

TESIS

**PREDIKSI KETEPATAN WAKTU KELULUSAN MAHASISWA  
MEGGUNAKAN METODE *NAIVE BAYES CLASSIFICATION* DAN  
*K-NEAREST NEIGHBORS***

**(Studi Kasus: Program Studi D3 Teknologi Laboratorium Medik  
STIKES Guna Bangsa)**



**Oleh:  
ANWARUDIN  
19/22/18/MTI-TSD/5**

**PROGRAM STUDI TEKNOLOGI INFORMASI  
PROGRAM MAGISTER  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA  
YOGYAKARTA  
2022**

**TESIS**

**PREDIKSI KETEPATAN WAKTU KELULUSAN MAHASISWA  
MEGGUNAKAN METODE *NAIVE BAYES CLASSIFICATION* DAN  
*K-NEAREST NEIGHBORS***

**(Studi Kasus: Program Studi D3 Teknologi Laboratorium Medik  
STIKES Guna Bangsa)**

**Diajukan sebagai salah satu syarat untuk menyelesaikan studi**



**Program Magister  
Program Studi Teknologi Informasi  
Fakultas Teknologi Informasi  
Universitas Teknologi Digital Indonesia  
Yogyakarta**

**Oleh:**

**ANWARUDIN  
19/22/18/MTI-TSD/5**

**PROGRAM STUDI TEKNOLOGI INFORMASI  
PROGRAM MAGISTER  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA  
YOGYAKARTA  
2022**

## DAFTAR ISI

	<b>Halaman</b>
<b>TESIS .....</b>	<b>ii</b>
<b>HALAMAN PERSETUJUAN UJIAN TESIS .....</b>	<b>iii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iv</b>
<b>PERNYATAAN KEASLIAN .....</b>	<b>v</b>
<b>DAFTAR ISI.....</b>	<b>vi</b>
<b>DAFTAR GAMBAR.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>ix</b>
<b>INTISARI .....</b>	<b>xii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Penelitian .....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>7</b>
2.1 Penelitian Terdahulu.....	7
2.2 Posisi Penelitian .....	13
<b>BAB III LANDASAN TEORI.....</b>	<b>19</b>
3.1. Data Mining.....	19
3.2. <i>Knowledge Discovery in Databases</i> .....	20
3.3. Klasifikasi.....	21
3.4. <i>Naïve Bayes Classification</i> .....	22
3.5. <i>K-Nearest Neighbors</i> .....	24
3.6. Evaluasi Performansi Metode Klasifikasi .....	26
3.7. RapidMiner.....	28
3.8. <i>Cross Validation</i> .....	29
<b>BAB IV METODOLOGI PENELITIAN .....</b>	<b>30</b>
4.1. Jenis penelitian .....	30
4.2. Alat dan Bahan .....	30
4.3. Prosedur Kerja .....	31

4.4.	Pengumpulan Data .....	38
4.5.	Identifikasi Variabel .....	42
<b>BAB V IMPLEMENTASI NAÏVE BAYES CLASSIFICATION.....</b>		<b>45</b>
5.1	Pengolahan dengan Data Mining .....	45
5.2	Prediksi Model <i>Naive Bayes Clasification</i> (NBC) .....	49
5.3	Evaluasi dan Validasi NBC .....	60
5.3.1	Hasil Kinerja Algoritma NBC Menggunakan <i>Tools Ms. Excel</i> .....	61
5.3.2	Hasil Kinerja Algoritma NBC Menggunakan <i>Tools Rapid Miner</i> ..	64
<b>BAB VI IMPLEMENTASI K-NEAREST NEIGHBORS .....</b>		<b>68</b>
6.1	Pengolahan dengan Data Mining .....	68
6.2	Prediksi Model <i>K-Nearest Neighbors</i> (KNN).....	72
6.3	Evaluasi dan Validasi KNN.....	78
6.3.1	Hasil Kierja Algoritma KNN Menggunakan <i>Tools Ms. Excel</i> .....	78
6.2.2	Hasil Kinerja Algoritma KNN Menggunakan <i>Tools Rapid Miner</i> .	81
<b>BAB VII UJI PERFORMA .....</b>		<b>86</b>
7.1.	Perbandingan Hasil Kinerja Algoritma .....	86
<b>BAB VIII PENUTUP .....</b>		<b>87</b>
8.1	Kesimpulan.....	87
8.2	Saran .....	88
<b>DAFTAR PUSTAKA .....</b>		<b>89</b>
<b>LAMPIRAN 1.....</b>		<b>92</b>
<b>LAMPIRAN 2.....</b>		<b>104</b>

