang bertanda tangan di bawah ini:

Nama : Kukuh Yudha Raharja

NIM : 1610651047

Program Studi : Teknik Informatika

Perguruan Tinggi : Universitas Muhammadiyah Jember

Menyatakan bahwa Tugas Akhir yang berjudul "**Perbandingan Kinerja Algoritma Gaussian Naive Bayes dan K-Nearest Neighbor (KNN) Untuk Mengklasifikasi Penyakit Hepatitis C Virus (HCV)**" bukan merupakan Tugas Akhir orang lain baik sebagian maupun keseluruhan kecuali yang diacu dalam Daftar Pustaka dan bentuk kutipan yang telah disebutkan sumbernya pada Tugas Akhir ini.

Demikian pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak sesuai, penulis bersedia mendapatkan sanksi dari akademik.

Jember, 15 Januari 2021

Yang Menyatakan,



Kukuh Yudha Raharja

NIM. 1610651047

## **UNGKAPAN TERIMA KASIH**

Puji syukur kepada Allah S.W.T atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir ini. Atas segala upaya, bimbingan, dan arahan dari semua pihak, penulis mengucapkan terimakasih sebesar-besarnya kepada:

1. Allah SWT berkat segala ridho, rahmat dan hidayah-Nya penulis dapat menyelesaikan segala urusan dalam menyusun laporan Tugas Akhir dan diberikan kesempatan mendapatkan gelar Sarjana Komputer.
2. Bapak Dr.Nanang Saiful Rizal, S.T., M.T selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
3. Bapak Triawan Adi Cahyanto, M.Kom selaku Kepala Program Studi Teknik Informatika di Universitas Muhammadiyah Jember.
4. Bapak Hadian Oktavianto S.Si., M.Kom selaku dosen pembimbing 1 tugas akhir yang telah memberikan bimbingan dari awal sampai akhir dengan penuh kesabaran sehingga tugasakhir ini dapat terselesaikan.
5. Ibu Reni Umilasari, S.Pd., M.Si selaku dosen pembimbing 2 tugas akhir yang telah memberikan bimbingan dari awal sampai akhir dengan penuh kesabaran sehingga tugas akhir ini dapat terselesaikan.
6. Bapak Ilham Saifudin, S.Pd, M.Si selaku dosen penguji 1 yang telah memberikan kritik, saran dan masukkan yang membangun dalam penelitian ini.
7. Inu Rosita Yanuarti, S.Kom., M.Cs selaku dosen penguji 2 yang telah memberikan kritik, saran dan masukkan yang membangun dalam penelitian ini.
8. Kedua orang tua saya Mujiono, M.Pd dan Kartini Titik Susanti, terimakasih atas *support* dan doa nya dalam penyusunan laporan tugas akhir ini.
9. Teman-teman seperjuangan yang senasib dan setanah air, baik “FAMILY STREET, SAMPE TUEK, PRAPATAN SUKERJO, KIKT” terimakasih atas doa dan dukungan nya selama ini, dan terima kasih atas kebersamaannya.

