

**PENERAPAN METODE DEEP LEARNING UNTUK  
MENENTUKAN STRATEGI PROMOSI DI PERGURUAN  
TINGGI**

**(STUDI KASUS : UNIVERSITAS ISLAM NAHDLATUL ULAMA JEPARA)**

**TESIS**

Disusun sebagai salah satu syarat untuk  
memperoleh gelar Magister Komputer dari  
Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI

Oleh :

**SANI NURMALASARI  
NIM: 2018210068**



**PROGRAM STUDI PASCASARJANA  
MAGISTER SISTEM INFORMASI  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI  
BANDUNG  
2020**

**PENERAPAN METODE DEEP LEARNING UNTUK  
MENENTUKAN STRATEGI PROMOSI DI PERGURUAN  
TINGGI**

**(STUDI KASUS : UNIVERSITAS ISLAM NAHDLATUL ULAMA JEPARA)**

Oleh :

**SANI NURMALASARI  
NIM: 2018210068**

Bandung, 2 Maret 2020

Menyetujui,

Dr. Hery Heryanto, S.Kom.,M.Kom.

Pembimbing

**PROGRAM STUDI PASCASARJANA  
MAGISTER SISTEM INFORMASI  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI  
BANDUNG  
2020**

## KATA PENGANTAR

Pertama- tama saya ingin memanjatkan puji syukur kehadiran Allah S.A.W yang telah melimpahkan rahmat dan hidayahnya sehingga penulis dapat menyelesaikan laporan **hasil penelitian yang berjudul “Penerapan Metode *Deep Learning* untuk Menentukan Strategi Promosi di Perguruan Tinggi (Studi Kasus : Universitas Islam Nahdlatul Ulama Jepara ) “**.

Laporan hasil penelitian ini tidak akan selesai tepat waktu tanpa adanya bantuan dari banyak pihak, untuk itu penulis ingin mengucapkan ucapan terima kasih kepada :

- 1) Kepala Rektor Universitas Islam Nahdlatul Ulama Jepara yang telah memberikan kesempatan untuk melakukan penelitian di Universitas ini.
- 2) Kepala UPT Universitas Islam Nahdlatul Ulama Jepara beserta Stafnya yang telah membantu dalam proses mengumpulkan data penelitian.
- 3) Dosen Pembimbing Dr. Hery Heryanto S.Kom.,M.Kom yang telah memberikan bimbingan selama pelaksanaan penelitian ini.
- 4) Suami dan Anakku yang telah memberikan semangat, bantuan dan motivasi untuk menyelesaikan penelitian ini.

Penulis menyadari bahwa masih banyak kekurangan yang terdapat dalam laporan hasil penelitian ini, maka dari itu kritik dan saran yang membangun sangat diharapkan oleh penulis dari para pembaca

Penulis sangat berharap agar laporan hasil penelitian ini dapat memberikan manfaat yang besar bagi penulis sendiri khususnya dan bagi para pembaca umumnya.

Bandung 2 Maret 2020

Sani Nurmalasari

## ABSTRAK

### PENERAPAN METODE DEEP LEARNING UNTUK PENENTUAN STRATEGI PROMOSI DI PERGURUAN TINGGI (STUDI KASUS : UNIVERSITAS ISLAM NAHDLATUL ULAMA JEPARA)

Oleh :  
Nama : Sani Nurmalasari  
NPM : 2018210068

Teknologi *data mining* kini telah berkembang sampai kepada sektor pendidikan hingga memunculkan bidang baru yaitu *Educational Data Mining (EDM)*. Ada banyak sekali pengolahan yang bisa dilakukan dengan menggunakan data yang didapat dari institusi pendidikan salah satunya adalah *data mining* untuk mengklasifikasikan data mahasiswa baru sebagai acuan untuk menentukan strategi promosi di perguruan tinggi. Pada penelitian ini akan dilakukan penerapan algoritma *deep learning* untuk bisa menentukan strategi promosi yang sesuai berdasarkan kebutuhan pelanggan yang berbeda karakteristiknya maka sebelumnya harus terlebih dahulu mempelajari karakteristik pelanggannya yang akan menjadi target promosi. Setelah target promosi ditentukan maka bisa diterapkan strategi promosi yang sesuai dengan target tersebut.

Sumber data yang akan diolah berasal dari data Penerimaan Mahasiswa Baru (PMB) di Universitas Islam Nahdlatul Ulama Jepara pada tahun 2014 sampai 2018 dengan menggunakan metode penelitian berbasis KDD (*Knowledge Discovery in Database*). Adapun langkah-langkah yang akan digunakan yaitu tahap persiapan, pembersihan data, data transformasi dan seleksi, pemodelan data, pengujian model, evaluasi hasil dan ekstraksi pengetahuan.

Proses pemodelan data dilakukan dengan menggunakan metode *deep learning* yaitu *multi layer perceptron (MLP)* dengan menggunakan 4 layer. Hasil pemodelan yang diperoleh kemudian diuji sampai akhirnya mendapat nilai akurasi yang tinggi. Setelah proses pengujian model kemudian dilakukan evaluasi hasil dan ekstraksi pengetahuan untuk mendapatkan hasil strategi promosi yang paling efektif dan dapat merekomendasikan perbaikan untuk strategi promosi yang kurang efektif.

Dari hasil penelitian ini dapat diketahui strategi promosi yang efektif dan strategi yang kurang efektif. Strategi promosi yang efektif adalah strategi promosi 2 yaitu potongan SPI 30% bagi calon mahasiswa yang berasal dari luar jepara dan strategi promosi 6 yaitu potongan SPI 40% bagi calon mahasiswa yang hasil nilai tesnya sangat baik. Sementara untuk strategi promosi yang kurang efektif yaitu strategi promosi 4 dan 5 dilakukan perbaikan melalui strategi promosi yang direkomendasikan dengan mengubah persyaratan untuk strategi promosi tersebut.

**Keyword : *data mining*, strategi promosi, *Educational Data Mining (EDM)*, KDD, *deep learning*, MLP.**

## ABSTRACT

### APPLICATION OF DEEP LEARNING METHOD FOR DETERMINING PROMOTIONAL STRATEGY IN HIGHER EDUCATION (CASE STUDY: ISLAMIC UNIVERSITY OF NAHDLATUL ULAMA JEPARA)

By:

Name: Sani Nurmalasari

NPM: 2018210068

Data mining technology has evolved to the education sector to bring up a new field, namely Educational Data Mining (EDM). There is a lot of processing that can be done using data obtained from educational institutions, one of which is data mining to classify new student data as a reference to determine the promotion strategy in institutions. In this study the deep learning method will be implemented to determine the promotion strategy according to the needs of customers with different characteristics, the first is to study the characteristics of their customers who will be the target of promotions. After the promotion target has been determined, a promotional strategy can be applied in accordance with the target.

The data source will be processed from the New Student Admission Data (PMB) at the Nahdlatul Ulama Islamic University in Jepara in 2014 to 2018 using the KDD-based research method (Knowledge Discovery in Database). The steps that will be use are the preparation stage, data cleaning, data transformation and selection, data modeling, model testing, results evaluation and knowledge extraction.

The data modeling process is done by using the deep learning method, namely multi-layer perceptron (MLP) using 4 layers. The modeling results obtained are then tested until finally it gets a high accuracy value. After the model testing process the results are evaluated and knowledge extracted to get the most effective promotional strategy results and can recommend improvements to the less effective promotional strategies.

The results of this study is to know which is the effective promotion strategies and strategies are less effective. An effective promotion strategy is promotion strategy 2 which is 30% SPI discount for prospective students who come from outside Jepara and promotion strategy 6 is 40% SPI discount for prospective students whose test results are very good. While for the less effective promotion strategy, which is promotion strategies 4 and 5, improvements were made through recommended promotion strategies by changing the requirements for the promotion strategy.

**Keyword:** *data mining, promotion strategy, Educational Data Mining (EDM), KDD, deep learning, MLP.*

## DAFTAR ISI

KATA PENGANTAR.....	i
ABSTRAK.....	ii
ABSTRACT.....	iii
DAFTAR ISI.....	iv
DAFTAR TABEL.....	vi
DAFTAR GAMBAR.....	vii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian.....	3
1.4. Ruang Lingkup Penelitian.....	3
1.5. Sistematika Penulisan.....	4
BAB II LANDASAN TEORI.....	5
2.1 Strategi <i>Marketing</i> dan Promosi.....	5
2.2 <i>Data Mining</i> dan Proses <i>Knowledge Discovery in Database</i> (KDD).....	7
2.2.1 Jenis Data.....	9
2.2.2 <i>Data cleaning and Integration</i> .....	10
2.2.3 <i>Data Selection</i> .....	11
2.2.4 <i>Data Preprocessing</i> .....	13
2.3 <i>Data Mining</i> untuk Promosi.....	15
2.4 Teknik Klasifikasi.....	17
2.5 Metode <i>Deep Learning Multy Layer Network (MLP)</i> .....	23
2.6 <i>Deep Learning</i> Menggunakan Rapidminer (H2O).....	25
2.7 Penelitian Terkait.....	27
BAB III OBJEK DAN METODOLOGI PENELITIAN.....	29
3.1 Profil Universitas Islam Nahdlatul Ulama Jepara.....	29
3.1.1 Visi,Misi dan Tujuan.....	29
3.1.2 Struktur Organisasi.....	30
3.1.3. Strategi Promosi UNISNU Jepara.....	31

