

**PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR*  
DAN *NAÏVE BAYES CLASSIFIER* DALAM MEMPREDIKSI  
KATEGORI INDEKS PRESTASI MAHASISWA**

Skripsi



oleh:  
**VERONICA HANNA ARUM SARI**  
**71130129**

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA

2017

**PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR*  
DAN *NAÏVE BAYES CLASSIFIER* DALAM MEMPREDIKSI  
KATEGORI INDEKS PRESTASI MAHASISWA**

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana  
Sebagai Salah Satu Syarat dalam Memperoleh Gelar  
Sarjana Komputer

Disusun oleh:

**VERONICA HANNA ARUM SARI**

**71130129**

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA

2017

# PERNYATAAN KEASLIAN TUGAS AKHIR

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### **PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR DAN NAÏVE BAYES CLASSIFIER DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI MAHASISWA**

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi keserjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar keserjanaan saya.

Yogyakarta, 2 Juni 2017



VERONICA HANNA ARUM SARI  
71130129

## HALAMAN PERSETUJUAN

### HALAMAN PERSETUJUAN

Judul Skripsi : PERBANDINGAN ALGORITMA K-NEAREST  
NEIGHBOR DAN NAÏVE BAYES CLASSIFIER  
DALAM MEMPREDIKSI KATEGORI INDEKS  
PRESTASI MAHASISWA

Nama Mahasiswa : VERONICA HANNA ARUM SARI

N I M : 71130129

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2016/2017

Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 2 Juni 2017

Dosen Pembimbing I



R. Gunawan Santosa, Drs. M.Si.

Dosen Pembimbing II



Antonius Rachmat C., S.Kom.,M.Cs.

# HALAMAN PENGESAHAN

## HALAMAN PENGESAHAN

### PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR DAN NAÏVE BAYES CLASSIFIER DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI MAHASISWA

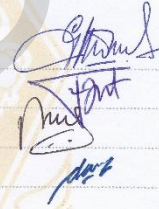
Oleh: VERONICA HANNA ARUM SARI / 71130129

Dipertahankan di depan Dewan Penguji Skripsi  
Program Studi Teknik Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana - Yogyakarta  
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Komputer  
pada tanggal 29 Mei 2017

Yogyakarta, 2 Juni 2017  
Mengesahkan,


Dewan Penguji:

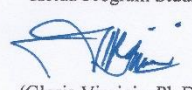
1. R. Gunawan Santosa, Drs. M.Si.
2. Antonius Rachmat C., S.Kom., M.Cs.
3. Nugroho Agus Haryono, M.Si
4. Danny Sebastian, S.Kom., M.M., M.T.



Dekan

Ketua Program Studi

  
(Budi Susanto, S.Kom., M.T.)

  
(Gloria Virginia, Ph.D.)



## UCAPAN TERIMAKASIH

Dalam proses penulisan tugas akhir ini, penulis mendapatkan banyak bantuan baik saran, kritik, serta bimbingan dari berbagai pihak. Oleh karena itu, sudah sepantasnya penulis mengantarkan ucapan terima kasih kepada:

1. Bapak Budi Susanto, S.Kom, M.T. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana
2. Ibu Gloria Virginia, S. Kom. MAI. selaku Ketua Program Studi Teknik Informatika Universitas Kristen Duta Wacana
3. Bapak R. Gunawan Santosa, Drs. M.Si. selaku pembimbing I yang telah membimbing dan mengarahkan penulis dalam penyusunan tugas akhir ini
4. Bapak Antonius Rachmat C., S.Kom.,M.Cs. selaku pembimbing II yang telah membimbing dan mengarahkan penulis dalam penyusunan tugas akhir ini.
5. Orangtua yang selalu mendoakan, memberikan motivasi dan pengorbanan baik segi moril dan materi kepada penulis sehingga dapat menyelesaikan tugas akhir ini dengan baik.
6. Teman-teman Sedulur yang selalu memberikan saran, semangat dan dorongan dalam penulisan tugas akhir ini.
7. Teman-teman mahasiswa/i Program Studi Teknik Informatika 2013 Universitas Kristen Duta Wacana yang telah memberikan motivasi menyelesaikan tugas akhir ini.
8. Semua pihak yang tidak dapat disebutkan satu per satu yang telah ikut memberikan dukungan baik secara langsung maupun tidak langsung.

Penulis menyadari bahwa masih banyak kekurangan , baik dalam penelitian ini maupun dalam penulisan laporan penelitian. Akhir kata penulis mengucapkan terima kasih kepada semua pihak yang telah membantu. Semoga tugas akhir ini dapat bermanfaat bagi kita semua dan menjadi bahan masukan bagi dunia pendidikan.