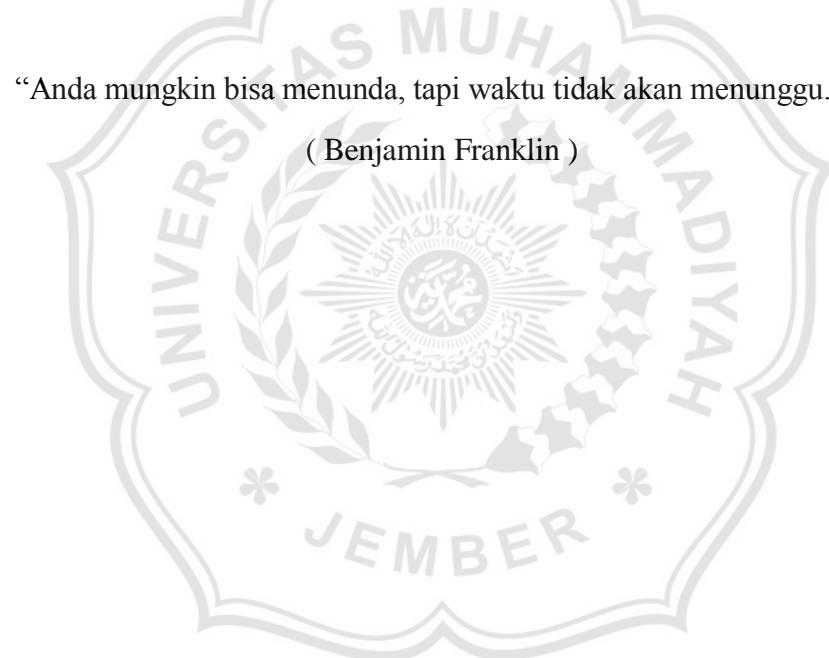
## **MOTTO**

“Apa yang kita pikirkan menentukan apa yang akan terjadi pada kita. Jadi jika kita ingin mengubah hidup, kita perlu sedikit mengubah pikiran kita.”

( Wayne Dyer )

“Kamu tidak perlu menjadi luar biasa untuk memulai, tapi kamu harus memulai untuk menjadi luar biasa.”

( Zig Ziglar )



**PERBANDINGAN KINERJA ALGORITMA GAUSSIAN NAIVE BAYES  
DAN K-NEAREST NEIGHBOR (KNN) DALAM MENKLASIFIKASI  
PENYAKIT HEPATITIS C VIRUS (HCV)**

Kukuh Yudha Raharja<sup>1</sup>, Hardian Oktavianto<sup>2</sup>, Reni Umilasari<sup>3</sup>.

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah  
Jember

[yudharaharja0@gmail.com](mailto:yudharaharja0@gmail.com)<sup>1</sup>, [hardian@unmuahjember.ac.id](mailto:hardian@unmuahjember.ac.id)<sup>2</sup>,

[Reni.umilasari@gmail.com](mailto:Reni.umilasari@gmail.com)<sup>3</sup>.

**ABSTRAK**

Hepatitis adalah penyakit peradangan pada hati, penyakit Hepatitis merupakan penyakit yang menular sehingga menjadi beban kesehatan bagi masyarakat karena penularannya yang begitu mudah. Jenis penyakit Hepatitis yang paling berbahaya adalah Hepatitis C yang disebabkan oleh Virus RNA. Pada penelitian ini dilakukan klasifikasi terhadap penyakit Hepatitis C Virus pada pasien di Jerman dikarenakan kasus hepatitis di Jerman termasuk tinggi dan di Jerman belum menerapkan strategi nasional untuk mengurangi beban Hepatitis C, Metode klasifikasi pada penelitian ini adalah membandingkan antara algoritma *Gaussian Naive Bayes* dan *K-Nearest Neighbor*. Hasil yang didapatkan dari penelitian ini adalah pada algoritma *Gaussian Naive Bayes* nilai akurasi sebesar 90,98%, presisi sebesar 69,91%, dan recall sebesar 61,57% sedangkan pada algoritma *K-Nearest Neighbor* (KNN) didapatkan nilai akurasi sebesar 91,80%, presisi sebesar 68,96%, dan recall sebesar 51,85%. Untuk akurasi, algoritma *K-Nearest Neighbor* (KNN) mendapatkan hasil lebih baik dibandingkan *Gaussian Naive Bayes*. Sedangkan untuk presisi, algoritma *Gaussian Naive Bayes* mendapatkan hasil lebih baik dibandingkan *K-Nearest Neighbor* (KNN). Dan untuk recall, algoritma *Gaussian Naive Bayes* mendapatkan hasil lebih baik dibandingkan *K-Nearest Neighbor* (KNN).

**Kata Kunci:** Klasifikasi penyakit, Hepatitis C Virus (HCV), *Gaussian Naive Bayes*, *K-Nearest Neighbor*.

**PERFORMANCE COMPARISON OF GAUSSIAN NAIVE BAYES AND K-  
NEAREST NEIGHBOR (KNN) ALGORITHMS IN CLASSIFYING  
HEPATITIS C VIRUS (HCV) DISEASE**

Kukuh Yudha Raharja<sup>1</sup>, Hardian Oktavianto<sup>2</sup>, Reni Umilasari<sup>3</sup>.