## DAFTAR TABEL

<b>Tabel 2. 2</b> Atribut berpengaruh (Miguéis et al, 2018).....	13
<b>Tabel 2. 3</b> Tinjauan pustaka .....	15
<b>Tabel 3.1</b> Akurasi Klasifikasi .....	26
<b>Tabel 4.1</b> Variabel dan definisi operasional variabel .....	41
<b>Tabel 4.2</b> Variabel masukan tanpa proses Konversi .....	43
<b>Tabel 4.3</b> Variabel Masukan dengan Proses Konversi.....	43
<b>Tabel 4.4</b> Variabel Kelurn Kelulusan Mahasiswa .....	44
<b>Tabel 5. 1</b> Atribut/Variabel.....	45
<b>Tabel 5. 2</b> Hasil Pembobotan Atribut .....	47
<b>Tabel 5. 3</b> Data sebelum dibersihkan .....	48
<b>Tabel 5. 4</b> Data setelah dibersihkan.....	49
<b>Tabel 5. 5</b> Data Training Mahasiswa (Dapat dilihat Lampiran).....	50
<b>Tabel 5. 6</b> Mean Atribut IPS .....	51
<b>Tabel 5. 7</b> Standar Deviasi Atribut IPS .....	51
<b>Tabel 5. 8</b> Mean Atribut TLM.....	51
<b>Tabel 5. 9</b> Standar Deviasi Atribut TLM .....	51
<b>Tabel 5. 10</b> Mean Atribut MLK .....	52
<b>Tabel 5. 11</b> Standar Deviasi Atribut MLK.....	52
Tabel 5. 12 Probabilitas Untuk Setiap Kategori Pada Status Kelulusan.....	53
Tabel 5. 13 Probabilitas Setiap Jenis Kelamin Pada Setiap Kategori .....	54
Tabel 5. 14 Probabilitas Setiap Asal Jurusan Pada Setiap Kategori .....	56
Tabel 5. 15 Probabilitas Setiap Asal Daerah Pada Setiap Kategori.....	57
Tabel 5. 16 Ilustrasi percobaan dengan <i>5-Fold Validation</i> . .....	61
Tabel 5. 17 Confusion Matrix 1 - fold cross validation .....	61
Tabel 5. 18 Perhitungan Akurasi, Presisi, dan Recall (1 - fold cross validation) .	61
Tabel 5. 19 Confusion Matrix 2 - fold cross validation .....	62
Tabel 5. 20 Perhitungan Akurasi, Presisi, dan Recall (2 - fold cross validation) .	62
Tabel 5. 21 Confusion Matrix 3 - fold cross validation .....	62
Tabel 5. 22 Perhitungan Akurasi, Presisi, dan Recall (3 - fold cross validation) .	62

