

TESIS

**ANALISIS PREDIKSI KELULUSAN MAHASISWA
MENGUNAKAN ALGORITMA *NAIVE BAYES* DAN
*BACKPROPAGATION ARTIFICIAL NEURAL NETWORK***



Disusun oleh:

Nama : Hartatik
NIM : 19.77.1182
Konsentrasi : Business Intelligence

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2021

TESIS

**ANALISIS PREDIKSI KELULUSAN MAHASISWA
MENGUNAKAN ALGORITMA *NAIVE BAYES* DAN
*BACKPROPAGATION ARTIFICIAL NEURAL NETWORK***

**ANALYSIS OF STUDENT GRADUATION PREDICTIONS
USING *NAÏVE BAYES* AND *BACKPROPAGATION ARTIFICIAL
NEURAL NETWORK* ALGORITHM**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Hartatlk
NIM : 19.77.1182
Konsentrasi : Business Intelligence

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2021

HALAMAN PENGESAHAN

**ANALISIS PREDIKSI KELULUSAN MAHASISWA
MENGUNAKAN ALGORITMA *NAIVE BAYES* DAN *BACKPROPAGATION*
*ARTIFICIAL NEURAL NETWORK***

**ANALYSIS OF STUDENT GRADUATION PREDICTIONS USING *NAÏVE BAYES*
AND *BACKPROPAGATION* ARTIFICIAL NEURAL NETWORK ALGORITHM**

Dipersiapkan dan Disusun oleh

Hartatik
19.77.1182

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Kamis, 7 Juli 2022

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 7 Juli 2022
Rektor

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

ANALISIS PREDIKSI KELULUSAN MAHASISWA MENGUNAKAN ALGORITMA *NAIVE BAYES* DAN *BACKPROPAGATION* *ARTIFICIAL NEURAL NETWORK*

ANALYSIS OF STUDENT GRADUATION PREDICTIONS USING *NAIVE* *BAYES* AND *BACKPROPAGATION* *ARTIFICIAL NEURAL NETWORK* ALGORITHM

Dipersiapkan dan Disusun oleh

Hartatik

19.77.1182

Telah Ditujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Kamis, 7 Juli 2022

Pembimbing Utama

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

Pembimbing Pendamping

Agung Budi Prasetyo, S.T., M.Eng.
NIK.

Anggota Tim Penguji

Prof. Dr. Ema Utami, S.Si., M.Kom.
NIK. 190302037

Anggota Tim Penguji

Dhani Ariatmanto, M.Kom., Ph.D.
NIK. 190302037

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 7 Juli 2022

Direktur Program Pascasarjana

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN

Yang bertandatangan di bawah ini,

Nama mahasiswa : Hartatik
NIM : 19.77.1182
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:
ANALISIS PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA NAIVE BAYES DAN BACKPROPAGATION ARTIFICIAL NEURAL NETWORK

Dosen Pembimbing Utama : Prof. Dr. Kusriani, M. Kom.
Dosen Pembimbing Pendamping : Agung Budi Prasetyo, S.T., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 7 Juli 2022
Yang Menyatakan,

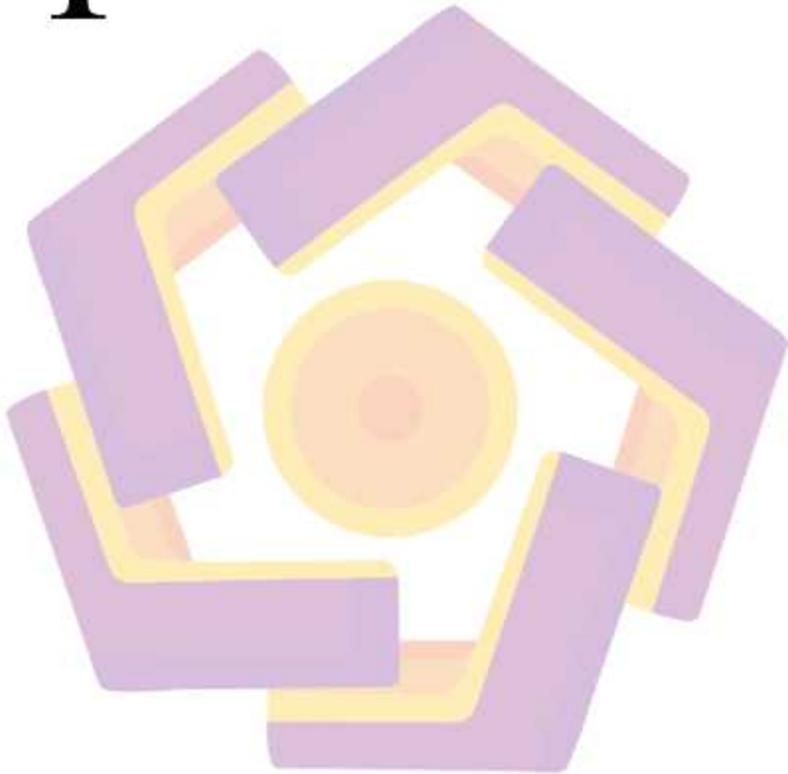


UNIVERSITAS AMIKOM
YOGYAKARTA
MEYERAN
TEMPEL
19771182

Hartatik

HALAMAN PERSEMBAHAN

Tesis ini saya persembahkan dan dedikasikan untuk perkembangan bidang teknologi data dan informasi.

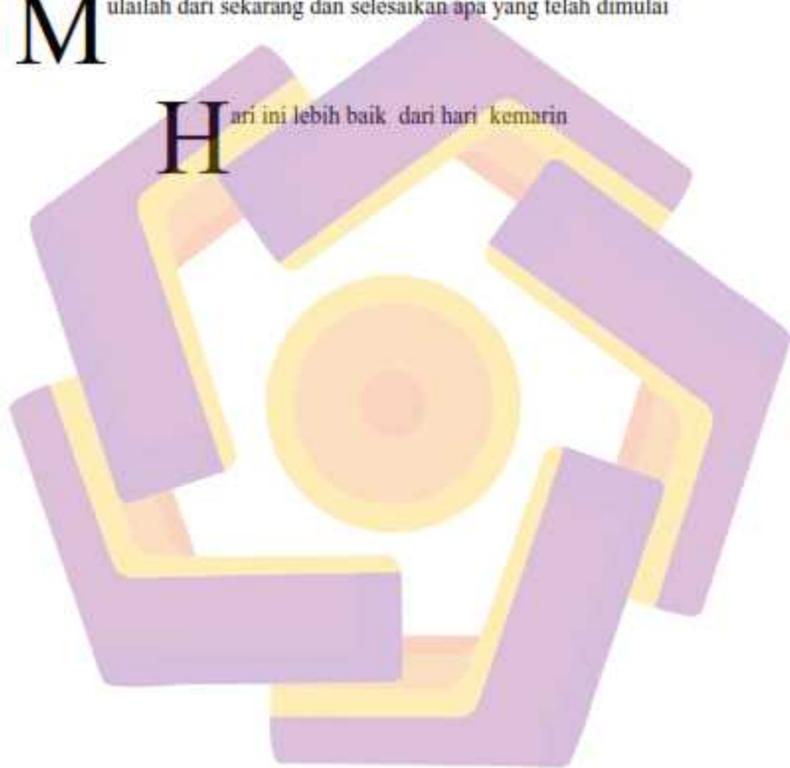


HALAMAN MOTTO

Fokus pada tujuan

Mulailah dari sekarang dan selesaikan apa yang telah dimulai

Hari ini lebih baik dari hari kemarin



KATA PENGANTAR

Puji syukur yang sedalam-dalamnya penulis panjatkan kehadirat Tuhan Yang Maha Esa atas segala berkat dan limpahan rahmat-Nya sehingga penulis dapat menyelesaikan proposal penelitian tesis dengan judul **“ANALISIS PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA NAIVE BAYES DAN BACKPROPAGATION ARTIFICIAL NEURAL NETWORK”**.

Tujuan dari penulisan tesis ini adalah untuk memenuhi syarat dalam mencapai derajat Magister Manajemen pada Program Studi Pasca Sarjana Universitas AMIKOM Yogyakarta.

Di dalam proses penulisan tesis ini, penulis banyak mendapatkan bimbingan dan dukungan dari berbagai pihak sehingga penulisan tesis ini dapat terselesaikan tepat waktu. Oleh karena itu, ucapan terimakasih yang sebesar-besarnya dan penghargaan setinggi-tingginya penulis sampaikan kepada :

1. Prof. Dr. Suyanto, M.M. selaku rektor Universitas AMIKOM Yogyakarta.
2. Prof. Dr. Kusriani, M.Kom. selaku Direktur Pascasarjana Universitas AMIKOM Yogyakarta dan juga Pembimbing utama yang telah mencurahkan perhatian bimbingan, semangat dan nasihat serta doa kepada penulis selama penelitian dan penyelesaian tesis.
3. Bapak Agung Budi Prasetyo, S.T., M.Eng. selaku pembimbing pendamping yang telah memberikan motivasi dan semangat dan nasihat serta doa kepada penulis selama penelitian dan penyelesaian tesis.

4. Bapak Ibu Dosen PJJ MTI Universitas AMIKOM Yogyakarta yang telah memberikan bimbingan.
5. Teman teman PJJ MTI Univeristas AMIKOM Yogyakarta yang telah memberikan support dan semangatnya selama ini.
6. Semua sahabat digital yang telah memberikan sharingnya selama penelitian ini.

Semoga karya ini dapat bermanfaat bagi kita semua.

Yogyakarta, 7 Juli 2022

Penulis



DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PERNYATAAN KEASLIAN TESIS	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah	5
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA	7
2.1. Tinjauan Pustaka	7
2.2. Keaslian Penelitian	11

2.3. Landasan Teori	15
2.3.1. Data Mining	15
2.3.2. Metode <i>Bayes</i> dan Algoritma <i>Naive Bayes</i>	16
2.3.3. <i>Artificial Neural Network</i> (ANN)	20
BAB III METODE PENELITIAN	26
3.1. Jenis, Sifat, dan Pendekatan Penelitian	26
3.2. Metode Pengumpulan Data	27
3.3. Metode Analisis Data	28
3.4. Alur Penelitian	29
3.4.1 Data Training	30
3.4.2 Pemodelan Data	33
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	39
4.1. Karakteristik dan Manajemen Data	39
4.1.1 <i>Action Planning</i>	39
4.1.1.1 Pengumpulan Data	39
4.1.1.2 Praproses Data	43
4.2. Implementasi	47
4.2.1 Implementasi Skenario 1: Algoritma <i>Naive Bayes</i>	48
4.2.2 Implementasi Skenario 2: Algoritma <i>Naive Bayes</i> dengan UN	52
4.2.3 Implementasi Skenario 3: Algoritma <i>Backpropagation – Artificial Neural Network</i>	56
4.2.4 Implementasi Skenario 4: Algoritma <i>Backpropagation – Artificial Neural Network</i> dengan UN	62



4.2.5 Evaluasi Prediksi dengan <i>Convusion Matrix</i>	64
4.3. Analisa dan Pembahasan	64
4.3.1 Hasil Implementasi Skenario 1: Algoritma <i>Naive Bayes</i>	64
4.3.2 Hasil Implementasi Skenario 2: Algoritma Naive Bayes dengan UN	66
4.3.3 Hasil Implementasi Skenario 3: Algoritma <i>Backpropagation – Artificial Neural Network</i>	78
4.3.4 Hasil Implementasi Skenario 3: Algoritma <i>Backpropagation – Artificial Neural Network</i> dengan UN	71
BAB V PENUTUP	77
5.1. Kesimpulan	77
5.2. Saran	77
DAFTAR PUSTAKA	79

DAFTAR TABEL

Tabel 2.1. Matriks literatur <i>review</i> dan posisi penelitian	11
Tabel 3.1. Transformasi data variabel	31
Tabel 4.1 Dataset mahasiswa diploma 3	40
Tabel 4.2 Data set yang digunakan	45
Tabel 4.3. Tabel Data set kelulusan mahasiswa	48
Tabel 4.4 Data training sebelum normalisasi	50
Tabel 4.5 Data training sesudah normalisasi	51
Tabel 4.6 Data testing sesudah normalisasi	51
Tabel 4.7. Tabel Data set kelulusan mahasiswa	52
Tabel 4.8 Data training sesudah normalisasi	54
Tabel 4.9 Data testing sesudah normalisasi	54
Tabel 4.10 Pemodelan dan komparasi hasil pemodelan <i>bayesian</i> skenario	65
Tabel 4.11 Confolution Matrix untuk hasil prediksi NB skenario 1	65
Tabel 4.12 Pemodelan dan komparasi hasil pemodelan NB skenario 2	66
Tabel 4.13 Confolution Matrix untuk hasil prediksi NB skenario 1	67
Tabel 4.14 Hasil prediksi Backpropagation –ANN skenario 3	69
Tabel 4.15 Hasil analisa dengan Backpropagation ANN sekanrio 3	70
Tabel 4.16 Hasil prediksi Backpropagation –ANN skenario 4	72
Tabel 4.17 Hasil analisa dengan Backpropagation ANN sekanrio 4	72
Tabel 4.18 Komparasi Hasil prediksi Bn dan Backpropagation-ANN	73
Tabel 4.19 komparasi nilai akurasi Model	74

DAFTAR GAMBAR

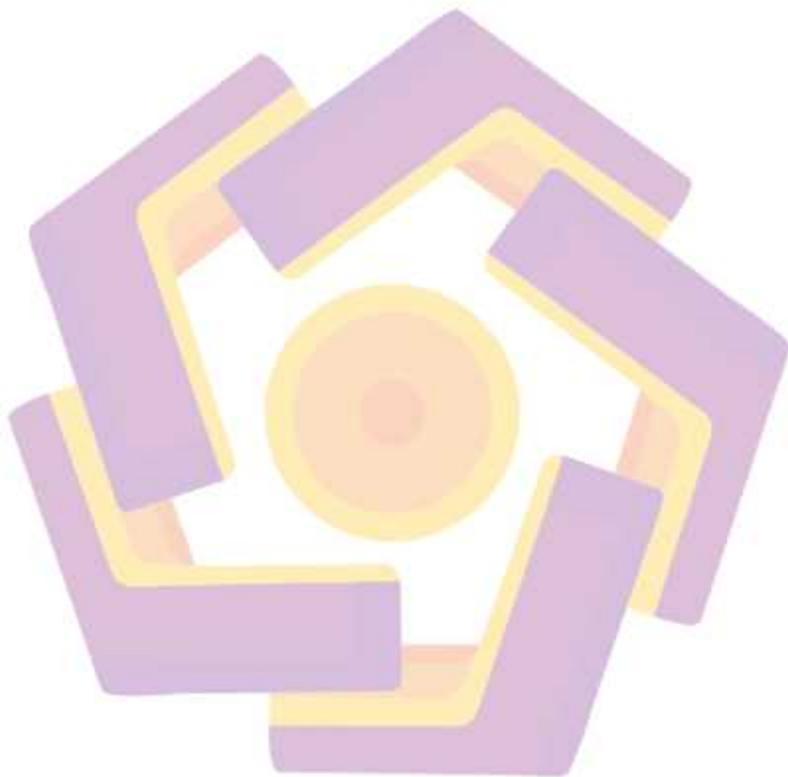
Gambar 2.1. Gambar ilustrasi <i>neuron</i> dalam ANN	21
Gambar 3.1. Langkah dalam penelitian	30
Gambar 3.2. Tahapan pemodelan dalam penelitian	32
Gambar 3.3. Algoritma <i>Naive Bayes</i>	34
Gambar 3.4 Flowchart algoritma B-ANN	35
Gambar 4.1 Potret ketepatan kelulusan mahasiswa oleh PA	40
Gambar 4.2 pola persebaran data mahasiswa	43
Gambar 4.3 Penanganan data kosong	44
Gambar 4.4 data IP kosong	44
Gambar 4.5. normalisasi data skenario1	49
Gambar 4.6. data training dan testing skenario1	50
Gambar 4.7 implementasi <i>naive bayes</i> untuk skenario 1	51
Gambar 4.8. implementasi normalisasi data skenario2	53
Gambar 4.9. data training dan testing skenario2	55
Gambar 4.10 implementasi <i>naive bayes</i> untuk skenario 2	55
Gambar 4.11. Langkah hasil Algoritma <i>Naive Bayes</i>	56
Gambar 4.12 prediksi untuk skenario 3	57
Gambar 4.13 Perancangan arsitektur <i>Backpropagation</i> -ANN skenario 4	58
Gambar 4.14 Prediksi dengan confusion matrix skenario 1 dan 2	63
Gambar 4.15 Prediksi dengan confusion matrix skenario 3 dan 4	64
Gambar 4.16 hasil arsitektur 4-4-1 <i>Backpropagation</i> -ANN	68

Gambar 4.17 hasil arsitektur 5-5-1 *Backpropagation*-ANN

71

Gambar 4.18 Rekomendasi penggunaan model

76



INTISARI

Data akademik mahasiswa yang terus bertambah dari tahun ke tahun tentunya berpotensi untuk menghasilkan informasi baru yang berguna bagi proses akademik selanjutnya bagi mahasiswa khususnya dan bagi institusi umumnya. Apabila banyaknya peserta didik yang indeks prestasi akademiknya buruk mengakibatkan masa studi menjadi lebih lama sehingga terjadi penumpukan dalam hal jumlah mahasiswa yang berakibat rasio dan keseimbangan kurang baik.

Dalam penelitian ini akan dilakukan prediksi kelulusan mahasiswa dengan menggunakan *Naive Bayes* dan akan dikomparasi dengan Algoritma *Backpropagation ANN* sehingga diharapkan bisa mendapatkan nilai prediksi yang baik. Algoritma ANN yang digunakan adalah *Backpropagation Artificial Neural Network* yaitu satu dari jaringan saraf tiruan (*Neural Network*) yang merupakan metode pelatihan yang terawasi (*Supervised Learning*) dengan jaringan multi layer dan memiliki ciri khusus meminimalkan *error* pada *output* yang dihasilkan oleh jaringan. Selain itu dalam penelitian ini dilakukan penambahan variable baru yaitu Nilai Ujian Nasional (UN) dan Status Tinggal Mahasiswa (status alamat asal mahasiswa) yang diharapkan ini bisa meningkatkan performa dari model yang didapatkan.

Berdasarkan analisa data, variabel yang digunakan untuk prediksi pada penelitian ini yaitu dengan algoritma *Naive Bayes* didapatkan hasil pengujian dengan penambahan nilai UN menghasilkan akurasi 77.35% dengan Algoritma *Naive Bayes* dan akurasi 78.04% untuk algoritma B-ANN.

Kata kunci : *Naive Bayes*, *Neural Network Backpropagation*, prediksi, MSE

ABSTRACT

Student academic data that continues to grow from year to year can certainly produce new information that is useful for the next academic process for students in particular and for institutions in general. If students who have poor academic achievements result in a longer study period so that it occurs in the number of students which results in poor ratios and balances.

In this study, student graduation predictions will be made using Naive Bayes and will be compared with the ANN Algorithm so that it is expected to get good predictions. The ANN Algorithm used is Backpropagation Artificial Neural Network, which is one of the Artificial Neural Networks (Neural Network) which is a supervised training (Supervised Learning) with a multi-layer network and has a special characteristic of interference with the output generated by the network. In addition, in this study, new variables were added, namely the National Examination Score (UN) and Student Residence Status which were expected to improve the performance of the model obtained.

