

SKRIPSI

**PENERAPAN METODE *NAÏVE BAYES CLASSFFIER*
DALAM MENGLASIFIKASIKAN UKT PADA
UNIVERSITAS SEMBILANBELAS NOVEMBER
KOLAKA**



**OLEH
NUR IDA
171210156**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS SEMBILANBELAS NOVEMBER
KOLAKA
2021**

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN METODE *NAÏVE BAYES CLASSFFIER*
DALAM MENGLASIFIKASIKAN UKT PADA
UNIVERSITAS SEMBILANBELAS NOVEMBER
KOLAKA**

Diusulkan oleh

**NUR IDA
171210156**

Telah disetujui
Pada tanggal 2021

Pembimbing I

Andi Tenri Sumpala, S.Kom.,M.Cs.

NIDN.0921058305

Pembimbing II

Yuwanda Purnamasari Parun, S.T.,M.Kom.

NIDN.0003089004

KATA PENGANTAR



Puji syukur kehadiran Allah SWT atas segala nikmat dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian yang berjudul “Penerapan Metode *Naïve Bayes Classifier* dalam mengklasifikasikan UKT di Universitas Sembilanbelas November Kolaka” sebagai salah satu persyaratan untuk menyelesaikan program Sarjana Strata Satu (S-1) Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Sembilanbelas November Kolaka.

Perjalanan panjang yang telah penulis lalui selama menyelesaikan penelitian ini. Banyak hambatan yang dialami dalam penyusunannya namun berkat dorongan, bimbingan, motivasi dan bantuan dari beberapa pihak. Oleh karena itu Penulis mengucapkan banyak terima kasih kepada:

1. ALLAH SWT. yang telah memberikan nikmat yang tak terhingga kepada penulis.
2. Kedua orang tua dan saudara-saudara tercinta atas doa yang tak pernah putus, selalu memberikan dorongan serta motivasi kepada penulis sehingga penulis dapat menyelesaikan proposal penelitian ini.
3. Bapak Dr.Azhari, S.STP, M.S.i, Selaku Rektor Universitas Sembilanbelas November Kolaka.
4. Ibu Noorhasanah. Z, S.Kom., M.Kom., Selaku Dekan Fakutlas Teknologi Informasi Universitas Sembilanbelas November Kolaka.
5. Bapak Anjar Pradipta, S.Kom., M.Kom., Selaku Ketua Program Studi Sistem Informasi Universitas Sembilanbelas November Kolaka dan selaku

dosen penasehat akademik yang telah memberikan bimbingan serta saran selama penulis menempuh studi.

6. Ibu Andi Tenri Sumpala, S.Kom., M.Cs. selaku pembimbing I dan Ibu Yuwanda purnamasari Pasrun, S.T., M.Kom selaku pembimbing II yang dengan ikhlas meluangkan waktunya dalam memberikan arahan dan masukan serta bimbingannya selama proses penyelesaian proposal ini.
7. Para Dosen Program Sistem Informasi dan Para Staf Fakultas Teknologi Informai Universitas Sembilanbelas November Kolaka
8. Teman-teman seperjuangan Sistem Informasi Angkatan 2017 atas segala dukungan, doa serta kerja samanya yang diberikan hingga sampai saat ini.
9. Dan Sahabat-sahabat yang selama ini selalu memberikan doa, dukungan, motivasi serta bantuan dalam penyusunan dan penyelesaian proposal ini.

Penulis menyadari bahwa terdapat banyak kekurangan dalam isi penelitian ini. Hal ini tidak terlepas dari kata-kata yang mungkin tidak berkenan dan perlu untuk diperhatikan kembali. Oleh karena itu Penulis sangat mengharapkan kepada bapak/ibu untuk mengoreksi demi kesempurnaan penulisan.

Kolaka, Oktober 2021

NUR IDA

**PENERAPAN METODE *NAÏVE BAYES CLASSIFIER* DALAM
MENGKLASIFIKASIKAN UKT PADA UNIVERSITAS
SEMILANBELAS NOVEMBER KOLAKA**

NUR IDA (171210156)

Program Studi S-1 Sistem Informasi, Fakultas Teknologi Informasi
Universitas Sembilanbelas November Kolaka
Alamat: Jln. Pemuda No.339 Kolaka Sulawesi Tenggara
Telp : 0405-2321132, Fax : 0405232440228
Email : **nuridasi99@gmail.com**

INTISARI

Uang Kuliah Tunggal (UKT) merupakan nama dari sebuah sistem pembayaran yang saat ini berlaku diseluruh PTN. Universitas Sembilanbelas November Kolaka membagi ke dalam 5 (lima) kelompok UKT. Dalam penelitian ini penulis menggunakan Algoritma *Naive Bayes Classifier* untuk mengklasifikasikan UKT menggunakan atribut antara lain jenis bantuan, jumlah saudara kandung, wilayah, kondisi orang tua, pendidikan ayah, pekerjaan ayah, pendapatan ayah, pendidikan ibu, pekerjaan ibu, dan pendapatan ibu. Berdasarkan perbandingan perhitungan manual dengan sistem, keduanya menunjukkan hasil yang sama sehingga Algoritma *Naive Bayes* terbukti dapat diimplementasikan ke dalam sistem. Berdasarkan hasil *confusion matrix* dimana nilai akurasi untuk tiga kali pengujian mengalami kenaikan. Pengujian pertama menggunakan 385 *data training* dan 100 *data testing* menghasilkan nilai akurasi sebesar 77%, pengujian kedua menggunakan 405 *data training* dan 80 *data testing* menghasilkan nilai akurasi sebesar 82,5%, pengujian ketiga menggunakan 425 *data training* dan 60 *data testing* menghasilkan nilai akurasi sebesar 85%, ini membuktikan bahwa semakin banyak *data training* maka nilai akurasi akan semakin tinggi.

Kata kunci : Klasifikasi, *Naive Bayes Classifier*, *Confusion Matrix*, Akurasi

**PENERAPAN METODE NAÏVE BAYES CLASSIFIER DALAM
MENGKLASIFIKASIKAN UKT PADA UNIVERSITAS
SEMILANBELAS NOVEMBER KOLAKA**

NUR IDA (171210156)

Program Studi S-1 Sistem Informasi, Fakultas Teknologi Informasi
Universitas Sembilanbelas November Kolaka
Alamat: Jln. Pemuda No.339 Kolaka Sulawesi Tenggara
Telp : 0405-2321132, Fax : 0405232440228
Email : **nuridasi99@gmail.com**

INTISARI

Single Tuition Fee (UKT) is the name of a payment system that currently applies to all PTNs. Nineteen November University Kolaka divided into 5 (five) UKT groups. In this study, the author uses the Naive Bayes Classifier Algorithm to classify UKT using attributes including the type of assistance, number of siblings, region, parental condition, father's education, father's occupation, father's income, mother's education, mother's occupation, and mother's income. Based on the comparison of manual calculations with the system, both show the same results so that the Naive Bayes Algorithm is proven to be able to be implemented into the system. Based on the results of the confusion matrix where the accuracy value for the three tests has increased. The first test uses 385 training data and 100 testing data produces an accuracy value of 77%, the second test uses 405 training data and 80 testing data produces an accuracy value of 82.5%, the third test uses 425 training data and 60 testing data produces an accuracy value of 85%, this proves that the more training data, the higher the accuracy value.

Keywords : Classification, Naive Bayes Classifier, Confusion Matrix, Accuracy

DAFTAR ISI

SKRIPSI.....	i
HALAMAN PERSETUJUAN.....	ii
KATA PENGANTAR	vii
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu	5
2.2 Landasan Teori.....	8
2.3 Perancangan Sistem	16
BAB III METODOLOGI PENELITIAN.....	23
3.1 Tempat dan Jadwal Penelitian.....	23
3.2 Waktu Penelitian.....	23
3.3 Teknik Pengumpulan Data.....	23
3.4 <i>Design</i> / Rancangan Sistem	24
3.5 <i>Coding</i> / Rancang bangun program	25
3.6 Sumber Data dan Jenis Penelitian.....	25
3.7 Alat dan Bahan.....	26
BAB IV HASIL DAN PEMBAHASAN	27
4.1 Hasil Penelitian	27
4.2 Rancangan Sistem.....	27
4.3 Pemodelan dan Struktur Data	32
4.4 Model Perhitungan Algoritma <i>Naive Bayes</i>	32
4.5 Implementasi Pemrograman	53
4.6 Pengujian Akurasi	59
BAB V PENUTUP.....	63
5.1 Kesimpulan	63
5.2. Saran	63
DAFTAR PUSTAKA	64
LAMPIRAN.....	66

DAFTAR GAMBAR

Gambar 2.1 Tahap-Tahap <i>Data mining</i>	8
Gambar 2.2 Tahapan Metode <i>Waterfall</i>	21
Gambar 3.1 Algoritma <i>Naïve Bayes Classffier</i>	24
Gambar 4. 1 Diagram Konteks.....	27
Gambar 4. 2 Data <i>Flow Diagram</i> Level 1.....	28
Gambar 4. 3 <i>Flowchart</i> Menu Utama Admin	29
Gambar 4. 4 <i>Flowchart</i> Menu Utama Pegawai.....	30
Gambar 4. 5 <i>Flowchart</i> Menu Klasifikasi.....	31
Gambar 4. 6 <i>Flowchart Dataset</i>	31
Gambar 4. 7 Tampilan Halaman <i>Login</i>	53
Gambar 4. 8 Tampilan Halaman Menu Utama	53
Gambar 4. 9 Tampilan <i>Dataset</i>	54
Gambar 4. 10 Tampilan Uji Akurasi.....	55
Gambar 4. 11 Tampilan Klasifikasi	56
Gambar 4. 12 Tampilan Hasil Klasifikasi.....	57
Gambar 4. 13 Hasil Implementasi Pada Sistem	61
Gambar 4. 14 Hasil Implementasi Pada Sistem	62

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	6
Tabel 2.4 Tabel <i>Confusion Matrix</i>	15
Tabel 2.7 Simbol-simbol <i>Flowchart</i>	17
Tabel 3.1 Jadwal Penelitian.....	23
Tabel 4. 1 Pemodelan dan Struktur Data	32
Tabel 4. 2 Data Awal	33
Tabel 4. 3 Hasil Transformasi Data	34
Tabel 4. 4 <i>Data Training</i>	35
Tabel 4. 5 Data Testing	36
Tabel 4. 6 Hasil Perhitungan.....	52
Tabel 4. 7 Hasil <i>Confusin Matrix</i> Uji ke-1.....	59
Tabel 4. 8 Hasil <i>Confusin Matrix</i> Uji ke-2.....	60
Tabel 4. 9 Hasil <i>Confusin Matrix</i> Uji ke-3.....	60

BAB I

PENDAHULUAN

1.1 Latar Belakang

Uang Kuliah Tunggal (UKT) merupakan nama dari sebuah sistem pembayaran yang saat ini berlaku diseluruh PTN. Ketentuan ini berdasarkan Permendibud No 55 Tahun 2013 pasal 1 ayat 3 yakni setiap mahasiswa hanya membayar satu komponen (Permendipud, 2013). Sistem UKT telah diterapkan mulai tahun akademik 2013/2014 bagi mahasiswa baru program S1 di seluruh PTN di Indonesia. Sistem ini meniadakan uang pangkal yang harus dibayar oleh setiap mahasiswa dengan semua biaya lain menjadi sebuah biaya tunggal. Tujuan dari penerapan sistem UKT yaitu mampu meringankan beban pembayaran pendidikan mahasiswa khususnya mahasiswa kurang mampu. Pemerintah menetapkan UKT dengan mensubsidi dan menanggung besaran UKT pada setiap mahasiswa/mahasiswi setiap satu semester. Pembayaran UKT yang ditanggung oleh setiap mahasiswa tidak sama, dikarenakan adanya sistem subsidi silang pada UKT yang didasarkan kondisi ekonomi dan sosial orang tua/wali mahasiswa.

Dalam menerapkan Uang Kuliah Tunggal, Universitas Sembilanbelas November Kolaka membagi ke dalam 5 (lima) kelompok UKT. Penentuan kelompok UKT dilakukan dengan mengisi biodata pada saat pendaftaran kembali lalu diverifikasi. Ini menjadi permasalahan yang baru ketika mulai memasuki masa penerimaan mahasiswa baru. Mengingat jumlah mahasiswa yang begitu banyak, pihak universitas kesulitan dan memerlukan waktu yang cukup lama untuk menentukan kategori yang tepat dalam penetapan UKT pada setiap mahasiswa. Untuk membantu penetapan kategori yang tepat pada setiap mahasiswa, dapat digunakan pendekatan *Data Mining* dengan algoritma tertentu untuk pengklasifikasian UKT mahasiswa. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat

dan pengetahuan yang terkait dari berbagai *database* besar (Turban & kk, 2005). Sedangkan klasifikasi merupakan salah satu metode dalam penambangan data (*data mining*) yang digunakan dalam menganalisis sekumpulan data penting. Metode-metode dalam klasifikasi dapat secara otomatis memprediksi kelas dari data lain yang belum diklasifikasikan (Sulistiawati & Utami, 2018).

Berdasarkan penelitian yang dilakukan oleh M.S. Chen, J.Han, and P.S. Yu menyatakan bahwa Analisis Klasifikasi mampu membuat deksripsi yang akurat atau membangun model yang akurat untuk setiap kategori (Sulistiawati & Utami, 2018). Pada Klasifikasi terdapat beberapa metode yang dapat digunakan, salah satunya yaitu metode *Naïve Bayes Classffier* yang bekerja sangat baik dibanding dengan metode *Classffier* lainnya. Hal ini dibuktikan pada peneitian (Nurhasan , Hikmah, & Utami, 2018) yang berjudul Perbandingan algoritma C4.5, KNN, dan *Naive Bayes* untuk Penentuan Model Klasifikasi Penanggung Jawab BSI *Entrepreneur Center* menyebutkan bahwa “*Naïve Bayes Classffier* memiliki nilai akurasi lebih tinggi dibanding model *classffier* lainnya”. Maka dari itu pihak Universitas memerlukan suatu sistem yang berbasis komputer yang mampu mengklasifikasikan penentuan UKT berdasarkan data *training* yang diambil dari data mahasiswa baru tahun 2020. Dalam penelitian ini ada beberapa atribut yang telah ditetapkan. Atribut yang digunakan yaitu antara lain jenis bantuan, jumlah saudara kandung, wilayah, kondisi orang tua, pendidikan ayah, pekerjaan ayah, pendapatan ayah, pendidikan ibu, pekerjaan ibu, dan pendapatan ibu.

Penelitian ini mengusulkan metode *Naive Bayes Clasifier* untuk klasifikasi pengelompokan UKT. Diharapkan mampu memberikan informasi yang berguna bagi pihak universitas dalam penentuan UKT mahasiswa/mahasiswi baru.

1.2 Rumusan Masalah

Berdasarkan pemaparan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah “Apakah penerapan metode *Naïve Bayes Classffier* dapat digunakan dalam mengklasifikasikan UKT pada Univeraitas Sembilanbelas November Kolaka?”.

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah

1. Data *set* yang digunakan sebanyak 485 dengan tiga kali pengujian. Pengujian pertama menggunakan 385 *data training* dan 100 *data testing*, pengujian kedua menggunakan 405 *data training* dan 80 *data testing* dan pengujian ketiga menggunakan 425 *data training* dan 60 *data testing*.
2. Atribut yang digunakan yaitu jenis bantuan, jumlah saudara kandung, wilayah, kondisi orangtua, pendidikan ayah, pekerjaan ayah, pendapatan ayah, pendidikan ibu, pekerjaan ibu dan pendapatan ibu.
3. Klasifikasi yang dihasilkan untuk mengklasifikasikan UKT ada 5 yaitu: “kelompok UKT I”, “kelompok UKT II”, “kelompok UKT III”, kelompok UKT IV”, kelompok UKT V”.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Menerapkan algoritma *Naïve Bayes Classffier* dalam klasifikasi Uang Kuliah Tunggal (UKT).
2. Mengetahui hasil dari akurasi sistem yang telah dibangun menggunakan algoritma *Naïve Bayes Classffier* dalam klasifikasi Uang Kuliah Tunggal (UKT).

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

1. Bagi pihak Universitas
Melalui penelitian ini diharapkan dapat memberikan sebuah rekomendasi hasil klasifikasi menggunakan metode *Naïve Bayes Classffier* dalam penentuan Uang Kuliah Tunggal (UKT).
2. Bagi Penulis

Melalui penelitian ini diharapkan dapat memperdalam pengetahuan penulis dalam membangun sistem klasifikasi penentuan Uang Kuliah Tunggal (UKT).

3. Bagi Pihak Lain

Penelitian ini diharapkan dapat memberikan sumbangsih berupa pemikiran terutama klasifikasi penentuan UKT serta dapat menjadi bahan rujukan untuk pengembangan penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Sebagai bahan rujukan dalam penelitian ini, penulis menggunakan beberapa penelitian terdahulu untuk menunjang penelitian.

Dalam penelitian yang berjudul Penerapan Naïve Bayes Classifier dengan Gaussian Function Untuk Menentukan Kelompok UKT. Tujuan penelitian ini adalah dapat membantu dalam menentukan kelompok UKT mahasiswa baru dalam menerapkan Uang Kuliah Tunggal, IAIN Zawiyah Cot Kala Langsa membagi kedalam 3 (Tiga) kategori kelompok UKT (Reza, 2018).

Dalam penelitian yang berjudul Klasifikasi Potensi Berdasarkan Kepribadian Siswa dengan Metode *Naïve Bayes*. Tujuan penelitian ini adalah untuk mengetahui dan mengklasifikasikan potensi sehingga potensi siswa dapat dikembangkan dan siswa menjadi sumber daya manusia yang bermutu dan berkualitas (Rochmawati, Istiadi, Marisa, & Nurdiansyah, 2020).

Dalam penelitian yang berjudul Pengembangan Sistem Rekomendasi Calon Penerima Beasiswa Dengan *Metode Naïve Bayes*. Tujuan penelitian ini adalah membangun sistem penentuan calon penerima beasiswa dengan metode *Naive Bayes Clasification* sesuai dengan kriteria-kriteria yang sudah ditentukan. Atribut kriteria yang digunakan dalam penelitian ini adalah nama, IPK, bekerja/tdk, anak ke, jumlah tanggungan, pekerjaan orangtua, penghasilan orangtua. Diharapkan dapat memudahkan pihak penyeleksi dalam merekomendasikan nama-nama calon penerima beasiswa sehingga diharapkan akan lebih tepat sasaran (Aditya, 2017).