3.2 Objek Penelitian .....	33
3.3 Strategi Promosi yang digunakan .....	36
3.4 Metodologi Penelitian .....	38
<b>BAB IV IMPLEMENTASI DAN PEMBAHASAN HASIL .....</b>	<b>41</b>
4.1 Tahap persiapan .....	41
4.2 Pembersihan Data .....	48
4.3 Seleksi Data dan Transformasi .....	51
4.4 Proses Pemodelan Data .....	57
4.5 Proses Pengujian Model .....	62
4.6 Evaluasi Hasil Pemodelan .....	65
4.7 Ekstraksi Pengetahuan .....	68
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>74</b>
5.1 Kesimpulan .....	74
5.2 Saran .....	75
<b>DAFTAR PUSTAKA .....</b>	<b>77</b>
<b>LAMPIRAN .....</b>	<b>80</b>

## DAFTAR TABEL

Tabel 2.1	Contoh <i>Fuzzy Set</i> .....	22
Tabel 2.2	Penelitian Terkait.....	27
Tabel 3.1	Daftar Atribut Data.....	33
Tabel 3.2	Strategi Promosi .....	37
Tabel 4.1.	Matrixs Pengujian .....	63
Tabel 4.2	Matrix Perbandingan Persentase Data Latih dan Durasi .....	64
Tabel 4.3	Perbandingan Hasil Akurasi Strategi Promosi .....	68
Tabel.4.4	Pengujian Korelasi Atribut Terhadap Akurasi.....	68
Tabel 4.5	Tabel <i>Confusion Matrix</i> Strategi Promosi .....	69
Tabel 4.6	Rekomendasi 1 Perbaikan Strategi Promosi.....	70
Tabel 4.7	Matriks Pengujian Terhadap Rekomendasi 1 Strategi Promosi.....	71
Tabel 4.8	Rekomendasi 2 Strategi Promosi.....	71
Tabel 4.9	Matriks Pengujian Terhadap Rekomendasi 2 .....	72

## DAFTAR GAMBAR

Gambar 2.1	langkah-langkah untuk melakukan data mining .....	8
Gambar 2.2	<i>Learning Step</i> .....	18
Gambar 2.3	<i>Classification Set</i> .....	18
Gambar 2.4	Pohon Keputusan.....	19
Gambar 2.5	Sel Syaraf Buatan .....	21
Gambar 2.6	<i>Deep Learning Multi layer Perceptrons</i> .....	24
Gambar 2.7.	Ilustrasi Proses Analisis data .....	26
Gambar 3.1.	Struktur Organisasi .....	31
Gambar 3.2	Visualisasi <i>Dataset</i> .....	35
Gambar 3.3.	Visualisasi Data Atribut Referensi .....	36
Gambar 3.4	Metode Penelitian .....	39
Gambar 4.1	Data PMB .....	41
Gambar 4.2	Perbandingan Atribut Nama Terhadap Atribut Jalur Masuk .....	42
Gambar 4.3	Visualisasi Atribut Alamat .....	42
Gambar 4.4	Visualisasi Atribut Jenis Kelamin .....	43
Gambar 4.5	Visualisasi Atribut Tahun Lahir.....	43
Gambar 4.6	Visualisasi Atribut jenis Asal Sekolah .....	44
Gambar 4.7	Visualisasi Atribut Pekerjaan Orang Tua.....	45
Gambar 4.8	Visualisasi Atribut Nilai Ujian .....	45
Gambar 4.9	Visualisasi Atribut NIM .....	46
Gambar 4.10	Visualisasi Atribut Program Studi .....	47
Gambar 4.11	Visualisasi Atribut Angkatan .....	47
Gambar 4.12	Visualisasi Atribut Referensi.....	48
Gambar 4.13	Langkah-langkah Pembersihan Data.....	48
Gambar 4.14	Data Atribut Jenis Sekolah yang Dihapus .....	49
Gambar 4.15	Atribut Asal Kota yang tidak Valid .....	49

Gambar 4.16	Data Atribut Tahun Lahir yang tidak Valid.....	50
Gambar 4.17	Ilustrasi Penggantian Atribut Pekerjaan Orang Tua .....	50
Gambar 4.18	Ilustrasi Penggantian Nilai Atribut Prodi .....	51
Gambar 4.19	Langkah-langkah seleksi dan transformasi data .....	51
Gambar 4.20	Hasil Pemilihan Data .....	52
Gambar 4.21	Pembuatan Atribut Re-Registrasi dan Jalur Masuk .....	53
Gambar 4.22	Pembuatan Atribut Usia Saat Mendaftar .....	54
Gambar 4.23	Penerapan Label Strategi Promosi .....	54
Gambar 4.24	Proses Pendiskritan.....	55
Gambar 4.25	Pengurangan Dimensi.....	56
Gambar 4.26	Pemilihan Atribut .....	57
Gambar 4.27	Automodel PraProses Data .....	58
Gambar 4.28	Automodel Spliting Data.....	59
Gambar 4.29	Automodel Deep Learning .....	62
Gambar 4.30	Grafik Perbandingan %Akurasi dan %Data Latih. ....	63
Gambar 4.31	Grafik Perbandingan %Akurasi dan <i>Epoch</i> .....	64
Gambar 4.32	Hasil Pemodelan <i>Deep Learning</i> .....	66
Gambar 4.33	Performance Vector .....	67
Gambar 4.34	Pembobotan Atribut .....	68

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada era Industri 4.0 ini hampir semua bidang mengalami digitalisasi, semua industri terpaksa harus melakukan perubahan di hampir semua bidang agar dapat terus berkembang mengikuti arus jaman. Pada era ini lalu lintas data yang diperoleh sangat tinggi mengakibatkan percepatan terhadap jumlah data yang masuk. Sayangnya tidak semua institusi yang sadar akan pentingnya data yang mereka miliki.

Dengan banyaknya data yang masuk serta timbulnya kebutuhan akan informasi yang berkualitas tinggi yang dapat digunakan sebagai panduan untuk pengambilan keputusan atau untuk menyelesaikan masalah-masalah bisnis maka digunakanlah teknik *data mining*. Dengan teknik *data mining* data-data tersebut bisa diolah untuk bisa mendapatkan suatu informasi baru yang berguna.

Di sektor pendidikan baru-baru ini telah muncul disiplin ilmu baru yang diakibatkan meningkatnya pertumbuhan data pada sektor pendidikan yaitu *Educational Data Mining (EDM)*. *Educational Data Mining* ini merupakan disiplin yang berusaha memanfaatkan data-data yang didapat di bidang pendidikan baik itu yang berasal dari siswa atau dari manajemen pendidikannya dengan cara mengembangkan metode dan mengeksplorasi data yang unik dan semakin besar yang dimiliki oleh institusi pendidikan.

Perkembangan perguruan tinggi di Indonesia semakin tahun semakin meningkat hal ini ditunjukkan dengan semakin banyaknya jumlah perguruan tinggi yang baru didirikan di Indonesia baik itu negeri maupun swasta. Menurut data PDDIKTI bahwa pada tahun 2019 tercatat ada lebih dari 4.500 jumlah perguruan tinggi yang tersebar di seluruh Indonesia.

Perguruan tinggi yang kini jumlahnya semakin banyak membuat persaingan antar perguruan tinggi semakin ketat membuat mahasiswa semakin lebih selektif dalam memilih tempat belajarnya untuk itu diperlukan penerapan strategi promosi tepat agar jumlah mahasiswa baru terus meningkat selain itu juga untuk menghindari pemborosan biaya yang dikeluarkan dari penerapan strategi promosi yang kurang efektif.

Universitas Islam Nahdlatul Ulama Jepara merupakan perguruan tinggi yang sedang berkembang, untuk terus meningkatkan mutu dan kualitasnya agar jumlah peminat yang mendaftar dari tahun ke tahun semakin meningkat dan juga untuk meningkatkan akreditasi perguruan tinggi maka salah satu caranya adalah dengan melakukan penentuan strategi promosi yang efektif dan tepat sasaran.

Penentuan strategi promosi yang tepat sangat penting untuk dapat di terapkan pada perguruan tinggi agar dapat meningkatkan daya saing dan pemerataan di setiap daerah. Pemerataan tersebut juga sangat penting karena merupakan salah satu poin yang berpengaruh dalam akreditasi program studi maupun institusi.( Prastyo. 2017)

Untuk bisa menentukan strategi promosi yang sesuai berdasarkan kebutuhan pelanggan yang berbeda karakteristiknya maka suatu institusi harus terlebih dahulu mempelajari karakteristik pelanggannya yang akan menjadi target promosi. Setelah target promosi ditentukan maka bisa diterapkan strategi promosi yang sesuai dengan target tersebut.

Berdasarkan data Penerimaan Mahasiswa Baru (PMB) yang didapat dari Universitas Islam Nahdlatul Ulama Jepara sejak tahun 2014 sampai 2018 jumlah pendaftar mahasiswa baru tidak mengalami peningkatan yang signifikan setiap tahunnya. Idealnya jumlah pendaftar senantiasa meningkat setiap tahun akan tetapi pada tahun 2014 jumlah pendaftar sekitar 1.800 pendaftar, pada tahun 2015 terjadi penurunan sehingga hanya 1.700 pendaftar, begitu juga pada tahun 2016 hanya 1.700 pendaftar, sedangkan pada tahun 2017 terjadi kenaikan 300 pendaftar menjadi total 2.000 pendaftar tetapi pada tahun 2018 turun kembali ke angka 1.918 pendaftar. Untuk itu diperlukan strategi promosi yang efektif untuk terus meningkatkan jumlah pendaftar setiap tahunnya.