Based on data analysis, the variables used for prediction in this study are the Naive Bayes algorithm, the test results with the addition of the UN value produce 77.35% accuracy with the Naive Bayes algorithm and 78.04% accuracy for the B-ANN Algorithm.

Keywords: Naive Bayes, Backpropagation Neural Network, prediction, accuracy

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Di masa sekarang ini teknologi basis data telah berkembang maka muncul kebutuhan untuk mendapatkan informasi yang lebih dari data yang tersimpan dalam basis data yaitu Pengetahuan (*Knowledge*). Informasi yang diinginkan tidak dapat diperoleh dengan mudah sebab volume data yang besar. Sehingga di butuhkan suatu metode untuk mendapatkan pengetahuan yang tidak terlihat di dalam data namun potensial untuk digunakan yaitu metode data mining.

Prestasi akademik menjadi hal yang utama yang menjadikan parameter keberhasilan suatu pendidikan. Salah satu indikator tercapainya tujuan tersebut adalah dari hasil prestasi akademik mahasiswa yang dinyatakan dengan Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK). Indeks Prestasi Semester merupakan nilai prestasi akademik mahasiswa dengan semua mata kuliah yang telah ditempuh untuk tiap semester tertentu. Indeks Prestasi Kumulatif merupakan prestasi akademik mahasiswa dengan menggabungkan semua mata kuliah yang telah ditempuh sampai suatu semester tertentu (Magdalena, 2012). Prestasi mahasiswa dalam setiap semester ini tentunya mempengaruhi masa studi mahasiswa sehingga berpengaruh terhadap kelulusan mahasiswa.

Data akademik mahasiswa yang terus bertambah dari tahun ke tahun tentunya berpotensi untuk menghasilkan informasi baru yang berguna bagi proses akademik selanjutnya bagi mahasiswa khususnya dan bagi institusi umumnya,

diantaranya adalah penelitian oleh Ni Komang Sri Julyantari dan I Ketut Dedy Suryawan (2014) tentang prestasi akademik yaitu berupa nilai indeks prestasi akademik yang baik dan masa studi yang tepat waktu. Akibatnya apabila banyaknya peserta didik yang indeks prestasi akademiknya buruk mengakibatkan masa studi menjadi lebih lama sehingga terjadi penumpukan dalam hal jumlah mahasiswa yang berakibat rasio dan keseimbangan kurang baik. Berdasarkan penelitian tersebut, jika prestasi akademik mahasiswa dan ketepatan kelulusan dapat diketahui lebih dini maka pihak program studi dapat melakukan tindakan yang diperlukan agar mahasiswa dapat mencapai prestasi akademik yang baik. Harapan akhirnya adalah semua mahasiswa dari berbagai latar belakang faktor dapat maksimal dalam meningkatkan prestasi akademiknya sehingga ketepatan kelulusan mahasiswa akan tinggi. Berdasarkan penjelasan di atas maka fokus penelitian ini adalah memprediksi ketepatan waktu kelulusan mahasiswa menggunakan metode data mining klasifikasi.

Studi prestasi akademik mahasiswa khususnya tentang ketepatan kelulusan mahasiswa sangatlah penting baik bagi mahasiswa, orang tua maupun program studi diantaranya adalah untuk mengukur kebaikan prestasi institusi juga guna perencanaan pembiayaan bagi orang tua mahasiswa. Telah banyak dikembangkan penelitian tentang prediksi kelulusan mahasiswa di antaranya penelitian yang dilakukan oleh Hendra, Mochammad Abdul Azis dan Suhardjono (2020) dengan judul Analisis Prediksi Kelulusan Mahasiswa Menggunakan *Decision Tree Berbasis Particle Swarm Optimization*. Atribut yang digunakan

dalam penelitian ini sebagai indikator prestasi mahasiswa adalah IPS semester 1, 2, 3, dan 4.

Aryasanti (2018) tentang prediksi kegagalan studi mahasiswa dengan menggunakan Algoritma *Naive Bayes* dan *Decision Tree s*. Metode yang digunakan melakukan pengklasifikasian studi mahasiswa dengan melihat kemungkinan/tingkat kegagalan studi mahasiswa. Hasil penelitian didapatkan bahwa Algoritma *Naive Bayes* memberikan hasil akurasi yang lebih baik dibandingkan *Decision Tree*.

Penelitian juga dilakukan Oleh Budiantara dkk (2020) yaitu berjudul Komparasi Algoritma *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbor* untuk Memprediksi Mahasiswa Lulus Tepat Waktu. Berdasarkan hasil penelitian didapatkan bahwa nilai akurasi pada algoritma C4.5 dan *Naive Bayes* hampir sama yaitu di atas 95%.

Berdasarkan beberapa penelitian bidang data mining sebelumnya, *Naive Bayes* merupakan 10 besar peringkat algoritma terbaik sehingga dapat digunakan dalam pembentukan suatu keputusan. Metode *Naive Bayes* merupakan salah satu metode yang dapat digunakan dalam hal pengambilan keputusan untuk mendapatkan hasil yang lebih baik pada suatu permasalahan klasifikasi (Ting, dkk, 2015). Sedangkan *Artificial Neural Network (ANN)* merupakan metode *deep learning* yang akhir-akhir ini banyak digunakan untuk prediksi dalam data mining.

Selain itu, di beberapa penelitian sebelumnya, dikatakan bahwa *Naive Bayes* dan ANN merupakan salah satu metode yang memberikan tingkat akurasi yang lebih baik dibandingkan algoritma pembandingnya. Untuk itu dalam

penelitian ini akan dilakukan prediksi kelulusan mahasiswa dengan menggunakan *Naive Bayes* dan akan dikomparasi dengan Algoritma ANN sehingga diharapkan bisa mendapatkan nilai prediksi yang baik. Algoritma ANN yang digunakan adalah *Backpropagation Artificial Neural Network* yaitu satu dari jaringan saraf tiruan (*Neural Network*) yang merupakan metode pelatihan yang terawasi (*Supervised Learning*) dengan jaringan *multi layer* dan memiliki ciri khusus meminimalkan *error* pada *output* yang dihasilkan oleh jaringan. Selain itu dalam penelitian ini dilakukan pengamatan dengan penambahan variable baru Nilai Ujian Nasional (UN) yaitu nilai akademik sebelum mahasiswa. Penambahan variabel baru dalam penelitian ini harapannya bisa menjadi salah satu faktor yang menjadi perhatian bagi pengelola suatu program studi khususnya dalam seleksi mahasiswa baru khususnya jenjang diploma.

1.2. Rumusan Masalah

Perumusan Masalah Berdasarkan latar belakang sebelumnya, maka dapat disusun rumusan masalah sebagai berikut :

- a. Berapa tingkat hasil akurasi klasifikasi yang telah dilakukan terhadap prediksi kelulusan mahasiswa yang ada di Program Studi Diploma 3 Teknik Informatika di salah satu perguruan tinggi di Surakarta khususnya Diploma 3 dengan *Naive Bayes*?
- b. Berapa tingkat hasil akurasi klasifikasi yang telah dilakukan terhadap prediksi kelulusan mahasiswa yang ada di Program Studi Diploma 3 Teknik Informatika di salah satu perguruan tinggi di Surakarta dengan *Bacpropagation ANN*?

1.3. Batasan Masalah

Beberapa hal yang menjadi batasan masalah dalam penulisan penelitian ini adalah :

- a. Sistem berlaku untuk data sesuai studi kasus yaitu salah satu perguruan Tinggi di Surakarta untuk data program studi Diploma 3 Teknik Informatika.
- b. Algoritma yang digunakan untuk membangun model prediksi adalah Algoritma *Naive Bayes* dan *Backpropagation Artificial Neural Network*.
- c. Dasar pengetahuan yang digunakan adalah nilai akademis mahasiswa yaitu variabel yang ada pada mahasiswa yaitu nilai Ujian Nasional (UN), jenis kelamin, Status Tinggal Mahasiswa (status alamat asal mahasiswa), IP Semester.
- d. Data dilakukan untuk mahasiswa angkatan 2012 - 2018 Program Studi Diploma 3 Teknik informatika pada salah satu perguruan Tinggi di Surakarta.
- e. Prediksi kelulusan dilakukan pada akhir semester yaitu prediksi ketepatan waktu lulusan mahasiswa.

1.4. Tujuan Penelitian

Dengan dilakukannya penelitian mengenai prediksi kelulusan mahasiswa ini diharapkan mencapai tujuan antara lain :

- a. Menghitung hasil akurasi klasifikasi yang telah dilakukan terhadap prediksi kelulusan mahasiswa yang ada di Program Studi Diploma 3 Teknik informatika pada salah satu perguruan Tinggi di Surakarta dengan *Naive Bayes*.

- b. Menghitung hasil akurasi klasifikasi yang telah dilakukan terhadap prediksi kelulusan mahasiswa yang ada di Program Studi Diploma 3 Teknik informatika pada salah satu perguruan Tinggi di Surakarta dengan *Backpropagation Artificial Neural Network*.

1.5. Manfaat Penelitian

Adapun manfaat yang didapat dengan dilakukannya penelitian ini antara lain sebagai berikut :

- a. Dapat digunakan sebagai acuan dan monitoring pembimbing akademik terhadap mahasiswa bimbingan sehingga bisa diketahui langkah antisipasi apabila ada suatu masalah.
- b. Dapat digunakan sebagai acuan dan monitoring Program Studi dalam hal prediksi kelulusan mahasiswa dan perkembangan prestasi mahasiswa sehingga bisa membantu manajemen program studi dalam menjaga kualitas penjaminan mutu akademik yaitu pantauan pada tahun pertama atau tahun kedua.
- c. Untuk mengetahui penerapan metode *Naïve Bayes* dan *Backpropagation Artificial Neural network* dalam prediksi kelulusan mahasiswa.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Syarlie dan Muin (2016) melakukan penelitian dengan judul Metode *Naive Bayes* Untuk Prediksi Kelulusan (Studi Kasus : Data Mahasiswa Baru Perguruan Tinggi). Penelitian ini akan menggunakan metode Algoritma *Naive Bayes* untuk melakukan prediksi peluang kelulusan mahasiswa baru. Kelulusan yang dimaksud adalah diterimanya seorang mahasiswa pada salah satu program studi di Perguruan Tinggi. Studi kasus dilakukan pada data mahasiswa Universitas Al Asyariah Mandar Sulawesi Barat.

Romadhona, S. Suprapedi, dan H. Himawan (2017) berjudul Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin dan Indeks Prestasi Menggunakan Algoritma *Decision Tree*. Penelitian ini bertujuan untuk menemukan pola untuk memprediksi kelulusan siswa tepat waktu menggunakan teknik dan model penambangan data untuk memprediksi periode penelitian yang panjang adalah algoritma pohon keputusan C4.5 hingga dibandingkan dengan algoritma *decision tree* (ID3) dan CHAID. Dari penelitian ini ditemukan bahwa prediksi siswa masa studi dipengaruhi oleh siswa yang masuk usia, jenis kelamin, IPK semester 1 sampai semester 4 dan yang paling berpengaruh adalah IPK mahasiswa semester 4 tepat waktu dengan nilai gain 0,340 dari semua atribut. Algoritma *Decision Tree* C4.5 mencapai akurasi tertinggi pada jumlah data 389 dengan nilai akurasi 91,51%.

Aryasanti (2018) tentang prediksi kegagalan studi mahasiswa dengan menggunakan Algoritma *Naive Bayes* dan *Decision Tree*. Penelitian ini bertujuan untuk mengetahui faktor apa saja yang dapat mempengaruhi tingkat kegagalan seseorang dalam menempuh studi di suatu Perguruan Tinggi. Penelitian ini menggunakan teknik klasifikasi data mining dengan dua metode sebagai komparasi untuk mendapatkan hasil yang terbaik yaitu dengan menggunakan metode *Naive Bayes* dan *Decision Tree*. Hasil penelitian menunjukkan bahwa metode *Decision Tree* dan *Naive Bayes* memiliki keakurasian yang hampir sama yaitu selisih 0.04%.

Muhamad Hanief Meninanda, dkk (2013) memprediksi masa studi sarjana dengan *Artificial Neural Network*. Variabel prediktor dari data akademis yang berpengaruh terhadap masa studi dan pembuatan model ANN untuk prediksi masa studi, model prediksi digunakan model *multiple regression* sebagai model pembanding. Data input 1289 mahasiswa dengan variabel ID mahasiswa, masa studi, kode mata kuliah, nama mata kuliah, pengambilan mata kuliah, nilai, bobot, sks, IPK, jumlah mata kuliah dan jumlah mengulang mata kuliah. Eksperimen penelitian menggunakan algoritma ANN Multilayer Perceptron, *Linear Regression* dan *Spearman Correlation*. Diperoleh uji beda menghasilkan nilai value sebesar 0.65 nilai tersebut lebih besar dari nilai kritis ($\alpha=0.05$) dengan tingkat kepercayaan 95% tidak ada perbedaan signifikan antara nilai masa studi aktual dengan nilai masa studi berdasarkan model prediksi.

Sri Widaningsih (2018) Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Program Studi Teknik Informatika Dengan Algoritma C4.5, *Naive Bayes*, Knn, Dan Svm. Tujuan dari penelitian ini

adalah bagaimana algoritma C4.5, *Naive Bayes*, Knn, Dan Svm dapat membantu memprediksi waktu dan nilai kelulusan mahasiswa Teknik Informatika dan dari keempat teknik tersebut manakah yang memberikan prediksi hasil yang paling baik. Dari hasil perbandingan terlihat bahwa Algoritma *Naive Bayes* memiliki nilai yang paling baik untuk semua kategori performansi dibandingkan dengan algoritma lainnya. Untuk nilai *accuracy* dan AUC nilai terbesar adalah yang terbaik, sedangkan untuk *error* adalah nilai yang terkecil. Nilai AUC untuk *Naive Bayes* dan C4.5 termasuk kedalam kategori "baik", sedangkan untuk algoritma SVM dan KNN termasuk kedalam kategori "cukup".

Penelitian juga dilakukan Oleh Agus Budiantara dkk (2020) yaitu berjudul Komparasi Algoritma *Decision Tree*, *Naive Bayes*. Prediksi ini bertujuan untuk menentukan faktor akademis yang berpengaruh terhadap masa studi dan membangun model prediksi terbaik dengan teknik Data Mining. Atribut yang digunakan untuk Klasifikasi Data Mining ada 11 atribut yaitu NPM, Jenis Kelamin, Usia, Jurusan, Kelas, Pekerjaan, Indeks Prestasi Semester 1, Indeks Prestasi Semester 2, Indeks Prestasi Semester 3, Indeks Prestasi Semester 4 dan Keterangan sebagai atribut hasil. Dari hasil evaluasi dan validasi yang telah dilakukan menggunakan *Tools Rapid Miner* diperoleh hasil akurasi dari Metode *Decision Tree* (C4.5) sebesar 98.04% pada pengujian ke 3. Akurasi Metode *Naive Bayes* sebesar 96.00% pada pengujian ke 4. Dan akurasi Metode *K-Nearest Neighbor* (K-NN) sebesar 90.00% pada pengujian ke 2.

Selanjutnya Eko Prasetyo Rohmawan (2018) melakukan penelitian untuk Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree*

dan *Artificial Neural Network*. Dalam penelitian ini fokus pada bagaimana prediksi kelulusan mahasiswa tepat waktu dengan data mining menggunakan Algoritma *Decision Tree* dan *Artificial Neural Network*. Penelitian ini menggunakan data mahasiswa yang sudah dinyatakan lulus sebagai data training dan mahasiswa yang belum lulus sebagai data testing. Penelitian ini bertujuan untuk mengetahui kelulusan mahasiswa tepat waktu atau tidak, yang diharapkan hasilnya dapat memberikan informasi dan masukan bagi pihak perguruan tinggi dalam membuat kebijakan demi perbaikan di masa yang akan datang. Penelitian ini membandingkan akurasi *Decision Tree* dan ANN dan didapatkan kesimpulan bahwa ANN memberikan nilai akurasi lebih baik dibandingkan *Decision Tree*.



2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur *review* dan posisi penelitian Analisis Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *Naive Bayes* dan *Backpropagation Artificial Neural Network*

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Metode <i>Naive Bayes</i> Untuk Prediksi Kelulusan (Studi Kasus : Data Mahasiswa Baru Perguruan Tinggi)	Syarlie dan Muin, Jurnal Ilmiah Ilmu Komputer, Vol. 2, No. 1, April 2016	Penelitian ini akan menggunakan metode Algoritma <i>Naive Bayes</i> untuk melakukan prediksi peluang kelulusan mahasiswa baru. Kelulusan yang dimaksud adalah diterimanya seorang mahasiswa pada salah satu program studi di Perguruan Tinggi. Studi kasus dilakukan pada data mahasiswa Universitas Al Asyariah Mandar Sulawesi Barat.	<i>Naive Bayes</i> dapat melakukan pengklasifikasian dengan metode probabilitas dan statistik, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Nilai presentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode <i>Naive Bayes Classification</i> . Implementasi <i>Naive Bayes</i> menggunakan aplikasi WEKA dapat menelusuri karakteristik atribut dari dataset dengan luaran Pilihan Lulus. Pengelompokkan Pilihan Lulus dilakukan berdasarkan atribut terpilih yaitu Program Studi, Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata.	Prestasi mahasiswa saat di bangku perkuliahan tentunya juga memegang peranan penting bagi hasil akhir prediksi kelulusan mahasiswa. Untuk itu, perlu dilakuakn penelitian adanya variable IP masuk dalam variable yang mempengaruhi nilai prestasi mahasiswa.	Berdasarkan penelitian Syarlie dan Muin ini mempertimbangkan variable Program Studi, Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata. Sebagai variable berpengaruh maka dalam penelitian saat ini akan dicoba untuk memasukkkn variable IP semester awal mahasiswa. Mengingat variabel karakteristik pembelajaran saat SMA agak berbeda dibandingkan dengan di perguruan tinggi yang menuntut mahasiswa aktif dan mandiri.

(Lanjutan) Tabel 2.1. Matriks literatur *review* dan posisi penelitian Analisis Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *Naive Bayes* dan *Backpropagation Artificial Neural Network*

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2	"Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin dan Indeks Prestasi Menggunakan Algoritma <i>Decision Tree</i> "	A. Romadhona, S. Suprapedi, and H. Himawan, Jurnal Teknologi Informasi Cyberku, 2017	Melakukan prediksi kelulusan mahasiswa tepat waktu dengan menggunakan <i>Decision Tree</i>	Didapatkan bahwa dengan metode <i>Decision Tree</i> nilai akurasi 91.51 % dengan prediksi kesalahan untuk mahasiswa yang lulus sebesar 389 data mahasiswa	Penelitian ini menerapkan Algoritma <i>Decision Tree</i> dan berdasarkan <i>record</i> data nilai akurasi masih bisa ditingkatkan maka selanjutnya perlu adanya peningkatan jumlah data <i>training</i> serta dibandingkan untuk algoritma yang lain sehingga bisa terlihat efektivitas/ akurasi dari algoritma	Sesuai dengan penelitian Romadhona dkk, ini didapatkan akurasi 91.51% maka dalam penelitian saat ini dilakukan upaya untuk meningkatkan akurasi yaitu dengan mengimplementasikan menggunakan algoritma <i>Naive Bayes</i> dan ANN yang diharapkan nanti bisa memberikan nilai akurasi yang lebih tinggi.
3	Sistem Komparasi <i>Naive Bayes</i> dan <i>Decision Tree</i> s untuk Menentukan Klasifikasi Kegagalan Studi Mahasiswa	A. Aryasanti, Jurnal TICOM, 2018	Melakukan komparasi Algoritma <i>Naive Bayes</i> dan <i>Decision Tree</i> pada prediksi kelulusan mahasiswa tepat waktu dengan menggunakan <i>Decision Tree</i>	Berdasarkan hasil penelitian diterapkan pada sarjana dan didapatkan bahwa nilai akurasi pada algoritma C4.5 dan <i>Naive Bayes</i> Hampir sama yaitu C4.5 sebesar 100% dan <i>baye</i> sebesar 99.56%.	Penelitian ini memberikan hasil bahwa <i>Naive Bayes</i> menghasilkan nilai akurasi yang hampir sama dengan <i>Decision Tree</i> . Dan faktor yang digunakan dalam penelitian ini adalah IPS 1, 2, 3, dan 4.	Penelitian sebelumnya menggunakan variabel berpengaruh yaitu IPS1, IPS2, IPS3 dan IPS4. Pada penelitian saat ini dengan memasukkan variabel rata rata UN/rapor dan status tinggal.