Dalam penelitian yang berjudul Aplikasi Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan Metode *Naive Bayes* di Bank BNI Syariah Surabaya. Tujuan penelitian ini adalah membantu pihak analisis kartu kredit dalam

mengklasifikasikan nasabah untuk mendapatkan kartu kredit yang tepat dengan menerapkan metode *Naive Bayes*. Metode *Naive Bayes* dimanfaatkan untuk mencari pola nasabah yang sudah diterima dalam pengajuan kartu kredit dengan variabel yang digunakan adalah jenis kelamin, status rumah, status, jumlah tanggungan, profesi, dan penghasilan pertahunnya (Antaristi & Kurniawan, 2017).

Dalam Penelitian yang berjudul Klasifikasi Status Gizi Balita Menggunakan *Naive Bayes Classification*. Tujuan dari Penelitian ini adalah untuk memudahkan dalam penentuan status gizi balita menggunakan *data mining* dengan algoritma *Naive Bayes Classification* (NBC). Sistem dibangun dengan bahasa pemrograman php dan *database* MySQL. Penelitian ini menggunakan data 17 balita dengan rentang waktu 2 tahun pengukuran. Total data yang digunakan berjumlah 408 data dengan dua kali pengujian data (Putra, 2018).

Secara rinci penelitian terdahulu yang termuat dalam penelitian ini disajikan pada Tabel 2.1.

Tabel 2.1 Penelitian Terdahulu

No	Peneliti	Judul	Metode	Hasil Penelitian
1.	Reza (2017)	Penerapan <i>Naive Bayes Classifier</i> dengan Gaussian Function Untuk Menentukan Kelompok UKT	<i>Naive Bayes Classifier</i>	Pengujian model klasifikasi NBC akurasi sebesar 86,67%.
2.	Rochmawati, Istiadi, Marisa, & Nurdiansyah (2020)	Klasifikasi Potensi Berdasarkan Kepribadian Siswa dengan Metode <i>Naive Bayes</i>	<i>Naive Bayes</i>	Pengujian yang telah dilakukan dengan metode <i>Naive Bayes</i> maka didapatkan akurasi sebesar 95,83% dari 24 data yang diuji.

Tabel 2.1 Lanjutan

No	Peneliti	Judul	Metode	Hasil Penelitian
3.	Aditya (2017)	Pengembangan Sistem Rekomendasi Calon Penerima Beasiswa Dengan Metode <i>Naïve Bayes</i>	<i>Naïve Bayes</i>	Pengujian yang telah dilakukan dengan metode <i>Naïve Bayes</i> menghasilkan tingkat akurasi 92% dengan menggunakan 25 <i>data training</i> .
4.	Antaristi & Kurniawan (2017)	Aplikasi Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan Metode <i>Naive Bayes</i> di Bank BNI Syariah Surabaya	<i>Naive Bayes</i>	Penelitian ini menunjukkan bahwa <i>accuracy</i> sebesar 98,67%. Dengan demikian, aplikasi ini sudah efektif untuk membantu pihak analisis kartu kredit dalam mengklasifikasikan nasabah untuk mendapatkan kartu kredit yang tepat dan terbaik sesuai kriteria.
5.	Putra (2018)	Klasifikasi Status Gizi Balita Menggunakan <i>Naïve Bayes Classification</i>	<i>Naïve Bayes Classification</i>	Pengujian pada perbandingan <i>data training</i> dan <i>data testing</i> sebesar 80:20 menghasilkan akurasi tertinggi dengan 95.121%.

Adapun Perbedaan penelitian yang dilakukan oleh penulis dengan penelitian yang dilakukan oleh Reza (2018) atribut klasifikasi yang digunakan yaitu penghasilan orang tua, rasio penghasilan terhadap jumlah tanggungan, dan tagihan rekening listik. Pada penelitian ini atribut klasifikasi yang digunakan yaitu jenis bantuan, jumlah saudara kandung, wilayah, kondisi orangtua, pendidikan ayah, pekerjaan ayah, pendapatan ayah, pendidikan ibu, pekerjaan ibu dan pendapatan ibu.

2.2 Landasan Teori

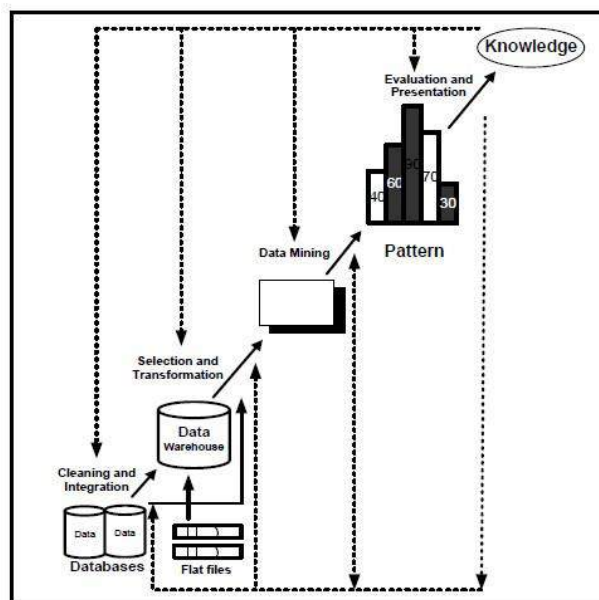
2.2.1 Data Mining

1. Pengertian Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Patut diingat bahwa kata *mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu *Data Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan *database* (Pramudiono, 2003).

2. Tahap-Tahap Data Mining

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap yang diilustrasikan di Gambar 2.1. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*.



Gambar 2.1 Tahap-Tahap *Data mining* (Han & Kamber, 2006)

Tahap – tahap *data mining* yaitu :

1. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak *valid* atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesis *data mining* yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau *file* teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi data (*data selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus *market basket analysis*, tidak perlu mengambil nama pelanggan, cukup dengan *id* pelanggan saja.

4. Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima *input* data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses *mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesis yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesis ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses *data mining*, mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining* (Han & Kamber, 2006).

2.2.2 Klasifikasi

Klasifikasi adalah suatu proses untuk mengelompokkan sejumlah data ke dalam kelas-kelas tertentu yang sudah diberikan berdasarkan kesamaan sifat dan pola yang terdapat dalam data-data tersebut. Secara umum, proses klasifikasi dimulai dengan diberikannya sejumlah data yang menjadi acuan untuk membuat aturan klasifikasi data. Data-data inilah yang biasa disebut dengan *training sets*. Dari *training sets* tersebut kemudian dibuat suatu model untuk mengklasifikasi data-data yang belum diketahui kelasnya dan biasa disebut sebagai *test sets*. (Rifqi, Maharani, & Shaufiah, 2011).

Algoritma klasifikasi akan menghasilkan pola atau aturan yang dapat digunakan untuk memprediksi kelas. Dalam klasifikasi terdapat juga target *variable* kelas. Tujuannya dari algoritma klasifikasi adalah untuk menemukan relasi antara *variable* yang tergolong dalam kelas yang sama.

2.2.3 Algoritma Naive Bayes Classifier

1. Definisi Algoritma

Menurut Suarga, algoritma adalah (1) teknik penyusunan langkah/tahap penyelesaian masalah dalam bentuk kalimat dengan jumlah kata yang terbatas, tetapi tersusun secara logis dan sistematis. (2) Suatu prosedur yang jelas untuk menyelesaikan suatu persoalan dengan menggunakan langkah-langkah tertentu dan terbatas jumlahnya. (3) Susunan langkah yang pasti, yang bila diikuti maka akan mentransformasi data *input* menjadi *output* yang berupa informasi. (Mustafa, Ramadhan, & Angelina, 2017)

2. Definisi Naïve Bayes Classffier

Naive Bayes Classifier adalah sebuah metode klasifikasi yang berdasar pada *Teorema Bayes*. Metode pengklasifikasian ini menggunakan metode probabilitas dan statistik yang pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga metode

ini dikenal sebagai *Teorema Bayes*. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian (Mustafa, Ramadhan, & Angelina, 2017).

Olson dan Delen dalam bukunya, menjelaskan bahwa *Naive Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar. Algoritma ini berdasar pada mengasumsikan bahwa atribut objek adalah sesuatu yang independen. Probabilitas yang terlibat dalam membuat perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan (Mustafa, Ramadhan, & Angelina, 2017).

Formula perhitungan *Naive Bayes Classifier* ditunjukkan sebagai berikut:

1. Menghitung nilai *class log prior*

$$\text{class log prior} = \log \left(\frac{N_{tic}}{n_c} \right) \quad (2.1)$$

Dimana

class log prior = probabilitas log empiris untuk setiap kelas

N_{tic} = berapa kali kategori t muncul dalam sampel yang termasuk kelas c

n_c = jumlah sampel dalam kelas c

2. Menghitung nilai *cat count* dan *class count*

$$\text{cat count} = n_i + \alpha \quad (2.2)$$

$$\text{class count} = \text{sum}(n_i + \alpha) \quad (2.3)$$

Dimana

cat count = jumlah sampel untuk setiap kelas dalam kategori fitur tertentu

class count = jumlah sampel untuk setiap kelas yang diberi nilai sesuai dengan jumlah sampel yang ada

n_i = jumlah kategori fitur dalam data

α = parameter pemulus/*laplace* untuk menangani masalah frekuensi nol

3. Menghitung nilai *smoothed cat count* dan *smoothed class count*

$$\text{smoothed cat count} = \log(n_i + \alpha) \quad (2.4)$$

$$\text{smoothed class count} = \log(\text{sum}(n_i + \alpha)) \quad (2.5)$$

dimana

smoothed cat count = pemulus untuk setiap kategori fitur yang ada dalam setiap kelas

smoothed class count = pemulus untuk setiap kelas yang diberi nilai sesuai dengan jumlah sampel yang ada

n_i = jumlah kategori fitur dalam data

α = parameter pemulus/*laplace* untuk menangani masalah frekuensi nol

4. Menghitung *feature log prob*

$$\text{feature log prob} = \log(n_i + \alpha) - \log(\text{sum}(n_i + \alpha)) \quad (2.6)$$

Dimana

feature log prob = probabilitas log empiris dari kategori fitur dalam masing-masing kelas

n_i = jumlah kategori fitur dalam data

α = parameter pemulus/*laplace* untuk menangani masalah frekuensi nol

5. Menghitung nilai *likelihood*

$$P(x|c) = P(t_1|c) + P(t_2|c) + \dots + P(t_n|c) \quad (2.7)$$

Dimana

P = probabilitas

x = variabel

c = kelas

t = kategori

3. Fase *Data Mining*

Dalam metode *Naive Bayes Classffier* terdapat dua fase yaitu fase *training* dan *testing*.

1. Data Percobaan (*testing*) proses ini merupakan proses perhitungan data yang mengacu pada *data training*. Pada proses ini ada beberapa tahapan, yaitu menghitung peluang kriteria terhadap golongan,

menghitung peluang setiap golongan, dan menentukan nilai maksimal pada masing-masing posterior (Sulistiawati & Utami, 2018).

2. Data Pelatihan (*training*) jika jenis data bersifat kategorial untuk sebagian data yang telah diketahui kelasnya diproses untuk membentuk model perkiraan. Namun jika jenis data bersifat numerik, maka proses *training* harus melewati beberapa tahapan yaitu perhitungan nilai *mean*, *variance*, dan *deviasi standart* pada tiap kriteria untuk masing-masing golongan.

5. Karakteristik *Naïve Bayes*

Klasifikasi dengan *Naïve Bayes* bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *Naïve Bayes* sebagai berikut (Laroussi, 2015) :

1. Metode *Naïve Bayes* bekerja teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outliner*) dan juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai kolerasi bisa mendegradasi kinerja klasifikasi *Naïve Bayes* karena asumsi independensi atribut tersebut sudah tidak ada.

2.2.4 Akurasi Sistem

Performa dari suatu model klasifikasi dapat diukur dengan tingkat akurasi berdasarkan *Confusion matrix*. *Confusion matrix* merupakan alat yang berguna untuk menganalisis seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda. *TP* dan *TN* memberikan informasi ketika *classifier* benar, sedangkan *FP* dan *FN* memberikan informasi ketika *classifier* salah (Elvianti, 2015). Tabel 2.2 adalah contoh dari *confusion matrix*.

Tabel 2.2 Tabel *Confusion Matrix* (Elvianti, 2015)

		<i>Actual Class</i>	
		Ya	Tidak
<i>Predictive Class</i>	Ya	<i>TP</i>	<i>FN</i>
	Tidak	<i>FP</i>	<i>TN</i>
	Total	<i>P'</i>	<i>N'</i>

Akurasi merupakan presentasi dari data yang diprediksi secara benar.

Perhitungan akurasi adalah:

$$Akurasi = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.8)$$

Keterangan :

TP = True Positives, merupakan jumlah data dengan kelas positif yang diklasifikasikan positif.

TN = True Negatives, merupakan jumlah data dengan kelas negatif yang diklasifikasikan negatif.

FP = False Positives, merupakan jumlah data dengan kelas positif yang diklasifikasikan negatif.

FN = False Negatives, merupakan jumlah data dengan kelas negatif yang diklasifikasikan positif.

2.2.5 Uang Kuliah Tunggal (UKT)

Uang Kuliah Tunggal atau disingkat UKT menurut Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 55 Tahun 2013, “Uang kuliah tunggal merupakan sebagian biaya kuliah tunggal yang ditanggung setiap mahasiswa berdasarkan kemampuan ekonominya” (Republik Indonesia, 2013). Uang kuliah tunggal sebagaimana yang dimaksud terdiri atas beberapa kelompok yang ditentukan berdasarkan kelompok kemampuan ekonomi masyarakat. Uang

kuliah tunggal ini ditetapkan berdasarkan biaya kuliah tunggal dikurangi biaya yang ditanggung oleh pemerintah.

Biaya kuliah tunggal adalah keseluruhan biaya operasional mahasiswa per semester pada program studi di perguruan tinggi negeri. Dimana, biaya kuliah tunggal ini digunakan untuk menetapkan biaya yang dibebankan kepada mahasiswa masyarakat dan pemerintah. Dengan diberlakukan UKT ini, perguruan tinggi negeri tidak dibolehkan memungut uang pangkal atau pungutan lainnya dari mahasiswa baru yang telah diterima diperguruan tinggi tersebut.

2.3 Perancangan Sistem

2.3.1 *Data Flow Diagram (DFD)*

Data Flow Diagram (DFD) merupakan model dari sistem untuk menggambarkan pembagian sistem ke modul yang lebih kecil. Al Bahra Bin Ladjamudin juga menjelaskan untuk memudahkan analisis dimulai dengan diagram-diagram sebagai berikut (Ningrum, 2017) :

a. Diagram Nol



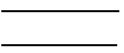

Diagram nol adalah diagram yang menggambarkan proses dari *data flow diagram*. Diagram nol memberikan gambaran secara menyeluruh mengenai sistem yang ditangani, menunjukan tentang fungsi-fungsi utama atau proses yang ada, aliran data dan eksternal *entity*.

b. Diagram Rinci

Diagram rinci adalah diagram yang menguraikan proses apa yang ada dalam diagram nol atau diagram level atasnya.

c. Simbol-simbol


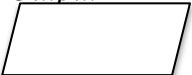
Tabel 2.3 Simbol-Simbol pada DFD (Ladjamudin, 2013)

Elemen DFD	Simbol	Fungsi
Proses		Menunjukkan pemrosesan data/informasi yang terjadi didalam sistem.
<i>Data Flow</i>		Menunjukkan arah aliran dokumen antar bagian yang terkait pada suatu sistem.
<i>Data Store</i>		Tempat menyimpan dokumen arsip.
Entitas		Menunjukkan entitas atau bagian yang telibat yang melakukan proses.

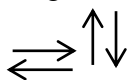
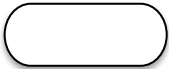
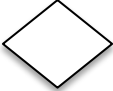

2.3.2 Flowchart

Flowchart (diagram alir) adalah bagan (*chart*) yang menunjukkan air (*Flow*) di dalam program atau prosedur sistem secara logika. *Flowchart* merupakan metode untuk menggambarkan tahap-tahap pemecahan masalah dapat mempresentasikan simbol-simbol tentu yang mudah dimengerti, mudah digunakan dan standar. Tujuan penggunaan *flowchart* adalah untuk menggambarkan suatu tahap penyelesaian masalah secara sederhana, terurai, rapi, dan jelas dengan menggunakan simbol-simbol yang standar. Tahapan penyelesaian masalah yang disajikan harus jelas, sederhana dan tepat (Jogiyanto, 2010). Tabel Berikut adalah simbol-simbol *flowchart*.

Tabel 2.4 Simbol-simbol *Flowchart* (Jogiyanto, 2010)

Simbol	Keterangan
Penghubung 	Simbol untuk keluar/masuk atau proses dalam lembar atau halaman berbeda
<i>Input Output</i> 	Simbol yang menyatakan proses <i>input</i> dan <i>output</i> tanpa tergantung dengan jenis peralatannya.

Tabel 2.7 Lanjutan

Simbol	Keterangan
Simbol garis alir 	Digunakan untuk menunjukkan arah selanjutnya yang akan dituju dari simbol-simbol dan <i>flowchart</i> .
Terminal 	Simbol yang menunjukkan untuk permulaan atau akhir suatu sistem.
Kondisi 	Simbol keputusan yang menunjukkan kondisi.
Proses 	Simbol yang menunjukkan pengolahan dilakukan oleh komputer.

2.3.3 Konsep Dasar Aplikasi Sistem

Dalam pengembangan penerapan metode *naïve bayes classffier* ini yang digunakan untuk klasifikasi Penurunan Sekolah Menengah Atas (SMA), digunakan beberapa alat pengembangan aplikasi sistem yaitu menggunakan bahasa pemrograman Visual Basic 2012, Xampp sebagai *server local* dan MySql untuk *database* dengan pengaksesan *database* menggunakan phpMyAdmin.

1. Python

Python adalah bahasa pemrograman yang fleksibel dan sederhana yang didefinisikan dalam dokumen-dokumennya sebagai berikut (Nosrati, 2011):

1. Python adalah bahasa pemrograman tujuan umum yang sangat tingkat tinggi, dinamis, berorientasi objek, yang umum digunakan
2. bisa digunakan dalam aplikasi yang luas.
3. Bahasa ini dapat mendukung berbagai gaya pemrograman termasuk struktural dan berorientasi objek. Gaya lain juga bisa digunakan.

4. Python sangat fleksibel, karena kemampuannya untuk menggunakan komponen modular yang dirancang dalam bahasa pemrograman lainnya. Sebagai contoh, Anda dapat menulis sebuah program di C ++ dan mengimpornya ke python sebagai modul.