Pada *dataset* penerimaan calon mahasiswa baru di Universitas Islam Nahdlatul Ulama Jepara yang berisi data pendaftar baik itu yang melakukan *Re-registrasi* ataupun yang tidak. Dari data tersebut akan dilakukan proses *data mining* untuk mengklasifikasikan calon mahasiswa baru menggunakan metode *Deep Learning*. Hasil dari proses klasifikasi tersebut akan dijadikan sebagai acuan untuk target promosi untuk selanjutnya diterapkan strategi promosi yang sesuai. Dari data yang telah didapat jumlah keseluruhan *dataset* adalah 9.341

*record* data, yang selanjutnya akan dianalisa melalui proses penerapan algoritma *deep learning* untuk menentukan strategi promosi di perguruan tinggi.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan sebelumnya maka rumusan masalah yang bisa disimpulkan yaitu :

1. Strategi promosi apa yang efektif sesuai dengan target promosi yang telah ditentukan pada Universitas Islam Nahdlatul Ulama Jepara ?
2. Bagaimana cara untuk mengklasifikasikan calon mahasiswa baru berdasarkan strategi promosi yang telah ditentukan pada Universitas Islam Nahdlatul Ulama Jepara (UNISNU) dengan metode *Deep Learning* ?.

## **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah disimpulkan maka tujuan penelitian yang didapat yaitu :

1. Menentukan strategi promosi yang sesuai dengan target promosi.
2. Mengklasifikasikan calon mahasiswa baru berdasarkan strategi promosi yang telah ditentukan.

## **1.4 Ruang Lingkup Penelitian**

Ruang lingkup penelitian dibatasi pada data mahasiswa di Universitas Nahdlatul Ulama Jepara pada tahun 2014-2018 yang akan digunakan sebagai dataset model tes dan uji.

## 1.5 Sistematika Penulisan

Pembuatan tesis ini dilakukan dengan pembagian bab sebagai berikut :

### BAB 1: PENDAHULUAN

Pada bab ini dibahas latar belakang, rumusan masalah, tujuan penelitian, ruang lingkup penelitian, metode penelitian, dan sistematika penulisan.

### BAB 2: TINJAUAN PUSTAKA

Pada bab ini akan dibahas tentang *data mining*, *data preprosesing* dan algoritma yang akan digunakan yaitu *Deep Learning*, adapun *literature* yang digunakan bersumber dari buku, maupun jurnal.

### BAB 3: OBJEK DAN METODOLOGI PENELITIAN

Bab ini menjelaskan tinjauan organisasi, tahap persiapan pemrosesan data untuk penentuan strategi promosi pada universitas Islam Nahdlatul Ulama Indonesia Jepara, dan metodologi penelitian yang akan digunakan yaitu metodologi KDD dengan penerapan metode klasifikasi *Deep Learning* .

### BAB 4: PENERAPAN METODOLOGI DAN PEMBAHASAN

Pada tahap ini menjelaskan langkah-langkap pengolahan data dan proses implementasi dari metodologi KDD dan uji coba dari model yang telah dibuat serta analisis hasil.

### BAB 5: KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dan saran dari hasil pembahasan tesis ini yang bermanfaat untuk penelitian selanjutnya.

## BAB II TINJAUAN PUSTAKA

### 2.1 Strategi *Marketing* dan Promosi

Marketing bukan hanya sekedar kegiatan pengiklanan dan promosi, tetapi istilah marketing sebenarnya meliputi semua hal mulai dari budaya perusahaan sampai positioning, riset pasar, pengembangan suatu proses bisnis atau produk, pengiklanan dan promosi, *public/press relation*, dan bisa dibilang semua fungsional *sales* dan *customer service*. (Whalley,2010)

Istilah *Marketing* mudah sekali dirancukan baik dalam dalam perusahaan besar **maupun perusahaan kecil, istilah ini sering kali hanya dianggap sebagai “penjualan” dan “periklanan”**. **Bagian pemasaran** dalam sebuah perusahaan sering kali hanya melakukan kegiatan penjualan dan penilaian kinerja bagian pemasaran sering kali hanya berdasarkan pencapaian target penjualan yang biasanya juga ditentukan secara intuitif. (Tjiptono,2017)

Kegiatan marketing merupakan bagian yang penting dalam pengembangan institusi untuk terus bisa menyelaraskan antara ketersediaan layanan dan permintaan pasar maka bagian manajemen marketing harus mengenal faktor-faktor dan situasi yang menciptakan kondisi pasar. Tujuan utama *marketing* pada institusi pendidikan tinggi adalah untuk menentukan kualitas sistem pendidikan, orientasi pasar, dan untuk menyediakan layanan spesifik yang sesuai dengan kebutuhan pelanggan.(Gajic, 2012)

Sebelum sebuah perusahaan atau institusi melakukan kegiatan marketing dengan baik maka bagian *marketing* tersebut harus memilih cara terbaik untuk melakukan promosi, maka dari itu perlu dibuat sebuah rencana atau strategi *marketing*, hal itu disebabkan karena promosi yang tidak berdasarkan pada strategi *marketing* yang baik maka hanya akan menjadi pemborosan pada institusi tersebut.

Secara garis besar konsep pemasaran adalah merupakan kunci untuk mencapai tujuan perusahaan. Kunci ini bisa dilihat dalam kemampuan sebuah perusahaan untuk menciptakan nilai, mengkomunikasikan nilai pelanggan dan memberikan nilai kepada target pasarnya dengan cara yang efektif dan lebih baik dibandingkan dengan pesaingnya.

Sementara itu menurut Fandy Tjiptono dan Gregorius Chandra (Tjiptono, 2017) bahwa keberhasilan suatu pemasaran yang dilakukan oleh suatu perusahaan sangat ditentukan oleh kemampuan perusahaan tersebut dalam membedakan 5 jenis kebutuhan pelanggan. Lima jenis kebutuhan pelanggan itu adalah :

1. *Stated needs* atau kebutuhan tercatat misalnya yaitu : masyarakat membutuhkan pendidikan tinggi yang terjangkau.
2. *Real needs* atau kebutuhan yang sebenarnya misalnya yaitu : masyarakat membutuhkan pendidikan tinggi untuk bisa memperoleh pekerjaan yang lebih baik.
3. *Unstated needs* atau kebutuhan yang tersirat misalnya yaitu : masyarakat mengharapkan layanan pendidikan yang baik dan optimal.
4. *Delight needs* misalnya yaitu : masyarakat mengharapkan mendapatkan nilai lebih, hadiah atau keuntungan lebih lainnya.
5. *Secret needs* atau kebutuhan tersembunyi atau rahasia misalnya : masyarakat ingin dipandang **sebagai konsumen yang “cerdas” dalam menentukan pilihannya oleh teman-teman** atau lingkungannya.

Adapun kegiatan-kegiatan yang bisa dilakukan dalam pemasaran suatu produk dan jasa untuk bisa meningkatkan permintaan terhadap produk dan jasa tersebut antara lain dengan cara *diferensiasi* harga, melakukan cara promosi khusus, melakukan kegiatan pengiklanan dan menentukan area pasar. Dalam praktik di dunia nyata sering kali beberapa bentuk pemasaran tersebut digabungkan atau dilakukan secara bersama-sama, tetapi ada juga yang karena keterbatasan anggaran biaya maka program pemasaran yang dilakukan harus dipilih terlebih dahulu menurut yang terbaik bagi perusahaan tersebut.

Sedangkan yang dimaksud dengan promosi adalah segala bentuk penawaran atau insentif jangka pendek yang ditujukan kepada konsumen, atau pelantara konsumen yang sengaja dibuat untuk mendapatkan hasil seperti yang diinginkan dan dalam jangka waktu yang cepat.

Ada beberapa jenis promosi yang bisa digunakan oleh perusahaan atau institusi untuk memperkenalkan produk dan jasanya kepada pelanggan diantaranya yaitu :

### 1. Promosi Konsumen

Kegiatan promosi konsumen memiliki tujuan untuk mendorong konsumen agar bersedia memilih produk baru, menjauhi produk pesaing, mempertahankan pelanggan yang sudah loyal serta menjalin hubungan baik dengan pelanggan, adapun contoh dari kegiatan promosi konsumen ini meliputi : pembagian kupon, sample produk gratis, hadiah, potongan harga, undian, garansi produk dan sebagainya.

### 2. Promosi Dagang

Tujuan dari promosi dagang ini adalah untuk membujuk pengecer atau pelantara produk agar bersedia memasarkan produk perusahaan, dan mendorong konsumen agar memilih produk dari perusahaan, jenis promosi ini meliputi : diskon kas, barang daganga (*merchandise*), bantuan peralatan, atau insentif lain untuk pengecer atau pelantara konsumen.