(Lanjutan) Tabel 2.1. Matriks literatur *review* dan posisi penelitian Analisis Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *Naive Bayes* dan *Backpropagation Artificial Neural Network*

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
4	Prediksi masa studi sarjana dengan <i>Artificial Neural Network</i>	Muhamad Hanief Meninanda, dkk, International Indonesian Journal, 2009	Penelitian ini bertujuan mengimplementasikan dan mengukur akurasi ANN pada data prediksi masa studi mahasiswa	Penelitian menggunakan algoritma ANN <i>Multilayer Perceptron</i> , <i>Linear Regression</i> dan <i>Spearman Correlation</i> . Diperoleh uji beda menghasilkan nilai p-value sebesar 0.65 nilai tersebut lebih besar dari nilai kritis (~ 0.05) dengan tingkat kepercayaan 95% tidak ada perbedaan signifikan antara nilai masa studi aktual dengan nilai masa studi berdasarkan model prediksi.	Pada penelitian ini memfokuskan pada penelitian untuk data level sarjana. Diharapkan hal ini juga berlaku pada tingkat diploma	Berdasarkan penelitian sebelumnya ini didapatkan bahwa ANN memberikan akurasi yang tinggi maka pada penelitian saat ini akan mengaplikasikan algoritma B-ANN dan diimplementasikan pada level diploma. Sehingga bisa memberikan pengembangan fokus penelitian.
5	Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Program Studi Teknik Informatika Dengan Algoritma C4.5, <i>Naive Bayes</i> , Knn, Dan Svm.	Sri Widaningsih. Jurnal techno Insentif, 2012	Melakukan komparasi Algoritma C4.5, <i>Naive Bayes</i> , Knn, Dan Svm pada prediksi kelulusan mahasiswa tepat waktu	Dari hasil perbandingan terlihat bahwa Algoritma <i>Naive Bayes</i> memiliki nilai yang paling baik untuk semua kategori performansi dibandingkan dengan algoritma lainnya.	Masih perlu ditambahkan variabel lain yang mempengaruhi prestasi mahasiswa. Sehingga diharapkan akan didapatkan nilai akurasi yang lebih tinggi.	Berdasarkan pada penelitian sebelumnya yaitu perlunya ditambahkan variabel yang lainnya yang berpengaruh terhadap nilai akurasi dari prediksi maka pada penelitian ini memasukkan variabel nilai rata rata UN/rapor dan status tinggal serta membandingkannya dengan ANN.

(Lanjutan) Tabel 2.1. Matriks literatur *review* dan posisi penelitian Analisis Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *Naive Bayes* dan *Backpropagation Artificial Neural Network*

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
6	Komparasi Algoritma <i>Decision Tree</i> , <i>Naive Bayes</i> Dan <i>K-Nearest Neighbor</i> Untuk Memprediksi Mahasiswa Lulus Tepat Waktu.	Agus Budiantara, Jurnal Ilmu Pengetahuan dan Teknologi Komputer, 2020.	Melakukan komparasi Algoritma <i>Bayes</i> dan <i>KNN</i> pada prediksi kelulusan mahasiswa tepat waktu	Berdasarkan hasil penelitian didapatkan bahwa nilai akurasi pada algoritma <i>C4.5</i> dan <i>Naive Bayes</i> Hampir sama yaitu di atas 95%.	Pada penelitian selanjutnya bisa diterapkan untuk komparasi dengan algoritma yang lain	Pada Penelitian ini menggunakan Algoritma <i>Bayesian</i> dan <i>NN</i> dan variabel yang digunakan adalah IP S1, 2, 3, dan 4 maka selanjutnya dikomparasikan nilai akurasi antara kedua algoritma tersebut dan ada tambahan variabel status tinggal dan jenis kelamin.
7	Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode <i>Decision Tree</i> Dan <i>Artificial Neural Network</i>	Eko Prasetyo Rohmawan, Jurnal Ilmiah Matrik, Vol.20 No.1, April 2018: 21-30	Penelitian membandingkan akurasi <i>Decision Tree</i> dan <i>NN</i>	<i>NN</i> memiliki akurasi lebih baik dibandingkan <i>Decision Tree</i>	Pada penelitian selanjutnya perlu dikomparasi dengan algoritma yang lainnya misal <i>Naive Bayesian</i> sebagai algoritma yang memberikan akurasi yang tinggi dalam penelitian yang lain	Dalam penelitian ini didapatkan <i>ANN</i> memiliki akurasi yang lebih baik dan pada penelitian sebelumnya <i>Naive Bayes</i> memberikan akurasi yang lebih baik dibandingkan algoritma terdahulu yang lain, maka dalam penelitian ini akan dibandingkan antara <i>Bayes</i> dan <i>ANN</i>

2.3. Landasan Teori

2.3.1. Data Mining

Dikenal sebagai *Knowledge Discovery in Database* (KDD), data mining merupakan sebuah proses untuk menemukan hubungan atau pola yang memiliki makna dengan cara mengamati data yang besar, dimana datanya disimpan dalam sebuah penyimpanan, dan mengelolanya dengan teknik statistik dan/ atau matematika guna mendapatkan manfaat data mining pada saat ini dapat mengatasi jumlah data sangat besar/ banyak, dimensi data yang sangat tinggi dan kumpulan data yang beragam. Metode-metode dalam pengolahan data mining yaitu :

1. *Predictive modelling* merupakan pengolahan data mining dengan melakukan prediksi. Tujuannya untuk membangun model prediksi suatu nilai dengan ciri-ciri tertentu. Contoh algoritmanya *Linear Regression*, *Neural Network*, *Support Vector Machine*, dan lain-lain.
2. *Association* (Asosiasi) merupakan teknik data mining yang mempelajari hubungan antar data. Contoh penggunaannya seperti untuk menganalisis perilaku mahasiswa yang datang terlambat. Contoh algoritmanya *FP-Growth*, *A Priori*, dan lain-lain.
3. *Clustering* (Klastering) atau pengelompokan merupakan teknik untuk mengidentifikasi data ke dalam suatu kelompok tertentu. Contoh algoritmanya *K-Means*, *K-Medoids*, *Self-Organisation Map (SOM)*, *Fuzzy C-Means*, dan lain-lain.
4. *Classification* merupakan teknik mengklasifikasikan data. Perbedaan mendasar dengan *clustering* adalah data, jika pada *clustering* tidak ada

variabel dependen, sedangkan *classification variable dependent* harus ada. Contoh algoritma yang menggunakan metode ini IDIPLOMA 3 dan *K-Nearest Neighbors*.

Prediksi merupakan suatu proses untuk memprediksi sesuatu yang belum terjadi atau memprediksi sesuatu yang akan terjadi di masa depan.

2.3.2. Metode Bayes dan Algoritma Naive Bayes

A. Probabilitas.

Definisi 2.1:

Bila kejadian B terjadi dalam m cara pada ruang sampel S yang terjadi dalam n cara, maka probabilitas kejadian B seperti pada persamaan 2.1 sebagai berikut:

$$P(B) = \frac{n(B)}{n(S)} \quad (\text{Persamaan 2.1})$$

Keterangan:

$n(B)$ = banyaknya anggota B

$n(S)$ = banyaknya anggota S

Probabilitas Bersyarat (conditional Probability)

Jika A dan B adalah dua peristiwa dari ruang sampel S, probabilitas bahwa B telah terjadi dinyatakan sebagai $P(A|B)$ atau P(A jika B), dan disebut sebagai probabilitas bersyarat (conditional probability) dari A jika B telah terjadi.

Definisi 2.2:

Jika A dan B adalah suatu peristiwa yang masing-masing merupakan himpunan bagian dari ruang sampel S, peluang bersyarat terjadinya A dengan syarat B ditulis dengan $P(A|B)$ yang dibaca peluang terjadinya peristiwa A jika diketahui peristiwa B telah terjadi, dan dapat dibaca singkat dengan peluang A dengan syarat B.

B. Teorema Bayes.

Banyak kejadian dalam kehidupan sehari-hari yang sulit diketahui dengan pasti apalagi kejadian di masa yang akan datang. Begitu juga dalam percobaan statistika, tidak bisa diketahui dengan pasti hasil-hasil yang akan muncul. Misalnya pada pelemparan sebuah dadu secara acak, tanpa rekayasa apa-apa maka ada derajat kepastian bahwa muka 1 dari dadu itu akan muncul. Derajat atau tingkat kepastian dari munculnya hasil percobaan statistik disebut probabilitas atau peluang. Suatu probabilitas dilambangkan dengan P. Setiap kasus pengambilan keputusan memerlukan informasi untuk menentukan peluang prior suatu peristiwa akan terjadi. Pengambilan keputusan dengan teorema Bayes, setiap informasinya mempunyai nilai tersendiri untuk menentukan peluang prior sebagai informasi baru. Peluang yang telah diperbaharui (devisi) disebut peluang posterior. Metode Bayes merupakan metode yang baik didalam mesin pembelajaran berdasarkan data training dengan menggunakan probabilitas bersyarat sebagai dasarnya. Teorema Bayes dikemukakan oleh seorang pendeta Presbyterian Inggris tahun 1763 yang bernama Thomas Bayes. Teorema Bayes ini kemudian disempurnakan oleh

Laplace. Teorema Bayes digunakan untuk menghitung probabilitas terjadinya suatu peristiwa berdasarkan pengaruh yang didapat dari hasil observasi. Teori ini menerangkan hubungan antara probabilitas terjadinya A dengan syarat peristiwa B telah terjadi dan probabilitas terjadinya peristiwa B dengan syarat peristiwa A telah terjadi.

C. Naive Bayes

Naive Bayes merupakan metode yang tidak memiliki aturan, *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data training. *Naive Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam data mining, algoritma ini juga dikenal dengan nama *Idiot's Bayes*, *Simple Bayes*, dan *Independence Bayes* (M Bramer, 2016). Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Klasifikasi *bayesian* didasarkan pada teorema *Bayes*, diambil dari nama seorang ahli matematika yang juga menteri Prebysterian Inggris, Thomas Bayes (1702-1761) (D. T. Larose and C. D. Larose, 2014). Klasifikasi *bayesian* memiliki kemampuan klasifikasi serupa dengan *Decision Tree* dan *Neural Network* (O. Maimon and L. Rokach, 2009).

Secara matematis *Naive Bayes* dapat dirumuskan seperti persamaan 2.2 berikut :

$$P(x|y) = P(y|x) P(x) P(y) \quad \text{(Persamaan 2.2)}$$

keterangan :

y : data dengan kelas yang belum diketahui

x : hipotesis data y merupakan suatu kelas spesifik

$P(x | y)$: probabilitas hipotesis x berdasar kondisi y (*posteriori probability*)

$P(x)$: probabilitas hipotesis x (*prior probability*)

$P(y | x)$: probabilitas y berdasarkan kondisi pada hipotesis x .

$P(y)$: probabilitas dari y .

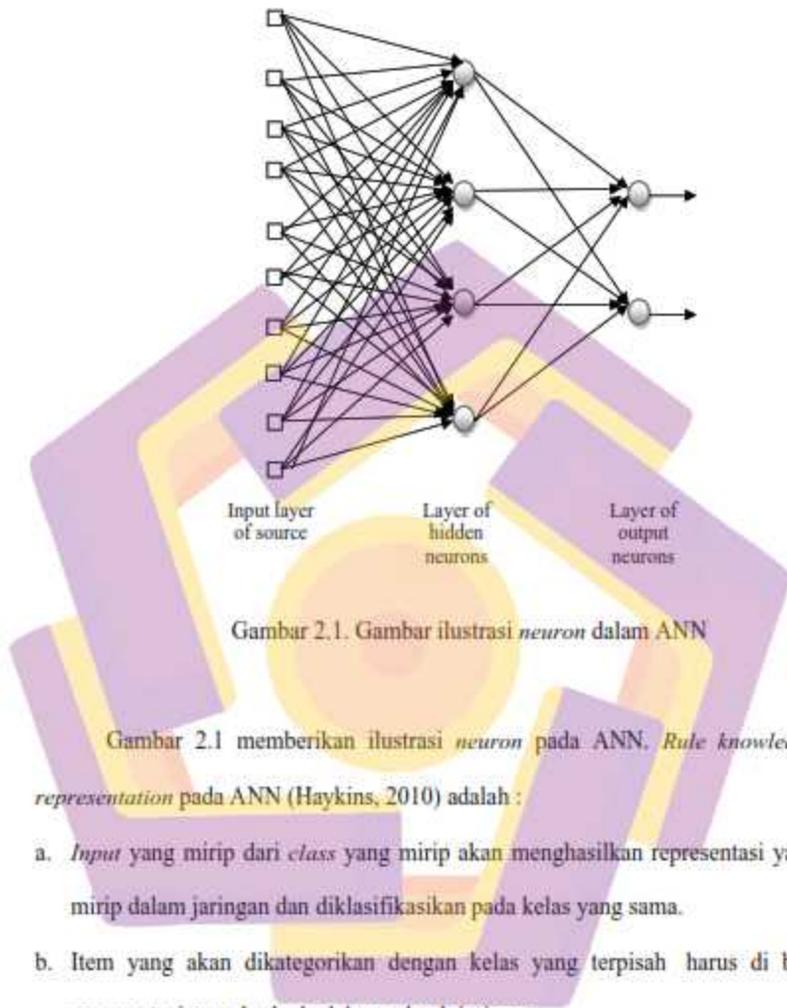
Data mining merupakan teknologi yang sangat berguna untuk membantu baik itu individu atau perusahaan tertentu dalam menemukan informasi yang sangat penting dari gudang data mereka. Data mining adalah perpaduan dari ilmu statistik, kecerdasan buatan, dan penelitian bidang *database* (J. Han, M. Kamber, and J. Pei, 2012). Data mining didefinisikan sebagai proses tentang memecahkan masalah dengan menganalisis data yang berada dalam *database* (I. H. Witten, E. Frank, and M. a. Hall, 2011). Nama data mining berasal dari kemiripan antara pencarian informasi yang bernilai dari *database* yang besar dengan menambang sebuah gunung untuk sesuatu yang bernilai (S. Sumathi and S. N. Sivanandam, 2006). Data mining didefinisikan sebagai proses pengenalan pola dalam data. Data mining, sering disebut *Knowledge Discovery in Database* (KDD), adalah suatu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menentukan keteraturan, pola atau hubungan dalam set data berukuran besar [I. H. Witten, E. Frank, and M. a. Hall, 2011). Keluaran dari data mining ini biasa dipakai untuk

memperbaiki dan membantu pengambilan keputusan di masa depan (S. Budi, 2007).

Salah satu cara untuk membantu dalam pengambilan keputusan adalah dengan adanya prediksi. Prediksi adalah sama dengan ramalan atau perkiraan. Menurut kamus besar bahasa Indonesia, prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan. Prediksi bisa berdasarkan metode ilmiah ataupun subjektif belaka. Sebagai contoh, prediksi cuaca selalu berdasarkan data dan informasi terbaru yang didasarkan pengamatan termasuk oleh satelit. Begitupun prediksi gempa, gunung meletus ataupun bencana secara umum. Namun, prediksi seperti pertandingan sepakbola, olahraga, dan lain-lain umumnya berdasarkan pandangan subjektif dengan sudut pandang sendiri yang memprediksinya (Puspitaningrum, 2006).

2.3.2. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) atau Jaringan Syaraf Tiruan merupakan sebuah teknik atau pendekatan pengolahan informasi yang terinspirasi oleh cara kerja sistem saraf biologis, khususnya pada sel otak manusia dalam memproses informasi. Elemen kunci dari teknik ini adalah struktur sistem pengolahan informasi yang bersifat unik dan beragam untuk tiap aplikasi. *Neural Network* terdiri dari sejumlah besar elemen pemrosesan informasi (*neuron*) yang saling terhubung dan bekerja bersama-sama untuk menyelesaikan sebuah masalah tertentu, yang pada umumnya adalah masalah klasifikasi ataupun prediksi (Fausett, 1994).



Gambar 2.1. Gambar ilustrasi *neuron* dalam ANN

Gambar 2.1 memberikan ilustrasi *neuron* pada ANN. *Rule knowledge representation* pada ANN (Haykins, 2010) adalah :

- Input* yang mirip dari *class* yang mirip akan menghasilkan representasi yang mirip dalam jaringan dan diklasifikasikan pada kelas yang sama.
- Item yang akan dikategorikan dengan kelas yang terpisah harus di beri representasi yang berbeda dalam sebuah jaringan.
- Jika ada fitur penting, maka harus disediakan jumlah *neuron* yang besar yang meliputi item pada sebuah jaringan.

- d. Informasi *prior* dan *invariances* harus di bangun ke dalam desain ANN ketika mereka tersedia, sehingga desain jaringan yang sederhana tidak harus mempelajarinya.

Cara kerja *Neural Network* dapat dianalogikan sebagaimana halnya manusia belajar dengan menggunakan contoh atau yang disebut sebagai *supervised learning*. Sebuah *Neural Network* dikonfigurasi untuk aplikasi tertentu, seperti pengenalan pola atau klasifikasi data, dan kemudian disempurnakan melalui proses pembelajaran. Proses belajar yang terjadi dalam sistem biologis melibatkan penyesuaian koneksi sinaptik yang ada antara neuron, dalam halnya pada *Neural Network* penyesuaian koneksi sinaptik antar *neuron* dilakukan dengan menyesuaikan nilai bobot yang ada pada tiap konektivitas baik dari *input*, *neuron* maupun *output*.

Neural Network memproses informasi berdasarkan cara kerja otak manusia. Dalam hal ini *Neural Network* terdiri dari sejumlah besar elemen pemrosesan yang saling terhubung dan bekerja secara paralel untuk memecahkan suatu masalah tertentu. Di sisi lain, komputer konvensional menggunakan pendekatan kognitif untuk memecahkan masalah, dimana cara pemecahan masalah haruslah sudah diketahui sebelumnya untuk kemudian dibuat menjadi beberapa instruksi kecil yang terstruktur. Instruksi ini kemudian dikonversi menjadi program komputer dan kemudian ke dalam kode mesin yang dapat dijalankan oleh komputer.