Beberapa karakteristik yang menjadi ciri khas atau pembeda python adalah:

1. Python cepat dan kuat
Python menyediakan semua fasilitas yang dibutuhkan untuk pemrograman dari operasi dasar hingga fungsi lanjutan. Alat pihak ketiga ini membuat segala sesuatunya dengan Python. Misalnya, dengan menulis hanya 3 baris kode sudah bisa membuat web server.
2. Python mendukung teknologi lainnya.
Hal ini dapat mendukung COM, .Net, objek lainnya. Selain itu, beberapa alternatif dan pelengkap dibuat untuk python yang memudahkan untuk bekerja dengan benda-benda ini dalam mode yang saling terhubung.
3. Python portabel Skrip Python dapat digunakan pada sistem operasi yang berbeda-beda seperti: Windows, Linux, UNIX, Amigo, Mac OS, dan OS lainnya, beberapa versi dirilis untuk .Net, Java Virtual Machine (JVM) dan Nokia S60.
4. Python sederhana dan indah
Python adalah bahasa tingkat tinggi yang memiliki banyak sumber untuk belajar. Selain itu, beragam perangkat pihak ketiga membuat bahasa ini menjadi mudah digunakan dan memotivasi pengguna untuk mengembangkan.
5. Python bersifat open source
Meskipun semua hak dari program ini untuk Python, namun bersifat open source dan tidak ada batasan dalam menggunakan, mengubah dan mendistribusikan

2. Django Web Framework

Django adalah framework untuk membangun aplikasi web yang ditulis dalam bahasa Python, menggunakan metode *Model-Template-View*

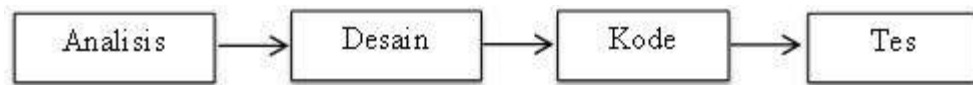
(Django (Version 1.9) (2015)). Pada akhir tahun 2003, pengembang situs web berita online "*World Online*" (Andrian dan Simon Willison) meninggalkan PHP dan mulai menggunakan Python untuk mengembangkan situs mereka. Setelah membangun situs yang intensif dan sangat interaktif seperti Lawrence.com, mereka mulai memisahkan *framework* web umum untuk pengembangan aplikasi web yang lebih cepat. Mereka mengubah dan menambahkan perbaikan pada *framework* itu selama dua tahun. Kemudian pada musim panas 2005, *World Online* memutuskan untuk menjadikan perangkat lunak Django sebagai *open-source*.

3. *Sublime Text*

Faridl & Miftah (2015) *sublime text* adalah teks editor berbasis Python, sebuah teks editor yang elegan, kaya fitur, cross platform, mudah, dan simpel yang cukup terkenal kalangan developer (pengembang), penulis, dan desainer. Para programmer biasanya menggunakan *sublime text* untuk menyunting *source code* yang sedang dikerjakan. Sampai saat ini *sublime text* sudah mencapai versi *Sublime text* mempunyai beberapa keunggulan-keunggulan yang dapat membantu dalam membuat sebuah web *development* antara lain *multiple selection*, *command palette*, *distraction free mode*, *find in project*, *plugin API switch*, *drag and drop*, *split editing*, dan *multi platform* (Rizqiani, Mulwinda, & Putri, 2017).

2.3.4 Model Pengembangan *Waterfall*

Model proses perangkat lunak merupakan gambaran dari proses pengembangan perangkat lunak. Menurut Pressman model *waterfall* (model air terjun) merupakan suatu model pengembangan secara sekuensial. Model *waterfall* bersifat sistematis dan berurutan dalam membangun sebuah perangkat lunak. Proses pembuatannya mengikuti alur dari mulai analisis, desain, kode, pengujian dan pemeliharaan.



Gambar 2.2 Tahapan Metode *Waterfall* (Pessman, 2002).

Tahap-tahap dari model pengembangan *Waterfall* ini, yaitu:

1) Analisis kebutuhan perangkat lunak.

Analisis kebutuhan perangkat lunak merupakan tahap pengumpulan kebutuhan yang diintensifkan dan difokuskan. Untuk mengetahui kebutuhan perangkat lunak, seperti apa yang dibutuhkan oleh pengguna. Spesifikasi kebutuhan perangkat lunak pada tahap ini perlu untuk didokumentasikan. Tujuan dari analisis kebutuhan yaitu merangkum hal-hal apa saja yang diinginkan pengguna dan mencari kebutuhan apa saja yang diperlukan dalam mengembangkan perangkat lunak. Kebutuhan untuk sistem maupun perangkat lunak didokumentasikan dan dilihat kembali oleh pengguna apakah sudah sesuai dengan yang diinginkan (Pessman, 2002).

2) Desain

Desain merupakan suatu tahapan yang berfokus pada desain untuk membuat perangkat lunak seperti struktur data, arsitektur perangkat lunak, *user interface* (antarmuka), dan prosedur pengkodean. Tahap desain dilakukan dengan menerjemahkan kebutuhan perangkat lunak berdasarkan dari hasil analisis kebutuhan ke dalam bentuk desain, sehingga dapat diimplementasikan menjadi program pada tahap implementasi. Desain didokumentasikan dan menjadi bagian dari konfigurasi perangkat lunak (Pessman, 2002).

3) Implementasi (Pembuatan Kode)

Desain merupakan tahap menerjemahkan desain sistem ke dalam perangkat lunak berdasarkan desain yang telah dibuat. Penerjemahan desain menggunakan kode bahasa program sehingga dapat berjalan dengan baik. Jika desain dilakukan dengan cara yang lengkap, pembuatan kode dapat diselesaikan secara mekanis. Hasil akhir dari tahap ini adalah

menghasilkan sistem informasi yang sesuai dengan desain yang sudah dibuat (Pessman, 2002).

Dalam tahap implementasi dilakukan evaluasi untuk mengurangi kesalahan dengan menjalankan kode yang dibuat, apakah sudah sesuai dengan spesifikasi yang ditentukan dan tidak terjadi kesalahan. Evaluasi berguna untuk mengurangi kesalahan yang terjadi sebelum semua bagian digabungkan menjadi satu kesatuan perangkat lunak.

4) Pengujian

Metode pengujian yang digunakan untuk mengukur kinerja klasifikator yaitu metode *hold-out*. Dalam metode *hold-out dataset* (yang sudah diketahui label kelas aslinya) di pecah menjadi dua bagian terpisah, yaitu set data latih dan set data uji. Model klasifikasi kemudian dibangun berdasarkan set data latih dan kemudian kinerja diukur berdasarkan set data uji (Prasetyo, 2014).

BAB III

METODOLOGI PENELITIAN

3.1 Tempat dan Jadwal Penelitian

Sesuai dengan judul penelitian ini, penulis memilih objek penelitian pada Universitas Sembilanbelas November Kolaka yang beralamat Jl. Pemuda, Tahoa, Kolaka, Kabupaten Kolaka, Sulawesi Tenggara.

3.2 Waktu Penelitian

Waktu penelitian yang dibutuhkan peneliti dimulai dari bulan juni sampai september 2021. Rencana jadwal penelitian dapat dilihat pada Tabel 3.1

Tabel 3.1 Jadwal Penelitian

No.	Keterangan	Bulan															
		Juni				Juli				Agustus				September			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengumpulan data																
2	Analisis																
2	<i>Design</i>																
3	<i>Code</i>																
4	<i>Testing</i>																

3.3 Teknik Pengumpulan Data

Dalam pelaksanaan penelitian diperlukan informasi dan data yang digunakan sebagai bahan dasar pengembangan sistem untuk mendukung pembahasan dalam laporan penelitian ini. Informasi dan data didapat dari observasi, dan wawancara, studi kepustakaan.

1. Pengamatan (observasi)

Pengamatan merupakan teknik pengumpulan data dimana peneliti melakukan pengamatan langsung di Universitas Sembilanbelas November Kolaka untuk mengumpulkan data primer berupa data UKT mahasiswa baru.

2. Wawancara (*interview*)

Penulis melakukan wawancara atau tanya-jawab dengan panitia penentuan UKT untuk mendapatkan informasi yang lebih mendalam dan lebih jelas.

3. Studi pustaka (*literature*)

Studi Pustaka dilakukan untuk mendapatkan landasan-landasan teori sebagai referensi untuk dijadikan pembahasan sesuai dengan judul penelitian, sumber *literature* berupa buku teks, *paper*, jurnal, karya ilmiah, dan situs-situs penunjang lainnya.

3.4 Design / Rancangan Sistem



Gambar 3.1 Algoritma *Naïve Bayes Classffier*

Berdasarkan Gambar 3.1, dapat dilihat bahwa proses pengklasifikasian *naïve bayes* dimulai dengan *input data testing*, membaca *data training*. selanjutnya masuk ke tahap perhitungan yaitu hitung *class log prior*, hitung *cat count* dan *class count*, hitung *smoothed cat count* dan *smoothed class count*, hitung *feature log prob* setelah itu akan dilanjutkan ke tahap perhitungan nilai likelihood dimana dari hasil perhitungan tersebut akan diperoleh solusi.

3.5 Coding / Rancang bangun program

Tahap ini merupakan tahap untuk mengimplementasikan desain yang telah dirancang ke dalam bentuk yang mudah dimengerti oleh mesin yaitu kedalam bahasa pemrograman melalui proses *coding*. Dalam penelitian ini pemrograman yang digunakan berbasis *desktop* menggunakan bahasa pemrograman Python dengan *Django Web Framework*.

3.6 Sumber Data dan Jenis Penelitian

3.6.1 Sumber data

- a. Data primer adalah data yang didapatkan secara langsung dari sumber data yang didapat dari hasil penelitian di Universitas Sembilanbelas November Kolaka.
- b. Data sekunder adalah data-data yang diperoleh dan digunakan untuk mendukung data, informasi data primer. Adapun data sekunder tersebut adalah dokumen, buku-buku, jurnal, internet, serta catatan-catatan yang berkaitan dengan judul penelitian.

3.6.2 Jenis Data Penelitian

- a. Data kualitatif adalah data yang berdasarkan kenyataan (fakta) yang diperoleh dengan berbagai macam teknik pengumpulan data seperti wawancara, observasi yang telah dilakukan ditempat penelitian. Berupa proses penentuan UKT di Universitas Sembilanbelas November Kolaka
- b. Data kuantitatif adalah data yang diperoleh dari data kualitatif yang bisa diubah dalam bentuk angka atau bilangan. Data kualitatif dapat diolah atau

dianalisis menggunakan teknik perhitungan matematis atau statistika. Berupa data nilai-nilai siswa baru.

3.7 Alat dan Bahan

a. Perangkat Keras (*Hardware*)

Laptop *Asus X441N* dengan spesifikasi sebagai berikut:

- a. *Processor Interl® Celeron® CPU N3350 @ 1.10GHz 1.10G Hz.*
- b. *RAM (memory) 4.00 GB (3.97 GB useble)*
- c. *Harddisk : 500GB*

b. Perangkat Lunak (*Software*)

- 1. *Microsoft Office 2010.*
- 2. *Python*
- 3. *Django web framework*
- 4. *Web browser*
- 5. *Sublime Text*

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Penelitian yang telah dilakukan di Universitas Sembilanbelas November Kolaka, diawali dengan melakukan pengambilan data sampel atau contoh data mahasiswa baru yang nantinya akan dijadikan sebagai data *training* dan data *testing*. Data *training* disini berupa data sampel mahasiswa baru. Data ini memiliki atribut jenis bantuan, jumlah saudara kandung, wilayah, kondisi orang tua, pendidikan ayah, pekerjaan ayah, pendapatan ayah, pendidikan ibu, pekerjaan ibu, dan pendapatan ibu. Data mahasiswa yang diambil sebanyak 485 data dengan tiga kali pengujian. Pengujian pertama menggunakan 385 *data training* dan 100 *data testing*, pengujian kedua menggunakan 405 *data training* dan 80 *data testing* dan pengujian ketiga menggunakan 425 *data training* dan 60 *data testing*. Berdasarkan hasil pengolahan data dan jumlah data yang digunakan tersebut dapat dibagi menjadi lima kelas, kelompok I sebanyak 49 data, kelompok II sebanyak 193 data, kelompok III sebanyak 169 data, kelompok IV sebanyak 54 data, dan kelompok V sebanyak 20 data. Penggunaan sistem klasifikasi Uang Kuliah Tunggal (UKT) diharapkan bisa memberikan hasil klasifikasi dengan akurasi terbaik sehingga hasilnya dapat diterima.

4.2 Rancangan Sistem

4.2.1 Data Flow Diagram

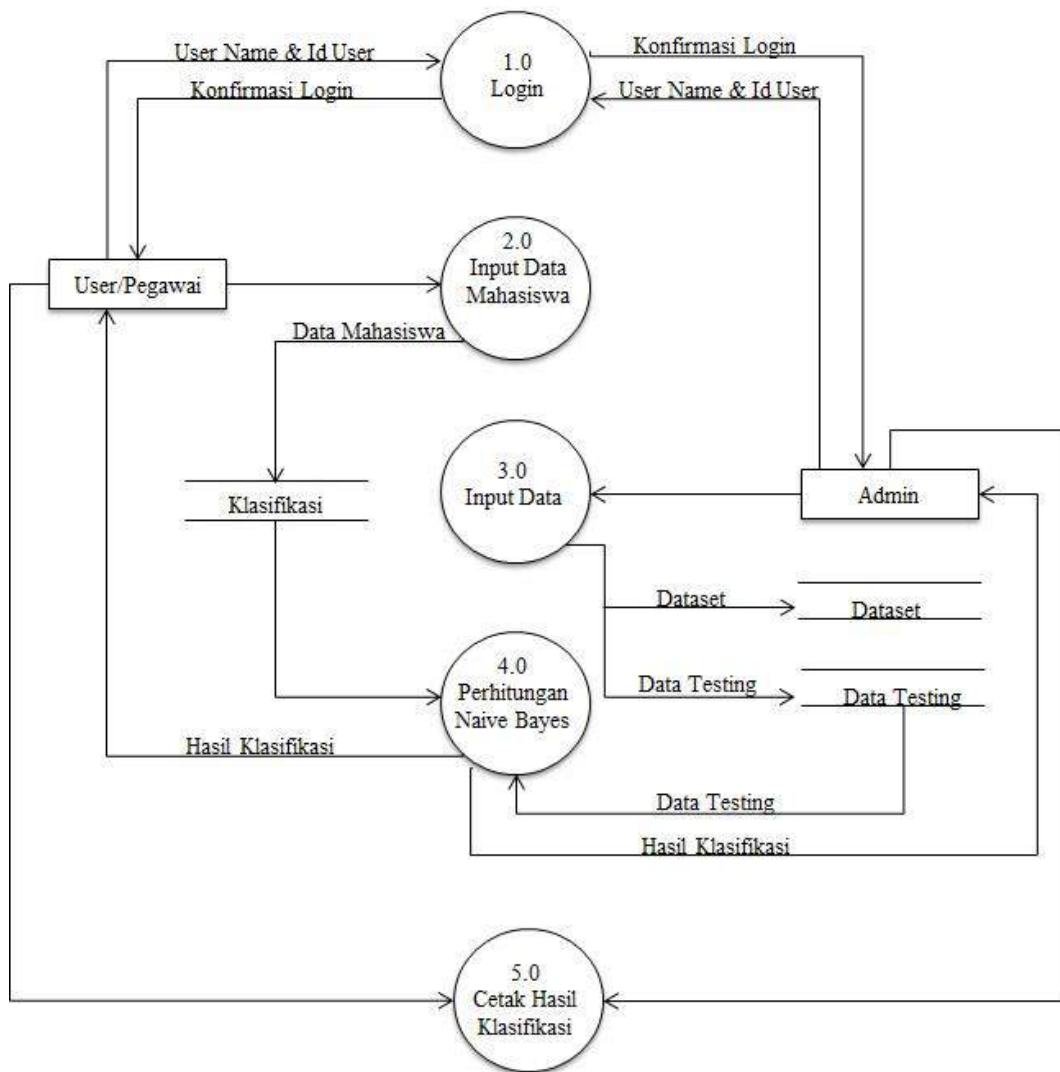
A. Diagram Konteks



Gambar 4. 1 Diagram Konteks

Gambar 4.1 menjelaskan bagaimana alur klasifikasi UKT Universitas Sembilanbelas November Kolaka menggunakan *Naive Bayes* yaitu: *User* menginput data mahasiswa baru dan sistem akan menampilkan hasil klasifikasi sedangkan admin menginput data training dan data uji akurasi.

B. Data Flow Diagram Level 1



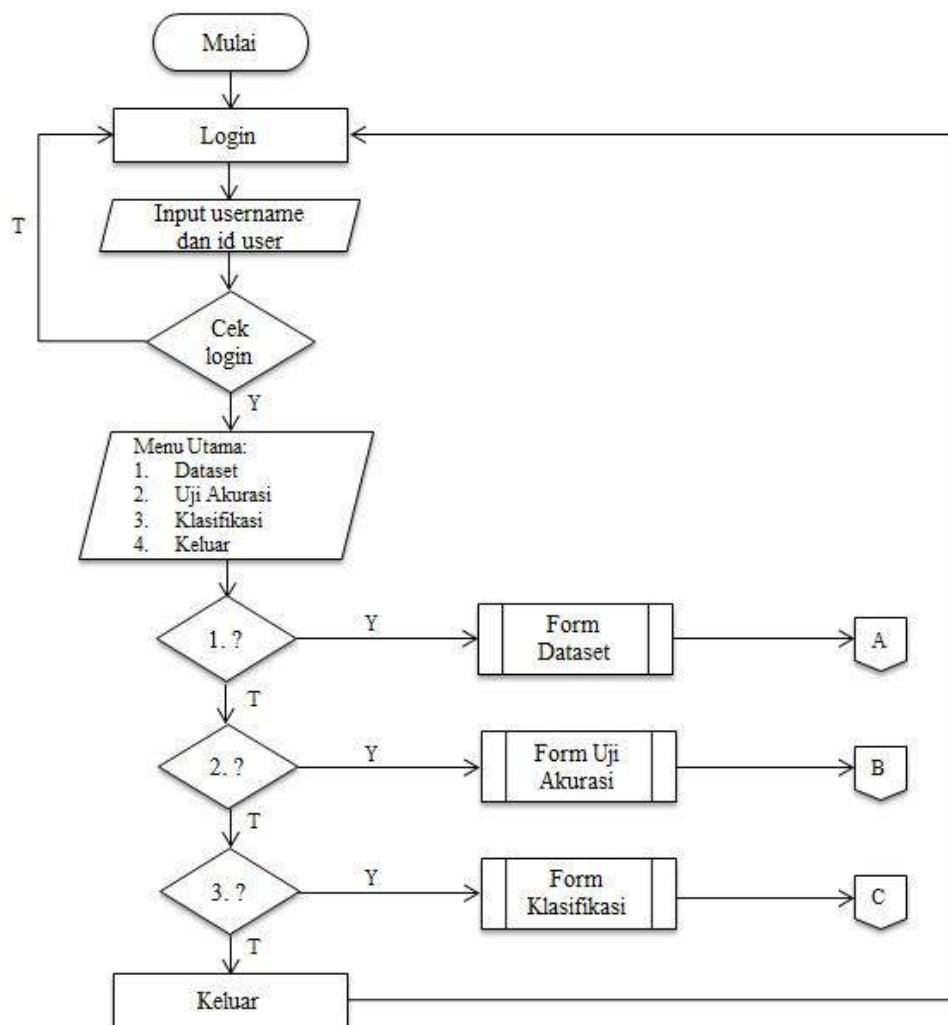
Gambar 4. 2 Data Flow Diagram Level 1

Pada DFD level 1 admin melakukan *login* dengan memasukkan *user name* dan *id user* serta konfirmasi *login* kemudian melakukan *input data dataset* dan *data testing*, sistem akan menghitung menggunakan algoritma *naive bayes* sehingga menampilkan hasil klasifikasi dan *admin* dapat mencetak hasil

klasifikasi sedangkan pegawai *login* dengan memasukan *user name* dan *id user* serta konfirmasi *login* kemudian melakukan *input* data mahasiswa baru sehingga sistem akan menampilkan hasil klasifikasi mahasiswa dan *user* dapat_mencetak hasil klasifikasi.