### 3. Promosi Wiraniaga atau *salesforce promotions*

Promosi Wiraniaga ini memiliki tujuan untuk mendorong terjadinya *bussines leads*, menstimulasi pembelian, memberikan *reward* bagi konsumen dan memotivasi wiraniaga agar lebih aktif menjual produk. Promosi jenis ini contohnya adalah dengan diadakannya pameran pendidikan atau kontes penjualan.(Tjiptono,2017)

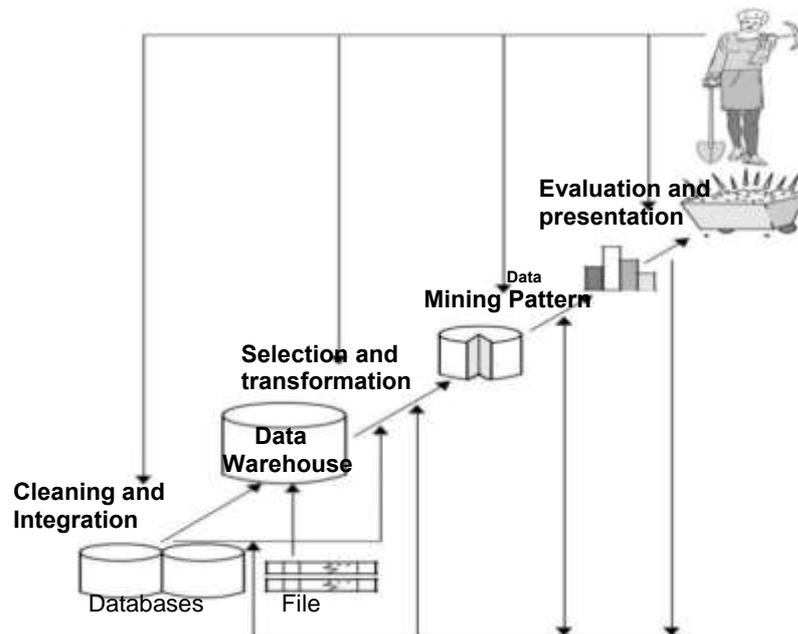
## 2.2 Data Mining dan Proses *Knowledge Discovery in Database (KDD)*

Pada era industry 4.0 sekarang ini kehidupan dikelilingi oleh data, dengan bantuan teknologi proses pengumpulan data menjadi mudah dan cepat. Maka dari itu diperlukan *data mining* untuk mengolah data-data tersebut agar dapat menemukan suatu informasi baru yang berguna. *Data mining* secara definisi bisa berarti seni dan ilmu untuk menemukan pengetahuan, wawasan, dan pola dalam data. Bisa juga dikatakan sebagai tindakan untuk mengekstraksi pola-pola yang berguna dari kumpulan data yang teroganisir. Pola yang didapat harus valid, baru, bermanfaat, dan dapat dimengerti. Sebagai contoh pola yang di dapat dari data tentang masa lalu bisa digunakan untuk mengungkapkan pola kegiatan yang dapat diproyeksikan ke masa depan. (Maheshwari,2015)

Selain untuk mengetahui pola tersembunyi yang terdapat dalam dataset, *data mining* juga bisa digunakan untuk menguraikan penemuan atau presentasi pengetahuan di dalam basis data dengan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait dari berbagai basis data besar (Kusrini, 2009)

*Data mining* merupakan pemberian cara bagi komputer untuk belajar bagaimana membuat keputusan dengan bersumber dari data. Keputusan ini dapat digunakan untuk memprediksi, memblokir *email spam*, mendeteksi bahasa dari situs web dan banyak aplikasi lain yang dapat dibangun berdasarkan hasil proses data mining menggunakan sumber data yang berbeda. (Layton,2015)

*Data mining* sering disebut juga dengan proses *Knowlegde discovery in database (KDD)* yaitu suatu proses untuk mencari yang terdiri dari langkah-langkah untuk melakukan *data mining* adalah sebagaimana yang terlihat pada gambar 2.1. berikut :



**Gambar 2.1**

langkah-langkah untuk melakukan *data mining* (Han,2012:7)

1. Pembersihan data

Sebelum proses *data mining* dapat dilakukan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan antara lain yaitu membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data.

2. Seleksi data

Seleksi data baru dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

3. Transformasi

Transformasi data adalah ketika data ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai untuk penambangan dengan melakukan operasi ringkasan atau agregasi.

4. *Data Mining*

Penambangan data adalah proses penting di mana metode cerdas diterapkan untuk mengekstraksi pola data. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan.

5. Evaluasi

Evaluasi pola untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan ukuran-ukuran ketertarikan.

6. Presentasi

Presentasi pengetahuan di mana teknik visualisasi dan representasi pengetahuan digunakan untuk menyajikan pengetahuan yang ditambang kepada pengguna.

### **2.2.1 Jenis Data**

Data adalah fakta yang sengaja dikumpulkan oleh seorang peneliti untuk menjawab pertanyaan atau untuk kepentingan penelitian. Data biasanya belum memiliki arti yang

banyak dan harus melalui proses pengolahan agar bisa didapat informasi yang berarti. Adapun beberapa jenis data dapat dibagi menjadi beberapa kelompok. (Siyoto,2015)

1. Data berdasarkan sumbernya antara lain data primer dan data sekunder
  - a. Data primer adalah data yang diperoleh dan dikumpulkan langsung oleh peneliti dari sumber datanya.
  - b. Data sekunder adalah data yang dikumpulkan dan diperoleh dari sumber yang telah ada sebelumnya.
2. Berdasarkan sifatnya yaitu data kualitatif dan data kuantitatif
  - a. Data kualitatif adalah data yang berbentuk kata-kata dan bukan angka.
  - b. Data kuantitatif adalah data yang berbentuk bilangan atau angka, data ini dapat dianalisis menggunakan teknik perhitungan matematika atau statistika.

### **2.2.2 Data Cleaning and Integration**

Data dapat berasal dari berbagai sumber sehingga akan cenderung memiliki sifat dan kualitas yang berbeda. Kualitas data yang masuk sangat bervariasi, data dari operasi internal cenderung berkualitas tinggi sementara data dari media social dan sumber public lainnya akan memiliki kualitas rendah. Untuk menjaga kualitas data ini sangat penting karena berpengaruh pada keberhasilan dan nilai hasil dari proses *data mining*, oleh karena itu hampir semua data perlu dibersihkan dan diubah sebelum dilakukan proses *data mining*. (Maheswari,2015)

Banyak cara yang bisa dilakukan untuk pembersihan dan persiapan data diantaranya yaitu dengan cara memperhatikan hal-hal berikut ini :

1. Data rangkap perlu dihapus. Data yang sama dapat diterima dari berbagai sumber. Saat menggabungkan kumpulan data, data harus tidak terduplikasi.
2. Nilai yang hilang perlu diisi, atau baris tersebut harus dihapus dari analisis. Nilai yang hilang dapat diisi dengan nilai rata-rata atau modal atau standar.
3. Elemen data mungkin perlu diubah dari satu unit ke unit lainnya. Sebagai contoh, total biaya perawatan kesehatan dan jumlah total pasien mungkin perlu dikurangi menjadi biaya / pasien untuk memungkinkan komparabilitas dari nilai itu.

4. Nilai kontinu mungkin perlu dimasukkan ke dalam beberapa tabel untuk membantu beberapa analisis. Misalnya, pengalaman kerja dapat dikelompokkan sebagai rendah, sedang, dan tinggi.
5. Elemen data mungkin perlu disesuaikan untuk membuatnya dapat dibandingkan dari waktu ke waktu. Misalnya, nilai mata uang mungkin perlu disesuaikan untuk inflasi, mereka perlu dikonversi ke tahun dasar yang sama untuk dapat diperbandingkan. Mereka mungkin perlu dikonversi ke mata uang umum.
6. Elemen data yang lebih awal perlu dihapus setelah peninjauan yang cermat, untuk menghindari kemiringan hasil. Sebagai contoh, satu donor besar dapat mengubah analisis donor alumni dalam lingkungan pendidikan.
7. Setiap bias dalam pemilihan data harus diperbaiki untuk memastikan data tersebut mewakili fenomena yang sedang dianalisis. Jika data termasuk lebih banyak anggota dari satu jenis kelamin daripada ciri khas yang diminati, maka penyesuaian perlu diterapkan pada data tersebut.
8. Data harus dibawa ke rincian yang sama untuk memastikan kompatibilitas. Data penjualan mungkin tersedia setiap hari, tetapi data kompensasi tenaga penjualan hanya tersedia setiap bulan. Untuk menghubungkan variabel-variabel ini, data harus dibawa ke penyebut umum terendah, dalam hal ini, bulanan.
9. Data mungkin perlu dipilih untuk menambah kerapatan informasi. Beberapa data mungkin tidak menunjukkan banyak variabilitas, karena tidak direkam dengan benar atau karena alasan lain. Data ini dapat memberikan efek menumpulkan data lain dan harus dihapus untuk meningkatkan kepadatan informasi data. (Maheswari,2015)

### **2.2.3 Data Selection**

Proses *Data Selection* (DS) adalah proses awal yang sangat penting sebelum memulai proses *data mining*. Proses DS sangat tergantung pada tujuan *data mining*, oleh karena itu proses seleksi data yang baik akan menghasilkan hasil *data mining* yang baik pula. (Han,2012)

Terkadang data yang akan ditambang berasal dari banyak sumber dan memiliki banyak format atau tipe data. Dari beberapa tipe data yang bisa dilakukan proses *data mining* diantaranya yaitu :

1. *Flat File*

Yaitu file yang berupa teks atau angka, tipe data ini yang paling sering digunakan dalam teknik *data mining*.

2. *Relational Database*

*Relational Database* adalah basis data yang terdiri dari table-tabel yang berisi nilai dan atribut yang masing-masing tabelnya saling berelasi.

3. *Data warehouse*

Sebuah tempat yang sengaja disediakan untuk menyimpan banyak data agar mempermudah proses analisa dan pembuatan laporan.