Penulis

## KATA PENGANTAR

Puji syukur dan terima kasih kepada Tuhan Yang Maha Esa karena atas berkat rahmatNya penulis dapat menyelesaikan tugas akhir yang berjudul “Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier dalam Memprediksi Kategori Indeks Prestasi Mahasiswa” dengan lancar.

Laporan tugas akhir ini diajukan guna melengkapi sebagai syarat dalam mencapai gelar sarjana strata satu (S1) di Fakultas Teknologi Informasi Program Studi Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta. Penulis menyadari meskipun telah berusaha untuk menyajikan pembahasan sebaik mungkin, namun masih terdapat kekurangan dalam tugas akhir ini. Hal ini terjadi dikarenakan masih terbatasnya kemampuan dan pengetahuan penulis. Penulis mengharapkan kritik dan saran yang membangun untuk menyempurnakan tugas akhir ini.

Dalam proses penyusunan tugas akhir ini, penulis banyak mengalami kendala, namun berkat bantuan, bimbingan, dan kerjasama dari berbagai piha serta berkah dari Tuhan Yang Maha Esa sehingga kendala-kendala yang dihadapi dapat teratasi. Oleh karena itu, penulis mengucapkan terima kasih dan penghargaan kepada Bapak R. Gunawan Santosa, Drs. M.Si. selaku pembimbing I dan Bapak Antonius Rachmat C., S.Kom.,M.Cs. selaku pembimbing II yang telah bersedia membimbing dengan sabar dan bersedia meluangkan waktu, tenaga dan pikiran dalam memberikan bimbingan, motivasi dan arahan serta saran-saran yang sangat berharga bagi penulis dalam menyusun tugas akhir ini.

Penulis

**INTISARI**  
**PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR* DAN *NAÏVE***  
***BAYES CLASSIFIER* DALAM MEMPREDIKSI KATEGORI INDEKS**  
**PRESTASI MAHASISWA**

Hasil belajar mahasiswa baru FTI UKDW dapat dilihat dari prestasi belajar mahasiswa. Tolak ukur keberhasilan prestasi belajar mahasiswa baru dapat dilihat dari indeks prestasi yang dicapai. Indeks Prestasi (IP) semester yang rendah akan memberikan dampak baik luar maupun dalam kampus. Baik buruknya indeks prestasi mahasiswa semester 1 dapat dipengaruhi oleh beberapa variabel baik eksternal maupun internal. Variabel eksternal meliputi kategori sekolah (SMA atau SMK), status sekolah (negeri atau swasta), lokasi sekolah (Jawa atau luar Jawa), dan kemampuan bahasa inggris yang dibagi dalam beberapa level. Variabel internal meliputi kemampuan spasial, kemampuan verbal, kemampuan numerik dan kemampuan analogi (Santosa & Rachmat, 2016). Proses penerimaan mahasiswa baru dibagi menjadi 2 jalur penerimaan yakni jalur prestasi dan jalur non prestasi. Jalur prestasi menggunakan variabel eksternal sedangkan jalur non prestasi menggunakan variabel eksternal dan internal. Pengolahan data-data baik variabel internal maupun eksternal, dalam *data mining* dapat menggunakan cara pengklasifikasian. Pada penelitian ini, penulis mengambil metode *Naïve Bayes Classifier* dan *K-Nearest Classifier* dalam melakukan pengklasifikasian.

Pada jalur prestasi, dari total 806 dengan menggunakan data *training* sebanyak 76% dan data *testing* sebanyak 24% didapatkan akurasi yang paling baik yaitu 61% menggunakan metode *K-Nearest Neighbor* dengan menggunakan nilai  $k = 2$ . Sedangkan pada jalur non prestasi, dari total 1211 data dengan menggunakan data *training* sebanyak 95% dan data *testing* sebanyak 5% didapatkan akurasi yang paling baik yaitu 48% menggunakan metode *K-Nearest Neighbor* dengan menggunakan nilai  $k = 8$ . Dari penelitian yang dilakukan, akurasi pada jalur prestasi lebih baik daripada jalur non prestasi.

Kata Kunci : , Akurasi, *Confusion Matrix*, *Data Mining*, *K-Nearest Neighbor*, *Naïve Bayes Classifier*



## DAFTAR ISI

PERNYATAAN KEASLIAN TUGAS AKHIR.....	iii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMAKASIH.....	v
KATA PENGANTAR .....	vii
INTISARI.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xii
BAB I.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Sistem .....	2
1.4 Tujuan Penelitian.....	3
1.5 Metodologi Penelitian .....	3
1.6 Sistematika Penulisan.....	4
BAB 2 .....	6
2.1 Tinjauan Pustaka .....	6
2.2 Landasan Teori .....	8
BAB 3 .....	24
3.1. Spesifikasi Kebutuhan.....	24
3.2. Perancangan Proses Sistem .....	33
BAB 4 .....	52
4.1. Implementasi Sistem .....	52
4.2. Analisis .....	72
BAB 5 .....	104
5.1. Kesimpulan .....	104
5.2. Saran.....	105
DAFTAR PUSTAKA .....	106
LAMPIRAN.....	1