4.2.2 Flowchart

1. Menu Utama Admin

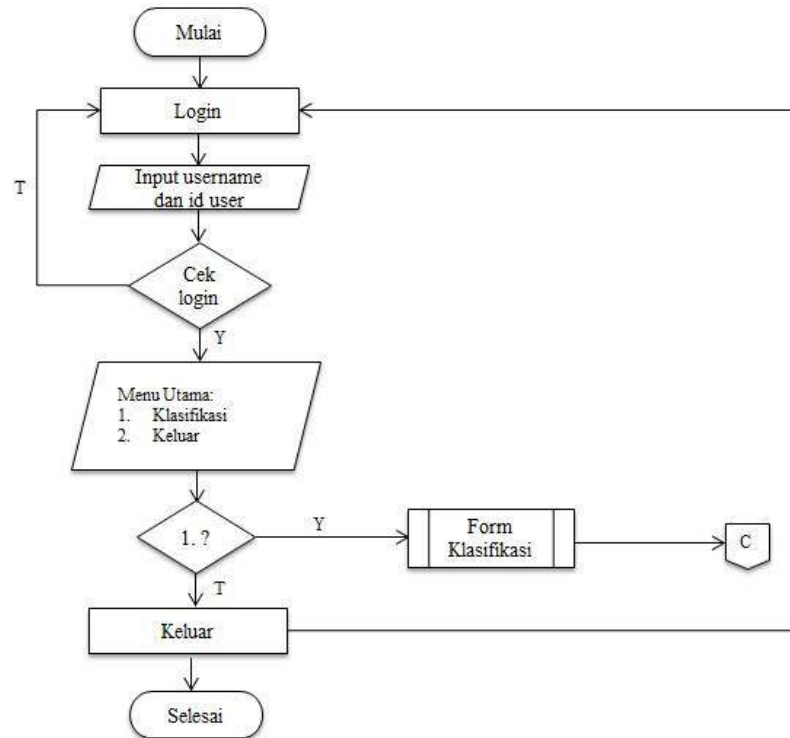


Gambar 4. 3 Flowchart Menu Utama Admin

Pada *Flowchart* Menu Utama Admin dapat dilihat ada 3 pilihan proses. Ketika memilih proses 1 akan ditampilkan *form dataset*. Ketika memilih proses 2 maka akan ditampilkan *form uji akurasi*. Ketika memilih proses 3 maka akan

ditampilkan *form* daftar klasifikasi. Dan jika keluar dari menu utama admin maka pilih *log out*.

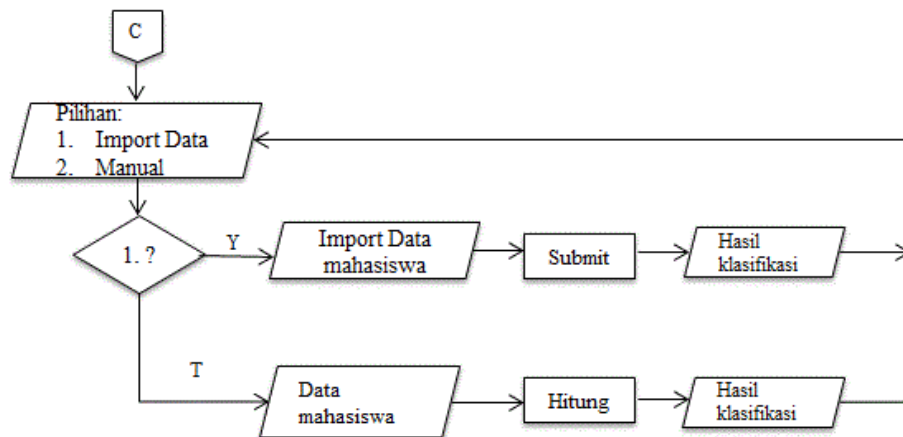
2. Menu Utama Pegawai



Gambar 4. 4 Flowchart Menu Utama Pegawai

Pada *Flowchart* Menu Utama Pegawai dapat dilihat ada 1 pilihan proses. Ketika memilih proses 1 akan ditampilkan *Form* Klasifikasi. Dan jika ingin keluar dari menu utama pegawai maka pilih *log out*.

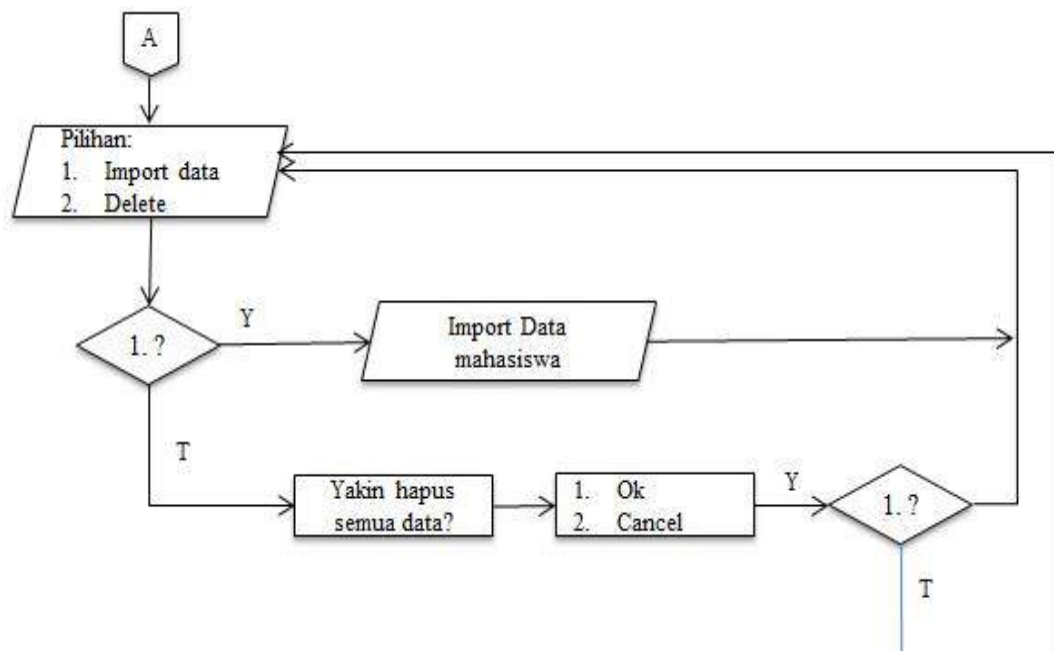
3. Menu Klasifikasi



Gambar 4. 5 Flowchart Menu Klasifikasi

Pada *Flowchart* Menu Utama Klasifikasi terdapat 2 proses, ketika memilih proses 1 maka akan melakukan proses *import* data mahasiswa kemudian akan tampil hasil klasifikasi. Ketika memilih proses 2 maka akan menginput data mahasiswa kemudian hitung sehingga akan tampil hasil klasifikasi.

4. Menu Dataset



Gambar 4. 6 Flowchart Dataset

Pada *flowchart dataset* terdapat 2 proses. Ketika memilih proses 1 maka akan melakukan perintah import data mahasiswa baru. Ketika memilih proses 2

akan muncul tampilan “yakin hapus semua data?” ketika memilih Ok maka di dalam *database* akan terhapus sedangkan ketika memilih *Cancel* Maka akan diarahkan ketampilan menu *dataset*.

4.3 Pemodelan dan Struktur Data

Berisi beberapa bagian yang ada pada *form dataset*, *form* uji akurasi dan *form* hasil klasifikasi diantaranya *field-field*, *length*, *type*, serta keterangan.

Tabel 4. 1 Pemodelan dan Struktur Data

No	<i>Field</i>	<i>Length</i>	<i>Type</i>	Keterangan
1.	jenis_bantuan	10	<i>varchar</i>	Jenis bantuan
2.	jumlah_saudara_kandung	5	<i>varchar</i>	Jumlah saudara kandung
3.	wilayah	5	<i>varchar</i>	Wilayah
4.	kondisi_orang_tua	5	<i>varchar</i>	Kondisi orangtua
5.	pendidikan_ayah	10	<i>varchar</i>	Pendidikan ayah
6.	pekerjaan_ayah	30	<i>varchar</i>	Pekerjaan ayah
7.	pendapatan_ayah	5	<i>varchar</i>	Pendapatan ayah
8.	pendidikan_ibu	10	<i>varchar</i>	Pendidikan ibu
9.	pekerjaan_ibu	30	<i>varchar</i>	Pekerjaan ibu
10.	pendapatan_ibu	5	<i>varchar</i>	Pendapatan ibu

4.4 Model Perhitungan Algoritma *Naive Bayes*

Langkah-langkah dalam menyelesaikan perhitungan dalam klasifikasi Uang Kuliah Tunggal (UKT) menggunakan algoritma *Naive Bayes* meliputi:

Tabel 4. 2 Data Awal

Nama	bantuan	saudara	Wilayah	Kondisi orangtua	Pendidikan ayah	Pekerjaan ayah	Pendapatan Ayah	Pendidikan Ibu	Pekerjaan Ibu	Pendapatan ibu	KL P
PIRDAUS	KIP	7	Bombana	Ayah meninggal	Tamat SD	Tidak Bekerja	. 0	Tamat SD	Lainnya	300.000	I
IRMASARI	KIS	3	Kolaka	Ibu meninggal	Tamat SD	Petani	300.000	Tamat SD	Tidak Bekerja	0	I
YULVIANI	PKH	4	Kolaka	Orangtua hidup	Tamat SMTP	Petani	250.000	Tamat SD	Tidak Bekerja	0	II
SUCIANTI	KIP	2	Kolaka	Orangtua hidup	Tamat SMTA	Lainnya	2.000.000	Tamat SMTA	Tidak Bekerja	0	IV
SUKMA	Tidak Ada	5	Buton	Orangtua hidup	Tamat SMTP	Buruh	500.000	Tamat SD	Tidak Bekerja	0	III
SULKIPLI	KIS	6	Kolaka	Orangtua hidup	Tamat SMTP	Buruh	250.000	Tamat SMTP	Tidak Bekerja	0	II
RITA	Tidak Ada	7	Bombana	Orangtua hidup	Tamat SMTP	Pedagang	500.000	Tamat SD	Lainnya	0	III
RIZALDIN	Tidak Ada	7	Kolaka	Orangtua hidup	Tidak Tamat SD	Lainnya	750.000	Tamat SD	Lainnya	0	III
RONI	KIS	3	Buton	Orangtua hidup	Tamat SD	Nelayan	750.000	Tamat SD	Tidak Bekerja	0	II
SABILA	PKH	3	Kolaka	Orangtua hidup	Tidak Tamat SD	Nelayan	500.000	Tidak Tamat SD	Tidak Bekerja	0	IV
SAENAL	Tidak Ada	2	Kolaka	Orangtua hidup	Tamat SD	Buruh	500.000	Tamat SMTP	Tidak Bekerja	0	III
SAFITRI	Tidak Ada	6	Kolaka	Orangtua hidup	Tamat SD	Nelayan	300.000	Tamat SD	Tidak Bekerja	0	II
SAIFUL	Tidak Ada	5	Konawe	Orangtua hidup	Tamat SD	Nelayan	500.000	Tamat SD	Tidak Bekerja	0	IV
SAIN	Tidak Ada	3	Kolaka	Orangtua hidup	Tamat SMTP	Wiraswasta	500.000	Tamat SMTA	Tidak Bekerja	0	III
SALMATI A	KIP	3	Bombana	Orangtua hidup	Tamat SMTA	Petani	250.000	Tamat SMTP	Petani	0	II
SANDI	KIS	6	Kolaka	Orangtua hidup	Tidak Tamat SD	Petani	250.000	Tidak Tamat SD	Tidak Bekerja	0	II
PUTRI	Tidak Ada	2	Kolaka	Orangtua hidup	Sarjana (S1)	Eksekutif	2.500.000	Diploma 3	Pedagang	2.000.000	V
ANJAR	KIP	3	Kolaka	Orangtua hidup	Sarjana (S1)	PNS Bukan Guru/Dosen	4.000.000	Tamat SD	Tidak Bekerja	0	V
ELMIANA	KIS	2	Kolaka	Orangtua hidup	Tamat SD	Lainnya	250.000	Tamat SMTA	Tidak Bekerja	0	II
SINDI	KIS	5	Kolaka	Orangtua hidup	Tamat SD	Nelayan	500.000	Tamat SMTA	Petani	1.000.000	IV

- **Transformasi Data**

Pada transformasi data atau proses perubahan data kedalam kategori tertentu yang sesuai untuk proses data *mining*. Hasil transformasi data dapat dilihat pada Tabel 4.6

Tabel 4. 3 Hasil Transformasi Data

Kategori		keterangan
Kelompok UKT	I	Rp. 500.000
	II	Rp. 750.000 – Rp. 1.000.000
	III	Rp. 1.000.000 – Rp. 1.800.000
	IV	Rp. 1.250.000 – Rp. 2.250.000
	V	Rp. 2.250.000 – Rp. 2.400.000
Jumlah saudara	I	Jumlah saudara 0 sampai dengan 5
	II	Jumlah saudara 6 sampai dengan 10
	III	Jumlah saudara lebih dari 10
Wilayah	I	Berasal dari kolaka
	II	Berasal dari luar kolaka
Kondisi Orangtua	I	Ayah meninggal
	II	Ibu meninggal
	III	Kedua orangtua hidup
	IV	Kedua orangtua cerai
Pendapatan Ayah/Ibu	I	Tidak mempunyai pendapatan
	II	lebih kecil dari Rp. 500.000 atau sama dengan Rp. 500.000
	III	lebih dari Rp. 500.000 - Rp. 1.000.000
	IV	lebih dari Rp. 1.000.000 - Rp. 1.500.000
	V	lebih dari Rp. 1.500.000 - Rp. 2.000.000
	VI	lebih dari Rp. 2.000.000 - Rp. 2.500.000
	VII	lebih dari Rp. 2.500.000 - Rp. 3.000.000
	VIII	lebih dari Rp. 3.000.000 - Rp. 4.000.000
	IX	lebih dari Rp. 4.000.000 - Rp. 5.000.000
	X	lebih dari Rp. 5.000.000 - Rp. 6.000.000

Dari hasil transformasi data dalam Tabel 4.3 didapatkan hasil dari transformasi data yang ditunjukkan pada Tabel 4.4 yaitu tabel *data training* yang diambil sebanyak 25 *data training* sebagai sampel perhitungan manual.

Tabel 4. 4 Data Training

Nama	bantuan	saudara	Wilayah	Kondisi orangtua	Pendidikan ayah	Pekerjaan ayah	Pendapatan Ayah	Pendidikan Ibu	Pekerjaan Ibu	Pendapatan ibu	Kelompok UKT
PIRDAUS	KIP	II	II	I	Tamat SD	Tidak Bekerja	I	Tamat SD	Lainnya	II	I
IRMASARI	KIS	I	I	II	Tamat SD	Petani/Nelayan	II	Tamat SD	Tidak Bekerja	I	I
YULVIANI	PKH	I	I	III	Tamat SMTP	Petani/Nelayan	II	Tamat SD	Tidak Bekerja	I	II
SUCIANTI	KIP	I	I	III	Tamat SMTA	Lainnya	V	Tamat SMTA	Tidak Bekerja	I	IV
SUKMA	Tidak Ada	I	II	III	Tamat SMTP	Buruh	II	Tamat SD	Tidak Bekerja	I	III
SULKIPLI	KIS	II	I	III	Tamat SMTP	Buruh	II	Tamat SMTP	Tidak Bekerja	I	II
RITA	Tidak Ada	II	II	III	Tamat SMTP	Wiraswasta/ Eksekutif/Pe dagang	II	Tamat SD	Lainnya	I	III
RIZALDIN	Tidak Ada	II	I	III	Tidak Tamat SD	Lainnya	III	Tamat SD	Lainnya	I	III
RONI	KIS	I	II	III	Tamat SD	Petani/Nelayan	III	Tamat SD	Tidak Bekerja	I	II
SABILA	PKH	I	I	III	Tidak Tamat SD	Petani/Nelayan	II	Tidak Tamat SD	Tidak Bekerja	I	IV
SAENAL	Tidak Ada	I	I	III	Tamat SD	Buruh	II	Tamat SMTP	Tidak Bekerja	I	III
SAFITRI	Tidak Ada	II	I	III	Tamat SD	Petani/Nelayan	II	Tamat SD	Tidak Bekerja	I	II
SAIFUL	Tidak Ada	I	II	III	Tamat SD	Petani/Nelayan	II	Tamat SD	Tidak Bekerja	I	IV
SAIN	Tidak Ada	I	I	III	Tamat SMTP	Wiraswasta/ Eksekutif/Pe dagang	II	Tamat SMTA	Tidak Bekerja	I	III
SALMATIA	KIP	I	II	III	Tamat SMTA	Petani/Nelayan	II	Tamat SMTP	Petani/Nelayan	I	II

Tabel 4.3 Lanjutan

SANDI	KIS	II	I	III	Tidak Tamat SD	Petani/Nelayan	II	Tidak Tamat SD	Tidak Bekerja	I	II
PUTRI	Tidak Ada	I	I	III	Sarjana (S1)	Wiraswasta/ Eksekutif/ Pedagang	VI	Diploma 3	0	VI	V
ANJAR	KIP	I	I	III	Sarjana (S1)	Wiraswasta/ Eksekutif/ Pedagang	VIII	Tamat SD	Tidak Bekerja	I	V
ELMIANA	KIS	I	I	III	Tamat SD	Lainnya	II	Tamat SMTA	Tidak Bekerja	I	II
SINDI	KIS	I	I	III	Tamat SD	Petani/Nelayan	II	Tamat SMTA	Petani/Nelayan	II	IV

Tabel 4. 5 Data Testing

Nama	bantuan	saudara	Wilayah	Kondisi orangtua	Pendidikan ayah	Pekerjaan ayah	Pendapatan Ayah	Pendidikan Ibu	Pekerjaan Ibu	Pendapatan ibu	Kelompok UKT
HARMA	KIS	II	I	I	Tamat SMTA	Tidak Bekerja	I	Tamat SD	Petani	II	?
SUCI	Tidak Ada	II	I	III	Tamat SMTP	Wiraswasta/ Eksekutif/ Pedagang	II	Tamat SD	Tidak Bekerja	I	?
SUHARMA	Tidak Ada	I	I	III	Tamat SD	Petani/Nelayan	II	Tamat SD	Tidak Bekerja	I	?
RAYHAN	Tidak Ada	I	I	III	Sarjana (S1)	Wiraswasta/ Eksekutif/ Pedagang	VIII	Tamat SMTA	Tidak Bekerja	I	?