Neural Network dengan kemampuannya dapat digunakan untuk memperoleh pengetahuan dari data yang rumit atau tidak tepat, serta juga dapat

digunakan untuk mengekstrak pola dan mendeteksi tren yang terlalu kompleks untuk diperhatikan baik oleh manusia atau teknik komputer lainnya. Sebuah *Neural Network* yang telah terlatih dapat dianggap sebagai “ahli” dalam kategori pemrosesan informasi yang telah diberikan untuk dianalisa. Ahli ini kemudian dapat digunakan untuk menyediakan proyeksi terkait kemungkinan kondisi di masa mendatang.

Keuntungan lainnya dari penggunaan *Neural Network* termasuk :

- a. Pembelajaran adaptif : Kemampuan untuk belajar dalam melakukan tugas-tugas berdasarkan data yang diberikan.
- b. *Self-Organization* : Sebuah *Neural Network* dapat membangun representasi dari informasi yang diterimanya selama proses pembelajaran secara mandiri.
- c. Operasi *Real-Time* : Penghitungan *Neural Network* dapat dilakukan secara paralel, sehingga proses komputasi menjadi lebih cepat.

Fungsi dari *Neural Network* diantaranya adalah :

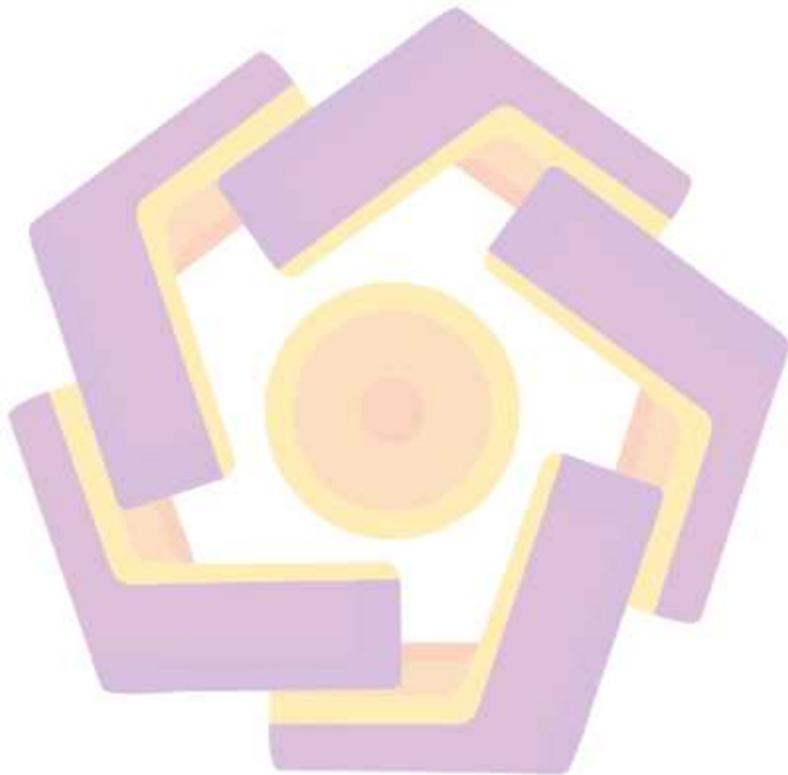
- a. Pengklasifikasian pola
- b. Memetakan pola yang didapat dari *input* ke dalam pola baru pada *output*
- c. Penyimpan pola yang akan dipanggil kembali
- d. Memetakan pola-pola yang sejenis
- e. Pengoptimasi permasalahan
- f. Prediksi

Seperti manusia yang belajar dari contoh, demikian juga dengan *Artificial Neural Network*. Sebuah *Artificial Neural Network* dikonfigurasi untuk aplikasi tertentu, seperti klasifikasi pengenalan pola atau data, melalui proses pembelajaran. Belajar dalam sistem biologis melibatkan penyesuaian terhadap koneksi sinaptik yang ada antara *neuron*. Hal ini berlaku juga pada *Artificial Neural Network* dimana memperoleh pengetahuan melalui pembelajaran, dan pengetahuan yang disimpan dalam koneksi antar-*neuron* yang kuat dan dikenal sebagai bobot sinaptik (Alyda,2006).

Metode pelatihan dibedakan menjadi 2 yaitu Pelatihan Terbimbing (*Supervised Training*) dan Pelatihan Tidak Terbimbing (*Unsupervised Training*). Metode pelatihan terbimbing dilakukan dengan memasukkan target *output* dalam data untuk proses *training*.

Salah satu algoritma pembelajaran yang dimiliki oleh *Artificial Neural Network* adalah *backpropagation*. *Backpropagation* merupakan algoritma pembelajaran yang digunakan oleh *perceptron* dengan banyak lapisan untuk mengubah setiap bobot yang terhubung dengan *neuron* pada *hidden layers*. Penggunaan *error* di dalam *backpropagation* bertujuan untuk mengubah nilai setiap bobot dalam arah mundur (*backward*). Sebelum mendapatkan *error* ini, terdapat tahap awal yang harus dilakukan yaitu tahap perambatan maju (*forward*). Pelatihan *backpropagation* meliputi tiga tahap. Tahap pertama adalah tahap maju (*forward*). tahap ini menghitung maju tahap *layer input* sampai tahap *layer output* dengan menggunakan fungsi aktivasi yang telah ditentukan. Tahap kedua adalah tahap mundur (*backward*), pada tahap ini selisih antara *output* jaringan dengan target

yang diinginkan merupakan kesalahan yang terjadi. Kesalahan tersebut dipropagasikan mundur, mulai dari garis yang terhubung langsung dengan setiap unit pada *layer output*. Kemudian tahap yang ketiga adalah tahap yang akan memodifikasi bobot untuk menurunkan tingkat kesalahan yang terjadi.



BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini merupakan penelitian *eksperimental comparative* yaitu menerapkan dan mengkomparasi nilai akurasi dari algoritma *Naive Bayes* dan ANN. Pada penelitian ini akan dilakukan menggunakan dua jenis data yaitu data kualitatif dan data kuantitatif.

a) Data kualitatif

Data kualitatif merupakan data yang berbentuk kata-kata, data kualitatif dalam data yang akan digunakan pada penelitian ini yaitu jenis kelamin, Status Tinggal Mahasiswa (status alamat asal mahasiswa) yang terkategori Solo atau di luar area Solo, waktu kelulusan tepat atau tidak tepat waktu.

b) Data kuantitatif

Data kuantitatif merupakan data yang berbentuk angka dan dapat dihitung nilainya dengan statistik ataupun matematika secara pasti. Data kuantitatif merupakan data yang menunjukkan probabilitas, dalam data yang digunakan untuk penelitian yaitu nilai sebelum masuk PT yaitu nilai UN/rapor, IP semester 1, IP semester 2.

3.1.1 Sifat Penelitian

Sifat dari penelitian ini adalah probabilistic yaitu meramalkan atau memprediksi suatu kejadian yang akan datang. Penelitian ini menggunakan data

yang tersedia sebagai dasar pengetahuan guna meramalakan keadaan bagi prestasi mahasiswa yaitu ketepatan kelulusan mahasiswa.

Dalam penelitian ini menggunakan dua metode yang digunakan tergantung dari permasalahan yang akan diselesaikan yaitu metode Naive Bayes dan Backpropagation Artificial Neural Network. Dalam penelitian ini bertujuan untuk membandingkan hasil prediksi yang baik antara prediksi dengan algoritma naïve bayes dan B-ANN. Penelitian dengan algoritma B-ANN diharapkan mampu membuktikan bahwa B-ANN memiliki tingkat akurasi yang lebih baik dibandingkan metode naïve bayes.

3.1.2 Pendekatan Penelitian

Dalam penelitian ini pendekatan yang dilakukan yaitu kuantitatif. Metode penelitian kuantitatif adalah sebagai metode penelitian yang berlandaskan pada filsafat *positivisme* digunakan untuk meneliti pada populasi atau sampel tertentu, teknik pengambilan sampel pada umumnya dilakukan secara *random*, pengumpulan data menggunakan instrumen penelitian analisis data bersifat kuantitatif atau statistik dengan tujuan untuk menguji hipotesis yang telah diterapkan.

3.2. Metode Pengumpulan Data

Dalam pelaksanaan penelitian mengenai prediksi kelulusan mahasiswa ini dibutuhkan berbagai komponen penelitian demi kelancaran penelitian. Salah satu komponen pendukung yang tidak bisa dipisahkan adalah data sebagai bahan ujicoba. Sebagai salah satu bahan ujicoba dalam penelitian ini digunakan dataset

data akademis mahasiswa di salah satu perguruan Tinggi di surakarta yaitu mahasiswa Program Studi Diploma 3 Teknik Informatika.

Data yang digunakan adalah data mahasiswa 2012-2018 yaitu data mahasiswa yang mengambil studi dan tercatat di sistem akademik , dalam penelitian diambil 859 data mahasiswa.

Tahap ini dilakukan pengambilan data mengenai data akademik mahasiswa Program Studi Diploma 3 Teknik informatika salah satu Perguruan Tinggi di Surakarta, dalam penelitian ini dilakukan dengan menggunakan 2 metode, yang dapat dijabarkan sebagai berikut :

a. Metode Wawancara (Interview)

Metode ini dilakukan dengan cara Penulis melakukan tanya jawab secara langsung kepada pihak pengelola Program Studi Diploma 3 teknik informatika Sekolah Vokasi di salah satu perguruan Tinggi di Surakarta khususnya bagian akademik.

b. Metode Observasi

Metode ini dilakukan dengan cara Penulis melakukan pengamatan atau observasi terhadap objek yang diteliti yaitu sistem akademik pada salah satu perguruan Tinggi di Surakarta.

3.3. Metode Analisis Data

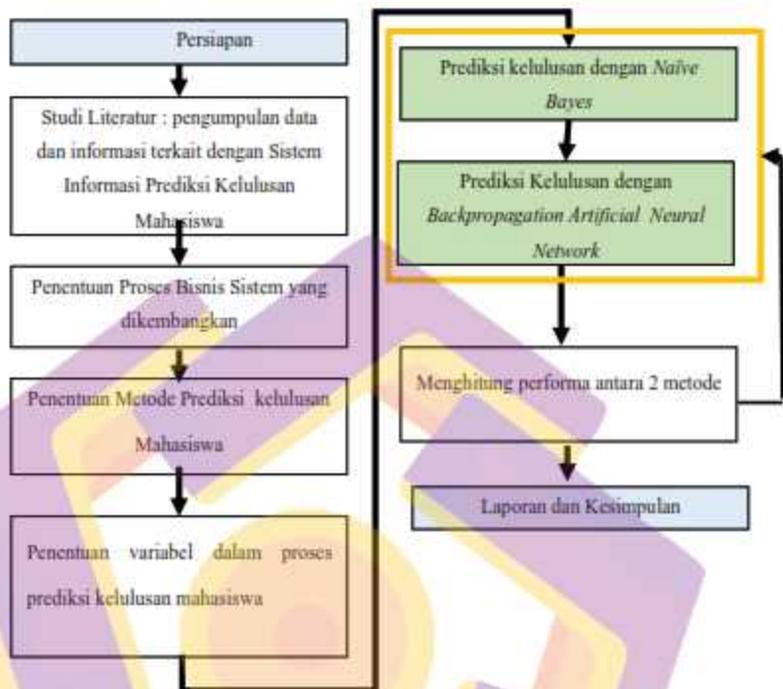
Ada beberapa tahapan dalam proses dalam melakukan analisa data guna proses prediksi. Pemodelan serta algoritma yang digunakan dalam penelitian ini mengambil referensi dari beberapa penelitian yang sudah ada sebelumnya dengan

tambahan variabel untuk meningkatkan tingkat akurasi dalam proses prediksi kelulusan mahasiswa sehingga model prediksi bisa digunakan di institusi terkait.

3.4. Alur Penelitian

Dalam penelitian ini menggunakan data kelulusan mahasiswa suatu Perguruan Tinggi Negeri di Surakarta. Data yang diolah sebanyak 859 data mahasiswa yaitu berasal dari satu jurusan yaitu jurusan Diploma 3 Teknik Informatika. Penelitian ini adalah penelitian *experiment* yang melibatkan penyelidikan tentang perlakuan pada parameter dan variabel yang semuanya ditentukan oleh peneliti. Dasar pengetahuan yang digunakan adalah nilai akademis mahasiswa yaitu variabel yang ada pada mahasiswa yaitu nilai Ujian sebelum menjadi mahasiswa yaitu nilai UN atau nilai rapor, jenis kelamin, Status Tinggal Mahasiswa (status alamat asal mahasiswa), IP Semester.

Model yang dibangun dengan prediksi apakah mahasiswa lulus tepat waktu atau tidak berdasarkan variabel inputan dalam penelitian ini. Alur penelitian terlihat pada gambar 3.1.



Gambar 3.1. Alur penelitian

3.4.1 DATA TRAINING

Data yang digunakan adalah data mahasiswa Program Studi Diploma 3 Teknik Informatika suatu Perguruan Tinggi Negeri di Surakarta. Berdasarkan data yang didapatkan untuk mahasiswa angkatan 2012-2018 dijadikan sebagai dataset yang selanjutnya digunakan untuk menentukan model prediksi untuk kelulusan mahasiswa.

Model yang digunakan dalam penelitian ini mengikuti tahapan yang ada pada proses *knowledge discovery in database* (KDD) (kusrini dan Luthfi, 2005) yaitu:

1. Selection

Pada tahap ini dilakukan seleksi data lulusan yang terdiri dari variabel-variabel prediktor dan satu target variabel. Variabel target yaitu klasifikasi lulusan yang lulus tepat waktu yaitu 3 tahun atau kurang dan memiliki nilai IPK minimal 3,00. Sedangkan variabel-variabel prediktor yaitu, jenis kelamin, nilai UN, Status Tinggal Mahasiswa (status alamat asal mahasiswa), indeks prestasi semester 1, 2. Data training diambil dari dataset sebanyak 859 data.

2. Preprocessing

Data yang diambil sesuai dengan banyaknya lulusan di Program Studi Diploma 3 Teknik Informatika. Dari data yang diambil dilakukan pembersihan data apabila terdapat data yang hilang, data ganda atau bersifat *outlier*.

3. Transformation

Setelah data bersih dari kesalahan, selanjutnya dilakukan transformasi pada data sesuai dengan jenis data pada tahapan transformasi dimana jenis akan dikelompokkan menjadi data yang bersifat kategori. Berikut pada Tabel 3.1 adalah kategori untuk variabel prediktor dan variabel target.

Tabel 3.1: Transformasi data variabel

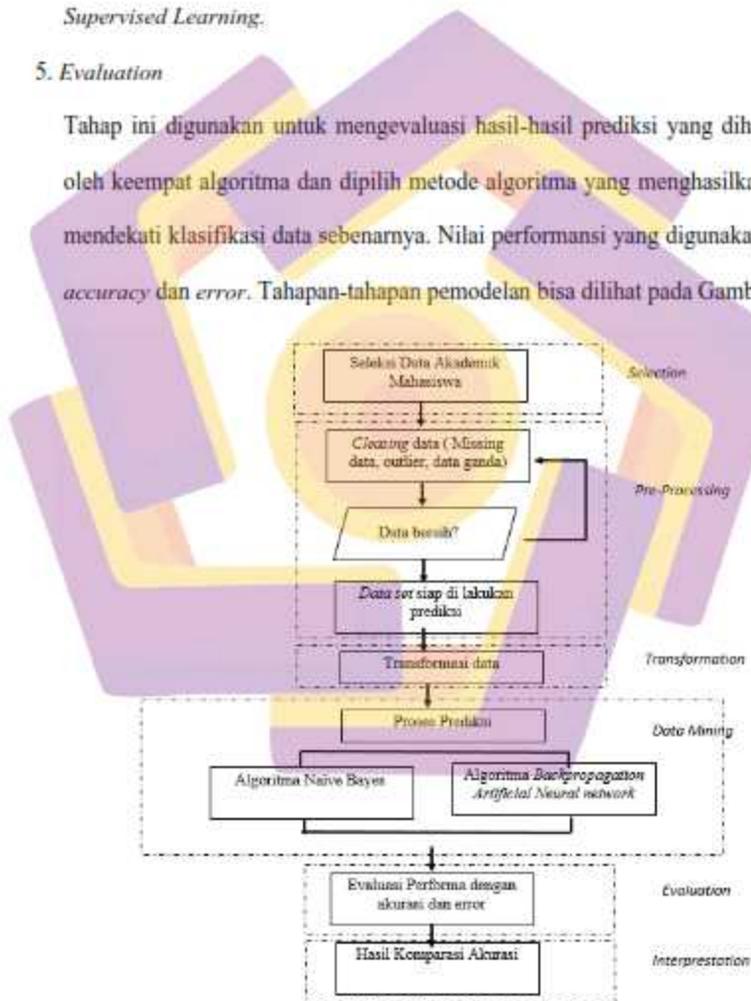
No	Variabel target	Kategori
1	Jenis Kelamin	Perempuan Laki Laki
2	Status Tinggal	Solo Luar Solo
3	Status Kelulusan Mahasiswa	Lulus tepat waktu Lulus tidak tepat waktu

4. Data Mining

Pada tahap ini dilakukan pemilihan teknik data mining yang sesuai. Untuk fungsi klasifikasi digunakan Algoritma *Naive Bayes* dan ANN. Karena klasifikasi merupakan *Supervised Learning* maka berikut ini adalah tahapan dalam model *Supervised Learning*.

5. Evaluation

Tahap ini digunakan untuk mengevaluasi hasil-hasil prediksi yang dihasilkan oleh keempat algoritma dan dipilih metode algoritma yang menghasilkan nilai mendekati klasifikasi data sebenarnya. Nilai performansi yang digunakan yaitu *accuracy* dan *error*. Tahapan-tahapan pemodelan bisa dilihat pada Gambar 3.2.



Gambar 3.2. Tahapan pemodelan dalam penelitian

Proses cleaning data atau pembersihan dilakukan terhadap data yang redundan atau ganda, inkonsisten, *missing value* dan *outlier data*. Hal tersebut perlu dilakukan agar tidak mempengaruhi performa proses klasifikasi yang dilakukan. Pembersihan data yang dilakukan juga menghapus fitur yang tidak diperlukan dalam proses klasifikasi nantinya.