4. *Transactional database*

*Transactional database* berisi data-data yang berasal dari proses transaksi seperti transaksi penjualan.

5. *Multimedia database*

Basis data ini bisa berupa video, gambar, dan atau suara.

6. *Spatial database*

*Spatial database* biasa digunakan sebagai tambahan data tentang lokasi suatu tempat seperti pada peta atau tempat lain di dunia.

7. *Word wide web data*

Data yang berasal dari *website*, bisa berupa dokumen teks, video, suara dan gambar atau bahkan sebuah aplikasi.

8. *Time-series database*

Basis data ini berisi data yang terikat dengan waktu seperti data pasar saham atau aktifitas *log*.

#### 2.2.4 Data Preprocessing

Dalam proses KDD data diolah beberapa kali sebelum diterapkan algoritma *data mining* terhadap data-data tersebut. Pengolahan ini dilakukan dengan tujuan untuk mendapatkan kumpulan data yang sesuai, dapat diolah dengan cepat dan menghasilkan kesimpulan yang tepat.

Beberapa proses dalam *data preprocessing* adalah proses pengumpulan data (*Aggregation*), pengurangan dimensi (*dimensionality reduction*), pemilihan fitur (*Feature subset selection*), pembuatan fitur (*Feature creation*), pendiskritan dan peminoran (*Discretization and Binarization*), dan transformasi atribut (*Attribute Transformation*). (Hermawati, 2013)

##### 1. Pengumpulan Data

Proses pengumpulan data dilakukan dengan tujuan untuk pengurangan jumlah atribut atau variable sehingga hanya memuat data yang penting untuk proses mining. Selain itu proses ini juga dilakukan untuk merubah ukuran data dari data yang bersifat global menjadi lebih spesifik. Proses pengumpulan data merupakan proses dimana beberapa atribut atau variable dikombinasikan untuk menjadi satu variable tunggal. Misalkan dalam sebuah dataset mahasiswa terdapat atribut nilai persemester, kita dapat melakukan proses *aggregation* terhadap data tersebut sehingga bisa mendapatkan informasi nilai hanya semester 4 saja. (Hermawati, 2013)

Menurut Suyanto (Suyanto, 2019) Agregasi dapat dilakukan pada saat keperluan data bersifat rangkuman, tidak terlalu detail, dalam melakukan analisis. Untuk keperluan data dapat direduksi menggunakan teknik agregasi. Caranya sangat sederhana yaitu dengan melakukan akumulasi data, misalnya data yang semula per bulan diakumulasi menjadi per tahun.

##### 2. Penarikan Contoh

Penarikan contoh digunakan sebelum proses data mining untuk mengurangi jumlah data, hal ini dikarenakan pengolahan data secara keseluruhan akan memakan banyak waktu dan biaya. Pada prinsipnya penarikan contoh yang efektif dapat dilakukan dengan menggunakan beberapa contoh dari data keseluruhan yang

nilainya dapat mewakili sebaik data secara keseluruhan. Untuk dapat mewakili data secara keseluruhan maka suatu contoh setidaknya harus memiliki sifat yang sama dengan himpunan data keseluruhan.

Beberapa teknik penarikan yaitu penarikan contoh secara acak, penarikan contoh tanpa penggantian, penarikan contoh dengan penggantian, dan penarikan contoh bertingkat dengan cara membagi data dalam beberapa bagian kemudian menarik contoh dari tiap bagian tersebut. (Hermawati,2013)

### 3. Pengurangan Dimensi

Kumpulan data yang didapat dari data warehouse biasanya akan berjumlah sangat besar dan memiliki banyak dimensi, untuk mengurangi waktu dan memori yang dibutuhkan untuk memproses data maka dilakukan proses pengurangan dimensi. Teknik ini digunakan untuk mendapatkan data yang lebih kecil volumenya tapi memiliki keterkaitan yang tinggi (Han, 2012)

Data yang memiliki banyak dimensi biasanya juga bermasalah ketika proses penambangan dan analisis untuk itulah perlu diperiksa apakah dimensi-dimensi tersebut dapat dikurangi dengan tetap mempertahankan sifat-sifat data yang penting. (Zaki, 2013)

### 4. Pemilihan Fitur

Pemilihan fitur merupakan pemilihan *sub-set* fitur atau proses pencarian terhadap semua kemungkinan fitur , proses ini dilakukan sebagai upaya untuk mengurangi dimensi dengan memilih fitur-fitur yang tepat atau hanya menggunakan atribut-atribut data yang diperlukan.(Hermawati,2013)

Teknik pemilihan fitur yang biasa digunakan diantaranya yaitu :

- a. *Brute-force approach* : mencoba semua kemungkinan fitur *sub-set*
- b. *Embed ded approach* : terjadi alami sebagai bagian dari algoritma dari proses *data mining*.
- c. *Filter approach* : fitur-fitur dipilih sebelum algoritma dijalankan
- d. *Wrapper approach* : menggunakan algoritma sebagai sebuah proses tersendiri untuk mendapatkan *sub-set* atribut yang paling baik.

## 5. Pembuatan Fitur

Pembuatan fitur adalah proses pembuatan fitur baru yang dapat mengungkapkan informasi-informasi penting dalam sebuah kumpulan data yang lebih baik dan efisien dari atribut-atribut yang telah ada sebelumnya. Metodologi yang biasa digunakan dalam pembuatan fitur ini diantaranya yaitu dengan ekstraksi fitur, pemetaan data ke ruang menggunakan wavelet transformation, dan dengan konstruksi fitur atau menggabungkan beberapa fitur menjadi satu. (Hermawati, 2013)

Secara sekilas proses pembuatan fitur terlihat kurang bermanfaat setelah melihat banyak informasi yang bisa didapat dalam dataset sendiri tanpa perlu menambah fitur. Namun dalam proses penerapan algoritma sering kali kesulitan dalam memproses fitur yang saling berkaitan atau fitur yang nilainya berlebihan. Untuk itulah pembuatan fitur ini dilakukan. (Layton, 2015)

## 6. Pendiskritan dan Pembineran

Pendiskritan adalah proses yang mengubah satu tipe data ke tipe data yang lain. Dalam banyak penelitian yang telah dilakukan sebelumnya, ada banyak variasi yang digunakan untuk menyebut pendiskritan. Variasi yang biasa digunakan antara lain mengubah tipe data dari kuantitatif ke kualitatif atau dari data kontinu ke diskrit atau dari data numerik ke kategorikal. (Yang, 2010)

Bukan hanya bentuk biner, perubahan nilai-nilai atribut bertipe numerik juga dapat menjadi nominal atau ordinal. Banyak metode diskritisasi data yang dapat digunakan, di antaranya adalah: binning, histogram, clustering, decision tree, dan analisis korelasi. (Suyanto, 2019)

### 2.3 **Data Mining untuk Promosi**

Penerapan teknik *data mining* dalam bisnis telah berkembang pesat dalam beberapa tahun terakhir ini. Hal ini disebabkan karena teknik ini telah menghasilkan keuntungan yang nyata terhadap perusahaan. Beberapa hal yang bisa dilakukan dengan menggunakan *data mining* adalah memprediksi jumlah penjualan barang, menemukan hubungan antara barang-

barang yang biasa dibeli oleh pelanggan, memprediksi kenaikan harga bahan baku, dan lain sebagainya. (Ohsawa, 2009)

Bidang promosi adalah salah satu bagian dari *marketing* di mana analisis *cross-selling* dan penerapan strategi penargetan dapat dilakukan. Dengan menggunakan teknik *data mining* maka analisis asosiasi untuk menentukan iklan tambahan yang akan ditawarkan dan penerapan strategi pemberian diskon terhadap pelanggan yang memenuhi kriteria tertentu dapat dilakukan dengan optimal. (Chiu, 2008)

Dalam beberapa tahun terakhir penelitian mengenai target promosi terfokus terhadap bagaimana mengidentifikasi pelanggan sebagai target promosi. Misalnya untuk memprediksi pelanggan mana yang memiliki prospek tinggi, atau untuk mencari pelanggan mana yang akan menjadi target dalam melakukan promosi secara spesifik, dan bagaimana memberikan rekomendasi kepada pelanggan terhadap jenis-jenis promosi yang tersedia dalam suatu produk/ jasa berdasarkan kepada data penjualan terhadap pelanggan dengan kriteria yang sama sebelumnya. (Reutterer, 2016)

Pola strategi promosi yang biasa dilakukan oleh perusahaan/ institusi adalah pola penjualan tradisional skala besar. Berdasarkan pola ini perusahaan membidik produksi, produk dan kemudian memberikan semua pelanggan promosi penjualan yang sama. Namun pola jenis tersebut mengabaikan perbedaan pada pelanggan mereka, pada sebagian besar kasus pola ini mengeluarkan banyak biaya tetapi hanya mendapatkan sedikit keuntungan yang nyata dari pelanggan. Itu berarti jenis promosi seperti ini kurang efektif karena hanya menghamburkan biaya. (Tjiptono, 2017)

Pada saat ini hampir semua pelaku bisnis menjalankan strategi marketing untuk mempromosikan produk atau jasanya. Para pelanggan sekarang tidak hanya sekedar menerima informasi suatu produk secara pasif, maka dari itu diperlukan suatu teknik dan strategi untuk bisa menarik minat pelanggan terutama pelanggan yang berprospek tinggi. (Wang, 2014)

Sementara itu semakin pesatnya perkembangan teknologi informasi, maka teknologi data mining pun semakin populer digunakan diberbagai bidang, hal ini dikarenakan teknologi

data mining dapat memecahkan banyak masalah komersial yang unik pada pemasaran, segmentasi pasar, klasifikasi pelanggan, analisis Churn dan masih banyak lagi. (Du 2006).