Tahapan dalam proses perhitungan *naive bayes* untuk lima data mahasiswa adalah sebagai berikut:

1) Menghitung nilai *class log prior*

I	= $\log(2/20)$	= -2,302
II	= $\log(7/20)$	= -1,049
III	= $\log(5/20)$	= -1,386
IV	= $\log(4/20)$	= -1.609
V	= $\log(2/20)$	= -2.302

2) Menghitung nilai *cat count* dan *class count*

- *cat count* jenis bantuan

Jenis bantuan = "KIP" | "I", "II", "III", "IV", "V"

I = 1+1=2, II = 1+1 = 2, III = 0+1 = 1, IV = 1+1 = 2, V = 1+1 = 2

Jenis bantuan = "KIS" | "I", "II", "III", "IV", "V"

I = 1+1=2, II = 4+1 = 5, III = 0+1 = 1, IV = 1+1 = 2, V = 0+1 = 1

Jenis bantuan = "PKH" | "I", "II", "III", "IV", "V"

I = 0+1=1, II = 1+1 = 2, III = 0+1 = 1, IV = 1+1 = 2, V = 0+1 = 1

Jenis bantuan = "Tidak ada" | "I", "II", "III", "IV", "V"

I = 0+1=1, II = 1+1 = 2, III = 5+1 = 6, IV = 1+1 = 2, V = 1+1 = 2

- *class count* jenis bantuan

I = 6, II = 11, III = 9, IV = 8, V = 6

- *cat count* jumlah saudara

Jumlah saudara = "I" | "I", "II", "III", "IV", "V"

I = 1+1=2, II = 5+1 = 6, III = 3+1 = 4, IV = 4+1 = 5, V = 2+1 = 3

Jumlah saudara = "II" | "I", "II", "III", "IV", "V"

I = 1+1=2, II = 4+1 = 3, III = 2+1 = 3, IV = 0+1 = 1, V = 0+1 = 1

- *class count* jumlah saudara

I = 4, II = 9, III = 7, IV = 6, V = 4

- *cat count* wilayah

wilayah = "I" | "I", "II", "III", "IV", "V"

I = 1+1=2, II = 5+1 = 6, III = 3+1 = 4, IV = 3+1 = 4, V = 2+1 = 3

wilayah = "II" | "I", "II", "III", "IV", "V"

$I = 1+1=2$, $II = 2+1 = 3$, $III = 2+1 = 3$, $IV = 1+1 = 2$, $V = 1+1 = 1$

- *class count* wilayah

$I = 4$, $II = 9$, $III = 7$, $IV = 6$, $V = 4$

- *cat count* kondisi orang tua

kondisi orang tua = "I" | "I", "II", "III", "IV", "V"

$I = 1+1=2$, $II = 0+1 = 1$, $III = 0+1 = 1$, $IV = 0+1 = 1$, $V = 0+1 = 1$

kondisi orang tua = "II" | "I", "II", "III", "IV", "V"

$I = 1+1=2$, $II = 0+1 = 1$, $III = 0+1 = 1$, $IV = 0+1 = 1$, $V = 0+1 = 1$

kondisi orang tua = "III" | "I", "II", "III", "IV", "V"

$I = 0+1=1$, $II = 7+1 = 8$, $III = 5+1 = 6$, $IV = 4+1 = 5$, $V = 2+1 = 3$

- *class count* kondisi orang tua

$I = 5$, $II = 10$, $III = 8$, $IV = 7$, $V = 5$

- *cat count* pendidikan ayah

pendidikan ayah = "sarjana (S1)" | "I", "II", "III", "IV", "V"

$I = 0+1=1$, $II = 0+1 = 1$, $III = 0+1 = 1$, $IV = 0+1 = 1$, $V = 2+1 = 3$

pendidikan ayah = "tamat SD" | "I", "II", "III", "IV", "V"

$I = 2+1=3$, $II = 3+1 = 4$, $III = 1+1 = 2$, $IV = 2+1 = 3$, $V = 0+1 = 1$

pendidikan ayah = "tamat SMTA" | "I", "II", "III", "IV", "V"

$I = 0+1=1$, $II = 1+1 = 2$, $III = 0+1 = 1$, $IV = 1+1 = 2$, $V = 0+1 = 1$

pendidikan ayah = "tamat SMTP" | "I", "II", "III", "IV", "V"

$I = 0+1=1$, $II = 2+1 = 3$, $III = 3+1 = 4$, $IV = 0+1 = 1$, $V = 0+1 = 1$

pendidikan ayah = "tidak tamat SD" | "I", "II", "III", "IV", "V"

$I = 0+1=1$, $II = 1+1 = 2$, $III = 1+1 = 2$, $IV = 1+1 = 2$, $V = 0+1 = 0$

- *class count* pendidikan ayah

$I = 7$, $II = 12$, $III = 10$, $IV = 9$, $V = 7$

- *cat count* pekerjaan ayah

pekerjaan ayah = "buruh" | "I", "II", "III", "IV", "V"

$I = 0+1=1$, $II = 1+1 = 2$, $III = 2+1 = 3$, $IV = 0+1 = 1$, $V = 0+1 = 1$

pekerjaan ayah = "lainnya" | "I", "II", "III", "IV", "V"

$I = 0+1=1$, $II = 1+1 = 2$, $III = 1+1 = 2$, $IV = 1+1 = 2$, $V = 0+1 = 1$

pekerjaan ayah = “petani/nelayan” | “I”, “II”, “III”, “IV”, “V”

$I = 1+1 = 2$, $II = 5+1 = 6$, $III = 0+1 = 1$, $IV = 3+1 = 4$, $V = 0+1 = 1$

pekerjaan ayah = “tidak bekerja” | “I”, “II”, “III”, “IV”, “V”

$I = 1+1=2$, $II = 0+1 = 1$, $III = 0+1 = 1$, $IV = 0+1 = 1$, $V = 0+1 = 1$

pekerjaan ayah = “wiraswasta/eksekutif/pedagang” | “I”, “II”, “III”, “IV”, “V”

$I = 0+1=1$, $II = 0+1 = 1$, $III = 2+1 = 3$, $IV = 0+1 = 1$, $V = 2+1 = 3$

- *class count* pendidikan ayah

$I = 7$, $II = 12$, $III = 10$, $IV = 9$, $V = 7$

- *cat count* pendapatan ayah

pendapatan ayah = “II” | “I”, “II”, “III”, “IV”, “V”

$I = 1+1=2$, $II = 6+1 = 7$, $III = 4+1 = 5$, $IV = 3+1 = 4$, $V = 0+1 = 1$

pendapatan ayah = “I” | “I”, “II”, “III”, “IV”, “V”

$I = 1+1=2$, $II = 0+1 = 1$, $III = 0+1 = 1$, $IV = 0+1 = 1$, $V = 0+1 = 1$

pendapatan ayah = “V” | “I”, “II”, “III”, “IV”, “V”

$I = 0+1 = 1$, $II = 0+1 = 1$, $III = 0+1 = 1$, $IV = 1+1 = 2$, $V = 0+1 = 1$

pendapatan ayah = “VI” | “I”, “II”, “III”, “IV”, “V”

$I = 0+1 = 1$, $II = 0+1 = 1$, $III = 0+1 = 1$, $IV = 0+1 = 1$, $V = 1+1 = 2$

pendapatan ayah = “VIII” | “I”, “II”, “III”, “IV”, “V”

$I = 0+1 = 1$, $II = 0+1 = 1$, $III = 0+1 = 1$, $IV = 0+1 = 1$, $V = 1+1 = 2$

pendapatan ayah = “III” | “I”, “II”, “III”, “IV”, “V”

$I = 0+1=1$, $II = 1+1 = 2$, $III = 1+1 = 2$, $IV = 0+1 = 1$, $V = 0+1 = 1$

- *class count* pendapatan ayah

$I = 8$, $II = 13$, $III = 11$, $IV = 10$, $V = 8$

- *cat count* pendidikan ibu

pendidikan ayah = “diploma 3” | “I”, “II”, “III”, “IV”, “V”

$I = 0+1=1$, $II = 0+1 = 1$, $III = 0+1 = 1$, $IV = 0+1 = 1$, $V = 1+1 = 2$

pendidikan ibu = “tamat SD” | “I”, “II”, “III”, “IV”, “V”

$I = 2+1=3$, $II = 3+1 = 4$, $III = 3+1 = 4$, $IV = 1+1 = 2$, $V = 1+1 = 2$

pendidikan ibu = “tamat SMTA” | “I”, “II”, “III”, “IV”, “V”

$I = 0+1=1$, $II = 1+1 = 2$, $III = 1+1 = 2$, $IV = 2+1 = 3$, $V = 0+1 = 1$

pendidikan ibu = “tamat SMTP” | “I”, “II”, “III”, “IV”, “V”

$$I = 0+1=1, II = 2+1 = 3, III = 1+1 = 2, IV = 0+1 = 1, V = 0+1 = 1$$

pendidikan ibu = “tidak tamat SD” | “I”, “II”, “III”, “IV”, “V”

$$I = 0+1=1, II = 1+1 = 2, III = 0+1 = 1, IV = 1+1 = 2, V = 0+1 = 0$$

- *class count* pendidikan ibu

$$I = 7, II = 12, III = 10, IV = 9, V = 7$$

- *cat count* pekerjaan ibu

pekerjaan ibu = “lainnya” | “I”, “II”, “III”, “IV”, “V”

$$I = 1+1=2, II = 0+1 = 1, III = 2+1 = 3, IV = 0+1 = 1, V = 0+1 = 1$$

pekerjaan ibu = “petani/nelayan” | “I”, “II”, “III”, “IV”, “V”

$$I = 0+1 = 2, II = 1+1 = 2, III = 0+1 = 1, IV = 1+1 = 2, V = 0+1 = 1$$

pekerjaan ibu = “tidak bekerja” | “I”, “II”, “III”, “IV”, “V”

$$I = 1+1=2, II = 6+1 = 7, III = 3+1 = 4, IV = 3+1 = 4, V = 1+1 = 2$$

pekerjaan ibu = “wiraswasta/eksekutif/pedagang” | “I”, “II”, “III”, “IV”, “V”

$$I = 0+1=1, II = 0+1 = 1, III = 0+1 = 1, IV = 0+1 = 1, V = 1+1 = 2$$

- *class count* pendidikan ibu

$$I = 6, II = 11, III = 9, IV = 8, V = 6$$

- *cat count* pendapatan ibu

- pendapatan ibu = “II” | “I”, “II”, “III”, “IV”, “V”

$$I = 1+1=2, II = 0+1 = 1, III = 0+1 = 1, IV = 1+1 = 2, V = 0+1 = 1$$

pendapatan ibu = “I” | “I”, “II”, “III”, “IV”, “V”

$$I = 1+1=2, II = 7+1 = 8, III = 5+1 = 6, IV = 3+1 = 4, V = 1+1 = 2$$

pendapatan ibu = “V” | “I”, “II”, “III”, “IV”, “V”

$$I = 0+1 = 1, II = 0+1 = 1, III = 0+1 = 1, IV = 0+1 = 1, V = 1+1 = 2$$

- *class count* pendidikan ibu

$$I = 5, II = 10, III = 8, IV = 7, V = 5$$

3. Menghitung nilai *smoothed cat count* & *smoothed class count*

- *smoothed cat count* jenis bantuan

Jenis bantuan = “KIP” | “I”, “II”, “III”, “IV”, “V”

$$I = \log(2) = 0,69, II = \log(2) = 0,69, III = \log(1) = 0, IV = \log(2) = 0,69,$$

$$V = \log(2) = 0,69$$

Jenis bantuan = “KIS” | “I”, “II”, “III”, “IV”, “V”

$I = \log(2) = 0,69$, $II = \log(5) = 1,60$, $III = \log(1) = 0$, $IV = \log(2) = 0,69$, $V = \log(1) = 0$

Jenis bantuan = "PKH" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(1) = 0,69$, $III = \log(1) = 10$, $IV = \log(2) = 0,69$, $V = \log(1) = 0$

Jenis bantuan = "Tidak ada" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(2) = 0,69$, $III = \log(6) = 1,79$, $IV = \log(2) = 0,69$, $V = \log(2) = 0,69$

- *class count* jenis bantuan

$I = \log(6) = 1,79$, $II = \log(11) = 2,39$, $III = \log(92) = 2,19$, $IV = \log(8) = 2,07$, $V = \log(6) = 1,79$

- *smoothed cat count* jumlah saudara

Jumlah saudara = "I" | "I", "II", "III", "IV", "V"

$I = \log(2) = 0,69$, $II = \log(6) = 1,79$, $III = \log(4) = 1,38$, $IV = \log(5) = 1,60$, $V = \log(3) = 1,09$

Jumlah saudara = "II" | "I", "II", "III", "IV", "V"

$I = \log(2) = 0,69$, $II = \log(3) = 1,09$, $III = \log(3) = 1,09$, $IV = \log(1) = 0$, $V = \log(1) = 0$

- *smoothed class count* jumlah saudara

$I = \log(4) = 1,38$, $II = \log(9) = 2,19$, $III = \log(7) = 1,94$, $IV = \log(6) = 1,79$, $V = \log(4) = 1,38$

- *smoothed cat count* wilayah

wilayah = "I" | "I", "II", "III", "IV", "V"

$I = \log(2) = 0,69$, $II = \log(6) = 1,79$, $III = \log(4) = 1,38$, $IV = \log(4) = 1,38$, $V = \log(3) = 1,09$

wilayah = "II" | "I", "II", "III", "IV", "V"

$I = \log(2) = 0,69$, $II = \log(3) = 1,09$, $III = \log(3) = 1,09$, $IV = \log(2) = 0,69$, $V = \log(1) = 0$

- *smoothed class count* wilayah

$I = \log(4) = 1,38$, $II = \log(9) = 2,19$, $III = \log(7) = 1,94$, $IV = \log(6) = 1,79$, $V = \log(4) = 1,38$

- *smoothed cat count* kondisi orang tua

kondisi orang tua = "I" | "I", "II", "III", "IV", "V"

$I = \log(2) = 0,69$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(1) = 0$

kondisi orang tua = "II" | "I", "II", "III", "IV", "V"

$I = \log(2) = 0,69$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(1) = 0$

kondisi orang tua = "III" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(8) = 2,07$, $III = \log(6) = 1,79$, $IV = \log(5) = 1,60$, $V = \log(3) = 1,09$

- *smoothed class count* kondisi orang tua

$I = \log(5) = 1,60$, $II = \log(10) = 2,30$, $III = \log(8) = 2,07$, $IV = \log(7) = 1,94$, $V = \log(5) = 1,60$

- *smoothed cat count* pendidikan ayah

pendidikan ayah = "sarjana (S1)" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(3) = 1,09$

pendidikan ayah = "tamat SD" | "I", "II", "III", "IV", "V"

$I = \log(3) = 1,09$, $II = \log(4) = 1,38$, $III = \log(2) = 0,69$, $IV = \log(3) = 1,09$, $V = \log(1) = 0$

pendidikan ayah = "tamat SMTA" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(2) = 0,69$, $III = \log(1) = 0$, $IV = \log(2) = 0,69$, $V = \log(1) = 0$

pendidikan ayah = "tamat SMTP" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(3) = 1,09$, $III = \log(4) = 1,38$, $IV = \log(1) = 0$, $V = \log(1) = 0$

pendidikan ayah = "tidak tamat SD" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(2) = 0,69$, $III = \log(2) = 0,69$, $IV = \log(2) = 0,69$, $V = \log(1) = 0$

- *smoothed class count* pendidikan ayah

$I = \log(7) = 1,94$, $II = \log(12) = 2,48$, $III = \log(10) = 2,30$, $IV = \log(9) = 1,19$, $V = \log(7) = 1,94$

- *smoothed cat count* pekerjaan ayah

pekerjaan ayah = “buruh” | “I”, “II”, “III”, “IV”, “V”

$I = \log(1) = 0$, $II = \log(2) = 0,69$, $III = \log(3) = 1,09$, $IV = \log(1) = 0$, $V = \log(1) = 0$

pekerjaan ayah = “lainnya” | “I”, “II”, “III”, “IV”, “V”

$I = \log(1) = 0$, $II = \log(2) = 0,69$, $III = \log(2) = 0,69$, $IV = \log(2) = 0,69$, $V = \log(1) = 0$

pekerjaan ayah = “petani/nelayan” | “I”, “II”, “III”, “IV”, “V”

$I = \log(2) = 0,69$, $II = \log(6) = 1,79$, $III = \log(1) = 0$, $IV = \log(4) = 1,38$, $V = \log(1) = 0$

pekerjaan ayah = “tidak bekerja” | “I”, “II”, “III”, “IV”, “V”

$I = \log(2) = 0,69$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(1) = 0$

pekerjaan ayah = “wiraswasta/eksekutif/pedagang” | “I”, “II”, “III”, “IV”, “V”

$I = \log(1) = 0$, $II = \log(1) = 0$, $III = \log(3) = 1,09$, $IV = \log(1) = 0$, $V = \log(3) = 1,09$

- *smoothed class count* pendidikan ayah

$I = \log(7) = 1,94$, $II = \log(12) = 2,48$, $III = \log(10) = 2,30$, $IV = \log(9) = 2,19$, $V = \log(7) = 1,94$

- *smoothed cat count* pendapatan ayah

pendapatan ayah = “II” | “I”, “II”, “III”, “IV”, “V”

$I = \log(2) = 0,69$, $II = \log(7) = 1,94$, $III = \log(5) = 1,60$, $IV = \log(4) = 1,38$, $V = \log(1) = 0$

pendapatan ayah = “I” | “I”, “II”, “III”, “IV”, “V”

$I = \log(2) = 0,69$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(1) = 0$

pendapatan ayah = “V” | “I”, “II”, “III”, “IV”, “V”

$I = \log(1) = 1$, $II = \log(1) = 1$, $III = \log(1) = 1$, $IV = \log(2) = 2$, $V = \log(1) = 1$

pendapatan ayah = “VI” | “I”, “II”, “III”, “IV”, “V”