## DAFTAR TABEL

Tabel 2. 1 Data Sampel Latih.....	12
Tabel 2. 2 Data Sampel Latih (sambungan horizontal).....	12
Tabel 2.3 Data Uji.....	12
Tabel 2. 4 Data Uji (sambungan horizontal).....	13
Tabel 2.5 Mencari jarak data latih.....	13
Tabel 2.6 Mencari jarak data latih (sambungan horizontal) .....	13
Tabel 2.7 Mencari jarak dan tetangga terdekat .....	13
Tabel 2.8 Mencari jarak dan tetangga terdekat (sambungan horizontal) .....	14
Tabel 2.9 Jarak terdekat dimana $k = 3$ .....	14
Tabel 2.10 Jarak terdekat dimana $k = 3$ (sambungan horizontal) .....	14
Tabel 2.11 Kelas yang bersesuaian .....	14
Tabel 2.12 Kelas yang bersesuaian (sambungan horizontal) .....	15
Tabel 2.13 Data mahasiswa FTI UKDW .....	16
Tabel 2. 14 Data mahasiswa FTI UKDW (sambungan horizontal) .....	16
Tabel 2.15 Data Uji .....	17
Tabel 2. 16 Data Uji (sambungan horizontal).....	17
Tabel 2.17 Data Sesungguhnya.....	22
Tabel 2. 18 Data Sesungguhnya (sambungan horizontal).....	22
Tabel 2.19 Data Hasil Pengujian.....	22
Tabel 2. 20 Data Hasil Pengujian (sambungan horizontal).....	22
Tabel 2. 21 Data Sesungguhnya yang disilangkan dengan Data Prediksi .....	23
Tabel 3. 1 Tabel testing jalur prestasi .....	39
Tabel 3. 2 Testing training jalur prestasi.....	40
Tabel 3. 3 Tabel testing jalur non prestasi .....	41
Tabel 3. 4 Tabel testing jalur non prestasi (sambungan vertical) .....	42
Tabel 3. 5 Tabel training jalur non prestasi.....	42
Tabel 3. 6 Tabel training jalur non prestasi (sambungan vertikal) .....	43
Tabel 3. 7.....	44
Tabel 4. 1 Confusion matrix metode Naïve Bayes Classifier jalur prestasi.....	73
Tabel 4. 2. Confusion matrix metode K-Nearest Classifier jalur prestasi dengan nilai $k = 2$ .....	74
Tabel 4. 3. Confusion matrix metode K-Nearest Neighbor jalur prestasi dengan nilai $k=3$ .....	75
Tabel 4. 4. Confusion matrix metode K-Nearest Neighbor jalur prestasi dengan nilai $k=4$ .....	76
Tabel 4. 5. Confusion matrix metode K-Nearest Neighbor jalur prestasi dengan nilai $k=5$ .....	78
Tabel 4. 6. Confusion matrix metode K-Nearest Neighbor jalur prestasi dengan nilai $k=6$ .....	79
Tabel 4. 7. Confusion matrix metode K-Nearest Neighbor jalur prestasi dengan nilai $k=7$ .....	80

Tabel 4. 8. Confusion matrix metode K-Nearest Neighbor pada jalur prestasi dengan nilai k=8.....	81
Tabel 4. 9. Confusion matrix metode K-Nearest Neighbor jalur prestasi dengan nilai k=9 .....	82
Tabel 4. 10. Confusion matrix metode Naïve Bayes Classifier jalur non prestasi	83
Tabel 4. 11. Confusion matrix metode K-Nearest Neighbor jalur non prestasi dengan nilai k=2.....	84
Tabel 4. 12. Confusion matrix metode K-Nearest Neighbor jalur non prestasi dengan nilai k=3.....	86
Tabel 4. 13. Confusion matrix metode K-Nearest Neighbor jalur non prestasi dengan nilai k=4.....	87
Tabel 4. 14. Confusion matrix metode K-Nearest Neighbor jalur non prestasi dengan nilai k=5.....	88
Tabel 4. 15. Confusion matrix metode K-Nearest Neighbor jalur non prestasi dengan nilai k=6.....	89
Tabel 4. 16. Confusion matrix metode K-Nearest Neighbor jalur non prestasi dengan nilai k=7.....	90
Tabel 4. 17. Confusion matrix metode K-Nearest Neighbor jalur non prestasi dengan nilai k=8.....	91
Tabel 4. 18. Confusion matrix metode K-Nearest Neighbor jalur non prestasi dengan nilai k=9.....	93
Tabel 4. 19 Hasil Akurasi Jalur Prestasi .....	94
Tabel 4. 20 Hasil Akurasi Jalur Non Prestasi .....	94
Tabel 4. 21 Hasil Akurasi Data Seimbang Jalur Prestasi.....	97
Tabel 4. 22 Hasil Akurasi Data Seimbang Jalur Non Prestasi.....	97
Tabel 4. 23 Hasil Akurasi Menambahkan 10% Data Training Jalur Prestasi.....	98
Tabel 4. 24 Hasil Akurasi Menambahkan 10% Data Testing Jalur Non Prestasi.	99
Tabel 4. 25 Hasil Akurasi Jalur Prestasi .....	99
Tabel 4. 26 Hasil Akurasi Jalur Prestasi (sambungan vertical) .....	100
Tabel 4. 27 Hasil Akurasi Menghilangkan Beberapa Variabel Jalur Non Prestasi .....	100
Tabel 4. 28 Hasil Akurasi Membagi Kelas Menjadi 5 Kategori Jalur Prestasi....	101
Tabel 4. 29 Hasil Akurasi Membagi Kelas Menjadi 5 Kategori Jalur Non Prestasi .....	101