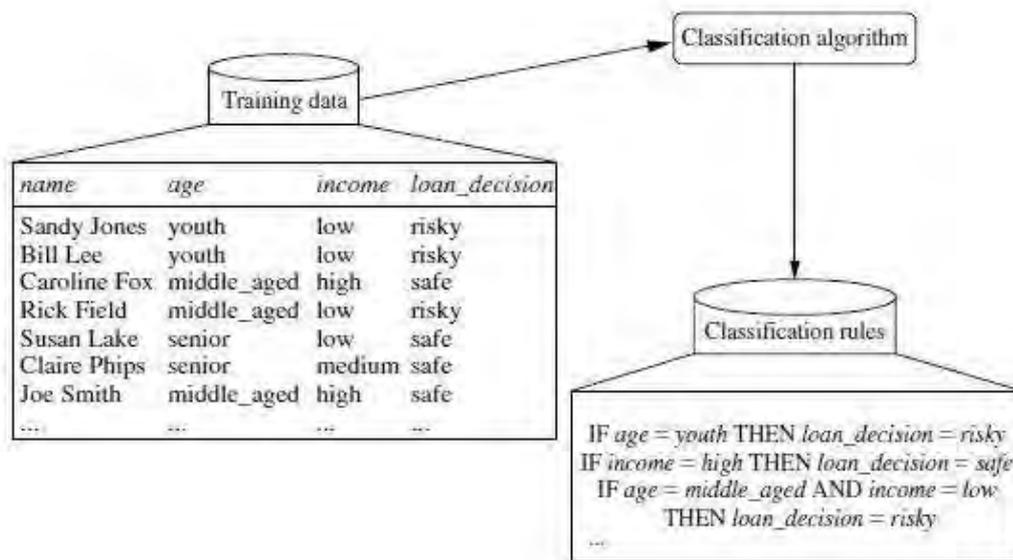
Oleh karena ini data mining sangat baik bila digunakan untuk menentukan strategi promosi perusahaan/institusi agar penerapan strategi promosi yang digunakan bisa lebih efektif dan efisien.

## 2.4 Teknik Klasifikasi

Dalam *data mining* teknik klasifikasi merupakan bagian yang sangat penting. Teknik ini bekerja dengan mempelajari sekumpulan data sehingga dihasilkan aturan yang bisa mengklasifikasikan atau mengenali data-data baru yang sebelumnya tidak diketahui. Klasifikasi juga bisa dijelaskan sebagai proses untuk menyatakan suatu objek data sebagai salah satu bagian dari kelas atau kategori yang sebelumnya telah ditentukan. (Zaki,2013)

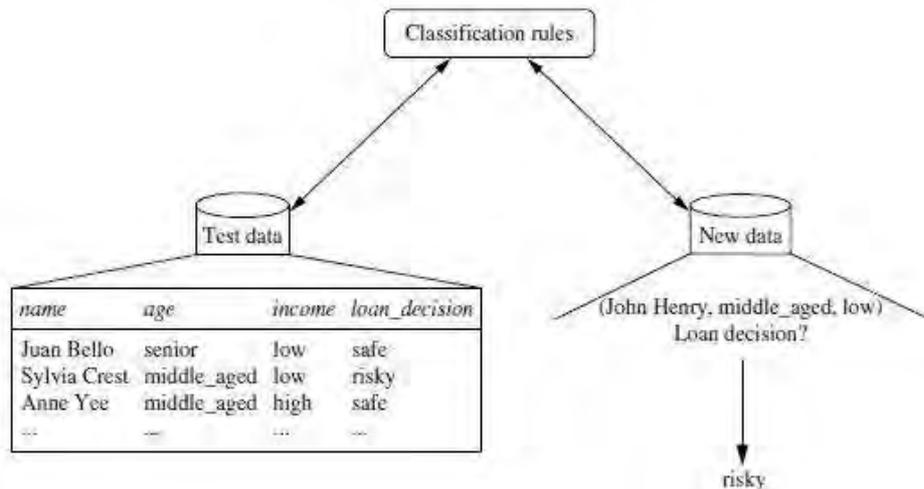
Untuk membuat suatu model dalam teknik klasifikasi dapat dilakukan dengan berdasarkan pengetahuan seorang pakar atau dengan bantuan *machine learning*. Dalam proses dengan bantuan *machine learning* atau secara otomatis terhadap sebuah *dataset* dapat dihasilkan model klasifikasi (target) yang memetakan suatu objek ke dalam salah satu kelas yang telah didefinisikan sebelumnya.

Teknik ini merupakan salah satu bentuk data analisis yang mengekstrak model dan menentukan kelas-kelas penting dalam data. Dalam klasifikasi data terdapat dua proses utama yang terdiri dari *learning step* yaitu proses dimana model klasifikasi dibentuk dan *classification step* yaitu proses menerapkan model yang telah dibuat untuk memprediksi label kelas ke dalam data. Seperti yang terlihat pada gambar 2.2. yaitu langkah pertama dalam proses klasifikasi yaitu pembuatan model (*learning step*) terhadap data peminjam yang memiliki atribut nama(*name*), usia(*age*), pendapatan(*income*) dan keputusan peminjam (*load decision*). Dari data tersebut keputusan peminjam adalah sebagai label atau target atribut, pada langkah pertama pembuatan model atau aturan-aturan berdasarkan dari data latihan (*training data*).



Gambar 2.2. *Learning Step* (Han,2013:329)

Selanjutnya pada tahap kedua yaitu penerapan model klasifikasi (*Classification step*), hasil dari aturan-aturan yang telah dibuat pada tahap sebelumnya kemudian diterapkan dalam data tes untuk memprediksi kategori kelas dari data-data tersebut untuk ilustrasinya bisa dilihat pada gambar 2.3.



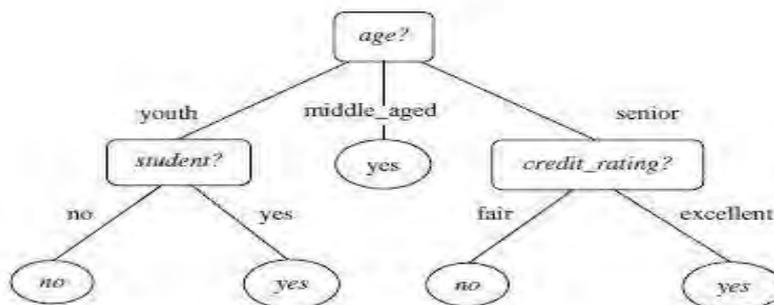
Gambar 2.3. *Classification Step* (Han,2013:329)

Dalam teknik klasifikasi terdapat beberapa metode yang biasa digunakan, diantaranya yaitu :

## 1. *Decision Tree*

Metode ini adalah metode yang paling populer dan sederhana dalam klasifikasi. Metode ini sangat baik untuk model klasifikasi yang harus tahan terhadap derau (*noise*). Beberapa jenis algoritma yang sering digunakan dalam metode ini antara lain yaitu ID3 dan C4.5. pada algoritma ID3 merupakan pembelajaran *decision tree* yang melakukan pencarian secara *greedy*, sehingga hasilnya belum tentu optimal. (Maheswari,2014)

Metode *Decision Tree* ini merupakan tampilan sederhana dari klasifikasi untuk menentukan kelas yang telah ditentukan, pada metode ini terdapat simpul internal dan simpul akar yang merupakan nama-nama atribut.. seperti pada gambar 2.4 yang merupakan contoh metode *decision Tree* terhadap prediksi pembelian computer dengan menggunakan atribut *age*, *student status* dan *credit\_rating*. (Han,2013)



Gambar 2.4 Pohon Keputusan (Han,2013:331)

## 2. Klasifikasi *Naïve Bayes*

Metode ini menggunakan teorema Bayes, yang ditemukan di abad ke 18 oleh Thomas Bayes. Dalam teorema ini, probabilitas atau peluang bersyarat dinyatakan sebagai

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Dimana X adalah bukti, H adalah Hipotesis, sehingga  $P(H|X)$  adalah peluang bahwa hipotesis H adalah benar untu bukti X atau dengan kata lain  $P(H|X)$  merupakan peluang posterior H dengan syarat X. (Zaki,2013)

### 3. *Support Vector Machine*

Metode *support vector machine (SVM)* ini merupakan metode yang bisa diterapkan terhadap data *linear* dan *non linear*. Algoritma ini pertama kali ditemukan oleh Vladimir Vapnik pada tahun 1982. Meskipun waktu yang diperlukan untuk melatih model dengan menggunakan algoritma ini sedikit lambat akan tetapi hasilnya cukup akurat. (Han,2012)

### 4. *K- Nearest Neighbour Rule*

Metode ini merupakan metode sederhana yang dapat digunakan untuk klasifikasi meskipun metode ini juga dapat digunakan untuk estimasi dan peramalan. Metode ini adalah jenis metode pembelajaran *instans* dengan cara menyimpan data pembelajaran kemudian mengklasifikasikan data baru dengan cara membandingkannya dengan data yang memiliki kemiripan paling besar dengan data baru tersebut. (Larose,2015)