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah  
Jember  
[yudharaharja0@gmail.com](mailto:yudharaharja0@gmail.com)<sup>1</sup>, [hardian@unmuahjember.ac.id](mailto:hardian@unmuahjember.ac.id)<sup>2</sup>,  
[Reni.umilasari@gmail.com](mailto:Reni.umilasari@gmail.com)<sup>3</sup>.

**ABSTRACT**

Hepatitis is an inflammatory disease of the liver, it is an infectious disease that becomes health burden for the community because of its easy to get infection. The most dangerous type of hepatitis is Hepatitis C caused by RNA virus. In this study, classification of Hepatitis C Virus disease in patient at Germany because cases of hepatitis in Germany were high and Germany had not implemented a national strategy to reduce the burden of Hepatitis C Virus, The classification method in this study is comparing between algorithm of the *Gaussian Naive Bayes* and *K-Nearest Neighbor* (KNN). The results obtained from this study are in *Gaussian Naive Bayes* algorithm accuracy value of 90.98%, precision by 69.91%, and recall by 61.57% while in *K-Nearest Neighbor* (KNN) algorithm obtained accuracy value of 91.80%, precision of 68.96%, and recall of 51.85%. For accuracy, *K-Nearest Neighbor* (KNN) algorithm gets better results than *Gaussian Naive Bayes*. As for precision, *Gaussian Naive Bayes* algorithm gets better results than *K-Nearest Neighbor* (KNN). And for recalls, *Gaussian Naive Bayes* algorithms get better results than *K-Nearest Neighbor* (KNN).

**Keywords:** Disease classification, Hepatitis C Virus (HCV), Gaussian Naive Bayes, K-Nearest Neighbor.

## KATA PENGANTAR

Puji syukur atas kehadirat Allah SWT yang telah ridho, rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir dengan judul **“Perbandingan Kinerja Algoritma Gaussian Naive Bayes dan K-Nearest Neighbor (KNN) Untuk Mengklasifikasi Penyakit Hepatitis C Virus (HCV)”**. Adapun tujuan penulisan laporan tugas akhir ini adalah untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan Sarjana pada Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember.

Penulis mengucapkan banyak terima kasih kepada bapak dan ibu pembimbing, serta bapak penguji yang telah memberikan bimbingan, kritik dan saran yang membangun sehingga tugas akhir ini dapat terselesaikan.

Penulis menyadari bahwa penyusunan skripsi ini masih banyak kekurangan dan jauh dari kata sempurna maka dari itu penulis sangat mengharapkan kritik dan saran yang membangun demi kesempurnaan tugas akhir ini. Semoga tulisan ini dapat bermanfaat tidak saja untuk penulis tetapi bagi pembaca.