## DAFTAR GAMBAR

Gambar 2.1. Langkah Proses Klasifikasi (Han & Kamber, 2006).....	10
Gambar 3. 1. Contoh dataset data training jalur non prestasi .....	25
Gambar 3. 2 Contoh Potongan Data Asli (Santosa & Rachmat, 2016) .....	26
Gambar 3. 3 Penyaringan Data Jalur Prestasi .....	27
Gambar 3. 4 Penyaringan Data Jalur Non Prestasi .....	27
Gambar 3. 5 Data Mahasiswa FTI UKDW Angkatan 2008-2014.....	27
Gambar 3. 6 Data Mahasiswa FTI UKDW Angkatan 2015 .....	27
Gambar 3. 7 Data Variabel yang Diperlukan.....	28
Gambar 3. 8 Data yang Siap Diolah untuk Jalur Prestasi .....	28
Gambar 3. 9 Data yang Siap Diolah untuk Jalur Non Prestasi .....	28
Gambar 3. 10. Use Case Diagram Sistem Klasifikasi.....	30
Gambar 3. 11. Diagram alir (flowchart) sistem .....	34
Gambar 3. 12. Diagram alir (flowchart) K-Nearest Neighbor .....	35
Gambar 3. 13. Diagram alir (flowchart) Naïve Bayes Classifier.....	36
Gambar 3. 14 Skema basis data sistem .....	37
Gambar 3. 15 Rancangan antarmuka form awal.....	45
Gambar 3. 16. Rancangan antarmuka KNN jalur prestasi .....	46
Gambar 3. 17. Rancangan antarmuka KNN jalur non prestasi .....	47
Gambar 3. 18. Rancangan antarmuka KNN jalur non prestasi (sambungan horizontal) .....	47
Gambar 3. 19. Rancangan antarmuka Naïve Bayes Classifier jalur prestasi .....	48
Gambar 3. 20. Rancangan antarmuka Naïve Bayes Classifier jalur non prestasi .	49
Gambar 3. 21. Rancangan antarmuka Naïve Bayes Classifier jalur non prestasi (sambungan).....	50
Gambar 3. 22. Rancangan antarmuka akurasi.....	50
Gambar 4. 1. Tampilan Halaman Login.....	52
Gambar 4. 2. Tampilan Halaman Jalur Penerimaan Mahasiswa .....	53
Gambar 4. 3. Tampilan Halaman Metode Naïve Bayes Classifier Jalur Prestasi .	54
Gambar 4. 4. Tampilan Halaman Naïve Bayes Classifier Jalur Prestasi .....	55
Gambar 4. 5. Halaman Kasus Baru Metode Naïve Bayes Classifier Jalur Prestasi .....	56
Gambar 4. 6. Tampilan Halaman Hasil Prediksi Kasus Baru Metode Naïve Bayes Classifier Jalur Prestasi .....	57
Gambar 4. 7. Tampilan Halaman Pertanyaan Kepada User.....	57
Gambar 4. 8. Tampilan Algoritma Kasus Baru Naïve Bayes Classifier Jalur Prestasi .....	58
Gambar 4. 9. Tampilan Halaman K-Nearest Neighbor Jalur Prestasi .....	59
Gambar 4. 10. Halaman Pengujian Akurasi Metode K-Nearest Neighbor Jalur Prestasi .....	59
Gambar 4. 11. Halaman Kasus Baru Metode K-Nearest Neighbor Jalur Prestasi	60