*Metode K-Nearest Neighbour pertama kali ditemukan pada tahun 1959an. Tujuan dari klasifikasi adalah untuk memprediksi suatu objek apakah objek tersebut masuk ke dalam satu kelas tertentu atau kelas yang lain.*

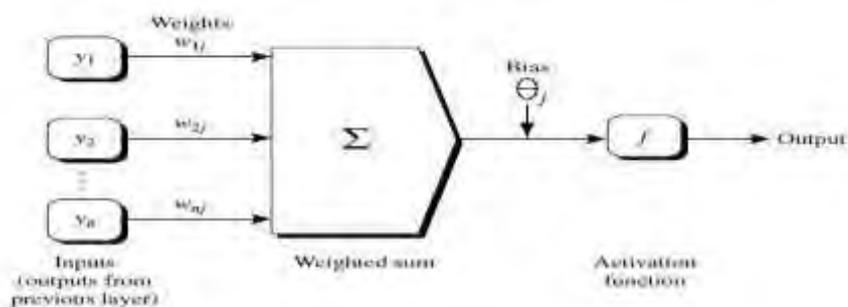
Langkah-langkah dalam algoritma ini cukup sederhana yaitu :

- a. Pilih  $K$  buah tetangga terdekat secara acak
- b. Petakan kumpulan data menjadi bentuk ruang vector
- c. Pisahkan kumpulan data menjadi dua yaitu data latih dan data uji
- d. Hitung jarak  $d$  , antara data latih dan data uji
- e. Lakukan pemilihan nilai dari  $d$  dari yang terkecil hingga terbesar
- f. Ambil dan pisahkan data hasil pemisahan sebanyak  $k$
- g. Perhatikan kelas mayoritas
- h. Klasifikasikan data uji berdasarkan mayoritas. (Primartha,2018)

### 5. *Artificial Neural Network*

Algoritma *Artificial Neural Network* adalah jenis algoritma yang didasari oleh kemampuan cara kerja otak manusia, sehingga bentuk pemodelannya pun mirip seperti bentuk jaringan sistem syaraf pada manusia. Dalam algoritma ini sel syaraf

manusia dimodelkan ke dalam bentuk sel syaraf buatan yang disebut *perceptron*, bentuk dari *perceptron* dapat dilihat pada gambar 2.5



Gambar 2.5. Sel syaraf buatan (*perceptron*)

(Sumber: Han,2013:402)

Dengan  $W_0, W_1, \dots, W_p$  adalah bobot-bobot sinaptik dimana  $W_0$  nilainya sama dengan nilai bias. Sementara itu nilai masukan input dinotasikan sebagai  $X_0, X_1, \dots, X_p$  sehingga dapat disimpulkan  $\sum$  bahwa nilai keluaran atau *output* dari *linear combiner* adalah :

$$V =$$

Secara umum arsitektur pada algoritma *artificial neural network* ini dapat dibagi kedalam 3 macam yaitu : *Single-layer Feedforward Network*, *Multi layer Perceptron* dan *Recurrent Network*. *Single Layer Feedforward Network* adalah jenis arsitektur yang paling sederhana dan hanya memiliki *input layer* dan node sumber yang terhubung ke dalam *output layer* tetapi tidak sebaliknya. (Han,2013)

## 6. *Fuzzy Logic*

Algoritma ini pertama kali muncul pada tahun 1965 dalam sebuah paper yang ditulis oleh Lotfi A. Zadeh yaitu seorang professor di bidang ilmu computer di University of California Berkeley. Dalam algoritma ini berlaku *fuzzy set* yang terdiri dari sebuah *dataset* yang memiliki *record* dan tiap *recordnya* memiliki fungsi keanggotaan. Untuk lebih jelasnya dapat dilihat pada tabel 2.1.

Dalam kenyataannya , derajat keanggotaan bisa berupa apa saja tidak harus bernilai sedikit, sedang atau banyak seperti yang terlihat pada tabel 2.1. Saat ini dengan bantuan *tools* , derajat keanggotaan bisa disederhanakan dengan menggunakan

fungsi keanggotaan dalam bentuk tertentu misalnya bentuk segitiga, trapezium, sigmoid, phi dan masih banyak lagi. (Suyanto, 2019)

Tabel 2.1 Contoh *Fuzzy Set*

<b>Crisp Set</b>	<b>Fuzzy Set</b>		
<b>Pulsa (ribu)</b>	<b>Sedikit</b>	<b>Sedang</b>	<b>Banyak</b>
50	1	0,2	0
100	0,9	0,5	0,1
200	0,1	0,4	0,8
400	0	0,1	1

#### 7. *Ensemble Learning*

Metode ini mulai dikenal disebabkan karena dalam metode-metode klasifikasi lain yang sebelumnya telah dibahas metode ini lebih stabil dan efektif untuk memecahkan masalah klasifikasi dalam dunia nyata. Oleh sebab itu metode ini telah mendapatkan perhatian dari para peneliti selama beberapa decade terakhir.

Selain itu ada banyak aplikasi di kehidupan nyata yang dibuat dengan menggunakan metode *ensemble learning* ini, beberapa diantaranya yaitu : *confidence estimation*, *Feature selection* dan *addressing missing feature*. Pada dasarnya prinsip kerja metode ini adalah mencari jalan atau solusi terbaik dari permasalahan yang ada sehingga dapat disimpulkan atau dipilih keputusan terbaik yang didasarkan pada keputusan-keputusan yang benar yang sebelumnya telah diketahui. (Zhang, 2012)

Adapun beberapa jenis algoritma populer yang biasa digunakan pada metode ini yaitu :

- a. Bagging
- b. Boosting
- c. Random Forest
- d. Stacking

#### 8. *Deep Learning*

Pada metode *neural network* yang telah dibahas sebelumnya yaitu dengan menggunakan *Single layer*. Dengan demikian perbedaan yang jelas antara metode

*deep learning* dengan *neural network* ada pada jumlah *neuron* atau *perceptron* yang dimilikinya.

Akan tetapi dengan jumlah *perceptron* yang cukup banyak maka akan menjadi tantangan tersendiri dalam implementasinya, karena proses implementasi yang buruk akan menyebabkan lambatnya proses pembelajaran jika dibandingkan dengan hasil proses implementasi yang baik. (Layton,2015)

## **2.5 Metode Deep Learning Multy Layer Perceptron (MLP)**

*Deep learning* merupakan jenis pembelajaran mesin dengan menggunakan jaringan saraf buatan. Ada banyak jenis jaringan saraf buatan dari yang sederhana sampai yang rumit. Jaringan saraf (*neural network*) yang sederhana hanya terdiri dari dua atau tiga *neuron* sedangkan yang memiliki jumlah lebih dari tiga dikategorikan rumit (*deep*). (Skansi,2018)

Istilah *Deep Learning* mulai muncul pertama kali pada sebuah paper yang memperkenalkan salah satu jenis *neural network* yaitu *deep belief nets* pada tahun 2006. Setelah itu mulai muncul istilah *deep learning* yang membedakan arsitektur *neural network single layer* dengan *neural network multi layer*.(Primartha,2018)

Teknik ini pada dasarnya adalah pengembangan dari metode *shallow learning* yang memerlukan model dan ekstraksi fitur untuk menghasilkan keakuratan yang tinggi. Metode *deep learning* ini memiliki kemampuan merekayasa fitur secara otomatis sehingga kita tidak kesulitan membangun model ekstraksi fitur (*feature engineering*). (Suyanto,2018)

Metode ini sangat sesuai untuk diterapkan dalam mengolah data besar (*big data*) oleh karena itu metode ini berkembang sangat cepat menjadi algoritma pilihan utama dengan akurasi prediksi yang tertinggi dibanding algoritma lainnya. Adapun model yang digunakan pada algoritma ini berdasarkan pada sel neuron otak manusia. Adapun model jaringan *deep learning* dapat dilihat pada gambar 2.6, model ini hanya menggambarkan 2 *hidden layers* akan tetapi pada kenyataannya jumlah *hidden layers* bisa sangat banyak.



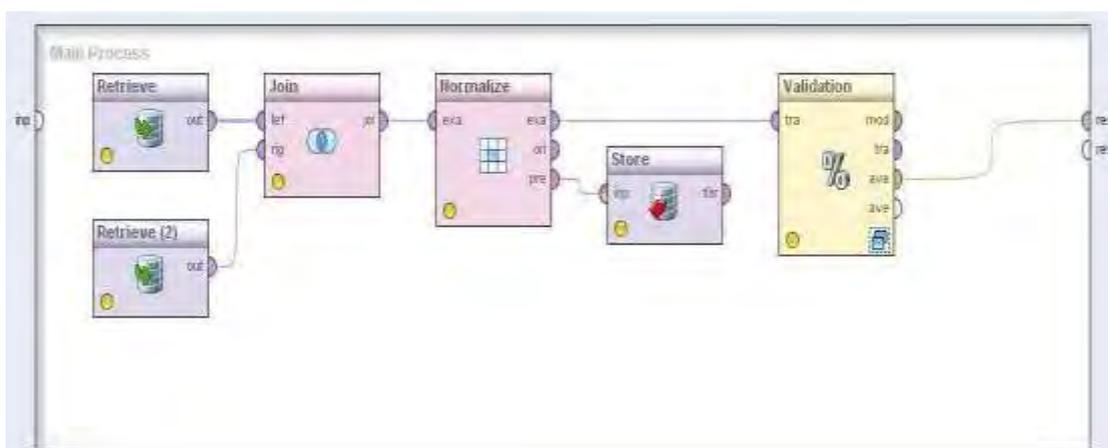
## 2.6 Deep Learning Menggunakan Rapidminer (H20)

Rapidminer merupakan sebuah *platform* yang biasa digunakan untuk pra proses data dan *data mining*. *Software* ini pertama kali dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa dan Simon Fischer pada tahun 2001 di Universitas Dortmund. *Platform* ini dapat digunakan secara komersial ataupun untuk pendidikan dan penelitian. (Land,2012)

Rapidminer ini dari tahun ke tahun semakin populer digunakan sebagai aplikasi untuk penambangan data. *Software* ini sangat mudah di *instal* dan dapat berjalan dalam sistem operasi Windows dan Linux. Selain itu *software* ini memiliki *library* algoritma yang cukup lengkap sehingga layak untuk dicoba. ( Primartha,2018)

*Software* ini memiliki lisensi *GNU Affero General Public License*. Selain nyaman digunakan *software* ini juga memiliki tampilan antarmuka yang sederhana. Dengan menggunakan konsep modelar, dimana setiap langkah-langkah analisis dapat diilustrasikan oleh operator dalam proses analisis.