Jember, 15 Januri 2021

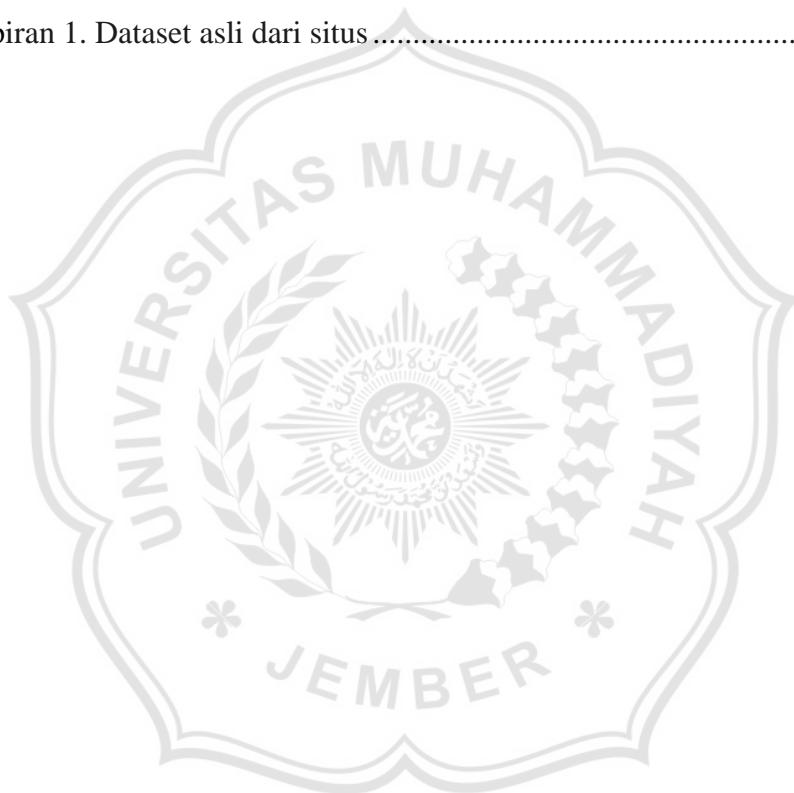
Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>iii</b>
<b>LEMBAR PERNYATAAN KEASLIAN .....</b>	<b>iv</b>
<b>UNGKAPAN TERIMA KASIH .....</b>	<b>v</b>
<b>MOTTO .....</b>	<b>vi</b>
<b>ABSTRAK .....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>viii</b>
<b>KATA PENGANTAR.....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xiii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.2 Latar Belakang .....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Maanfaat Penelitian .....	3
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>4</b>
2.1 Hepatitis C Virus.....	4
2.2 Data Mining .....	6
2.2.1. Definisi Data Mining.....	6
2.2.2. Tahapan Data Mining.....	7
2.2.3. Teknik Data mining.....	8
2.3 Klasifikasi .....	9
2.4 <i>Naive Bayes</i> .....	9
2.4.1 Gaussian Naive Bayes.....	10

2.4.2 Mean.....	11
2.4.3 Standar Deviasi .....	11
2.5 Algoritma K- <i>Nearest Neighbor</i> (KNN) .....	11
2.5.1 Euclidean Distance.....	12
2.6 Confusion Matrix.....	12
2.7 Phyton .....	15
2.7.1 Python Jupyter Notebook.....	16
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>17</b>
3.1 Tahapan Penelitian.....	17
3.2 Studi Literatur .....	18
3.3 Pengumpulan Data.....	18
3.4 Praproses Data .....	18
3.5 Implementasi Gaussian Naive Bayes.....	19
3.5.1 Menghitung Nilai Probabilitas .....	22
3.5.2 Mencari Nilai Mean.....	23
3.5.3 Mencari Standar Deviasi .....	24
3.5.4 Mencari Nilai Probabilistik .....	26
3.5.5 Mecari Hasil Klasifikasi .....	27
3.6 Implementasi K-Nearest Neigbor .....	28
3.6.1 Menghitung jarak Euclidean Data Uji dengan Data Latih .....	32
3.6.2 Menentukan Mayoritas Data .....	34
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>36</b>
4.1 Pengumpulan Data.....	36
4.2 Pre-processing Data .....	36
4.3 Hasil Klasifikasi.....	38

4.3.1 Hasil Klasifikasi Gaussian Naive Bayes .....	39
4.3.2 Hasil Klasifikasi K- <i>Nearest Neighbor</i> (KNN) .....	43
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>51</b>
5.1 Kesimpulan .....	51
5.2 Saran .....	51
<b>DAFTAR PUSTAKA .....</b>	<b>53</b>
<b>LAMPIRAN.....</b>	<b>55</b>
Lampiran 1. Dataset asli dari situs .....	53