Gambar 4. 12. Tampilan Halaman Hasil Prediksi Kasus Baru Metode K-Nearest Neighbor Jalur Prestasi .....	61
Gambar 4. 13. Tampilan Halaman Pertanyaan Kepada User.....	61
Gambar 4. 14. Tampilan Algoritma Kasus Baru K-Nearest Neighbor Jalur Prestasi .....	62
Gambar 4. 15 Tampilan Halaman Metode Naïve Bayes Classifier Jalur Non Prestasi .....	63
Gambar 4. 16. Halaman Pengujian Akurasi Metode Naïve Bayes Classifier Jalur Non Prestasi .....	64
Gambar 4. 17. Halaman Kasus Baru Metode Naïve Bayes Classifier Jalur Non Prestasi .....	65
Gambar 4. 18. Tampilan Halaman Hasil Prediksi Kasus Baru Metode Naïve Bayes Classifier Jalur Non Prestasi.....	66
Gambar 4. 19. Tampilan Halaman Pertanyaan Kepada User.....	66
Gambar 4. 20. Tampilan Algoritma Kasus Baru Naïve Bayes Classifier Jalur Non Prestasi .....	67
Gambar 4. 21. Tampilan Halaman K-Nearest Neighbor Jalur Non Prestasi .....	68
Gambar 4. 22. Tampilan Halaman Pengujian Akurasi Metode K-Nearest Neighbor Jalur Non Prestasi.....	69
Gambar 4. 23. Halaman Kasus Baru Metode K-Nearest Neighbor Jalur Non Prestasi .....	70
Gambar 4. 24. Tampilan Halaman Hasil Prediksi Kasus Baru Metode K-Nearest Neighbor Jalur Non Prestasi .....	71
Gambar 4. 25. Tampilan Halaman Pertanyaan Kepada User.....	71
Gambar 4. 26. Tampilan Algoritma Kasus Baru K-Nearest Neighbor Jalur Non Prestasi .....	72
Gambar 4. 27 Grafik Akurasi K-Nearest Neighbor .....	95
Gambar 4. 28 Grafik Akurasi Naïve Bayes Classifier .....	96

**INTISARI**  
**PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR* DAN *NAÏVE***  
***BAYES CLASSIFIER* DALAM MEMPREDIKSI KATEGORI INDEKS**  
**PRESTASI MAHASISWA**

Hasil belajar mahasiswa baru FTI UKDW dapat dilihat dari prestasi belajar mahasiswa. Tolak ukur keberhasilan prestasi belajar mahasiswa baru dapat dilihat dari indeks prestasi yang dicapai. Indeks Prestasi (IP) semester yang rendah akan memberikan dampak baik luar maupun dalam kampus. Baik buruknya indeks prestasi mahasiswa semester 1 dapat dipengaruhi oleh beberapa variabel baik eksternal maupun internal. Variabel eksternal meliputi kategori sekolah (SMA atau SMK), status sekolah (negeri atau swasta), lokasi sekolah (Jawa atau luar Jawa), dan kemampuan bahasa inggris yang dibagi dalam beberapa level. Variabel internal meliputi kemampuan spasial, kemampuan verbal, kemampuan numerik dan kemampuan analogi (Santosa & Rachmat, 2016). Proses penerimaan mahasiswa baru dibagi menjadi 2 jalur penerimaan yakni jalur prestasi dan jalur non prestasi. Jalur prestasi menggunakan variabel eksternal sedangkan jalur non prestasi menggunakan variabel eksternal dan internal. Pengolahan data-data baik variabel internal maupun eksternal, dalam *data mining* dapat menggunakan cara pengklasifikasian. Pada penelitian ini, penulis mengambil metode *Naïve Bayes Classifier* dan *K-Nearest Classifier* dalam melakukan pengklasifikasian.

Pada jalur prestasi, dari total 806 dengan menggunakan data *training* sebanyak 76% dan data *testing* sebanyak 24% didapatkan akurasi yang paling baik yaitu 61% menggunakan metode *K-Nearest Neighbor* dengan menggunakan nilai  $k = 2$ . Sedangkan pada jalur non prestasi, dari total 1211 data dengan menggunakan data *training* sebanyak 95% dan data *testing* sebanyak 5% didapatkan akurasi yang paling baik yaitu 48% menggunakan metode *K-Nearest Neighbor* dengan menggunakan nilai  $k = 8$ . Dari penelitian yang dilakukan, akurasi pada jalur prestasi lebih baik daripada jalur non prestasi.

Kata Kunci : , Akurasi, *Confusion Matrix*, *Data Mining*, *K-Nearest Neighbor*, *Naïve Bayes Classifier*

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Penerimaan mahasiswa baru merupakan proses yang selalu dilakukan setiap tahunnya oleh Universitas Kristen Duta Wacana Yogyakarta (UKDW). Mahasiswa baru merupakan mahasiswa transisi dari Sekolah Menengah Atas (SMA) menuju perkuliahan. Perbedaan akademis yang tinggi tentu dirasakan oleh mahasiswa tahun pertama khususnya mahasiswa Fakultas Teknologi Informasi (FTI). Sebagai contoh, tugas yang diberikan di perkuliahan umumnya menuntut mahasiswa untuk mencari literatur yang lain. Selanjutnya mahasiswa mengembangkan sendiri pola pikirnya dalam menyelesaikan tugas yang diberikan. Hasil belajar mahasiswa dapat dilihat dari prestasi belajar mahasiswa tersebut. Tolak ukur keberhasilan prestasi belajar mahasiswa baru dapat dilihat dari indeks prestasi semester 1 yang dicapai.