3.4.2 PEMODELAN DATA

Berdasarkan Gambar 3.2, langkah perhitungan dalam Algoritma *Naive Bayes* dan *Backpropagation Artificial Neural Network* adalah :

a. Algoritma *Naive Bayes*

Klasifikasi *Bayes* secara matematis dapat dirumuskan seperti persamaan 3.1 berikut :

$$P(x|y) = P(y|x) P(x) P(y) \quad (\text{Persamaan 3.1})$$

keterangan :

y : data dengan kelas yang belum diketahui

x : hipotesis data y merupakan suatu kelas spesifik

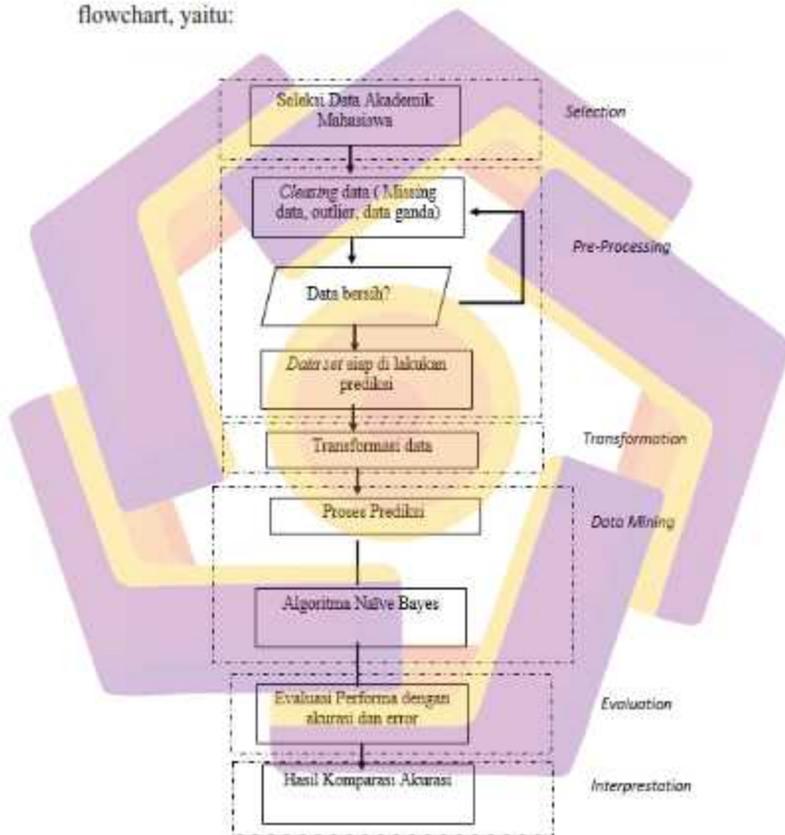
$P(x | y)$: probabilitas hipotesis x berdasar kondisi y (*posteriori probability*)

$P(x)$: probabilitas hipotesis x (*prior probability*)

$P(y | x)$: probabilitas y berdasarkan kondisi pada hipotesis x

$P(y)$: probabilitas dari y

Setelah melakukan semua tahap pemrosesan data, maka akan dihasilkan data training. Data training merupakan data yang akan digunakan untuk melakukan proses perhitungan menggunakan metode *Naive Bayes*. Berikut gambaran proses perhitungan menggunakan metode *Naive Bayes* dalam sebuah flowchart, yaitu:

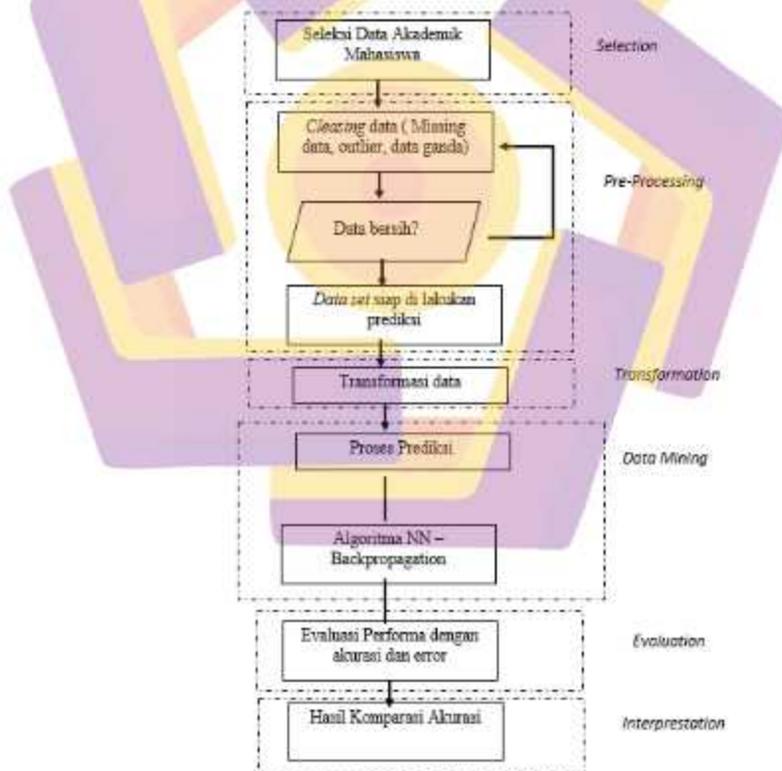


Gambar 3.3. Algoritma *Naive Bayes*

Dari tahapan pada gambar 3.3 tersebut, dilakukan perhitungan Algoritma *Naive Bayes* menggunakan data training dan data testing sebagai data uji.

b. Algoritma *Backpropagation Artificial Neural Network*

Algoritma *Backpropagation Artificial Neural Network* (B-ANN) merupakan algoritma pelatihan terbimbing yang mempunyai banyak lapisan. B-ANN menggunakan *error output* untuk mengubah nilai bobot-bobotnya dalam arah mundur (*backward*). Gambar 3.4 merupakan alur untuk algoritma B-ANN.



Gambar 3.4 Flowchart algoritma B-ANN

Berikut langkah langkah dalam algoritma *Backpropagation Artificial Neural Network* (Cynthia, E. P., dan Ismanto, E., 2017) :

Tahap Pertama : Propagasi maju (*Forward*)

Langkah 0 : **Instalasi semua bobot** dengan bilangan acak kecil

Langkah 1 : Jika kondisi penghentian belum terpenuhi, lakukan langkah 2-9

Langkah 2 : Untuk setiap pasangan data pelatihan lakukan langkah 3-8

Langkah 3 : Tiap unit masukan menerima sinyal dan meneruskannya ke unit tersembunyi di atasnya

Langkah 4 : Hitunglah semua *output* di unit tersembunyi tersebut z_j , ($j=1,2,\dots,p$). Perhatikan rumus pada persamaan 3.2 dan persamaan 3.3 dibawah ini :

$$z_netj = Vj0 + \sum xjvj1 \quad (\text{Persamaan 3.2})$$

$$zj = f(z_netj) = \frac{1}{1+e^{-z_netj}} \quad (\text{Persamaan 3.3})$$

Langkah 5 : Hitunglah semua keluaran jaringan di unit y_k ($k = 1,2, \dots, m$), perhatikan rumus pada persamaan 3.4 dan persamaan 3.5 dibawah ini :

$$y_netk = W_{k0} + \sum_{j=1}^p Z_j W_{kj} \quad (\text{Persamaan 3.4})$$

$$y_k = f(z_netk) = \frac{1}{1+e^{-y_netk}} \quad (\text{Persamaan 3.5})$$

Tahap Kedua : Propagasi Mundur

Langkah 6 : Hitung faktor δ unit keluaran berdasarkan error pada setiap y_k ($k=1,2, \dots, m$) Perhatikan rumus berikut ini pada persamaan 3.6:

$$\delta_k = (t_k - y_k) f'(y_{netk}) = (t_k - y_k) y_k (1 - y_k) \quad (\text{Persamaan 3.6})$$

δ_k merupakan unit kesalahan yang akan dipakai dalam perubahan bobot layer di bawahnya (langkah 7). Hitung suku perubahan bobot w_{kj} dengan laju percepatan α . Perhatikan rumus persamaan 3.7 dibawah ini:

$$\Delta w_{kj} = \alpha \delta_k z_j; \quad (\text{Persamaan 3.7})$$

Keterangan $k = 1,2, \dots, m$; $j = 0,1,2, \dots, p$

Langkah 7 : Hitung faktor δ unit tersembunyi berdasarkan *error* pada setiap unit tersembunyi z_j ($j = 1,2, \dots, p$). Perhatikan rumus persamaan 3.8 berikut:

$$\delta_{netj} = \sum_{k=1}^m \delta_j w_{kj} \quad (\text{Persamaan 3.8})$$

faktor δ unit tersembunyi :

$$\delta_j = \delta_{netj} f'(z_{netj}) = \delta_{netj} z_j (1 - z_j)$$

Hitung suku perubahan bobot v_{ji} (yang di pakai nanti untuk merubah bobot v_{ji}) persamaan 3.9.

$$\Delta v_{kj} = \alpha \delta_j x_i ; j = 1, 2, \dots, p ; i = 0, 1, 2, \dots, n \text{ (Persamaan 3.9)}$$

- **Tahap Ketiga : Perubahan bobot**

Langkah 8 : Hitunglah semua perubahan pada bobot. Rumus perubahan bobot garis yang menuju ke unit keluaran sebagai berikut pada persamaan 3.10 :

$$w_{kj}(\text{baru}) = w_{kj}(\text{lama}) + \Delta w_{kj} \text{ (Persamaan 3.10)}$$

dengan $(k = 1, 2, \dots, m ; j = 0, 1, 2, \dots, p)$

Perubahan bobot garis yang menuju ke unit tersebut pada persamaan 3.11 :

$$v_{ji}(\text{baru}) = v_{ji}(\text{lama}) + \Delta v_{ji} \text{ (Persamaan 3.11)}$$

dengan $(j = 1, 2, \dots, p ; i = 0, 1, 2, \dots, n)$

Langkah 9 : Menguji apakah kondisi berhenti sudah terpenuhi.

Kondisi berhenti ini terpenuhi jika nilai kesalahan yang dihasilkan lebih kecil dari nilai kesalahan yang dihasilkan.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Karakteristik dan Manajemen Data

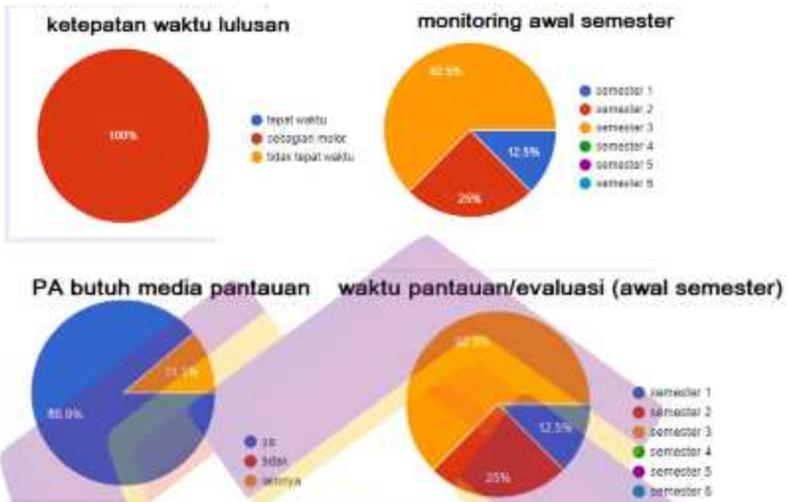
Tahapan yang dilakukan dalam penelitian ini meliputi action planning, pengumpulan data, preprosesing data. Pada penelitian ini yang dilakukan adalah dengan mencari data yang dibutuhkan.

4.1.1 Action Planning

Tahapan yang dilakukan pada *action planning* yaitu penyusunan rencana tindakan yang tepat untuk menyelesaikan masalah. Pada penelitian ini yang dilakukan adalah dengan mencari data yang dibutuhkan, praproses data, menentukan parameter dan algoritma yang digunakan.

4.1.1.1 Pengumpulan Data

Data yang digunakan dalam implementasi algoritma *Naive Bayes* dan *Backpropagation - Artificial Neural Network* ini adalah data mahasiswa pada Program Studi Diploma 3 Teknik Informatika Di salah satu Perguruan Tinggi Negeri di Surakarta, yaitu sebanyak 859 data pada angkatan 2012-2018 seperti dalam Tabel 4.1.



Gambar 4.1 Potret ketepatan kelulusan mahasiswa oleh PA

Tabel 4.1 Dataset mahasiswa diploma 3

ID	TG_LUS	ANGK ATAN	TAHU N	SEM ESTE R	IP_S EM	Mhs_stt	mhsab tanilai
2		2012	2012	A	2.85	KAB SUKOHARJO	49.55
2		2012	2012	B	3.05	KAB SUKOHARJO	49.55
3	5 Agustus 2015	2012	2012	A	3.2	KAB NGAWI	50.6
3	5 Agustus 2015	2012	2012	B	3.33	KAB NGAWI	50.6
3	5 Agustus 2015	2012	2013	A	2.92	KAB NGAWI	50.6
3	5 Agustus 2015	2012	2013	B	3.05	KAB NGAWI	50.6
3	5 Agustus 2015	2012	2014	A	3.82	KAB NGAWI	50.6
3	5 Agustus 2015	2012	2014	B	2.67	KAB NGAWI	50.6
4	5 Agustus 2015	2012	2012	A	3.3	KAB NGAWI	53.15
4	5 Agustus 2015	2012	2012	B	3.76	KAB NGAWI	53.15
4	5 Agustus 2015	2012	2013	A	3.46	KAB NGAWI	53.15
4	5 Agustus 2015	2012	2013	B	3.33	KAB NGAWI	53.15
4	5 Agustus 2015	2012	2014	A	3.73	KAB NGAWI	53.15
4	5 Agustus 2015	2012	2014	B	3.8	KAB NGAWI	53.15
5	5 Agustus 2015	2012	2012	A	3.35	KAB KLATEN	48.1
5	5 Agustus 2015	2012	2012	B	3.33	KAB KLATEN	48.1
5	5 Agustus 2015	2012	2013	A	3	KAB KLATEN	48.1

(Lanjutan) Tabel 4.1 Dataset mahasiswa diploma 3

ID	TG_LUS	ANGKAT AN	TAH UN	SEMESTER	IP_S EM	Mhs_ett	mhsebta nilai
5	5 Agustus 2015	2012	2013	B	3.19	KAB KLATEN	48.1
5	5 Agustus 2015	2012	2014	A	3.5	KAB KLATEN	48.1
5	5 Agustus 2015	2012	2014	B	2.87	KAB KLATEN	48.1
6		2012	2012	A	3.4	KAB SUKOHARJO	43.6
6		2012	2012	B	3.24	KAB SUKOHARJO	43.6
6		2012	2013	A	3	KAB SUKOHARJO	43.6
6		2012	2013	B	3.38	KAB SUKOHARJO	43.6
6		2012	2014	A	2.27	KAB SUKOHARJO	43.6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
883		2018	2018	A	3.41	KAB BEKASI	25.1
883		2018	2018	B	0	KAB BEKASI	25.1
883		2018	2019	A	0	KAB BEKASI	25.1
883		2018	2020	A	1.5	KAB BEKASI	25.1

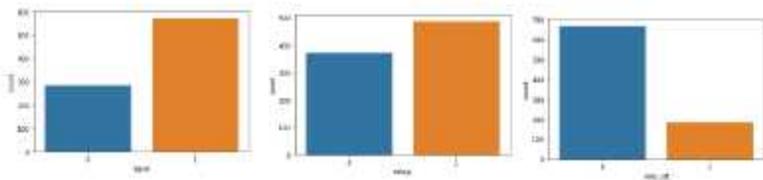
Dataset yang digunakan dalam penelitian ini adalah dataset salah satu program studi Diploma 3 salah satu PTN di Surakarta sebanyak 859 Record data. Data

Atribut yang digunakan adalah :

1. Id mahasiswa
2. Jenis Kelamin
3. Status tinggal
4. Nilai semester IPS1
5. Nilai Semester IPS2
6. Nilai UN/rapor mahasiswa
7. Status Kelulusan

Mahasiswa Diploma 3 Teknik Informatika menempuh studi 3 tahun yang mana ketepatan waktu lulusan dihitung berdasarkan waktu tempuh mahasiswa selama 6 semester yaitu sampai dengan bulan agustus di tahun kalender akademik berjalan atau 37 bulan. Ketepatan waktu lulusan dikategorikan dalam 2 kategori yaitu tepat waktu dan terlambat. Sedangkan IP semester dicatat untuk IP per semester mahasiswa.

Dalam penelitian ini dilakukan tambahan variabel yaitu Nilai sebelum masuk perguruan tinggi yaitu nilai UN atau nilai rapor. Data didapatkan dari data akademik Program Studi Diploma 3 Teknik Informatika Sekolah Vokasi Di Salah Satu Perguruan Tinggi Negeri Di Surakarta . Nilai UN adalah nilai yang dihasilkan dari Ujian Nasional yang diselenggarakan secara nasional pada tingkat akhir sekolah dasar, sekolah menengah pertama, dan sekolah menengah atas. Status Tinggal merupakan variabel yang mengidentifikasi tempat tinggal asal mahasiswa yaitu Solo atau Luar Solo. Variabel ini dimasukkan dalam penelitian akan dilihat apakah daerah mempengaruhi prestasi mahasiswa dalam ketepatan studi. Berdasarkan hasil observasi didapatkan pola data dari dataset seperti terlihat dalam Gambar 4.2 yaitu mayoritas mahasiswa lulus tepat waktu dan sekitar 30% mahasiswa tidak lulus tepat waktu. Hal ini tentunya menjadi perhatian khusus bagi program studi untuk bisa meningkatkan kualitas lulusan agar bisa tepat waktu. Sedangkan dari jenis kelamin terlihat bahwa mayoritas mahasiswa berjenis kelamin laki laki.



Gambar 4.2 pola persebaran data mahasiswa

Pengumpulan data didapatkan dari data yaitu sistem akademik mahasiswa Program Studi Diploma 3 Teknik Informatika di Salah Satu Perguruan Tinggi Negeri di Surakarta.

4.1.1.2 Praproses Data

Analisa dan prediksi model data diawali dengan persiapan data, yaitu pengecekan data, integrasi data.

a. *Missing Value*

Pada tahap menghilangkan *missing value*, data akademik yang jauh dari data seharusnya akan cek atau dihapus secara manual atau diberikan nilai 0, misalnya data IPS-1, IPS-2, nilai UN seperti pada gambar 4.3.

```

## cek data kosong pada masing-masing kolom
print("Seluruh data: ", df_mhs.ID.count())
df_mhs.isnull().sum()
## banyak data kosong di kolom ip semester 2 ke atas

```

```

Seluruh data: 859
ID              0
mhsjk          0
mhsebtanilai  4
kos            0
durasi lulus   0
tepat         0
ip1           0
ip2          45
ip3          70
ip4          81
ip5          89
ip6          93
dtype: int64

```

Gambar 4.3 Penanganan data kosong



Gambar 4.4 data IP kosong

Berdasarkan deskripsi data, Gambar 4.3 dan 4.4 terlihat bahwa mulai semester 2 ada data IP yang bernilai kosong dan ada kenaikan nilai IP kosong dari semester ke semester berikutnya. Hal ini menggambarkan bahwa adanya potensi mahasiswa tidak menyelesaikan studi dengan tepat waktu atau bahkan berhenti di tengah semester.

b. Manajemen data

Setelah data bersih dari *missing value* maka dilakukan integrasi data, yaitu menggabungkan data hasil *cleaning* data, sehingga selanjutnya data sudah siap untuk digunakan dalam analisa data. Manajemen data juga dilakukan terhadap data yang belum terkategori misla pada status tinggal, pengolahan data yang belum terformat menjadi data tabular.