$I = \log(1) = 0$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(2) = 0,69$

pendapatan ayah = "VIII" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(2) = 0,69$

pendapatan ayah = "III" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(2) = 0,69$, $III = \log(2) = 0,69$, $IV = \log(1) = 0$, $V = \log(1) = 0$

- *smoothed class count* pendapatan ayah

$I = \log(8) = 2,07$, $II = \log(13) = 2,56$, $III = \log(11) = 2,39$, $IV = \log(10) = 2,30$, $V = \log(8) = 2,07$

- *smoothed cat count* pendidikan ibu

pendidikan ibu = "diploma 3" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(2) = 0,69$

pendidikan ibu = "tamat SD" | "I", "II", "III", "IV", "V"

$I = \log(3) = 1,09$, $II = \log(4) = 1,38$, $III = \log(4) = 1,38$, $IV = \log(2) = 0,69$, $V = \log(2) = 0,69$

pendidikan ibu = "tamat SMTA" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(2) = 0,69$, $III = \log(2) = 0,69$, $IV = \log(3) = 1,09$, $V = \log(1) = 0$

pendidikan ibu = "tamat SMTP" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(3) = 1,09$, $III = \log(2) = 0,69$, $IV = \log(1) = 0$, $V = \log(1) = 0$

pendidikan ibu = "tidak tamat SD" | "I", "II", "III", "IV", "V"

$I = \log(1) = 0$, $II = \log(2) = 0,69$, $III = \log(1) = 0$, $IV = \log(2) = 0,69$, $V = \log(1) = 0$

- *smoothed class count* pendidikan ibu

$I = \log(7) = 1,94$, $II = \log(12) = 2,48$, $III = \log(10) = 2,30$, $IV = \log(9) = 2,19$, $V = \log(7) = 1,94$

- *smoothed cat count* pekerjaan ibu

pekerjaan ibu = "lainnya" | "I", "II", "III", "IV", "V"

$I = \log(2) = 0,69$, $II = \log(1) = 0$, $III = \log(3) = 1,09$, $IV = \log(1) = 0$, $V = \log(1) = 0$

pekerjaan ibu = “petani/nelayan” | “I”, “II”, “III”, “IV”, “V”

$I = \log(2) = 0,69$, $II = \log(2) = 0,69$, $III = \log(1) = 0$, $IV = \log(2) = 0,69$, $V = \log(1) = 0$

pekerjaan ibu = “tidak bekerja” | “I”, “II”, “III”, “IV”, “V”

$I = \log(2) = 0,69$, $II = \log(7) = 1,94$, $III = \log(4) = 1,38$, $IV = \log(4) = 1,38$, $V = \log(2) = 0,69$

pekerjaan ibu = “wiraswasta/eksekutif/pedagang” | “I”, “II”, “III”, “IV”, “V”

$I = \log(1) = 0$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(2) = 0,69$

- *smoothed class count* pekerjaan ibu

$I = \log(6) = 1,79$, $II = \log(11) = 2,39$, $III = \log(9) = 2,19$, $IV = \log(8) = 2,07$, $V = \log(6) = 1,79$

- *smoothed cat count* pendapatan ibu

pendapatan ibu = “II” | “I”, “II”, “III”, “IV”, “V”

$I = \log(2) = 0,69$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(2) = 0,69$, $V = \log(1) = 0$

pendapatan ibu = “I” | “I”, “II”, “III”, “IV”, “V”

$I = \log(2) = 0,69$, $II = \log(8) = 2,07$, $III = \log(6) = 1,79$, $IV = \log(4) = 1,38$, $V = \log(2) = 0,69$

pendapatan ibu = “V” | “I”, “II”, “III”, “IV”, “V”

$I = \log(1) = 0$, $II = \log(1) = 0$, $III = \log(1) = 0$, $IV = \log(1) = 0$, $V = \log(2) = 0,69$

- *class count* pendapatan ibu

$I = \log(5) = 1,60$, $II = \log(10) = 2,30$, $III = \log(8) = 2,07$, $IV = \log(7) = 1,94$, $V = \log(5) = 1,60$

4. Menghitung nilai *feature log prob*

- Data *test* dengan nama Harma

Jenis bantuan = “KIS” | “I”, “II”, “III”, “IV”, “V”

$I = 0,69 - 1,79 = -1,09$, $II = 1,60 - 2,39 = -0,78$, $III = 0 - 2,19 =$

$$-2,19, IV = 0,69 - 2,07 = -1,38, V = 0 - 1,79 = -1,79$$

$$\text{Jumlah saudara} = \text{"II"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 0,69 - 1,38 = -0,69, II = 1,09 - 2,19 = -1,09, III = 1,09 - 1,94 = -0,84, IV = 0 - 1,79 = -1,79, V = 0 - 1,38 = -1,38$$

$$\text{wilayah} = \text{"I"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 0,69 - 1,38 = -0,69, II = 1,79 - 2,19 = -0,40, III = 1,38 - 1,94 = -0,55, IV = 1,38 - 1,79 = -0,40, V = 1,09 - 1,38 = -0,28$$

$$\text{kondisi orang tua} = \text{"I"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 0,69 - 1,60 = -0,91, II = 0 - 2,30 = -2,30, III = 0 - 2,07 = -2,07, IV = 0 - 1,94 = -1,94, V = 0 - 1,60 = -1,60$$

$$\text{pendidikan ayah} = \text{"tamam SMTA"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 0 - 1,94 = -1,94, II = 0,69 - 2,48 = -1,79, III = 0 - 2,30 = -2,30, IV = 0,69 - 2,19 = -1,50, V = 0 - 1,94 = -1,94$$

$$\text{pekerjaan ayah} = \text{"tidak bekerja"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 0,69 - 1,94 = -1,25, II = 0 - 2,48 = -2,48, III = 0 - 2,30 = -2,30, V = 0 - 2,19 = -2,19, V = 0 - 1,94 = -1,94$$

$$\text{pendapatan ayah} = \text{"I"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 0,69 - 2,07 = -1,38, II = 0 - 2,56 = -2,56, III = 0 - 2,39 = -2,39, IV = 0 - 2,30 = -2,30, V = 0 - 2,07 = -2,07$$

$$\text{pendidikan ibu} = \text{"tamam SD"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 1,09 - 1,94 = -0,84, II = 1,38 - 2,48 = -1,09, III = 1,38 - 2,30 = -0,91, IV = 0,69 - 2,19 = -1,50, V = 0,69 - 1,94 = -1,25$$

$$\text{pekerjaan ibu} = \text{"petani/nelayan"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 0,69 - 1,79 = -1,79, II = 0,69 - 2,39 = -1,70, III = 0 - 2,19 = -2,19, IV = 0,69 - 2,07 = -1,38, V = 0 - 1,79 = -1,79$$

$$\text{pendapatan ibu} = \text{"II"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 0,69 - 1,60 = -0,91, II = 0 - 2,30 = -2,30, III = 0 - 2,07 = -2,07, IV = 0,69 - 1,94 = -1,25, V = 0 - 1,60 = -1,60$$

b. Data *test* dengan nama Suci

$$\text{Jenis bantuan} = \text{"Tidak ada"} \mid \text{"I"}, \text{"II"}, \text{"III"}, \text{"IV"}, \text{"V"}$$

$$I = 0 - 1,79 = -1,79, II = 0,69 - 2,39 = -1,70, III = 1,79 - 2,19 = -0,40, IV = 0,69 - 2,07 = -1,38, V = 0,69 - 1,79 = -1,09$$

$$\text{Jumlah saudara} = \text{"II"} \mid \text{"I", "II", "III", "IV", "V"}$$

$$I = 0,69 - 1,38 = -0,69, II = 1,09 - 2,19 = -1,09, III = 1,09 - 1,94 = -0,84, IV = 0 - 1,79 = -1,79, V = 0 - 1,38 = -1,38$$

$$\text{wilayah} = \text{"I"} \mid \text{"I", "II", "III", "IV", "V"}$$

$$I = 0,69 - 1,38 = -0,69, II = 1,79 - 2,19 = -0,40, III = 1,38 - 1,94 = -0,55, IV = 1,38 - 1,79 = -0,40, V = 1,09 - 1,38 = -0,28$$

$$\text{kondisi orang tua} = \text{"III"} \mid \text{"I", "II", "III", "IV", "V"}$$

$$I = 0 - 1,60 = -1,60, II = 2,07 - 2,30 = -0,22, III = 1,79 - 2,07 = -0,28, IV = 1,60 - 1,94 = -0,33, V = 1,09 - 1,60 = -0,51$$

$$\text{pendidikan ayah} = \text{"tamat SMTP"} \mid \text{"I", "II", "III", "IV", "V"}$$

$$I = 0 - 1,94 = -1,94, II = 1,09 - 2,48 = -1,38, III = 1,38 - 2,30 = -0,91, IV = 0 - 2,19 = -2,19, V = 0 - 1,94 = -1,94$$

$$\text{pekerjaan ayah} = \text{"wiraswasta/eksekutif/pedagang"} \mid \text{"I", "II", "III", "IV", "V"}$$

$$I = 0 - 1,94 = -1,94, II = 0 - 2,48 = -2,48, III = 1,09 - 2,30 = -1,20, IV = 0 - 2,19 = -2,19, V = 1,09 - 1,94 = -0,84$$

$$\text{pendapatan ayah} = \text{"Lebih kecil sama dengan Rp. 500.000"} \mid \text{"I", "II", "III", "IV", "V"}$$

$$I = 0,69 - 2,07 = -1,38, II = 1,94 - 2,56 = -0,61, III = 1,60 - 2,39 = -0,78, IV = 1,38 - 2,30 = -0,91, V = 0 - 2,07 = -2,07$$

$$\text{pendidikan ibu} = \text{"tamat SD"} \mid \text{"I", "II", "III", "IV", "V"}$$

$$I = 1,09 - 1,94 = -0,84, II = 1,38 - 2,48 = -1,09, III = 1,38 - 2,30 = -0,91, IV = 0,69 - 2,19 = -1,50, V = 0,69 - 1,94 = -1,25$$

$$\text{pekerjaan ibu} = \text{"tidak bekerja"} \mid \text{"I", "II", "III", "IV", "V"}$$

$$I = 0,69 - 1,79 = -1,09, II = 1,94 - 2,39 = -0,45, III = 1,38 - 2,19 = -0,81, IV = 1,38 - 2,07 = -0,69, V = 0,69 - 1,79 = -1,09$$

$$\text{pendapatan ibu} = \text{"Rp. 0 / Tidak Punya Penghasilan"} \mid \text{"I", "II", "III", "IV", "V"}$$

$$I = 0,69 - 1,94 = -0,91, II = 2,07 - 2,48 = -0,22, III = 1,79 - 2,30 = -0,28, IV = 1,38 - 2,19 = -0,55, V = 0,69 - 1,94 = -0,91$$

c. Data *test* dengan nama Suharma

Jenis bantuan = “Tidak ada” | “I”, “II”, “III”, “IV”, “V”

$$I = 0 - 1,79 = -1,79, \text{ II} = 0,69 - 2,39 = -1,70, \text{ III} = 1,79 - 2,19 = -0,40, \text{ IV} = 0,69 - 2,07 = -1,38, \text{ V} = 0,69 - 1,79 = -1,09$$

Jumlah saudara = “I” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 1,38 = -0,69, \text{ II} = 1,79 - 2,19 = -0,40, \text{ III} = 1,38 - 1,94 = -0,55, \text{ IV} = 1,60 - 1,79 = -0,18, \text{ V} = 1,09 - 1,38 = -0,28$$

wilayah = “I” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 1,38 = -0,69, \text{ II} = 1,79 - 2,19 = -0,40, \text{ III} = 1,38 - 1,94 = -0,55, \text{ IV} = 1,38 - 1,79 = -0,40, \text{ V} = 1,09 - 1,38 = -0,28$$

kondisi orang tua = “III” | “I”, “II”, “III”, “IV”, “V”

$$I = 0 - 1,60 = -1,60, \text{ II} = 2,07 - 2,30 = -0,22, \text{ III} = 1,79 - 2,07 = -0,28, \text{ IV} = 1,60 - 1,94 = -0,33, \text{ V} = 1,09 - 1,60 = -0,51$$

pendidikan ayah = “tamat SD” | “I”, “II”, “III”, “IV”, “V”

$$I = 1,09 - 1,94 = -0,84, \text{ II} = 1,38 - 2,48 = -1,09, \text{ III} = 0,69 - 2,30 = -1,60, \text{ IV} = 1,09 - 2,19 = -1,09, \text{ V} = 0 - 1,94 = -1,94$$

pekerjaan ayah = “petani/nelayan” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 1,94 = -1,25, \text{ II} = 1,79 - 2,48 = -0,69, \text{ III} = 0 - 2,30 = -2,30, \text{ IV} = 1,38 - 2,19 = -0,81, \text{ V} = 0 - 1,94 = -1,94$$

pendapatan ayah = “II” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 2,07 = -1,38, \text{ II} = 1,94 - 2,56 = -0,61, \text{ III} = 1,60 - 2,39 = -0,78, \text{ IV} = 1,38 - 2,30 = -0,91, \text{ V} = 0 - 2,07 = -2,07$$

pendidikan ibu = “tamat SD” | “I”, “II”, “III”, “IV”, “V”

$$I = 1,09 - 1,94 = -0,84, \text{ II} = 1,38 - 2,48 = -1,09, \text{ III} = 1,38 - 2,30 = -0,91, \text{ IV} = 0,69 - 2,19 = -1,50, \text{ V} = 0,69 - 1,94 = -1,25$$

pekerjaan ibu = “tidak bekerja” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 1,79 = -1,09, \text{ II} = 1,94 - 2,39 = -0,45, \text{ III} = 1,38 - 2,19 = -0,81, \text{ IV} = 1,38 - 2,07 = -0,69, \text{ V} = 0,69 - 1,79 = -1,09$$

pendapatan ibu = “I” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 1,94 = -0,91, \text{ II} = 2,07 - 2,48 = -0,22, \text{ III} = 1,79 - 2,30 = -0,28, \text{ IV} = 1,38 - 2,19 = -0,55, \text{ V} = 0,69 - 1,94 = -0,91$$

d. Data *test* dengan nama Rayhan

Jenis bantuan = “Tidak ada” | “I”, “II”, “III”, “IV”, “V”

$$I = 0 - 1,79 = -1,79, \text{ II} = 0,69 - 2,39 = -1,70, \text{ III} = 1,79 - 2,19 = -0,40, \text{ IV} = 0,69 - 2,07 = -1,38, \text{ V} = 0,69 - 1,79 = -1,09$$

Jumlah saudara = “I” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 1,38 = -0,69, \text{ II} = 1,79 - 2,19 = -0,40, \text{ III} = 1,38 - 1,94 = -0,55, \text{ IV} = 1,60 - 1,79 = -0,18, \text{ V} = 1,09 - 1,38 = -0,28$$

wilayah = “I” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 1,38 = -0,69, \text{ II} = 1,79 - 2,19 = -0,40, \text{ III} = 1,38 - 1,94 = -0,55, \text{ IV} = 1,38 - 1,79 = -0,40, \text{ V} = 1,09 - 1,38 = -0,28$$

kondisi orang tua = “III” | “I”, “II”, “III”, “IV”, “V”

$$I = 0 - 1,60 = -1,60, \text{ II} = 2,07 - 2,30 = -0,22, \text{ III} = 1,79 - 2,07 = -0,28, \text{ IV} = 1,60 - 1,94 = -0,33, \text{ V} = 1,09 - 1,60 = -0,51$$

pendidikan ayah = “sarjana (S1)” | “I”, “II”, “III”, “IV”, “V”

$$I = 0 - 1,94 = -1,94, \text{ II} = 0 - 2,48 = -2,48, \text{ III} = 0 - 2,30 = -2,30, \text{ IV} = 0 - 2,19 = -2,19, \text{ V} = 1,09 - 1,94 = -0,84$$

pekerjaan ayah = “wiraswasta/eksekutif/pedagang” | “I”, “II”, “III”, “IV”, “V”

$$I = 0 - 1,94 = -1,94, \text{ II} = 0 - 2,48 = -2,48, \text{ III} = 1,09 - 2,30 = -1,20, \text{ IV} = 0 - 2,19 = -2,19, \text{ V} = 1,09 - 1,94 = -0,84$$

pendapatan ayah = “VIII” | “I”, “II”, “III”, “IV”, “V”

$$I = 0 - 2,07 = -2,07, \text{ II} = 0 - 2,56 = -2,56, \text{ III} = 0 - 2,39 = -2,39, \text{ IV} = 0 - 2,30 = -2,30, \text{ V} = 0,69 - 2,07 = -1,38$$

pendidikan ibu = “tamat SMTA” | “I”, “II”, “III”, “IV”, “V”

$$I = 0 - 1,94 = -1,94, \text{ II} = 0,69 - 2,48 = -1,79, \text{ III} = 0,69 - 2,30 = -1,60, \text{ IV} = 1,09 - 2,19 = -1,09, \text{ V} = 0 - 1,94 = -1,94$$

pekerjaan ibu = “tidak bekerja” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 1,79 = -1,09, \text{ II} = 1,94 - 2,39 = -0,45, \text{ III} = 1,38 - 2,19 = -0,81, \text{ IV} = 1,38 - 2,07 = -0,69, \text{ V} = 0,69 - 1,79 = -1,09$$

pendapatan ibu = “I” | “I”, “II”, “III”, “IV”, “V”

$$I = 0,69 - 1,94 = -0,91, \quad II = 2,07 - 2,48 = -0,22, \quad III = 1,79 - 2,30 = -0,28, \\ IV = 1,38 - 2,19 = -0,55, \quad V = 0,69 - 1,94 = -0,91$$