Tabel 5. 23 Confusion Matrix 4 - fold cross validation .....	63
Tabel 5. 24 Perhitungan Akurasi, Presisi, dan <i>Recall</i> (4 - fold cross validation) .	63
Tabel 5. 25 Confusion Matrix 5- fold cross validation .....	63
Tabel 5. 26 Perhitungan Akurasi, Presisi, dan Recall (5 - fold cross validation) .	63
Tabel 5. 27 Rata-rata 5- <i>Fold Cross Validation</i> .....	63
<b>Tabel 5. 28</b> Akurasi algoritma NBC.....	67
<b>Tabel 5. 29</b> Presisi algoritma NBC.....	67
<b>Tabel 5. 30</b> Recall Algoritma NBC .....	67
<b>Tabel 6. 1</b> Atribut/Variabel.....	69
<b>Tabel 6. 2</b> Hasil Pembobotan Atribut .....	71
<b>Tabel 6. 3</b> Data sebelum dibersihkan .....	72
<b>Tabel 6. 4</b> Data setelah dibersihkan.....	72
<b>Tabel 6. 5</b> Dataset algoritma KNN .....	73
<b>Tabel 6. 6</b> Transformasi Data .....	74
<b>Tabel 6. 7</b> Data training .....	74
<b>Tabel 6. 8</b> Data testing .....	74
<b>Tabel 6. 9</b> Nilai max dan min pada data training .....	75
<b>Tabel 6. 10</b> Nilai max dan min pada data testing .....	75
<b>Tabel 6. 11</b> Normalisasi Data Training .....	75
<b>Tabel 6. 12</b> Normalisasi Data Testing .....	75
<b>Tabel 6. 13</b> Hasil perhitungan Euclidean Distance pada data uji .....	76
<b>Tabel 6. 14</b> Menentukan nilai k.....	76
<b>Tabel 6. 15</b> Tabel Hasil pengujian dengan KNN .....	77
Tabel 6. 16 <i>Confusion Matrix</i> .....	77
Tabel 6. 17 Perhitungan Akurasi, Presisi, dan Recall .....	77
Tabel 6. 18 Ilustrasi percobaan dengan 5-Fold Validation. ....	78
Tabel 6. 19 Confusion Matrix 1- fold cross validation .....	79
Tabel 6. 20 Perhitungan Akurasi, Presisi, dan Recall (1 - fold cross validation) ..	79
Tabel 6. 21 Confusion Matrix 2 - fold cross validation .....	79
Tabel 6. 22 Perhitungan Akurasi, Presisi, dan Recall (2 - fold cross validation) ..	79
Tabel 6. 23 Confusion Matrix 3 - fold cross validation .....	79

Tabel 6. 24 Perhitungan Akurasi, Presisi, dan Recall (3 - fold cross validation) ..	80
Tabel 6. 25 Confusion Matrix 4 - fold cross validation .....	80
Tabel 6. 26 Perhitungan Akurasi, Presisi, dan Recall (4 - fold cross validation) ..	80
Tabel 6. 27 Confusion Matrix 5 - fold cross validation .....	80
Tabel 6. 28 Perhitungan Akurasi, Presisi, dan Recall (5 - fold cross validation) ..	80
Tabel 6. 29 Rata-rata 5- Fold Cross Validation .....	81
<b>Tabel 6. 30</b> Akurasi algoritma KNN .....	84
<b>Tabel 6. 31</b> Presisi algoritma KNN .....	84
<b>Tabel 6. 32</b> Recall algoritma KNN.....	84
<b>Tabel 7. 1</b> Confusion Matrix 5 - fold cross validation (NBC) .....	86
<b>Tabel 7. 2</b> Confusion Matrix 5 - fold cross validation (KNN) .....	86
<b>Tabel 7. 3</b> Perbandingan Performance Algoritma Menggunakan Rapid Miner ...	86