## **DAFTAR TABEL**

Tabel 2. 1 Confusion Matrix Dua Kelas .....	13
Tabel 3. 1 Inisialisasi Atribut .....	18
Tabel 3. 2 Data Latih Gaussian Naive Bayes.....	20
Tabel 3. 3 Data Uji Gaussian Naive Bayes .....	22
Tabel 3. 4 Probabilitas Kategori.....	23
Tabel 3. 5 Probabilitas Jenis Kelamin.....	23
Tabel 3. 6 Hasil pencarian Nilai Mean.....	23
Tabel 3. 7 Total Jarak Mean Usia KategoriDonor Darah .....	24
Tabel 3. 8 Total Jarak Mean Usia Kategori Hepatitis .....	24
Tabel 3. 9 Total Jarak Mean Usia Kategori Fibrosis .....	25
Tabel 3. 10 Total Jarak Mean Usia Kategori Cirrhosis.....	25
Tabel 3. 11 Data Latih K-Nearest Neighbor .....	29
Tabel 3. 12 Data Uji K-Nearest Neighbor .....	31
Tabel 3. 13 Jarak Euclidean antar Data Latih dan Data Uji .....	33
Tabel 3. 14 Majoritas Data.....	34
Tabel 4. 1 Data Hepatitis C Virus .....	36
Tabel 4. 2 Data Pre-processing Hepatitis C Virus .....	37
Tabel 4. 3 Confusion Matrix Akurasi Pada Algoritma Gaussian Naive Bayes ..	39
Tabel 4. 4 Confusion Matrix Presisi Pada Algoritma Gaussian Naive Bayes ..	40
Tabel 4. 5 Confusion Matrix Recall Pada Algoritma Gaussian Naive Bayes...	41
Tabel 4. 6 Confusion matrix f1 score .....	42
Tabel 4. 7 Confusion Matrix Akurasi pada Algoritma K-Nearest Neighbor.....	43
Tabel 4. 8 Confusion Matrix Presisi Pada Algoritma K-Nearest Neighbor.....	44
Tabel 4. 9 Confusion Matrix Recall Pada Algoritma K-Nearest Neighbor .....	46
Tabel 4. 10 Confusion Matrix f1-score .....	47
Tabel 4. 11 Nilai Akurasi, Presisi, Recall, F1-Score Pada Percobaan k .....	48
Tabel 4. 12 Perbandingan Akurasi, Presisi, Recall, Dan F1-Score .....	49

## DAFTAR GAMBAR

<b>Gambar 3. 1</b> Tahapan Penelitian.....	17
<b>Gambar 3. 2</b> <i>Flowchart Gaussian Naive Bayes</i> .....	19
<b>Gambar 3. 3</b> <i>Flowchart K-Nearest Neighbor</i> .....	28
<b>Gambar 4. 1</b> Input pemanggilan data.....	36
<b>Gambar 4. 2</b> Input <i>Pre processing</i> Kategori.....	37
<b>Gambar 4. 3</b> Input <i>Preprocessing Sex</i> .....	37
<b>Gambar 4. 4</b> Input <i>import library Gaussian naive bayes dan k-nearest neighbour</i> .....	38
<b>Gambar 4. 5</b> Input klasifikasi <i>Gaussian Naive Bayes</i> .....	38
<b>Gambar 4. 6</b> Input klasifikasi <i>k-nearest neighbor</i> .....	39
<b>Gambar 4. 7</b> <i>Source code</i> Input Grafik.....	42
<b>Gambar 4. 8</b> Grafik <i>Gaussian naïve bayes</i> .....	43
<b>Gambar 4. 9</b> <i>Source code</i> Input Grafik Akurasi .....	44
<b>Gambar 4. 10</b> Grafik Akurasi <i>K-Nearest Neighbor</i> .....	44
<b>Gambar 4. 11</b> <i>Source code</i> Input Grafik Presisi .....	45
<b>Gambar 4. 12</b> Grafik Presisi <i>K-nearest Neighbor</i> .....	45
<b>Gambar 4. 13</b> <i>Source code</i> Input Grafik Recall.....	46
<b>Gambar 4. 14</b> Grafik recall <i>k-nearest neighbor</i> .....	47
<b>Gambar 4. 15</b> <i>Source code</i> Input Grafik F1 Score .....	48
<b>Gambar 4. 16</b> Grafik f1 score <i>k-nearesst neighbor</i> .....	48