5. Hitung nilai *likelihood*

a. Data *test* dengan nama Harma

$$I = (-1,09)+(-0,69)+(-0,69)+(-0,91)+(-1,94)+(-1,25)+(-1,38)+(-0,84)+(-1,79)+(-0,91)+(2,30) = \mathbf{-13,84}$$

$$II = (-0,79)+(-1,09)+(-0,40)+(-2,30)+(-1,79)+(-2,48)+(-2,56)+(-1,09)+(-1,70)+(-2,30)+(-1,04) = \mathbf{-16,54}$$

$$III = (-2,19)+(-0,84)+(-0,55)+(-2,07)+(-2,20)+(-2,30)+(-2,39)+(-0,91)+(-2,19)+(-2,07)+(-1,38) = \mathbf{-17,87}$$

$$IV = (-1,38)+(-1,79)+(-0,40)+(-1,94)+(-1,50)+(-2,19)+(-2,30)+(-1,50)+(-1,38)+(-1,25)+(1,60) = \mathbf{-15,67}$$

$$V = (-1,79)+(-1,38)+(-0,28)+(-1,60)+(-1,94)+(-1,94)+(-3,07)+(-1,25)+(-1,79)+(-1,60)+(2,30) = \mathbf{-15,70}$$

b. Data *test* dengan nama Suci

$$I = (-1,79)+(-0,69)+(-0,69)+(-1,94)+(-1,94)+(-1,25)+(-1,38)+(-0,84)+(-1,09)+(-0,91)+(2,30) = \mathbf{-15,23}$$

$$II = (-1,70)+(-1,09)+(-0,40)+(-0,22)+(-1,38)+(-2,48)+(-0,61)+(-1,09)+(-0,45)+(-0,22)+(-1,04) = \mathbf{-9,69}$$

$$III = (-0,40)+(-0,84)+(-0,55)+(-0,28)+(-0,91)+(-1,20)+(-0,78)+(-0,91)+(-0,81)+(-0,28)+(-1,38) = \mathbf{-7,02}$$

$$IV = (-1,38)+(-1,79)+(-0,40)+(-1,33)+(-2,19)+(-2,19)+(-0,91)+(-1,50)+(-0,69)+(-0,55)+(1,60) = \mathbf{-11,98}$$

$$V = (-1,09)+(-1,38)+(-0,28)+(-0,51)+(-1,94)+(-0,84)+(-2,07)+(-1,25)+(-1,09)+(-0,91)+(2,30) = \mathbf{-11,42}$$

c. Data *test* dengan nama Suharma

$$I = (-1,79)+(-0,69)+(-0,69)+(-1,60)+(-0,84)+(-1,25)+(-1,38)+(-0,84)+(-1,09)+(-0,91)+(2,30) = \mathbf{-13,43}$$

$$II = (-1,70)+(-0,40)+(-0,40)+(-0,22)+(-1,09)+(-0,69)+(-0,61)+(-1,09)+(-0,45)+(-0,22)+(-1,04) = \mathbf{-6,92}$$

$$\text{III} = (-0,40)+(-0,55)+(-0,55)+(-0,28)+(-1,60)+(-2,30)+(-0,78)+(-0,91)+(-0,81)+(-0,28)+(-1,38) = \mathbf{-8,52}$$

$$\text{IV} = (-1,38)+(-0,18)+(-0,40)+(-0,33)+(-1,09)+(-0,81)+(-0,91)+(-1,50)+(-0,69)+(-0,55)+(1,60) = \mathbf{-7,89}$$

$$\text{V} = (-1,09)+(-0,28)+(-0,28)+(-0,51)+(-1,94)+(-1,94)+(-2,07)+(-1,25)+(-1,09)+(-0,91)+(2,30) = \mathbf{-11,42}$$

d. Data *test* dengan nama Rayhan

$$\text{I} = (-1,79)+(-0,69)+(-0,69)+(-1,60)+(-1,94)+(-1,94)+(-2,07)+(-1,94)+(-1,09)+(-0,91)+(2,30) = \mathbf{-17,02}$$

$$\text{II} = (-1,70)+(-0,40) +(-0,40)+(-0,22)+(-2,48)+(-2,48)+(-2,56)+(-1,79)+(-0,45)+(-0,22)+(-1,04) = \mathbf{-12,74}$$

$$\text{III} = (-0,40)+(-0,55)+(-0,55)+(-0,28)+(-2,30)+(-1,20)+(-2,39)+(-1,60)+(-0,81)+(-0,28)+(-1,38) = \mathbf{-10,42}$$

$$\text{IV} = (-1,38)+(-0,18)+(-0,40)+(-0,33)+(-2,19)+(-2,19)+(-2,30)+(-1,09)+(-0,69)+(-0,55)+(1,60) = \mathbf{-11,35}$$

$$\text{V} = (-1,09)+(-0,28)+(-0,28)+(-0,51)+(-1,84)+(-1,84)+(-1,38)+(-1,95)+(-1,09)+(-0,91)+(2,30) = \mathbf{-9,22}$$

6. Menghitung nilai terbesar dari setiap kelas

- a. Data *test* dengan nama Harma : hasil perhitungan menghasilkan jumlah *class "I"* dengan nilai terbesar.
- b. Data *test* dengan nama Suci : hasil perhitungan menghasilkan jumlah *class "III"* dengan nilai terbesar.
- c. Data *test* dengan nama Suharma : hasil perhitungan menghasilkan jumlah *class "I"* dengan nilai terbesar.
- d. Data *test* dengan nama Rayhan : hasil perhitungan menghasilkan jumlah *class "V"* dengan nilai terbesar.

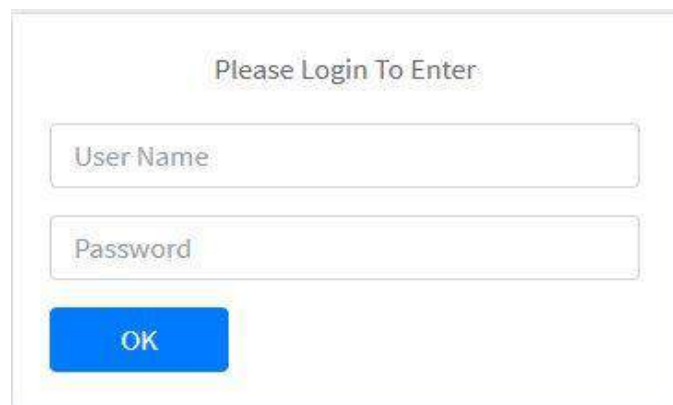
Tabel 4. 6 Hasil Perhitungan

Nama	bantuan	saudara	Wilayah	Kondisi orangtua	Pendidikan ayah	Pekerjaan ayah	Pendapatan Ayah	Pendidikan Ibu	Pekerjaan Ibu	Pendapatan ibu	Kelompok UKT
HARMA	KIS	II	I	I	Tamat SMTA	Tidak Bekerja	I	Tamat SD	Petani	II	I
SUCI	Tidak Ada	II	I	III	Tamat SMTP	Wiraswasta/ Eksekutif/Pedagang	II	Tamat SD	Tidak Bekerja	I	III
SUHARMA	Tidak Ada	I	I	III	Tamat SD	Petani/Nelayan	II	Tamat SD	Tidak Bekerja	I	II
RAYHAN	Tidak Ada	I	I	III	Sarjana (S1)	Wiraswasta/ Eksekutif/Pedagang	VIII	Tamat SMTA	Tidak Bekerja	I	V

4.5 Implementasi Pemrograman

4.5.1 Tampilan Halaman *Login*

Tampilan halaman *login* dari *username* dan *password* untuk melakukan *login* sebelum masuk kehalaman utama. Ketika pengguna mengisi *user name* dan *Password* dan memilih *OK* maka akan muncul tampilan Menu Utama. Tampilan *login* terdapat pada Gambar 4.6

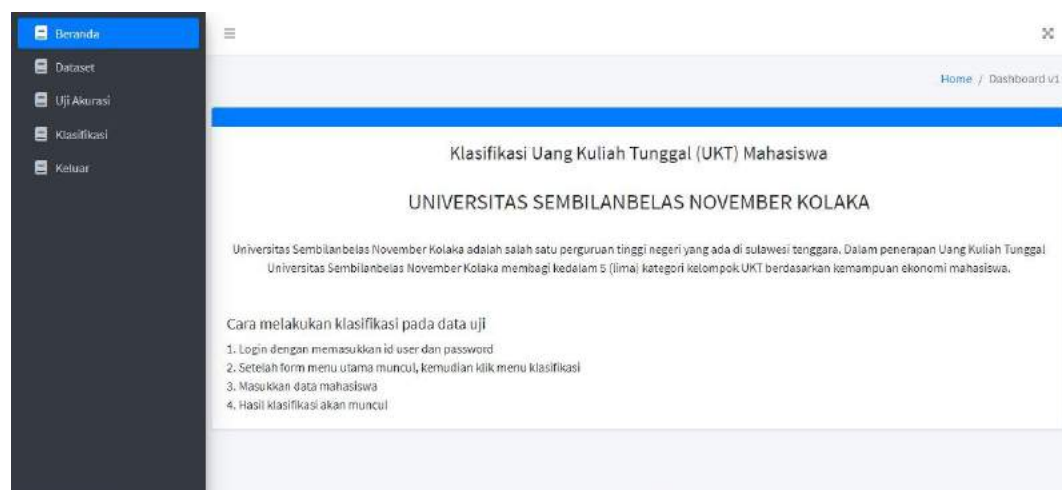


The image shows a login form with a title 'Please Login To Enter'. Below the title are two text input fields. The first field is labeled 'User Name' and the second is labeled 'Password'. Below these fields is a blue button with the text 'OK' in white.

Gambar 4. 7 Tampilan Halaman *Login*

4.5.2. Tampilan Menu Utama

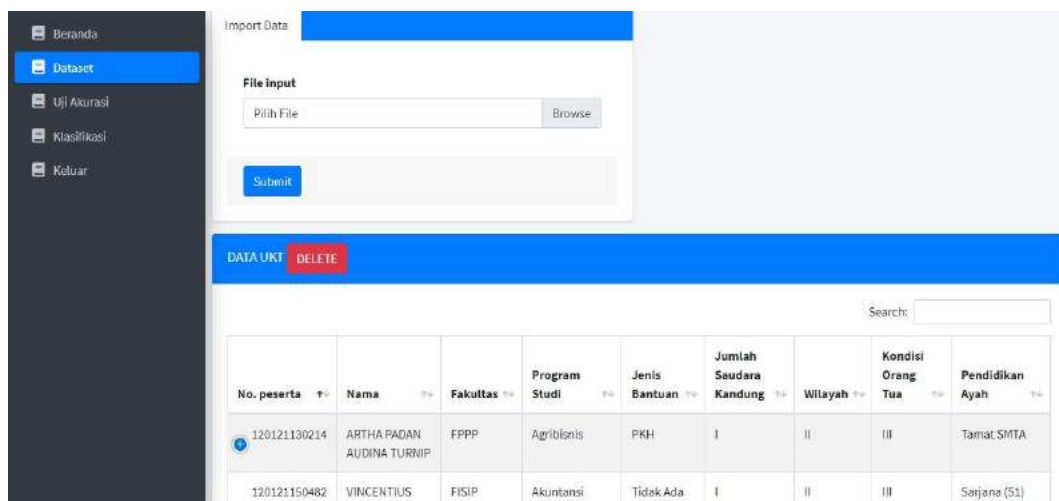
Tampilan Menu Utama terdiri dari informasi dan tata cara melakukan klasifikasi pada data uji. Tampilan Menu Utama terdapat pada Gambar 4.7



Gambar 4. 8 Tampilan Halaman Menu Utama

4.5.3. Tampilan *Dataset*

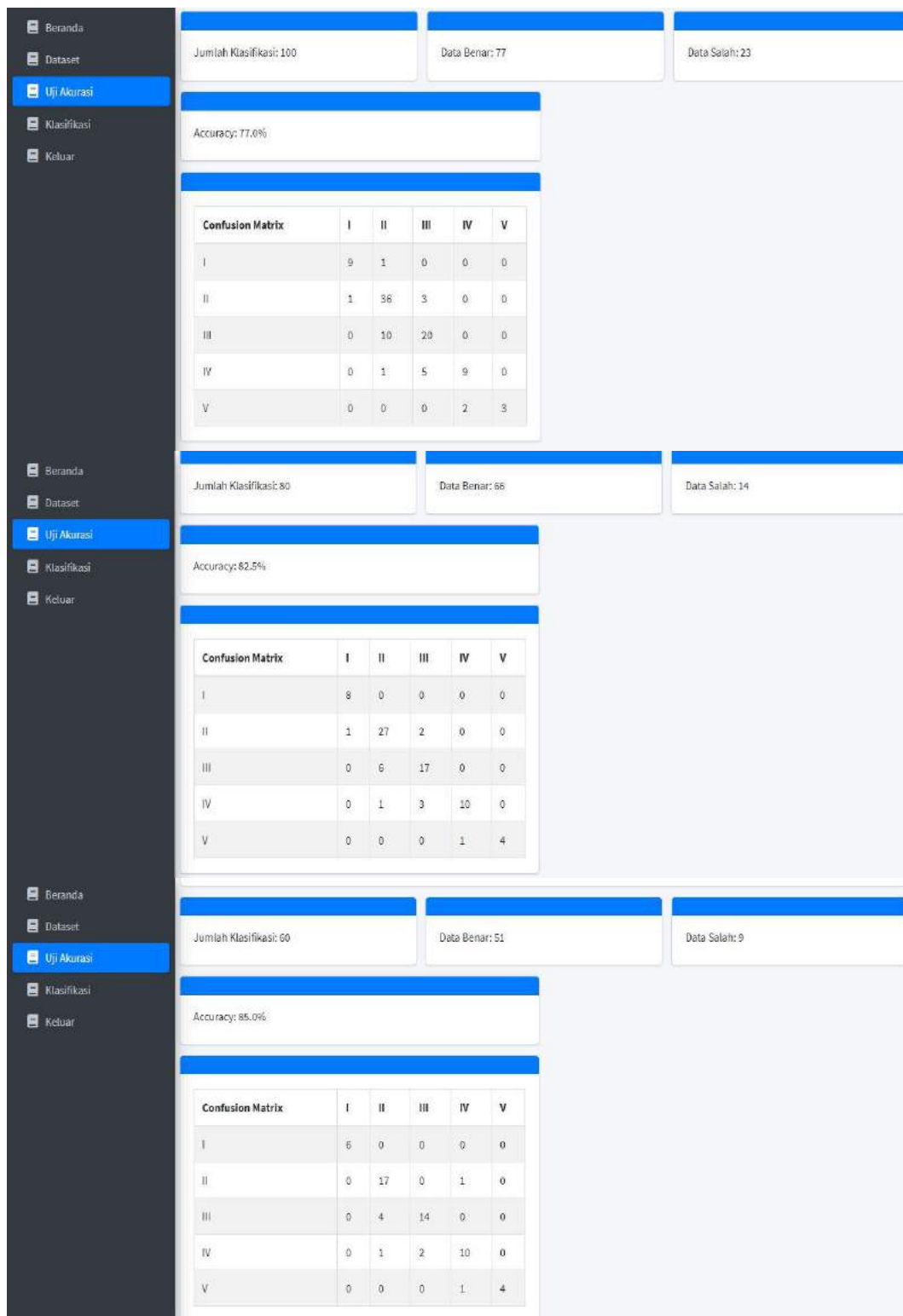
Tampilan *Data Training* terdiri dari tampilan tabel *data training*, tombol *delete* untuk menghapus jumlah *data training*, dan *import data* untuk menambahkan *data training* dengan format *excel*. Halaman ini menampilkan *data training*. Tampilan *Data Training* terdapat pada Gambar 4.8



Gambar 4. 9 Tampilan *Dataset*

4.5.4. Tampilan Uji Akurasi

Tampilan uji akurasi dengan tiga kali pengujian. Pengujian pertama *data testing* yang digunakan sebanyak 100 data dengan jumlah data benar sebanyak 77 dan jumlah data salah sebanyak 23 data sehingga menghasilkan nilai akurasi sebesar 77%. Pengujian kedua *data testing* yang digunakan sebanyak 80 data dengan jumlah data benar sebanyak 66 dan jumlah data salah sebanyak 14 data sehingga menghasilkan nilai akurasi sebesar 82,5%. Pengujian kedua *data testing* yang digunakan sebanyak 60 data dengan jumlah data benar sebanyak 51 dan jumlah data salah sebanyak 9 data sehingga menghasilkan nilai akurasi sebesar 82,5%. Tampilan uji akurasi terdapat pada Gambar 4.9



Gambar 4. 10 Tampilan Uji Akurasi

4.5.5. Tampilan Klasifikasi

Tampilan klasifikasi terdiri dari tampilan *import* dan manual. Dimana untuk *import* menggunakan data dengan format *xlsx* dan manual dengan mengisi *combo box* dengan data mahasiswa yaitu no peserta, nama, fakultas, program studi, jenis bantuan, wilayah, kondisi orangtua, pendidikan ayah, pekerjaan ayah, pendidikan ibu, pekerjaan ibu serta pendapatan ibu. Tampilan Klasifikasi terdapat pada Gambar 4.10

Gambar 4. 11 Tampilan Klasifikasi

4.5.6. Tampilan Hasil Klasifikasi

Berisi tampilan tabel hasil klasifikasi dan nilai *naive bayes* serta mencetak hasil klasifikasi dalam bentuk format Excel atau PDF. Tampilan hasil klasifikasi dapat dilihat pada Gambar 4.11

Home / Dashboard v1

HASIL KLASIFIKASI

Excel PDF Print Column visibility Search:

No. peserta	Nama	Fakultas	Program Studi	Jenis Bantuan	Jumlah Saudara Kandung	Wilayah	Kondisi Orang Tua	Pendidikan Ayah	Pekerjaan Ayah
112	Nurida	FTI	Sistem Informasi	Tidak Ada	I	II	III	Tamat SMTA	Buruh

Showing 1 to 1 of 1 entries

Previous 1 Next

Nilai Naive Bayes: [-7.28800526
-2.48046834 -0.12816674 -3.33188044
-8.62395395]

Gambar 4. 12 Tampilan Hasil Klasifikasi

Penerapan algoritma *naive bayes* untuk klasifikasi Uang Kuliah Tunggal (UKT) Universitas Sembilanbelas November Kolaka dapat dilihat pada *listing* program sebagai berikut :

1) Menghitung *class log prior*

```
log_class_count = np.log(class_count_)
n_classes = len(model.classes_)
class_log_prior_ = (log_class_count -
np.log(class_count_.sum( )))
```

2) Menghitung *category cat count* dan *class count*

```
cat_count = category_count_(i) + alpha
class_count = cat_count.sum(axis=1)
```

3) Menghitung *smoothed cat count* dan *smoothed class count*

```
Smoothed_cat_count = np.log(smoothed_cat_count)
smoothed_class_count =
np.log(smoothed_clas_conut.reshape(-1, 1))
```

4) Menghitung *feature log prob*

```
def_update_feature_log_prob(alpha)
    feature_log_prob = [ ]
    for i in range(model.n_features_);
        Smoothed_cat_count = model.
            category_count_(i) +alpha
        Smoothed_class_count =
            Smoothed_cat_count.sum(axis=1)
        Feature_log_prob.append(
            np.log(smoothed_cat_count)-
            np.log(smoothed_clas_conut.reshape(-1,
                1))

    self.feature_log_prob = feature_log_prob
```

5) Menghitung *likelihood*

```
def joint_log_likelihood(X):
    jll=np.zeros((X.shape[0],model.class_count_.shape[0]))
    for i in range(model.n_features_):
        indices = X[:, i]
        jll += model.feature_log_prob_[i][:, indices].T
    total_ll = jll + class_log_prior_
    return total_ll
```

6) Menghitung nilai terbesar dari setiap kelas

```
def predict(X):
    jll = self._joint_log_likelihood(X)
    return model.classes_[np.argmax(jll, axis=1)]
```

4.6 Pengujian Akurasi

4.6.1 Hasil *Counfusion Matrix*

Pada tahap ini menampilkan hasil *counfision matrix* dari klasifikasi UKT mahasiswa dengan tiga kali pengujian. Pengujian pertama menggunakan 385 *data training* dan 100 *data testing*, pengujian kedua menggunakan 405 *data training* dan 80 *data testing* dan pengujian ketiga menggunakan 425 *data training* dan 60 *data testing*. Hasil *counfision matrix* dari penyeleksian program klasifikasi UKT mahasiswa untuk data uji pertama dapat dilihat pada Tabel 4.6

Tabel 4. 7 Hasil *Confusin Matrix* Uji ke-1

	<i>Actual</i>				
<i>Prediction</i>	I	II	III	IV	V
I	9	1	0	0	0
II	1	36	3	0	0
III	0	10	20	0	0
IV	0	1	5	9	0
V	0	0	0	2	3

Pada Tabel 4.6 didapatkan hasil *counfision matrix* dari penyeleksian klasifikasi UKT mahasiswa. Jumlah *true* I adalah 9, jumlah *true* II adalah 36, jumlah *true* III adalah 20, jumlah *true* IV adalah 9, jumlah *true* V adalah 3, jumlah *false* I adalah 1, jumlah *false* II adalah 5, jumlah *false* III adalah 8, jumlah *false* IV adalah 6, dan jumlah *false* V adalah 2.