Tabel. 4.2 Data set yang digunakan

	ID	m hs jk	mhs ebta nilai (UN/rapor)	statu sting gal	ip1	ip2	kelulusan
0	2	1	49.55	1	2.85	3.05	TEPAT
1	3	1	50.6	0	3.2	3.33	TERLAMBAT
2	4	1	53.15	0	3.3	3.76	TERLAMBAT
3	5	1	48.1	0	3.35	3.33	TERLAMBAT
4	6	1	43.6	1	3.4	3.24	TEPAT
5	7	0	52.5	0	3.1	3.48	TERLAMBAT
6	8	1	47	0	3.6	3.38	TERLAMBAT
7	9	1	53.15	0	2.85	3.1	TERLAMBAT
8	10	0	49	0	3.1	3.14	TERLAMBAT
9	11	0	48	0	2.85	3.14	TERLAMBAT
10	12	1	0	0	3.2	3.33	TERLAMBAT
11	13	1	35.35	1	3.1	3.24	TERLAMBAT
12	14	1	47	0	3.35	3.24	TERLAMBAT

(Lanjutan) Tabel. 4.2 Data set yang digunakan

13	15	1	50.75	0	3.6	3.9	TERLAMBAT
14	16	1	36.8	0	2.85	2.47	TEPAT
15	17	1	55	0	0.62		TEPAT
16	18	1	48.55	0	3.17	3.33	TERLAMBAT
17	19	1	48.5	0	3.35	3.19	TERLAMBAT
18	20	0	51.65	0	3.35	3.29	TERLAMBAT
19	21	0	51.75	0	3.4	3.24	TERLAMBAT
20	22	0	50.05	0	2.4	2.57	TERLAMBAT
21	24	1	51	0	3.25	3.14	TERLAMBAT
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
859	882	1	31.25	1	3.65	3.86	TERLAMBAT

Data yang digunakan adalah data mahasiswa 2012-2018 yaitu mahasiswa aktif maupun saat ini statusnya sudah alumni, dalam penelitian nanti diambil data yaitu data mahasiswa yang sudah lulus sebagai data training dan sebagian sebagai data testing. Sebelum dataset digunakan sebagai umpan masukan dalam tahap *training* maka dilakukan tahap awal yaitu *preprocessing* data. Tahapan data *preparation* merupakan tahapan untuk membuat dataset dalam format yang tepat untuk digunakan dalam analisis pembuatan model yang mana merupakan tahapan sebelum masuk tahap *modelling* (Nisbet, 2009: 40). Tahapan *preprocessing* dapat mencakup kegiatan transformasi data, penghapusan objek yang mengandung *missing value*, dan reduksi data untuk mengurangi jumlah data yang digunakan (Nisbet, 2009: 40). Tahapan ini ditujukan untuk mengolah dataset agar bisa digunakan dalam proses analisa dengan Algoritma *Backpropagation*. Beberapa langkah yang dilakukan diantaranya adalah :

- a. *missing value*
- b. integrasi data

- c. inisialisasi
- d. normalisasi

4.2 Implementasi

Dalam penelitian ini dilakukan beberapa skenario pemodelan yaitu dalam dua skenario. Skenario yang pertama model yang dibangun dengan mengasumsikan bahwa nilai sebelum masuk PT (nilai UN/rapor) dan IP semester awal bisa berpengaruh terhadap ketepatan waktu lulusan, variabel yang digunakan meliputi model dengan variabel JK, status Tinggal, IP semester 1,2 Skenario yang kedua model yang dibangun dengan mengasumsikan bahwa variabel JK, status Tinggal, nilai UN/rapor dan IP semester awal bisa berpengaruh terhadap ketepatan waktu lulusan, variabel yang digunakan meliputi model dengan variabel nilai UN/rapor, IP semester 1,2. Kedua skenario itu diimplementasikan dengan dua algoritma yaitu Algoritma Naïve Bayes dan Algoritma *back propagation - Artificial neural Network*. dalam penelitian ini dilakukan empat skenario atau kondisi yang dijadikan sasaran data dalam pengolahan datanya agar diperoleh hasil yang lebih variatif dan lebih presisi dengan tingkat akurasi prediksi tertinggi, yaitu:

1. Skenario 1 : Algoritma Naive Bayes dengan menggunakan IP sem 1,2 tanpa nilai UN/rapor(mhsebtanilai)
2. Skenario 2 : Algoritma Naive Bayes dengan menggunakan IP sem 1,2 dengan nilai UN/rapor(mhsebtanilai)
3. Skenario 3 : Algoritma Backpropagation –Artificial Neural Network dengan menggunakan IP sem 1,2 tanpa nilai UN/rapor(mhsebtanilai)

4. Skenario 4 : Algoritma Backpropagation –Artificial Neural Network dengan menggunakan IP sem 1,2 dengan nilai UN/rapor(mhsebtanilai)

4.2.1 Implementasi Skenario 1: Algoritma *Naive Bayes*

Implementasi pengembangan model dengan skenario 1 yaitu dengan memasukkan komponen atau faktor saat sudah menjadi mahasiswa yaitu jenis kelamin, status tinggal mahasiswa, IP semester 1,2. Tabel 4.3 menunjukkan data set untuk pengembangan model prediksi kelulusan dengan skenario 1 yaitu sejumlah 859 record data.

Tabel 4.3. Tabel Data set kelulusan mahasiswa

	ID	mhsjk	S. tinggal	ip1	ip2
0	2	1	1	2.85	3.05
1	3	1	0	3.20	3.33
2	4	1	0	3.30	3.76
3	5	1	0	3.35	3.33
4	6	1	1	3.40	3.24
⋮	⋮	⋮	⋮	⋮	⋮
859	882	1	1	36.5	38.5

Perhitungan yang dilakukan dibagi menjadi dua, yaitu perhitungan tipe numerik dan nominal. Langkah langkah dalam perhitungan bayesian adalah:

1. Mencari *mean* data (μ) dengan membagi jumlah data keseluruhan dengan banyaknya data.
2. Mencari standart deviasi (σ).

Perintah Excel yang digunakan untuk mencari standart deviasi dari indeks prestasi semester 1 dengan status kelulusan tepat yaitu :

3. Melakukan perhitungan distribusi *gaussian*.

Perintah Excel yang digunakan untuk mencari probabilitas dengan melakukan perhitungan distribusi gaussian dari indeks prestasi semester I dengan status kelulusan tepat.

Pada tahap ini yaitu menentukan status kelulusan mahasiswa dengan mengimplementasikan algoritma *Naive Bayes*. Variabel yang dimasukkan pada skenario 1 adalah variabel jenis kelamin, status tinggal, dan IPS 1 dan IPS 2. Pelatihan dilakukan pada seluruh data mahasiswa angkatan 2012 sampai dengan 2018 dengan tahun kelulusan 2015 sampai 2021 dengan total data 859 dimana penentuan data training 572 dan data testing 287 data. Variabel yang digunakan untuk melakukan prediksi adalah jenis kelamin, status tinggal, nilai UN/rapor serta IP semester berdasarkan data set IP semester terbagi dalam 6 nilai yaitu IPS1, 2. Pada tahapan ini dilakukan normalisasi data, berdasarkan pada data set tabel 4.3 maka didapatkan nilai normalisasi berdasarkan kepada data uji, hasil implementasi dapat dilihat pada Gambar 4.5 sebagai berikut.

```

## pecah data Input(X) dan output(y)
## dihidupkan/ diaktifkan kemenera untuk memilih data semester 1-2 target nilai 0N dan dengan nilai 0N
y2 = df_ahs.tepat

## semester 1 + 2
X2 = df_ahs.drop(["tepat", "status_lulus","ID","merotondia","ip1","ip2","ip3","ip4"], axis=1) # target nilai 0N
print(X2)

## melakukan normalisasi (mengubah skala data menjadi rata-rata = 0 dan standar deviasi = 1)
from sklearn.preprocessing import StandardScaler
sc2 = StandardScaler()
X_sc2 = sc2.fit_transform(X2)
print(X_sc2)

```

Gambar 4.5. normalisasi data skenario1

Setelah dilakukan normalisasi maka didapatkan data training dan data testing dapat dilihat pada Gambar 4.6 sebagai berikut:

```

## perbandingan sebelum dan sesudah
## sebelum
print("Data set tanpa nilai 0")
print(X2.head())
## sesudah
pd.DataFrame(X2_s12).head()
## split data, jumlah data testing 1/3 dari total data
from sklearn.model_selection import train_test_split
## data 2
X2_train, X2_test, y1_train, y2_test = train_test_split(X2_s12, y2, test_size = 1/3, random_state = 1)

## tampilkan data training ()
print("data training")
print(pd.DataFrame(X2_train))
## tampilkan data testing (88 baris)
print("data testing")
print(pd.DataFrame(X2_test))

## membandingkan jumlah data training dan testing
print("jumlah data training : ", X2_train.shape)
print("jumlah data testing : ", X2_test.shape)

```

Gambar 4.6. data training dan testing skenario 1

Berikut pada tabel 4.4 dan tabel 4.5 adalah data training sebelum dan sesudah dilakukan normalisasi yang digunakan untuk keseluruhan data pada skenario 1.

Tabel 4.4 Data training sebelum normalisasi

	mhsjk	Mhs_stt	ip1	ip2
0	1	1	2.85	3.05
1	1	0	3.20	3.33
2	1	0	3.30	3.76
3	1	0	3.35	3.33
4	1	1	3.40	3.24
⋮	⋮	⋮	⋮	⋮
882	1	1	3.65	3.85

Tabel 4.5 merupakan tabel data training sesudah normalisasi yaitu sebanyak 572 data.

Tabel 4.5 Data training sesudah normalisasi

	0	1	2	3
0	-1.146893	1.889219	0.693490	0.335377
1	-1.146893	-0.529319	0.317215	0.441148
2	-1.146893	-0.529319	0.875139	0.610381
3	0.871921	-0.529319	0.706465	0.398839
4	-1.146893	-0.529319	0.096640	0.462302
⋮	⋮	⋮	⋮	⋮
571	-1.146893	1.889219	-0.097985	-0.193479

Pada tabel 4.6 terlihat tabel data testing setelah dilakukan normalisasi data yaitu sebanyak 287 data.

Tabel 4.6 Data testing sesudah normalisasi

	0	1	2	3
0	-1.146893	-0.529319	0.369115	0.176720
1	-1.146893	-0.529319	0.096640	0.462302
2	-1.146893	1.889219	0.654565	0.546919
3	0.871921	-0.529319	-0.162859	-0.447329
4	-1.146893	-0.529319	0.446965	0.546919
⋮	⋮	⋮	⋮	⋮
286	0.871921	1.889219	0.226390	0.261337

Selanjutnya terlihat pada Gambar 4.7 yaitu pembangunan model (tambahan gambaran model) dengan skenario 1.

```
## buat model
from sklearn.naive_bayes import GaussianNB
nb2 = GaussianNB()

## latih model
nb2.fit(X2_train, y2_train)
## prediksi data testing
pred_nb2 = nb2.predict(X2_test)
print(pred_nb2)
```

Gambar 4.7 implementasi naïve bayes untuk skenario 1

4.2.2 Implementasi Skenario 2: Algoritma *Naive Bayes* dengan UN

Implementasi pengembangan model dengan skenario 2 yaitu dengan memasukkan komponen atau faktor saat sudah menjadi mahasiswa yaitu nilai UN/rapor, jenis kelamin, status tinggal mahasiswa, IP semester 1,2. Tabel 4.7 menunjukkan data set untuk pengembangan model prediksi kelulusan dengan skenario 1 yaitu sejumlah 859 record data.

Tabel 4.7. Tabel Data set kelulusan mahasiswa

	ID	mhseb tanilai	mhsjk	S. tinggal	ip1	ip2
0	2	49.55	1	1	2.85	3.05
1	3	50.60	1	0	3.20	3.33
2	4	53.15	1	0	3.30	3.76
3	5	48.10	1	0	3.35	3.33
4	6	43.60	1	1	3.40	3.24
⋮	⋮	⋮	⋮	⋮	⋮	⋮
859	882	31.25	1	1	36.5	38.6

Perhitungan yang dilakukan dibagi menjadi dua, yaitu perhitungan tipe numerik dan nominal. Langkah langkah dalam perhitungan bayesian adalah:

1. Mencari *mean* data (μ) dengan membagi jumlah data keseluruhan dengan banyaknya data.
2. Mencari standart deviasi (σ).

Perintah Excel yang digunakan untuk mencari standart deviasi dari indeks prestasi semester 1 dengan status kelulusan tepat.

3. Melakukan perhitungan distribusi *gaussian*.

Perintah Excel yang digunakan untuk mencari probabilitas dengan melakukan

perhitungan distribusi gaussian dari indeks prestasi semester I dengan status kelulusan tepat.

Pada tahap ini yaitu menentukan status kelulusan mahasiswa dengan mengimplementasikan algoritma *Naive Bayes*. Variabel yang dimasukkan pada skenario 2 adalah variabel nilai UN/rapor, jenis kelamin, status tinggal, dan IPS 1 dan IPS 2. Pelatihan dilakukan pada seluruh data mahasiswa angkatan 2012 sampai dengan 2018 dengan tahun kelulusan 2015 sampai 2021 dengan total data 859 dimana penentuan data training 572 dan data testing 287 data. Variabel yang digunakan untuk melakukan prediksi adalah nilai UN/rapor, jenis kelamin, status tinggal, nilai UN/rapor serta IP semester berdasarkan data set IP semester terbagi dalam 2 nilai yaitu IPS1, 2. Pada tahapan ini dilakukan normalisasi data, berdasarkan pada data set tabel 4.3 maka didapatkan nilai normalisasi berdasarkan kepada data uji, hasil implementasi dapat dilihat pada Gambar 4.8 sebagai berikut:

```
## pecah data input(X) dan output(y)
## dihindarkan/ diabaikan koefisien untuk memilih data semester 1-2 dengan nilai IM
y = df_aha.tepat

## semester 1+2
X = df_aha.drop(["tepat", "durasi_lulus", "IM", "IM2", "IM3", "IM4", "IM5", "IM6"], axis=1) # dengan nilai IM 1, 2, 3, 4

## melakukan normalisasi (mengubah skala data menjadi rata-rata = 0 dan standar deviasi = 1)
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_sc = sc.fit_transform(X)
```

Gambar 4.8. Implementasi normalisasi data skenario2

Setelah dilakukan normalisasi maka data training dan data testing terlihat pada tabel 4.7 dan tabel 4.8 berikut.

Tabel 4.8 Data training sesudah normalisasi

	0	1	2	3	4
0	-1.146893	-0.670182	1.889219	0.693490	0.335377
1	-1.146893	0.310393	-0.529319	0.317215	0.441148
2	-1.146893	0.921935	-0.529319	0.875139	0.610381
3	0.871921	0.046798	-0.529319	0.706465	0.398839
4	-1.146893	-0.424160	-0.529319	0.096640	0.462302
⋮	⋮	⋮	⋮	⋮	⋮
571	-1.146893	-0.308177	1.889219	-0.097985	-0.103479

Pada tabel 4.9 terlihat tabel data testing setelah dilakukan normalisasi data.

Tabel 4.9 Data testing sesudah normalisasi

	0	1	2	3	4
0	-1.146893	0.261189	-0.529319	0.369115	0.176720
1	-1.146893	0.566960	-0.529319	0.096640	0.462302
2	-1.146893	0.559931	1.889219	0.654565	0.546919
3	0.871921	0.914906	-0.529319	-0.162859	-0.447320
4	-1.146893	0.147316	-0.529319	0.446965	0.546919
⋮	⋮	⋮	⋮	⋮	⋮
286	0.871921	0.060856	1.889219	0.226390	0.261337

Setelah dilakukan normalisasi maka didapatkan data training dan data testing dapat dilihat pada Gambar 4.9 sebagai berikut.

```

## perbandingan sebelum dan sesudah
## sebelum
X.head()
## sesudah
pd.DataFrame(X_sc).head()
## split data, jumlah data testing 1/3 dari total data
from sklearn.model_selection import train_test_split
## data 1
X_train, X_test, y_train, y_test = train_test_split(X_sc, y, test_size = 1/3, random_state = 1)

## tampilkan data training ()
print("data training")
print(pd.DataFrame(X_train))
## tampilkan data testing (10 baris)
print("data testing")
print(pd.DataFrame(X_test))

## sebandingkan jumlah data training dan testing
print("jumlah data training : ", X_train.shape)
print("jumlah data testing : ", X_test.shape)

```

Gambar 4.9. data training dan testing skenario2

Selanjutnya terlihat pada Gambar 4.10 yaitu pembangunan model dengan skenario 2.

```

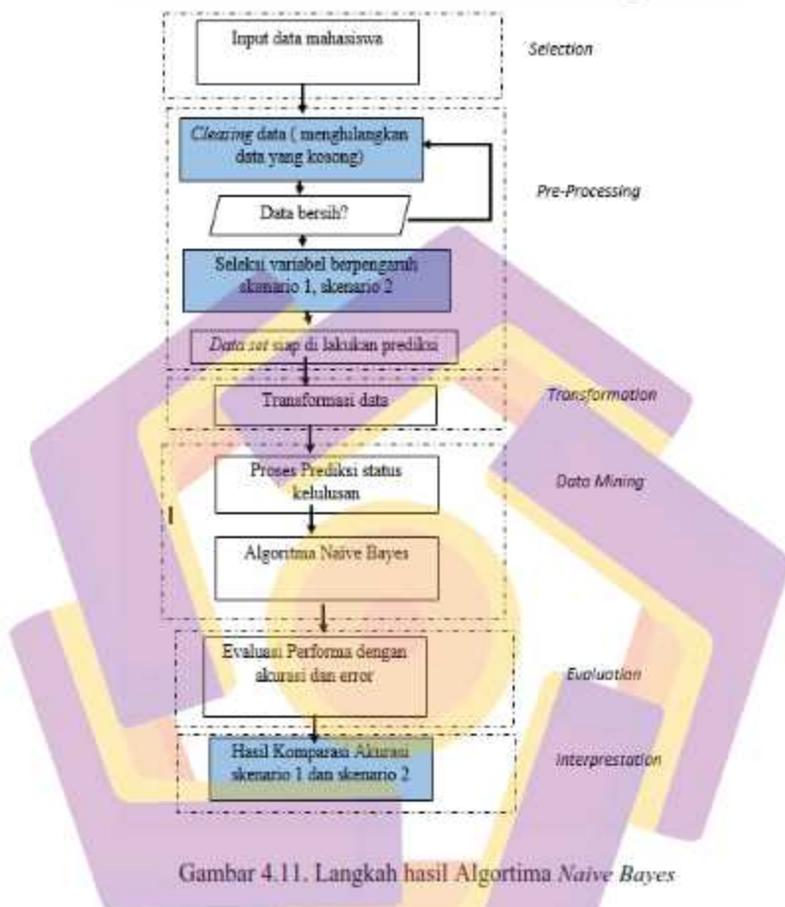
## buat model
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()

## latih model
nb.fit(X_train, y_train)
## prediksi data testing
pred_nb = nb.predict(X_test)

```

Gambar 4.10 implementasi naïve bayes untuk skenario 2

Berdasarkan skenario 1 dan 2, hasil pemodelan dengan algoritma naïve bayes dapat digambarkan sebagai berikut pada gambar 4.11.



4.2.3 Implementasi Skenario 3 : Algoritma *Backpropagation* – *Artificial Neural Network*

4.2.3.1 Parameter dan Arsitektur ANN skenario 3

Pemodelan skenario 3 yaitu dengan memasukkan variabel IP 1,2 dan staus tinggi serta gender mahasiswa dengan menggunakan algoritma B-ANN> Pemodelan dengan Backpropagation-ANN bertujuan untuk menentukan arsitektur

jaringan yang optimal. Dalam penelitian ini dilakukan pemilihan arsitektur dengan mencari kombinasi dari *input* dan jumlah *hidden layer*. Teknik dilakukan dengan langkah langkah sebagai berikut :

1. Menentukan arsitektur awal dengan jumlah *input* sebanyak variabelnya
jumlah variabel yang digunakan ada 4 yaitu jenis kelamin, status tinggal ,
nilai IPS1, 2
2. Menentukan jumlah *hidden layer* yaitu sejumlah 4
3. Menentukan fungsi aktivasi dalam hal ini menggunakan fungsi *rule*, *sigmoid*
4. Mengevaluasi pemilihan model terbaik yang merupakan kombinasi antara
input dan *neuron* dari langkah 2.