Indeks Prestasi (IP) yang rendah akan memberikan dampak baik luar maupun dalam kampus. Dampak tersebut diantaranya kepuasan orangtua dalam menitipkan anak dan masalah pada proses belajar mengajar serta suasana akademik yang ada. Baik buruknya indeks prestasi mahasiswa semester 1 dapat dipengaruhi oleh beberapa variabel baik eksternal maupun internal. Variabel eksternal meliputi kategori sekolah (SMA atau SMK), status sekolah (negeri atau swasta), lokasi sekolah (Jawa atau luar Jawa), dan kemampuan bahasa Inggris yang dibagi dalam beberapa level. Variabel internal meliputi kemampuan spasial, kemampuan verbal, kemampuan numerik dan kemampuan analogi (Santosa & Rachmat, 2016).

Variabel-variabel tersebut disajikan ke bentuk data yang kompleks. Oleh karena itu, diperlukan *data mining* agar data-data variabel baik internal maupun eksternal dapat diolah dan digali dengan baik.

Pengolahan data-data baik variabel internal maupun eksternal, dalam *data mining* dapat menggunakan cara pengklasifikasian. Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han, 2006). Algoritma yang digunakan

dalam mengklasifikasi indeks prestasi semester 1 mahasiswa FTI UKDW adalah algoritma *K-Nearest Neighbor* yang akan dibandingkan keakuratannya dengan *Naive Bayes Classifier*.

Dalam penelitian ini, data yang diambil merupakan data penelitian yang dilakukan oleh Santosa & Rachmat (2016). Data yang akan digunakan dalam penelitian ini adalah data mahasiswa FTI UKDW angkatan 2008-2015. Data indeks prestasi yang akan diprediksi pengkategoriannya adalah data mahasiswa FTI UKDW angkatan 2015 dengan menggunakan data sampel mahasiswa FTI UKDW angkatan 2008-2014.

## 1.2 Rumusan Masalah

Permasalahan yang akan diteliti dalam memprediksi pengklasifikasian ini adalah

- Berapa besar akurasi algoritma *K-Nearest Neighbor* dalam sistem untuk memprediksi kategori IP semester 1 mahasiswa FTI UKDW angkatan 2015?
- Berapa besar akurasi algoritma *Naive Bayes Classifier* dalam sistem untuk memprediksi kategori IP semester 1 mahasiswa FTI UKDW angkatan 2015?
- Bagaimana perbandingan keakuratan algoritma *K-nearest neighbor* dengan *Naive bayes classifier* dalam mengklasifikasi IP Semester 1 mahasiswa FTI UKDW angkatan 2015?

## 1.3 Batasan Sistem

Dalam penelitian ini, penulis memiliki beberapa batasan masalah agar tidak menyimpang jauh dari permasalahan, diantaranya:

- Data yang akan dijadikan sebagai sampel data latih adalah data indeks prestasi semester 1 mahasiswa FTI UKDW angkatan 2008-2014. Sedangkan data yang akan dijadikan pengujian adalah data indeks prestasi



semester 1 mahasiswa FTI UKDW angkatan 2015. Data ini diambil dari penelitian yang dilakukan oleh Santosa & Rachmat, 2016

- Metode *K-Nearest Neighbor* diterapkan dalam sistem untuk mengkategorikan IP Semester 1 mahasiswa FTI UKDW angkatan 2015 dengan nilai k yaitu 2, 3, 4, 5, 6, 7, 8, dan 9.
- Metode *Naive Bayes Classifier* diterapkan dalam sistem untuk mengkategorikan IP Semester 1 mahasiswa FTI UKDW angkatan 2015
- Variabel yang akan digunakan sebagai bahan pengolahan data adalah kategori sekolah (SMA atau SMK), status sekolah (negeri atau swasta), lokasi sekolah (Jawa atau luar Jawa), kemampuan bahasa Inggris yang dibagi dalam beberapa level, kemampuan spasial, kemampuan verbal, kemampuan numerik dan kemampuan analogi.
- Kategori IP semester 1 mahasiswa FTI UKDW angkatan 2015 akan dibagi menjadi 3 kelas. Untuk pengkategorian 3 kelas terdiri kelas rendah (0 – 1.33), sedang (1.34 – 2.66), dan tinggi (2.67 – 4.00)

#### **1.4 Tujuan Penelitian**

Adapun tujuan yang mendasari penelitian ini adalah mengetahui perbandingan keakuratan metode *K-Nearest Neighbor* dan *Naive Bayes Classifier* dalam mengkategorikan IP Semester 1 mahasiswa FTI UKDW angkatan 2015.