Setelah itu, menghitung nilai *Accuracy* yaitu dengan menggunakan rumus:

$$\begin{aligned}
 Accuracy &= \frac{TI + TII + TIII + TIV + TV}{TI + FI + TII + FII + TIII + FIII + TIV + FIV + TV + FV} \\
 &= \frac{9 + 36 + 20 + 9 + 3}{9 + 1 + 36 + 4 + 20 + 10 + 9 + 6 + 3 + 2} = \frac{77}{100} = 0,77 = 77\%
 \end{aligned}$$

Hasil *counfision matrix* dari penyeleksian program klasifikasi UKT mahasiswa untuk data uji kedua dapat dilihat pada Tabel 4.7

Tabel 4. 8 Hasil Confusin Matrix Uji ke-2

	<i>Actual</i>				
<i>Prediction</i>	I	II	III	IV	V
I	8	0	0	0	0
II	1	27	2	0	0
III	0	6	17	0	0
IV	0	1	3	10	0
V	0	0	0	1	4

Pada Tabel 4.7 didapatkan hasil *counfision matrix* dari penyeleksian klasifikasi UKT mahasiswa. Jumlah *true* I adalah 8, jumlah *true* II adalah 27, jumlah *true* III adalah 17, jumlah *true* IV adalah 10, jumlah *true* V adalah 4, jumlah *false* I adalah 0, jumlah *false* II adalah 3, jumlah *false* III adalah 6, jumlah *false* IV adalah 4, dan jumlah *false* V adalah 1.

Setelah itu, menghitung nilai *Accuracy* yaitu dengan menggunakan rumus:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TI + TII + TIII + TIV + TV}{TI + FI + TII + FII + TIII + FIII + TIV + FIV + TV + FV} \\
 &= \frac{8 + 27 + 17 + 10 + 4}{8 + 0 + 27 + 3 + 17 + 6 + 10 + 4 + 4 + 1} = \frac{66}{80} = 0,825 = 82\%
 \end{aligned}$$

Hasil *counfision matrix* dari penyeleksian program klasifikasi UKT mahasiswa untuk data uji ketiga dapat dilihat pada Tabel 4.8

Tabel 4. 9 Hasil Confusin Matrix Uji ke-3

	<i>Actual</i>				
<i>Prediction</i>	I	II	III	IV	V
I	6	0	0	0	0
II	0	17	0	0	0
III	0	4	14	0	0
IV	0	1	2	10	0
V	0	0	0	1	4

Pada Tabel 4.8 didapatkan hasil *counfision matrix* dari penyeleksian klasifikasi UKT mahasiswa. Jumlah *true I* adalah 6, jumlah *true II* adalah 17, jumlah *true III* adalah 14, jumlah *true IV* adalah 10, jumlah *true V* adalah 4, jumlah *false I* adalah 0, jumlah *false II* adalah 0, jumlah *false III* adalah 4, jumlah *false IV* adalah 3, dan jumlah *false V* adalah 1.

Setelah itu, menghitung nilai *Accuracy* yaitu dengan menggunakan rumus:

$$Accuracy = \frac{TI + TII + TIII + TIV + TV}{TI + FI + TII + FII + TIII + FIII + TIV + FIV + TV + FV}$$

$$= \frac{6 + 17 + 14 + 10 + 4}{6 + 0 + 17 + 0 + 14 + 4 + 10 + 3 + 4 + 1} = \frac{51}{60} = 0,85 = 85\%$$

Dari tiga kali pengujian disimpulkan bahwa apabila *data training* semakin banyak maka nilai akurasi akan semakin tinggi.

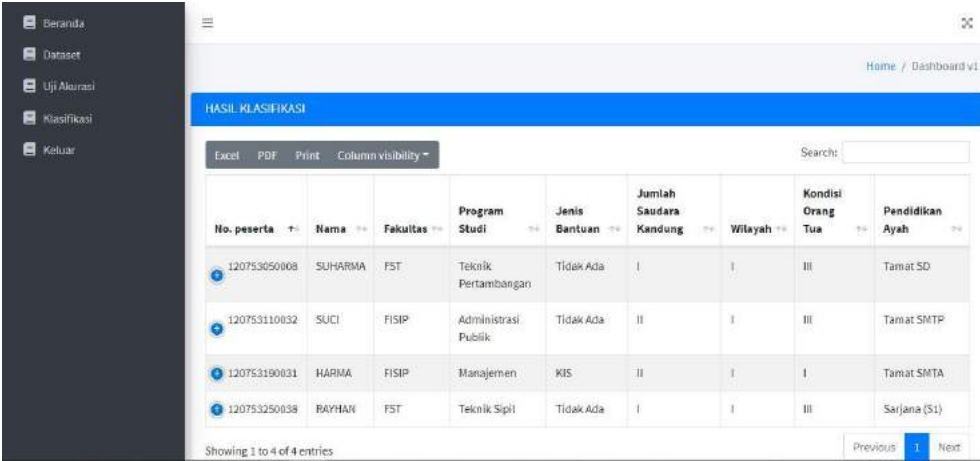
4.7. Perbandingan Manual dan Implementasi pada Sistem

Berdasarkan perhitungan manual yang terdapat pada Tabel 4.6 menghasilkan nilai yang sama dengan hasil implementasi pada sistem yang telah dibuat. Gambar hasil impementasi sistem dapat dilihat pada Gambar 4.11



No. peserta	Nama	Fakultas	Program Studi	Jenis Bantuan	Jumlah Saudara Kandung	Wilayah	Kondisi Orang Tua	Pendidikan Ayah
120711051120	PIRDAUS	FST	Teknik Mesin	KIP	II	II	I	Tamat SD
120751110388	RONI	FTI	Ilmu Komputer	KIS	I	II	III	Tamat SD
120751270054	SAIFUL	FST	Teknik Mesin	Tidak Ada	I	II	III	Tamat SD
120753030015	IRMASARI	FST	Farmasi	KIS	I	I	II	Tamat SD
120753110010	SAENAL	FTI	Sistem Informasi	Tidak Ada	I	I	III	Tamat SD
120753110049	SAFITRI	FH	Hukum	Tidak Ada	I	I	III	Tamat SD
120753110009	SANDI	FST	Teknik Sipil	KIS	II	I	III	Tidak Tamat SD
120753130050	SULKIPLI	FPPP	Agribisnis	KIS	II	I	III	Tamat SMTP
120753130055	SAIN	FPPP	Agribisnis	Tidak Ada	I	I	III	Tamat SMTP
120753150038	ANJAR	FISIP	Manajemen	KIP	I	I	III	Sarjana (S1)

Gambar 4. 13 Hasil Implementasi Pada Sistem



Home / Dashboard v1

HASIL KLASIFIKASI

Excel PDF Print Column visibility Search:

No. peserta	Nama	Fakultas	Program Studi	Jenis Bantuan	Jumlah Saudara Kandung	Wilayah	Kondisi Orang Tua	Pendidikan Ayah
120753050008	SUHARMA	FST	Teknik Pertambangan	Tidak Ada	I	I	III	Tamat SD
120753110032	SUCI	FISIP	Administrasi Publik	Tidak Ada	II	I	III	Tamat SMTP
120753190031	HARMA	FISIP	Manajemen	KIS	II	I	I	Tamat SMTA
120753250038	RAYHAN	FST	Teknik Sipil	Tidak Ada	I	I	III	Sarjana (S1)

Showing 1 to 4 of 4 entries Previous 1 Next

Gambar 4. 14 Hasil Implementasi Pada Sistem

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian Klasifikasi Uang Kuliah Tunggal (UKT) Universitas Sembilanbelas November Kolaka Menggunakan Algoritma *Naïve Bayes* ini dapat diambil kesimpulan sebagai berikut:

- 1) Berdasarkan perbandingan perhitungan manual dan sistem, keduanya menunjukkan hasil yang sama, sehingga Algoritma *Naïve Bayes* terbukti dapat diimplementasikan ke dalam sistem.
- 2) Berdasarkan hasil *counfision matrix* dimana nilai akurasi untuk tiga kali pengujian mengalami kenaikan. Pengujian pertama menggunakan 385 *data training* dan 100 *data testing* menghasilkan nilai akurasi sebesar 77%, pengujian kedua menggunakan 405 *data training* dan 80 *data testing* menghasilkan nilai akurasi sebesar 82,5%, pengujian ketiga menggunakan 425 *data training* dan 60 *data testing* menghasilkan nilai akursi sebesar 85%, ini membuktikan bahwa semakin banyak *data training* maka nilai akurasi akan semakin tinggi. Sehingga Algoritma *Naïve Bayes* dapat diterapkan untuk menentukan Klasifikasi Uang Kuliah Tunggal (UKT) Universitas Sembilanbelas November Kolaka.

5.2. Saran

Sebagai akhir dari penelitian ini, terdapat beberapa saran yang dapat berguna bagi peneliti lainnya, antara lain sebagai berikut:

- 1) Penelitian ini dapat dikembangkan dengan menggabungkan atau membandingkan dengan algoritma klasifikasi lain untuk mendapatkan hasil prediksi yang lebih baik.
- 2) Sebaiknya jumlah data ditambah, sehingga dapat diperoleh hasil akurasi yang lebih baik.

DAFTAR PUSTAKA

- Aditya, Y. (2017). Pengembangan Sistem Rekomendasi Calon Penerima Beasiswa Dengan Metode Naïve Bayes.
- Antaristi, M., & Kurniawan, I. Y. (2017). Aplikasi Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan Metode Naive Bayes di Bank BNI Syariah Surabaya. *Jurnal Teknik Elektro*.
- Elvianti. (2015). Penerapan Metode Modified k-Nearest Neighbor untuk Klasifikasi Penderita Penyakit Liver [Skripsi].
- Han, J., & Kamber, M. (2006). *Data mining Concept and Techniques Second Edition*. San Francisco: Morgan Kaufmann.
- Jogiyanto, H. M. (2010). *Teori Portofolio dan Analisis Investasi*. Yogyakarta: Edisi ketiga :BPFE.
- Ladjamudin, B. A.-B. (2013). *Analisis dan Desain Sistem Informasi*. Yogyakarta: Graha Ilmu.
- Laroussi, M. H. (2015). „Implementasi algoritma naïve bayes sebagai proses seleksi penerima beasiswa libyan embassy berbasis web.
- Mustafa, S. M., Ramadhan, R. M., & Angelina, T. P. (2017). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Cite c Journal*, 151-162.
- Ningrum, K. P. (2017). Sistem Pendukung Keputusan Pemberian Kredit Motor Menggunakan Metode Simple Additive Weighting (SAW) pada FIFGROUP Cabang Bandar Lampung.
- Nosrati, M. (2011). Python : An approroriate language for real world programming. *World Applied Programming*, 110-117.
- Nurhasan , F., Hikmah, N., & Utami, D. Y. (2018). Perbandingan algoritma C4.5, KNN, dan Naive Bayes untuk Penentuan Model Klasifikasi Penanggung Jawab BSI Entrepreneur Center. *Jurnal Pilar Nusa Mandiri*, 169-174.
- Pessman, R. S. (2002). *Rekayasa Perangkat Lunak Pendekatan Praktisi (Buku Satu)*. Yogyakarta: Andi.
- Pramudiono, J. (2003). Pengantar Data Mining : Menambang Permata Pengetahan di Gunung Data.

- Prasetyo, M. (t.thn.). DATA MINING - Mengolah Data Menjadi Informas Menggunakan Matlab. 2014.
- Putra, A. S. (2018). Klasifikasi Status Gizi Balita Menggunakan Naive Bayes Classification (Studi Kasus Posyandu Ngudi Luhur).
- Republik Indonesia. (2013). *Peraturan Menteri Pendidikan dan Kebudayaan No. 55 Tahun 2013 Tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal pada Perguruan Tinggi Negeri di Lingkungan Kementerian Pendidikan dan Kebudayaan. Kementerian Pendidikan dan Kebudayaan Republik Indonesia*. Jakarta.
- Reza, P. A. (2018). Penerapan Naïve Bayes Classifier dengan Gaussian Function Untuk Menentukan Kelompok UKT . *JURNAL ILMIAH INFORMATIKA GLOBAL* , 112-117.
- Rifqi, N., Maharani, W., & Shaufiah. (2011). Analisis dan Implementasi Klasifikasi Data Mining Menggunakan Jaringan Syaraf Tiruan dan Evalution Strategis. *Konferensi Nasional Sistem dan Informatika*, 183-191.
- Rizqiani, v., Mulwinda , A., & Putri. (2017). Klasifikasi Judul Buku dengan Algoritma Naive Bayes dan Pencarian Buku pada Perpustakaan Jurusan Teknik Elektro. *9(2)*, 60-65.
- Rochmawati, N. S., Istiadi, Marisa, F., & Nurdiansyah, F. (2020). Klasifikasi Potensi Berdasarkan Kepribadian Siswa dengan Metode Naïve Bayes. *Conference on Innovation and Application of Science and Technology* , 327-334.
- Sulistiawati, H., & Utami, T. Y. (2018). Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa. *Universitas Trisakti*, 300-305.
- Turban, E., & kk. (2005). *Decision Support System and Intelligent System*. Yogyakarta: andi.

LAMPIRAN

Lampiran I Surat Permohonan Rekomendasi Penelitian



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET DAN TEKNOLOGI
UNIVERSITAS SEMBILANBELAS NOVEMBER KOLAKA
**LEMBAGA PENELITIAN, PENGABDIAN MASYARAKAT, DAN
PENJAMINAN MUTU PENDIDIKAN (LP2M-PMP)**
Jl. Pemuda No. 339 Kab. Kolaka-Sulawesi Tenggara
Telp. (0405) 2321132 Fax. 2324028 Kolaka 93517e-Mail : lp2m@usn.ac.id

Nomor : 1292/UN56D/PN.0100/2021
Lampiran :-
Perihal : Rekomendasi untuk izin Penelitian

Kepada,

Yth. Wakil Rektor I Bidang Akademik Kemahasiswaan USN Kolaka
Di

Tempat

Dalam rangka penyelesaian studi mahasiswa Universitas Sembilanbelas November Kolaka, salah satu syarat yang harus ditempuh adalah melaksanakan penelitian yang sesuai dengan bidang ilmu dan ruang lingkup permasalahan yang diteliti, baik penelitian lapangan maupun penelitian pustaka

Oleh karena itu, Lembaga Penelitian, Pengabdian Masyarakat, dan Penjaminan Mutu Pendidikan (LP2M-PMP USN Kolaka) memberikan rekomendasi kepada mahasiswa tersebut kiranya yang bersangkutan dapat diberi izin penelitian pada dinas dan badan yang terkait sesuai dengan nama yang tercantum dibawah ini:

Nama : NUR IDA
NIM : 171210156
Prog. Studi : SISTEM INFORMASI
Fakultas : TEKNOLOGI INFORMASI
Alamat : DESA POPALIA KECAMATAN TANGGETADA KABUPATEN KOLAKA
Lokasi Penelitian : KAMPUS UNIVERSITAS SEMBILANBELAS NOVEMBER KOLAKA
Waktu Penelitian : SAMPAI SELESAI
Judul Penelitian : PENERAPAN METODE NAÏVE BAYES CLASSIFIER DALAM PENENTUAN UKT DI UNIVERSITAS SEMBILANBELAS NOVEMBER KOLAKA.
Pembimbing I : ANDI TENRI SUMPALA, S.Kom., M.Cs
Pembimbing II : YUWANDA PURNAMASARI PASRUN S.T., M.Kom

Demikian surat rekomendasi ini dibuat, atas perhatian dan kerja sama yang baik diucapkan terima kasih.

05 Juli 2021
Ketua LP2M-PMP

Dr. Wayan Pageyasa, M.Pd
NIDN. 09050375501

Tembusan :
1. Wakil Rektor I Bidang Akademik
2. Arsip

Lampiran II Surat Izin Penelitian



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET, DAN TEKNOLOGI
UNIVERSITAS SEMBILANBELAS NOVEMBER KOLAKA
Jl. Pemuda No. 339, Kabupaten Kolaka, Sulawesi Tenggara, 93517
Telp (0405) 2321132 Fax (0405) 2324028
E-mail : rektorat@usn.ac.id ; Website : <http://usn.ac.id>

SURAT IZIN PENELITIAN

Nomor : 165 /UN56/KM/2021

Menindaklanjuti surat Lembaga Penelitian, Pengabdian Masyarakat, dan Penjaminan Mutu Pendidikan (LP2M-PMP) Nomor : 1292/UN56D/PN.01.00/2021 perihal Rekomendasi Izin Penelitian, maka dengan ini kami memberikan izin penelitian kepada mahasiswa :

Nama : Nur Ida
NIM : 171210156
Program Studi : Sistem Informasi
Fakultas : Teknologi Informasi
Lokasi Penelitian : Universitas Sembilanbelas November Kolaka
Judul Penelitian : Penerapan Metode Naïve Bayes Claaiifier dalam Penentuan UKT di Universitas Sembilanbelas November Kolaka

Waktu Penelitian : 07 Juli sampai dengan 6 Agustus 2021

Demikian surat izin penelitian ini dibuat untuk dipergunakan sebagaimana mestinya.

Kolaka, 6 Juli 2021
Wakil Rektor Bidang Akademik

Prof. Dr. Ruslin Hadanu, M.Si
NIP 19721206 2000121001

Tembusan:

1. Rektor USN Kolaka
2. Ketua LP2M-PMP USN Kolaka
3. Arsip

Lampiran III Dokumentasi