Berikut pada gambar 4.12 adalah pemodelan ANN dengan menggunakan python.

```

## ANN
import keras
import tensorflow as tf
from keras.layers import Dense, Dropout
from keras.models import Sequential

# merancang arsitektur ANN

### semester 1, 2
# ## model 2
ann2 = Sequential()
ann2.add(Dense(units= 4, activation = 'relu', input_dim=4))
ann2.add(Dense(units= 1, activation = 'sigmoid'))      ## output 1 dimensi
ann2.compile(optimizer="adam", loss='mean_squared_error', metrics=['accuracy'])

```

Gambar 4.12 Perancangan arsitektur *Backpropagation-ANN*

Selanjutnya pada gambar 4.13 dilakukan prediksi untuk model dengan skenario 3 dengan menggunakan B-ANN.

```
# training ANN
ann.fit(X_train,y_train, batch_size=100, epochs=300,verbose= 0)

# prediksi
pred_ann = ann.predict(X_test)
pred_ann = [ 1 if y>0.5 else 0 for y in pred_ann] ## bulatkan output jadi 0 dan 1
pred_ann
```

Gambar 4.13 prediksi untuk skenario 3

Berikut merupakan gambaran dari arsitektur jaringan syaraf tiruan dalam penelitian ini. Pada tahapan training data dilakukan metode *Backpropagation Neural Network*. Data akan dilatih hingga mencapai batas *error* yang diinginkan dengan menentukan jumlah *neuron hidden layer*, *learning rate*, *maximum iteration*.

4.2.3.2 Prototype Prediksi B-ANN Skenario 3

Arsitektur metode ANN *Backpropagation* yang menghasilkan prediksi yang optimal, maka dilakukan penentuan atau penginputan jumlah nilai parameter-parameter yang digunakan, seperti jumlah *hidden layer*, laju pemahaman (*learningrate*), maksimum iterasi (*epoch*), dan momentum dengan menggunakan proses *trial and error*. Setelah menentukan jenis data dan parameter, prediksi dilakukan dengan metode ANN *Backpropagation* berdasarkan langkah-langkah arahan Laurene Fausett [18].

Rumus untuk mengupdate bobot menggunakan *backpropagation* adalah sebagai berikut persamaan 4.1.

$$w_{new} = w_{old} - \alpha \frac{\partial E}{\partial w} \quad (\text{Persamaan 4.1})$$

Jadi kita perlu menghitung rumus tersebut untuk semua w . Sebagai awalan kita coba tuliskan rumus $\frac{\partial E}{\partial w}$ untuk semua nilai w berdasarkan perhitungan di persamaan sebelumnya persamaan 4.2.

$$\frac{\partial E}{\partial w_{ij}} = -(t_i - y_i) \cdot y_i(1 - y_i) \cdot x_j \quad (\text{Persamaan 4.2})$$

Dengan rincian untuk $i=1,2,\dots$ dan $j=1,2,\dots$ didapatkan hasil sebagai berikut:

$$\frac{\partial E}{\partial w_{11}} = -(t_1 - y_1) \cdot y_1(1 - y_1) \cdot x_1$$

$$\frac{\partial E}{\partial w_{21}} = -(t_1 - y_1) \cdot y_1(1 - y_1) \cdot x_2$$

$$\frac{\partial E}{\partial w_{31}} = -(t_1 - y_1) \cdot y_1(1 - y_1) \cdot x_3$$

$$\frac{\partial E}{\partial w_{12}} = -(t_2 - y_2) \cdot y_2(1 - y_2) \cdot x_1$$

$$\frac{\partial E}{\partial w_{22}} = -(t_2 - y_2) \cdot y_2(1 - y_2) \cdot x_2$$

$$\frac{\partial E}{\partial w_{32}} = -(t_2 - y_2) \cdot y_2(1 - y_2) \cdot x_3$$

Nilai α pada rumus di atas adalah *learning rate*, sebuah konstanta (biasanya antara 0-1) yang menentukan seberapa cepat proses pembelajaran model dilakukan. Dari rumus-rumus tersebut akan terlihat ada pola yang bisa digunakan, sehingga bisa

dibuat perulangan sebagai berikut (sudah termasuk update bobot) :

```
alpha = 0.5
for c in range(2): # neuron output
    for r in range(3): # neuron input
        # menghitung turunan
        delta = -1 * (T[c] - Y[c]) * Y[c] * (1 - Y[c]) * X[r]
        # update bobot
        W[r][c] = W[r][c] - alpha * delta
```

Dengan cara yang sama, kita bisa adaptasi perulangan tersebut untuk menghitung update dari bias. Perlu diingat, hanya ada satu bias untuk satu neuron *output* sehingga tidak perlu ada iterasi *r*.

```
alpha = 0.5
for c in range(2): # neuron output
    delta = -1 * (T[c] - Y[c]) * Y[c] * (1 - Y[c]) * 1
    b[c] = b[c] - alpha * delta
```

Setelah dua sintaks di atas dijalankan, maka nilai bobot *W* dan *b* akan sudah di-*update* berdasarkan *error* yang muncul.

Berikut langkah dalam Algoritma *Backpropagation* :

Langkah 0 : **Installsasi semua bobot** dengan bilangan acak kecil

Langkah 1 : Jika kondisi penghentian belum terpenuhi, lakukan langkah 2-9

Langkah 2 : Untuk setiap pasangan data pelatihan lakukan langkah 3-8

Langkah 3 : Tiap unit masukan menerima sinyal dan meneruskannya ke unit tersembunyi di atasnya

Langkah 4 : Hitunglah semua *output* di unit tersembunyi tersebut z_j , ($j=1,2,\dots,p$). Perhatikan rumus pada persamaan 3.2 dan persamaan 3.3.

Langkah 5 : Hitunglah semua keluaran jaringan di unit y_k ($k = 1,2, \dots, m$), perhatikan rumus pada persamaan 3.4 dan persamaan 3.5.

- **Tahap Kedua : Propagasi Mundur**

Langkah 6 : Hitung faktor δ unit keluaran berdasarkan *error* pada setiap y_k ($k=1,2, \dots, m$). Data dimasukkan sesuai dengan perhitunagn pada persamaan 3.6. Dimana δ_k merupakan unit kesalahan yang akan dipakai dalam perubahan bobot *layer* di bawahnya (langkah 7). Hitung suku perubahan bobot w_{kj} dengan laju percepatan α . Peratikan rumus berikut ini pada persamaan 4.3 :

$$\Delta w_{kj} = \alpha \delta_k z_j ; k = 1,2, \dots, m ; j = 0,1,2, \dots, p \quad (\text{Persamaan 4.3})$$

Langkah 7 : Hitung faktor δ unit tersembunyi berdasarkan *error* pada setiap unit tersembunyi z_j ($j = 1,2, \dots, p$) sesuai persamaan 3.8 dengan faktor δ unit tersembunyi :

$$\delta_j = \delta_{netj} f'(z_{netj}) = \delta_{netj} z_j (1-z_j)$$

- **Tahap Ketiga : Perubahan bobot**

Langkah 8 : Hitunglah semua perubahan pada bobot. Rumus perubahan bobot garis yang menuju ke unit keluaran berdasarkan persamaan 3.10 dan persamaan 3.11.

Langkah 9 : Menguji apakah kondisi berhenti sudah terpenuhi.

Kondisi berhenti ini terpenuhi jika nilai kesalahan yang dihasilkan lebih kecil dari nilai kesalahan yang dihasilkan.

4.2.4 Implementasi Skenario 4: Algoritma *Backpropagation* – *Artificial Neural Network* dengan UN

Berdasarkan data diatas, dalam pengujian perhitungan diperlukan data uji untuk melihat kesesuaian hasil dari pengujian. Data uji yang digunakan dapat dilihat pada tabel 4.3.

4.2.3.1 Parameter dan Arsitektur ANN Skenario 4

Pemodelan skenario 3 yaitu dengan memasukkan variabel IP 1,2 dan status tinggi serta gender mahasiswa dan nilai UN dengan menggunakan algoritma B-ANN. Pemodelan dengan *Backpropagation*-ANN bertujuan untuk menentukan arsitektur jaringan yang optimal. Dalam penelitian ini dilakukan pemilihan arsitektur dengan mencari kombinasi dari *input* dan jumlah *hidden layer*. Teknik dilakukan dengan langkah langkah sebagai berikut :

1. Menentukan arsitektur awal dengan jumlah *input* sebanyak variabelnya
jumlah variabel yang digunakan ada 5 yaitu jenis kelamin, status tinggal, nilai UN, IPS1, 2

2. Menentukan jumlah *hidden layer* yaitu sejumlah 5
3. Menentukan fungsi aktivasi dalam hal ini menggunakan fungsi *rule, sigmoid*
4. Mengevaluasi pemilihan model terbaik yang merupakan kombinasi antara *input* dan *neuron* dari langkah 2.

Berikut pada gambar 4.14 adalah pemodelan ANN dengan menggunakan python untuk skenario 4.

```
## ANNs
import keras
import tensorflow as tf
from keras.layers import Dense, Dropout
from keras.models import Sequential

# merancang arsitektur ANN

### semester 1, 2
## model 1
ann = Sequential()
ann.add(Dense(units= 5, activation = 'relu', input_dim=5))
ann.add(Dense(units= 1, activation = 'sigmoid'))      ## output 1 dimensi
ann.compile(optimizer="adam", loss='mean_squared_error', metrics=['accuracy'])
```

Gambar 4.14 Perancangan arsitektur *Backpropagation*-ANN skenario 4

Berikut merupakan gambaran dari arsitektur jaringan syaraf tiruan dalam penelitian ini. Pada tahapan training data dilakukan metode *Backpropagation Neural Network*. Data akan dilatih hingga mencapai batas *error* yang diinginkan dengan menentukan jumlah *neuron hidden layer, learning rate, maximum iteration*,

4.2.5 Evaluasi Prediksi dengan *Confusion Matrix*

Pengukuran kinerja klasifikasi dilakukan dengan confusion matrix [17]. Confusion matrix merupakan alat pengukuran yang dapat digunakan untuk menghitung kinerja atau tingkat kebenaran proses klasifikasi.

Berikut gambar 4.15 menggambarkan implementasi perhitungan confusion matrix untuk skenario 1,2,3,4.

```

## buat fungsi prediksi ANN 1
def prediksi_ann(input):
    ## normalisasi
    print(input)
    input_sc = sc.transform([input])
    ## prediksi
    hasil_ann = ann.predict(input_sc)
    hasil = [ "tepat" if y>=0.5 else "tidak tepat" for y in hasil_ann]
    return hasil[0]

## aplikasikan ke data
kolom = X.columns
kolom2 = X2.columns
df_hasil["pred_nb"] = df_mhs[kolom].apply(prediksi_ann, axis=1)

## cek hasil
df_hasil.head()

```

Gambar 4.15 Prediksi dengan confusion matrix skenario 3 dan 4

4.3 Anallsa dan Pembahasan

4.3.1 Hasil Implementasi Skenario 1: Algoritma *Naive Bayes* tanpa Nilai UN/Rapor

Berdasarkan data implementasi pada sub ba 4.1 , pemodelan untuk skenario 1 yaoutu dengan algoritma naive bayes untuk variabel yang masuk ke model prediski adalah jenis kelamin, Status tinggal, dan IP semester 1, 2 atau semester 1, 2, terlihat pada tabel 4.10.

Tabel 4.10 Pemodelan dan komparasi hasil pemodelan *bayesian* skenario 1

	ID	mhsj k	Mhs_stt	durasi lulus	ip1	ip2	Actual kelulusan	pred_nb 2
0	2	1	1	9999	2.8 5	3.0 5	Tidak tepat	Tidak tepat
1	3	1	0	37	3.2 0	3.3 3	tepat	tepat
2	4	1	0	37	3.3 0	3.7 6	tepat	Tepat
3	5	1	0	37	3.3 5	3.3 3	tepat	Tepat
4	6	1	1	9999	3.4 0	3.2 4	Tidak tepat	Tepat

Didapatkan hasil *convolustion matrik* sebagai berikut pada Tabel 4.11 menyatakan hasil pengukuran kinerja prediksi model.

Tabel 4.11 *Confolution Matrix* untuk hasil prediksi NB skenario 1

		Prediksi		Akurasi
Status\prediksi		Tepat waktu	Terlambat	
Actual	Tepat Waktu	185 (TP)	61 (FN)	76.65%
	Terlambat	7 (FP)	34 (TN)	

Keterangan :

- TP (True Positive) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas positif.
- FN (False Negative) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas negatif.

- c. FP (False Positive) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas positif.
- d. TN (True Negative) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas negatif.

Berdasarkan penelitian didapatkan nilai akurasi untuk model prediksi ketepatan kelulusan dengan bayesian dengan mempertimbangkan variabel saat mahasiswa di perkuliahan yaitu variabel JK, S Tinggal, IP S1,2 adalah sebesar 76.65%.

4.3.2 Hasil Implementasi Skenario 2: Algoritma *Naive Bayes* dengan UN

Berdasarkan pemodelan dengan *bayesian* dengan variabel jenis kelamin, Status tinggal, nilai UN, dan IP semester 1, 2 atau semester 1, 2, 3, dan 4 selanjutnya dilakukan pemodelan dan komparasi hasil pemodelan berdasarkan nilai UN, dan IPS seperti pada tabel 4.12.

Tabel 4.12 Pemodelan dan komparasi hasil pemodelan *NB* skenario 2

	ID	mhsj k	Mhs ebta nilai	Mh s_s tt	durasi lulus	ip1	ip2	Actual kelulusan	pred _nb
0	2	1	49.55	1	9999	2.85	3.0 5	Tidak tepat	Tidak tepat
1	3	1	50.6	0	37	3.20	3.3 3	tepat	tepat
2	4	1	53.15	0	37	3.30	3.7 6	tepat	Tepa t
3	5	1	48.1	0	37	3.35	3.3 3	tepat	Tepa t
4	6	1	43.6	1	9999	3.40	3.2 4	Tidak tepat	Tepa t

Tabel 4.13 *Confusion Matrix* untuk hasil prediksi NB skenario 2

	Status\ prediksi	Prediksi		Akurasi
		Tepat waktu (TP)	Terlambat (FN)	
Actual	Tepat Waktu	184 (TP)	57 (FN)	77.35%
	Terlambat	8 (FP)	38 (TN)	

Keterangan :

- TP (True Positive) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas positif yaitu sebanyak 184 data.
- FN (False Negative) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas negatif yaitu sebanyak 57 data.
- FP (False Positive) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas positif yaitu sebanyak 8 data.
- TN (True Negative) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas negatif yaitu sebanyak 38 data.

Dari perhitungan probabilitas kelulusan diatas, dapat disimpulkan bahwa berdasarkan Algoritma *Bayesian* untuk prediksi ketepatan waktu kelulusan mahasiswa nilai UN dan tanpa nilai UN kurang begitu ada pengaruh signifikan terhadap nilai akurasi.

Akurasi dari model yaitu :

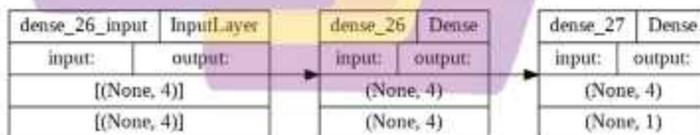
$$akurasi = \frac{38 + 184}{859} = 77,35\%$$

$$Error = \frac{57 + 8}{859} = 32,65\%$$

Berdasarkan penelitian didapatkan nilai akurasi untuk model prediksi ketepatan kelulusan dengan bayesian dengan variabel JK, S Tinggal, Nilai UN, IP S1,2 adalah 77.35% yang mana selanjutnya akan dilakukan pemodelan dengan algoritma B-ANN.

4.3.3 Hasil Implementasi Skenario 3: Algoritma *Backpropagation – Artificial Neural Network* tanpa Nilai UN/Rapor.

Arsitektur Backpropagation-ANN meliputi 5 neuron *input* , 5 neuron lapisan tersembunyi dan 1 lapisan *output* atau arsitektur 4-4-1 seperti pada gambar 4.16.



Gambar 4.16 hasil arsitektur 4-4-1 *Backpropagation-ANN*

Berdasarkan data diatas, dalam pengujian perhitungan diperlukan data uji untuk melihat kesesuaian hasil dari pengujian. Data set yang digunakan dapat dilihat pada tabel 4.3.

Menggunakan model arsitektur pada Gambar 4.12 dengan fungsi aktivasi yang digunakan adalah fungsi *sigmoid* pada lapisan tersembunyi dan fungsi *rule* pada lapisan *output*, pelatihan dilakukan dengan mencoba beberapa kali *epoch* 300.

Berikut hasil prediksi perhitungan *backpropagation* untuk data kelulusan mahasiswa. Berdasarkan hasil ANN didapatkan bahwa dari 859 data set maka dihasilkan prediksi sebagai berikut pada Tabel 4.14 :

Tabel 4.14 Hasil prediksi *Backpropagation* –ANN skenario 3

	ID	mhs jk	mhs ebta nilai	kos	durasi lulus	ip1	ip2	kelulu san	pred_ ann2
0	2	1	49.55	1	9999	2.85	3.05	Tidak tepat	Tidak tepat
1	3	1	50.80	0	37	3.20	3.33	tepat	Tidak tepat
2	4	1	53.15	0	37	3.30	3.70	tepat	Tepat
3	5	1	48.10	0	37	3.33	3.33	tepat	Tidak tepat
4	6	1	43.80	1	9999	3.40	3.24	Tidak tepat	Tepat

Berikut hasil akurasi dari perhitungan model B-ANN seperti terlihat pada tabel 4.15.

Tabel 4.15 Hasil analisa dengan *Backpropagation* ANN sekenario 3

	Status\ prediksi	Prediksi		Akurasi
		Tepat waktu	Terlambat	
Actual	Tepat Waktu	185 (TP)	59 (FN)	77.00%
	Terlambat	7 (FP)	36 (TN)	

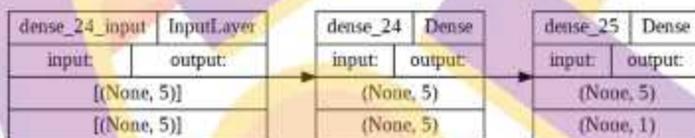
Keterangan :

- TP (True Positive) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas positif yaitu sebanyak 185 data.
- FN (False Negative) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas negatif yaitu sebanyak 59 data.
- FP (False Positive) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas positif yaitu sebanyak 7 data.
- TN (True Negative) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas negatif yaitu sebanyak 36 data.

Dilakukan analisis lebih lanjut dari hasil yang telah diperoleh untuk menemukan model yang paling terbaik diantara kandidat yang ada pada tiap neuron *input layer*. Analisis dilakukan dengan mempresentasikan kandidat tiap neuron *input layer* pada suatu tabel yang ditampilkan pada Tabel 4.15.