## DAFTAR GAMBAR

<b>Gambar 1. 1</b>	Rasio mahasiswa putus kuliah .....	2
<b>Gambar 2. 1</b>	Perbandingan klasifikasi (Hussain et al., 2018) .....	9
<b>Gambar 2. 2</b>	Grafik Perbandingan Jumlah Mahasiswa Baru dan Lulusan Perguruan Tinggi di Indonesia .....	12
<b>Gambar 3. 1</b>	Tahapan proses KDD.....	20
<b>Gambar 3. 2</b>	Cross Validation (sumber: <a href="https://academy.rapidminer.com">https://academy.rapidminer.com</a> ) .....	29
<b>Gambar 4. 1</b>	Kerangka tahapan peneitian.....	32
<b>Gambar 4. 2</b>	Diagram Proses Pengolahan data dengan algoritma NBC .....	35
<b>Gambar 4. 3</b>	Diagram Proses Pengolahan data dengan algoritma KNN .....	36
<b>Gambar 4. 4</b>	Alur Algoritma NBC .....	37
<b>Gambar 4. 5</b>	Alur Algoritma KNN.....	38
<b>Gambar 5.1</b>	Fiture Selesction .....	47
<b>Gambar 5. 2</b>	Visualisasi Hasil Pembobtan Atribut.....	48
<b>Gambar 5. 3</b>	Validasi algoritma NBC .....	65
<b>Gambar 5. 4</b>	Kinerja algoritma NBC.....	66
<b>Gambar 6. 1</b>	Fiture Selesction .....	70
<b>Gambar 6. 2</b>	Visualisasi Hasil Pembobtan Atribut.....	71
<b>Gambar 6. 3</b>	Validasi algoritma KNN .....	82
<b>Gambar 6. 4</b>	Kinerja algoritma KNN .....	83



## **INTISARI**

### **PREDIKSI KETEPATAN WAKTU KELULUSAN MAHASISWA MEGGUNAKAN METODE *NAIVE BAYES CLASSIFICATION* DAN *K-NEAREST NEIGHBORS***

**(Studi Kasus: Program Studi D3 Teknologi Laboratorium Medik  
STIKES Guna Bangsa)**

Oleh

**ANWARUDIN  
19/22/18/MTI/-TSD/5**

Mutu perguruan tinggi dapat dilihat dari tingkat kelulusan mahasiswa tepat waktu. Prediksi kelulusan mahasiswa ini dapat digunakan sebagai salah satu penunjang keputusan evaluasi kinerja mahasiswa. Saat ini, program studi (prodi) D3 Teknologi Laboratorium Medik STIKES Guna Bangsa Yogyakarta belum memiliki *tools* untuk memprediksi tingkat ketepatan waktu lulus mahasiswa tersebut. Tujuan penelitian itu yaitu mengevaluasi penerapan algoritma *Naive Bayes Classification* dan *K-Nearest NEIGHBORS* dalam pemodelan prediksi ketepatan waktu lulus mahasiswa dengan mengukur akurasi hasil prediksi menggunakan teknik skenario uji *n-Folds Cross Validation*. Penelitian ini menggunakan data akademik berupa sampel data mahasiswa program studi D3 Teknologi Laboratorium Medik Tahun Akademik (TA) 2015/2016 sampai dengan 2018/2019. Penelitian ini menggunakan pendekatan eksperimen (*Experimental*) dengan membandingkan metode algoritma *Naive Bayes Classification* dengan *K-Nearest NEIGHBORS*. Pelatihan dan pengujian sistem dilakukan dengan metode *5-fold Cross Validation* dengan mengukur hasil akurasi, presisi dan *recall*. Hasil penelitian menunjukkan bahwa prediksi dengan *Naive Bayes Classification* pada kasus ini dengan tingkat *performance* yaitu akurasi sebesar 96,11 % dengan tingkat presisi sebesar 82,11% dan *Recall* sebesar 100,00% dan prediksi dengan menggunakan algoritma KNN yaitu akurasi sebesar 97,68% dengan tingkat presisi sebesar 100,00% dan *Recall*

sebesar 86,11%. Dengan demikian, KNN merupakan algoritma dengan tingkat akurasi lebih baik untuk menyelesaikan kasus prediksi ketepatan waktu lulus mahasiswa Program Studi D3 Teknologi Laboratorium Medik STIKES Guna Bangsa Yogyakarta.

**Kata Kunci:** *Prediksi ketepatan waktu lulus mahasiswa, Naive Bayes Classification, K-Nearest Neighbors, n-Folds Cross Validation*