Dengan adanya penelitian ini, diharapkan menjadi evaluasi bagi fakultas teknologi informasi UKDW dalam seleksi mahasiswa baru bagi Fakultas Teknologi Informasi UKDW di masa yang akan datang serta berkurangnya tingkat mahasiswa FTI UKDW yang pindah ke prodi lain.

#### **1.5 Metodologi Penelitian**

Metodologi yang dilakukan dalam penelitian ini adalah:

##### a) Studi Pustaka

Studi Pustaka dilakukan dengan mempelajari teori-teori melalui internet, buku, artikel, jurnal dan bahan lain yang mendukung yang berhubungan dengan *data mining*, *K-Nearest Neighbor* dan *Naive Bayes Classifier*.

b) Pengumpulan data

Dalam tahap ini, penulis melakukan pengambilan data dari penelitian yang dilakukan oleh Santosa & Rachmat, 2016. Penelitian ini menggunakan dataset sebanyak 2017. Data tersebut merupakan data mahasiswa FTI UKDW angkatan 2008-2015.

c) Perancangan Sistem

Tahap ini merupakan tahapan perancangan basis data dan perancangan antarmuka untuk sistem yang akan dibangun. Perancangan sistem ini meliputi perancangan struktur pengambilan data untuk pelatihan dan pengujian. Data latih yang digunakan adalah data mahasiswa FTI UKDW angkatan 2008-2014. Sedangkan data uji yang digunakan adalah data mahasiswa FTI UKDW angkatan 2015.

d) Pengujian Sistem

Tahap ini akan menggunakan data indeks prestasi semester 1 mahasiswa FTI UKDW angkatan 2015 sebagai data yang akan diuji. *User* akan menguji kategori kelas mahasiswa FTI UKDW 2015 dengan data-data yang telah tersedia yaitu NIM dan beberapa variabel (kategori sekolah, status sekolah, lokasi sekolah, kemampuan Bahasa Inggris, nilai numerik, nilai verbal, nilai spasial, dan nilai analogi). Setelah *user* menginputkan variabel-variabel tersebut, *user* dapat mengetahui kategori kelas indeks prestasi mahasiswa FTI UKDW 2015. Selanjutnya *user* dapat mengetahui akurasi dari *K-Nearest Neighbor* yang dibandingkan dengan *Naive Bayes Classifier*.

## 1.6 Sistematika Penulisan

Dalam mendapatkan gambaran yang lengkap dan jelas mengenai penelitian yang akan dilakukan, penulis membagi laporan Tugas Akhir ini menjadi lima bab, yaitu:

Bab 1 Pendahuluan yang menguraikan hal-hal seperti latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode/pendekatan yang digunakan serta sistematika penulisan laporan Tugas Akhir.

Bab 2 Tinjauan Pustaka yang berisi berisi dua bagian utama yaitu Tinjauan Pustaka dan Landasan Teori yang diperlukan untuk memecahkan masalah dalam riset yang dilakukan. Tinjauan Pustaka ini berisi tentang penelitian-penelitian yang telah dilakukan sebelumnya yang berkaitan dengan algoritma *K-Nearest Neighbor* dan *Naive Bayes Classifier*. Landasan Teori dalam penelitian ini berisi penjelasan mengenai *data mining*, *K-Nearest Neighbor*, dan *Naive Bayes Classifier*.

Bab 3 Analisis dan Perancangan Sistem akan menjelaskan analisis teori yang digunakan, menerjemahkan ke dalam sistem, dan membangun rancangan tampilan dasar dari sistem yang akan dibuat

Bab 4 Implementasi dan Analisis Sistem berisi tentang hasil penelitian atau implementasi serta pembahasan dan analisis dari penelitian yang telah dilakukan dan dijelaskan secara terpadu.

Bab 5 Kesimpulan dan Saran merupakan uraian singkat dari penelitian yang dilakukan serta saran yang akan berguna untuk pengembangan sistem selanjutnya

## BAB 5 KESIMPULAN DAN SARAN

### 5.1. Kesimpulan

Dari penelitian yang telah dilakukan oleh penulis, terdapat beberapa kesimpulan yang diperoleh yaitu sebagai berikut:

1. Pada proses penerimaan mahasiswa FTI UKDW untuk jalur prestasi dengan menggunakan data *training* sebanyak 613 dan data *testing* sebanyak 193, metode akurasi yang paling baik pada metode *K-Nearest Neighbor* dengan nilai  $k$  yaitu 2 dengan tingkat akurasi 61%. Sedangkan pada metode *Naïve Bayes Classifier*, memiliki tingkat akurasi sebesar 55%.
2. Pada proses penerimaan mahasiswa baru jalur non prestasi, dengan menggunakan data *training* sebanyak 1150 dan data *testing* sebanyak 61, metode akurasi yang paling baik pada metode *K-Nearest Neighbor* dengan nilai  $k = 8$  sebesar 48%. Sedangkan pada metode *Naïve Bayes Classifier*, memiliki tingkat akurasi sebesar 41%.
3. Dari kedua proses penerimaan mahasiswa baru, metode *K-Nearest Neighbor* memiliki nilai akurasi yang lebih baik daripada metode *Naïve Bayes Classifier*.
4. Berdasarkan metode *K-Nearest Neighbor* yang dilakukan, nilai  $k$  yang semakin besar tidak menjamin bahwa akurasi itu akan semakin baik ataupun buruk.
5. Dari kedua proses penerimaan mahasiswa baru, proses yang melalui jalur prestasi memiliki rata-rata nilai akurasi 53,67% sedangkan jalur non prestasi memiliki rata-rata nilai akurasi 41,89%. Jadi, jalur prestasi memiliki nilai akurasi yang lebih baik daripada jalur non prestasi.
6. Dari beberapa percobaan yang dilakukan, penulis menyimpulkan bahwa dengan menggunakan eksperimen dengan menyeimbangkan

data menjadi proporsi yang sama besar, akan meningkatkan akurasi menjadi 64% untuk jalur prestasi dengan metode *Naïve Bayes Classifier*. Sedangkan untuk jalur non prestasi pada percobaan ini memiliki akurasi sebesar 48% dengan metode *K-Nearest Neighbor* yang menggunakan nilai  $k=5$ .

## 5.2. Saran

Untuk pengembangan sistem lebih lanjut, saran yang dapat diberikan oleh penulis adalah:

1. Langkah pengurutan / *sorting* pada metode *K-Nearest Neighbor* dapat menggunakan cara lain agar proses pengurutan tidak membutuhkan waktu yang lama dalam memprediksi data dalam jumlah banyak.
2. Dengan nilai akurasi yang telah didapatkan pada penelitian ini, perlu adanya teknik pengklasifikasian menggunakan metode lain seperti *C4.5*, *K-Means*, *Decision Tree*, dll agar mendapatkan nilai akurasi yang lebih baik.

## DAFTAR PUSTAKA

- Amin, K. R., Indwiarti, & Sibaroni, Y. (2015). IMPLEMENTASI KLASIFIKASI DECISION TREE DENGAN ALGORITMA C4.5 DALAM PENGAMBILAN KEPUTUSAN PERMOHONAN KREDIT OLEH DEBITUR.
- Banjarsari, A. M., Budiman, I., & Farmadi, A. (2015). Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4.
- Chitupe, A., & Joshi, S. (2013). Data Classification Algorithm Using k-Nearest Neighbour Method Applied to ECG Data. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 5.
- Han, J., & Kamber, M. (2006). *Data mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.
- Hastuti, K. (2012). ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING. 4-7.
- Indriani, A. (2014). Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier.
- Iskandar, D., & Suprpto, Y. K. (2013). Perbandingan akurasi klasifikasi tingkat kemiskinan antara algoritma C4.5 dan Naïve Bayes Clasifier. *JAVA Journal of Electrical and Electronics Engineering*, 2.
- Jananto, A. (2013). Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa.
- Larose, D. (2005). *Discovering Knowledge in Data*.
- Leidiyana, H. (2013). PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK PENENTUAN RESIKO KREDIT KEPEMILIKAN KENDARAAN BEMOTOR. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*.
- Liu, B. (2007). *Web Data mining: Exploring Hyperlinks, Contents, dan Usage Data*.
- Moertini, V. S. (2002). *Data mining sebagai solusi bisnis*.
- Nursalim, Suprapedi, & Himawan. (2014). KLASIFIKASI BIDANG KERJA LULUSAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR.

- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier.
- Saito, T., & Rehmsmeier, M. (2015, October). *Basic evaluation measures from the confusion matrix*. Retrieved from <https://classeval.wordpress.com/introduction/basic-evaluation-measures/>
- Santosa, R. G., & Rachmat, A. (2016). *Regresi Logistik Untuk Prediksi Kategori IP Mahasiswa Fakultas Teknologi Informasi UKDW*. Yogyakarta.
- Teknomo, K. (2017). *KNN Numerical Example*. Retrieved from [http://people.revoledu.com/kardi/tutorial/KNN/KNN\\_Numerical-example.html](http://people.revoledu.com/kardi/tutorial/KNN/KNN_Numerical-example.html)
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. New York: CRC Press.
- Yunanto, W., Hariadi, M., & Purnomo, M. H. (n.d.). PEMETAAN KECELAKAAN LALU LINTAS BERBASIS KLASIFIKASI NAIVE BAYES DENGAN PARAMETER INFRASTRUKTUR JALAN.