4.3.4 Hasil Implementasi Skenario 4: Algoritma *Backpropagation* – *Artificial Neural Network* dengan UN

Dalam penelitian ini dilakukan perhitungan untuk model B-ANN dengan arsitektur : Arsitektur *Backpropagation-ANN* meliputi 5 neuron *input* , 5 neuron lapisan tersembunyi dan 1 lapisan *output* atau arsitektur 4-4-1 seperti pada gambar 4.17



Gambar 4.17 hasil arsitektur 5-5-1 *Backpropagation-ANN*

Berdasarkan data diatas, dalam pengujian perhitungan diperlukan data uji untuk melihat kesesuaian hasil dari pengujian. Data set yang digunakan dapat dilihat pada tabel 4.3.

Menggunakan model arsitektur pada Gambar 4.13 dengan fungsi aktivasi yang digunakan adalah fungsi *sigmoid* pada lapisan tersembunyi dan fungsi *rule* pada lapisan *output*, pelatihan dilakukan dengan mencoba beberapa kali *epoch* 300.

Berikut hasil prediksi perhitungan backpropagation untuk data kelulusan mahasiswa. Berdasarkan hasil ANN didapatkan bahwa dari 859 data set maka dihasilkan prediksi sebagai berikut pada Tabel 4.16 :

Tabel 4.16 Hasil prediksi *Backpropagation* –ANN skenario 4

	ID	mhs_jk	mhs_ebta_nilai	Mhs_stt	durasi_lulus	ip1	ip2	kelulusan	pred_ann
0	2	1	49.55	1	9999	2.85	3.05	Tidak tepat	Tidak tepat
1	3	1	50.00	0	37	3.20	3.33	tepat	tepat
2	4	1	53.15	0	37	3.30	3.70	tepat	tepat
3	5	1	48.10	0	37	3.35	3.33	tepat	tepat
4	6	1	43.00	1	9999	3.40	3.24	Tidak tepat	tepat

Berikut hasil akurasi dari perhitungan model B-ANN seperti terlihat pada tabel 4.17.

Tabel 4.17 Hasil analisa dengan *Backpropagation* ANN skenario 4

	Status\prediksi	Prediksi		Akurasi
		Tepat waktu	Terlambat	
Actual	Tepat Waktu	186 (TP)	56 (FN)	78.39%
	Terlambat	1 (FP)	39 (TN)	

Dari total keseluruhan data testing yang digunakan sebanyak 287 data testing, 186 diantaranya mendapat hasil *true-positive* (prediksi tepat waktu dan kenyataan tepat waktu), 39 diantaranya *true-negative* (prediksi terlambat dan kenyataan terlambat), 56 diantaranya *false-negative* (prediksi terlambat namun kenyataan tepat waktu), dan 1 diantaranya *false-positive* (prediksi telat waktu dan kenyataan terlambat) pada tabel 4.17.

Hasil akurasi pada Tabel 4.17 dapat didapatkan dengan cara membagikan jumlah prediksi yang sesuai kenyataan (*true-positive+false-positive*) dengan

jumlah data keseluruhan (287) kemudian dikalikan 100, terlihat dalam tabel 4.17.

Dilakukan analisis lebih lanjut dari hasil yang telah diperoleh untuk menemukan model yang paling terbaik diantara kandidat yang ada pada tiap neuron *input layer*. Analisis dilakukan dengan mempresentasikan kandidat tiap neuron *input layer* pada suatu tabel yang ditampilkan pada Tabel 4.17.

Prediksi dengan B-ANN

Berikut hasil prediksi perhitungan *backpropagation* untuk data kelulusan mahasiswa. Berdasarkan hasil ANN didapatkan bahwa dari 859 data set maka dihasilkan prediksi. Dilakukan analisis lebih lanjut dari hasil yang telah diperoleh untuk menemukan model yang paling terbaik diantara kandidat yang ada pada tiap neuron *input layer*. Analisis dilakukan dengan mempresentasikan kandidat tiap neuron *input layer* pada suatu tabel yang ditampilkan pada Tabel 4.18 .

Tabel 4.18 Komparasi Hasil prediksi Bn dan *Backpropagation*-ANN

	ID	mhs_jk	mhs_ebta_nilai	Mhs_stt	ip1	ip2	Actual_kelulusan	Prediksi_skenario1	Prediksi_skenario2	Prediksi_skenario3	Prediksi_skenario4
0	2	1	49.55	1	2.83	3.05	Tidak tepat	Tidak tepat	Tidak tepat	Tidak tepat	Tidak tepat
1	3	1	50.00	0	3.20	3.33	tepat	tepat	tepat	Tidak tepat	tepat
2	4	1	53.15	0	3.30	3.70	tepat	Tepat	Tepat	Tepat	tepat
3	5	1	48.10	0	3.35	3.33	tepat	Tepat	Tepat	Tidak tepat	tepat
4	0	1	43.00	1	3.40	3.24	Tidak tepat	Tepat	Tepat	Tepat	tepat

Berikut didapatkan komparasi hasil akurasi untuk empat skenario pengujian terlihat pada tabel 4.19.

Tabel 4.19 komparasi nilai akurasi Model

No	Skenario pengujian	Hasil Akurasi(%)	keterangan
1	Skenario 1	76.30	Algoritma naïve bayes
2	Skenario 2	77.35	Algoritma naïve bayes dengan UN
3	Skenario 3	77.00	Algoritma B-ANN
4	Skenario 4	78.04	Algoritma B-ANN dengan UN

Berdasarkan hasil perhitungan didapatkan nilai akurasi untuk pemodelan prediksi kelulusan dengan variabel yang masuk ke pemodelan adalah variabel saat mahasiswa yaitu pada skenario 1 (prediksi dengan algoritma naïve bayes) dan skenario 2 (prediksi dengan algoritma B-ANN) ada kenaikan akurasi yaitu dari 77.35% menjadi 78.04%, dengan variabel yang masuk ke model adalah variabel JK, Stinggal, IPS1 dan IPS 2.

Sedangkan apabila memasukkan variabel akademik sebelum menjadi mahasiswa yaitu n pada skenario 2 (prediksi dengan algoritma naïve bayes) dan skenario 4 (prediksi dengan algoritma b-ANN) ada kenaikan akurasi yaitu dari 77.35% menjadi 78.04%, dengan variabel yang masuk ke model adalah variabel JK, Stinggal, UN/rapor, IPS1 dan IPS 2.

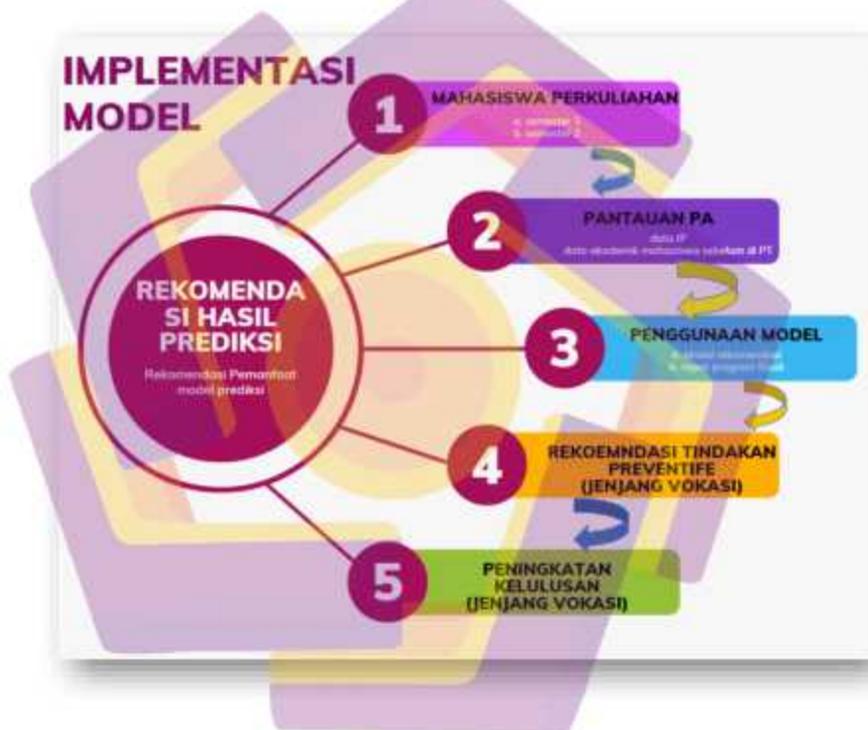
Penelitian ini memberikan penambahan informasi bagi penelitian sebelumnya yaitu adanya penekanan bahwa algoritma B-ANN memiliki nilai akurasi lebih tinggi dibandingkan naïve bayes dalam memprediksi waktu kelulusan mahasiswa nilai UN/rapor berpengaruh terhadap model yang dibangun baik dengan algoritma naïve bayes maupun B-ANN.

Rekomendasi bagi PA atau manajemen sekolah bahwa model ini bisa digunakan untuk memprediksi kelulusna bagi mahasiswa. Dan untuk mahasiswa

yang diprediksikan kelulusan tidak tepat waktu dapat dilakukan tindakan preventive misalnya saja dengan dilakukannya tindakan pembimbingan dan monitoring secara intensif. Fakta di lapangan menunjukkan bahwa pembimbing akademik membutuhkan adanya perkiraan akan status kelulusan mahasiswa bimbingannya berdasarkan status tempuh mahasiswa pada semester berjalan. Sebagaimana dalam wawancara melalui kuisioner pada pembimng akademik didapatkan bahwa pembimbing akadmeik padaawal semester 3 membutuhkan adanya monitoring akan konsisi akadmeik mahasiswa. Monitoring ini salah satunya adalah berupa data estimasi status kelulusan mahasiswa, sehingga apabila estimasi/prediksi status mahasiswa dikatakan "terlambat" maka pembimbing akademik bisa melaksanakan pembimbingan secara intens dengan salah satunya peningkatan pembimbingan berkala.

Pemodelan yang dikembangkan dari penelitian ini menghasilkan model estimasi dengan kategori status kelulusan mahasiswa tepat waktu atau tidak tepat waktu. Pemodelan yang dikembangkan dengan algoritma Naive Bayes dan algoritma Back propagation neural network memberikan petunjuk sebagai rekomendasi bagi manajemen program studi dalam hal ini pembimbing akademik di dalam memberikan pembimbingan akademik terhadap mahasiswa dibawah bimbingannya. Hasil implementasi dari estimasi /prediksi status kelulusan sesuai harapan pembimbing akademik akan memberikan early warning model (EWD) dalam pelaksanaan manajemen pembimbingan mahasiswa khususnya bagi mahasiswa dengan jenjang vokasional. Pendidikan tinggi vokasi diarahkan untuk menghasilkan lulusan yang terampil dan unggul sesuai dengan kebutuhan dunia

kerja. Oleh karena itu, salah satu ciri pendidikan tinggi vokasi adalah kemitraan yang erat dengan dunia kerja untuk membangun link and match, melalui pengembangan kurikulum, magang industri, penerapan pembelajaran yang berorientasi pada project-based learning atau case study atau pembelajaran diluar kampus lainnya sebagai salah satu wujud implementasi Merdeka Belajar.



Gambar 4.18 Skenario implementasi penggunaan model

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan dari hasil penelitian yang telah dilakukan maka dapat disimpulkan bahwa :

- a. Pada penelitian ini terdapat beberapa faktor yang mempengaruhi ketepatan kelulusan mahasiswa yaitu jenis kelamin, status tinggal, IPS 1, 2 dan nilai UN. Penambahan nilai UN pada algoritma *bayes* memberikan dampak akurasi yaitu penambahan nilai akurasi sebesar 0.7 % pada *Naive Bayes*.
- b. Pada penelitian ini terdapat beberapa faktor yang mempengaruhi ketepatan kelulusan mahasiswa yaitu *jenis kelamin*, status tinggal, IPS 1, 2. Penambahan nilai UN pada memberikan nilai akurasi 78,04% pada Algoritma *Backpropagation NN*.

Berdasarkan data maka *NN-Backpropagation* memiliki akurasi yang lebih tinggi dibandingkan algoritma *naive Bayes*.

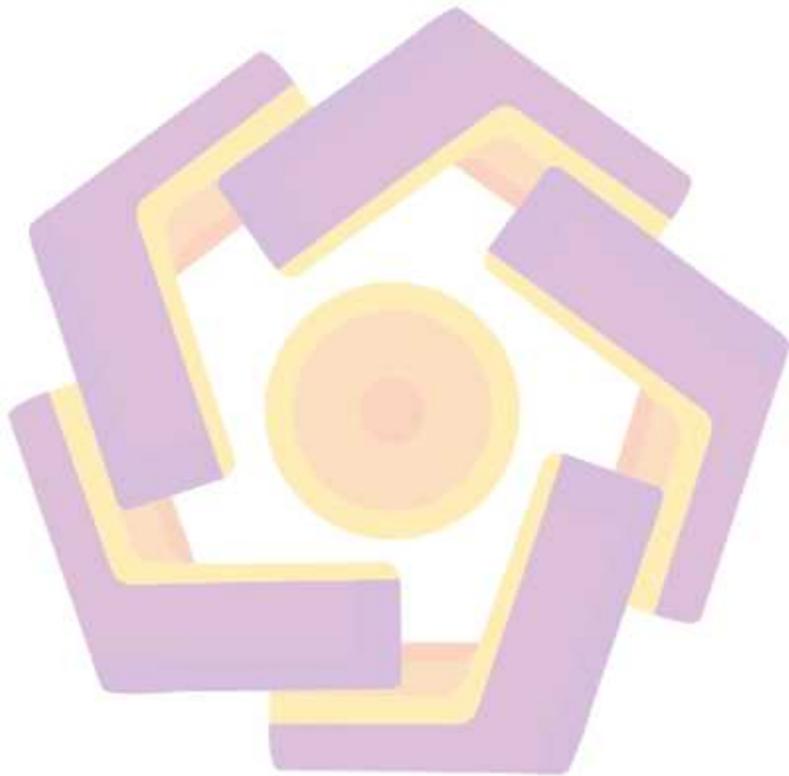
5.2. Saran

Adapun beberapa saran bagi pengembangan penelitian ini yaitu :

- a. Data yang digunakan sebagai data latih belum memperhitungkan jika mahasiswa cuti. Maka untuk penelitian selanjutnya disarankan melakukan penelitian jika mahasiswa ambil cuti.
- b. Menggunakan variabel *input* yang lebih banyak dan karakteristik

mahasiswa yang berbeda beda.

Selain dengan analisis dari hasil wawancara, penentuan variabel *input* dapat dilakukan terlebih dahulu dengan menganalisis faktor-faktor yang mempengaruhi kelulusan mahasiswa menggunakan metode statistik ataupun metode lainnya.



DAFTAR PUSTAKA

- Budi, S., 2007, Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. 2007
- Bramer,M., 2016, Principles of Data Mining, vol. 98, no. 461. London: Springer London, 2016.
- Fausett,L, 1994. Fundamental of Neural networks, New Jersey: Prentice Hall
- Han,J., M. Kamber, and J. Pei, 2012, Data Mining: Concepts and Techniques. 2012.
- Haykin, S. 2009. *Neural Networks and Learning Machines*. United State of America: Pearson.
- Jong jek siang.2009, “ Jaringan Syaraf Tiruan & Pemrograman Menggunakan Matlab”, penerbit andi, yogyakarta.
- Larose, D.T. and C. D. Larose, 2014, Discovering Knowledge in Data, vol. 190, no. 5. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2014
- Maimon,O. and L. Rokach, 2009, “Introduction to Knowledge Discovery and Data Mining,” in Data Mining and Knowledge Discovery Handbook, 2009.
- Puspitaningrum,D., 2006, Pengantar Jaringan Saraf Tiruan”, Penerbit Andi. Yogyakarta.
- Sumathi and S. N. 2006, Sivanandam, Introduction to Data Mining and its Applications. 2006.
- Witten, I. H. , E. Frank, and M. a. Hall, 2011, Data Mining: Practical Machine Learning Tools and Techniques, Third Edition.2011.
- Aryasanti,A., 2018, Sistem Komparasi *Naive Bayes* dan *Decision Tree s* untuk Menentukan Klasifikasi Kegagalan Studi Mahasiswa,” Jurnal TICOM, Vol. 6, No. 3, 2018, pp. 103-111.
- Bernard renaldy suteja, 2007, “Penerapan Jaringan Saraf Tiruan Propagasi Balik Studi Kasus Pengenalan Jenis Kopi”, viewed juni 2007.
- Budiantara,A.,dkk 2020, Komparasi Algoritma *Decision Tree* , *Naive Bayes* Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu. Jurnal Ilmu Pengetahuan dan teknologi Komputer.
- Eko Prasetyo Rohmawan , 2018, Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree* Dan Artificial Neural Network , Jurnal Ilmiah MATRIK Vol.20 No.1, April 2018: 21-30

- Hendra, Mochammad Abdul Azis dan Suhardjono, 2020, Analisis Prediksi Kelulusan Mahasiswa Menggunakan *Decision Tree* Berbasis *Particle Swarm Optimization*, Jurnal SISFOKOM (Sistem Informasi dan Komputer), Volume 09, Nomor 01, PP 102 - 107 .
- Hidayat,V, 2015, "Sistem Klasifikasi Penentuan Keputusan Pendanaan Beasiswa Menggunakan Metode *Naive Bayes*," pp. 1-4, 2015.
- Hsieh, L.F., Huang, C.J. & Huang, C.L. 2007. Applying Particle Swarm Optimization To Schedule Order Picking Routes In A Distribution Center. *Asian Journal of Management and Humanity Sciences*. Vol. 1, No. 4. pp. 558-576.
- Cynthia, E. P., & Ismanto, E. (2017). Jaringan Syaraf Tiruan Backpropagation dalam Memprediksi Ketersediaan Komoditi Pangan Provinsi Riau. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9*, 271-282.
- Magdalena. H .2012, Sistem Pendukung Keputusan Untuk Menentukan Mahasiswa Lulusan Terbaik Di Perguruan Tinggi. *Seminar Nasional Teknologi Informasi dan komunikasi*. ISSN : 2089-9815
- Muhamad Hanief Meinanda, dkk., 2009, "Prediksi masa studi sarjana dengan Artificial Neural Network". *International Indoensian Journal*, Volume 1 No.2. Institut Teknologi Bandung, Bandung.
- Ni Komang Sri Julyantari dan I Ketut Dedy Suryawan , 2014, Data Mining Prestasi Akademik Dengan *Naive Bayes* Berdasarkan Attribut Importance (AI), *Jurnal Sistem Dan Informatika (JSI)*, 9(1).
- Kusnanto and A. M. Nur Azizah, 2016, "Sistem Pendukung Keputusan Penyeleksian Mahasiswa Penerima Beasiswa Menggunakan Metode *Naive Bayes* Berbasis Web," *Jurnal Ilmiah dan Teknik Informatika*, vol. 10, no. 1, pp. 6-8, Mei 2016.
- Romadhona,A., S. Suprapedi, and H. Himawan, 2017, "Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin dan Indeks Prestasi Menggunakan Algoritma *Decision Tree* ," *Jurnal Teknologi Informasi Cyberku*, Vol. 13, No. 1, pp. 69–83, 2017. Google Scholar
- Sri Widaningsih. 2012, Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Program Studi Teknik Informatika Dengan Algoritma C4.5, *Naive Bayes*, Knn, Dan Svm. *Jurnal Tekno Insentif* ISSN(p): 1907-4964 | ISSN (e): 2655-089X

Syarli,S and A. A. Muin, 2018, "Metode *Naive Bayes* Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2018.

Ting,S, W. Ip and A. H. Tsang, 2011, "Is *Naive Bayes* a Good Classifier for Document Classification?," *International Journal of Software Engineering and Its Applications*, vol. 5, no. 3, 3 July 2011.