Sebagaimana yang terlihat pada gambar 2.7 yang memperlihatkan proses sederhana dengan contoh operator *loading*, *preprocessing* dan produksi model. Operator input dan output ini dapat saling berkomunikasi dengan operator lainnya sehingga dapat menerima data input atau meneruskan perubahan data sehingga diperoleh sebuah model yang sesuai dengan operator tersebut.



Gambar 2.7. Ilustrasi Proses Analisis data

(Sumber : Land,2012)

Disamping banyak komponen inti yang dimiliki oleh *software* ini, Rapidminer juga memiliki ekstensi tambahan yang bisa digunakan untuk fungsi-fungsi yang lebih lanjut seperti *platform* R, WEKA, dan fitur H2O yang dapat diintegrasikan secara bebas pada tampilan antarmuka pengguna hal ini membuat Rapidminer memiliki fungsi yang cukup lengkap dan mudah digunakan. (Land,2012)

Penerapan metode *deep learning* dapat dijalankan menggunakan *platform* Rapidminer dengan menggunakan algoritma *deep learning H2O*. Model *deep learning H2O* adalah *multi layer feedforward neural network* yang memiliki fungsi sebagai berikut :

1. Protokol pelatihan *supervised learning* untuk mengerjakan tugas-tugas klasifikasi dan regresi.
2. Cepat dan memori efisien.
3. perhitungan paralel atau multi-threaded dan didistribusikan yang dapat dijalankan pada kluster tunggal atau multi-node.
4. Otomatis, per-neuron dan pembelajaran yang adaptif dan cepat
5. Tambahan parameter ahli untuk proses *tunning*
6. *Deep autoencoder* untuk fitur *unsupervised learning* dan *anomaly detection*.
7. Otomatis mengisi nilai bila terdapat *missing value*.
8. Tampilan antarmuka yang elegan dan intuitif. (Candel,2019)

H2O memiliki banyak jenis algoritma pembelajaran mesin seperti generalisasi pemodelan linier (regresi linier, regresi logistik, dll.), *Naive Bayes*, *principal components analysis*, *k-means clustering*, dan *word2vec*. H2O dapat mengimplementasikan algoritma yang paling baik dikelasnya, seperti pada algoritma *distributed random forest*, *gradient boosting*, and *deep learning*. (Candel,2019)

## 2.7 Penelitian Terkait

Penelitian sebelumnya yang terkait dengan penelitian ini diantaranya yaitu seperti yang terlihat pada tabel 2.2.

Tabel 2.2. Penelitian Terkait

No	Penulis	Tahun	Judul	Teknik	Kesimpulan
1	Sadiq Hussain	2019	<i>Prediction Model on Student Performance based on Internal Assessment using Deep Learning</i>	<i>Deep learning</i>	Penerapan Teknik Pembelajaran Mesin dalam memprediksi kinerja siswa terbukti bermanfaat untuk mengidentifikasi orang yang berkinerja buruk dan dapat memungkinkan tutor untuk mengambil langkah-langkah perbaikan pada tahap awal.
2	Sultana Jabeen, Rani M. Usha, M.A.H. Farquad	2019	<i>Student's Performance Prediction using Deep Learning and Data Mining Methods</i>	<i>Deep learning</i>	Dalam penelitian ini didapat hasil yang dapat disimpulkan bahwa teknik MLP lebih efisien dibandingkan dengan teknik lain dalam prediksi kinerja siswa.
3	Fox Wilton W.T. et all	2018	<i>Prediction Model for Students' Future Development by Deep Learning and Tensorflow Artificial Intelligence Engine</i>	<i>Deep learning</i>	Hasil dari penelitian ini menunjukkan bahwa metode <i>deep learning</i> bisa menjadi alat yang efektif untuk memprediksi kinerja siswa. Itu hasilnya berkisar antara 80% hingga 91%.
4	Choelho, Orlando Bisacchi dan Silveria, Ismar Frango	2017	<i>Deep Learning applied to Learning Analytics and Educational Data Mining</i>	<i>Deep learning</i>	Model yang paling banyak diterapkan dalam <i>educational data mining</i> adalah menggunakan <i>deep learning</i> .
5	Patil Akhilesh P, et all	2017	<i>Effective Deep Learning Model to Predict Student Grade Point Averages</i>	<i>Deep learning</i>	Dari berbagai model <i>deep learning</i> prediksi yang digunakan Model <i>Long Short-Term Memory Network</i> adalah yang paling baik dari segi akurasi untuk memprediksi nilai akhir seorang siswa.

Dalam penelitian sebelumnya (Patil,2017) penerapan metode *deep Learning* digunakan untuk membuat model prediksi nilai akhir siswa mendapatkan hasil akurasi yang lebih tinggi dari pada penerapannya terhadap metode *machine learning* lainnya. Menurut penelitian yang dilakukan oleh Choelho *et all*, dalam beberapa tahun terakhir metode *deep learning* telah menjadi pilihan utama dalam penerapan dan pengolahan *educational data mining*.(Choelho,2017)

Sebelumnya penelitian yang dilakukan oleh Fox, Sultana dan Shadiq mengenai penerapan metode *deep learning* untuk prediksi performa mahasiswa, pada penelitian ini akan mencoba penerapan metode yang sama tetapi untuk masalah klasifikasi profil calon

pendaftar mahasiswa baru. Penerapan metode *deep learning* dapat dilakukan dengan menggunakan berbagai macam alat (*tools*) seperti tensorflow, R, Rapidminer, maupun jenis *tools* yang lain. Sebelumnya Fox (Fox, 2018) dan shadiq , (Shadiq,2019) menggunakan Tensorflow, sementara itu penelitian yang dilakukan oleh Sultana, (Sultana, 2019) menggunakan WEKA. Pada penelitian ini *tools* yang dipakai adalah Rapidminer.

## BAB III OBJEK DAN METODOLOGI PENELITIAN

### 3.1 Profil Universitas Islam Nahdlatul Ulama Jepara

Universitas Islam Nahdlatul Ulama Jepara terletak di kabupaten Jepara yaitu sebuah kota di pesisir utara pulau Jawa. Universitas ini merupakan penggabungan dari INISN, STEINU dan STTDDNU berdasarkan keputusan Mendikbud Nomor 149/E/0/2013. Adapun profil lengkap dari Universitas Islam Nahdlatul Ulama Jepara akan dijelaskan pada bagian selanjutnya

Universitas Islam Nahdlatul Ulama Jepara memiliki 18 program studi strata 1 dan satu program studi jenjang strata 2. Universitas ini telah mendapatkan akreditasi B dari BAN-PT dengan nomor 3598/SK/BAN-PT/Akred/PT/X/2017. Akan tetapi meskipun telah memiliki akreditasi peringkat B, Universitas Islam Nahdlatul Ulama Jepara selalu berkomitmen untuk semakin meningkatkan kualitasnya.

Dalam perkembangannya Universitas Islam Nahdlatul Ulama Jepara telah melakukan kerjasama dengan berbagai perguruan tinggi dalam negeri maupun luar negeri melalui program pertukaran pelajar. Diantaranya kerjasama yang dilakukan dengan University Brunei Darussalam pada tahun 2017, sebanyak 21 mahasiswa dari Universitas tersebut melakukan program pertukaran pelajar di Universitas Islam Nahdlatul Ulama Jepara selama 50 hari. Selain itu Universitas Islam Nahdlatul Ulama Jepara juga melakukan kerjasama dengan Universitas lain yang berada di Malaysia dan Mesir.

#### 3.1.1 Visi , Misi dan Tujuan

Adapun visi yang dimiliki oleh Universitas Islam Nahdlatul Ulama Jepara yaitu :  
"Menjadi Universitas Islam Terdepan dalam Melahirkan Sumber Daya Manusia Unggul Yang Berilmu, Cendekia, Berakhlakul Karimah Berlandaskan Nilai–**nilai Ahlussunnah Wal Jama’ah** dalam Pengembangan Iptek dan Seni Budaya pada tahun 2028" . Sedangkan misi yang dimiliki antara lain yaitu :

1. Mengembangkan sumber daya akademik universitas untuk meningkatkan kualitas pembelajaran, penelitian dan pengembangan masyarakat yang efektif dan efisien.
2. Mengembangkan sistem manajemen yang bermutu,sehat, harmonis, yang berbasis pada nilai-nilai islam Ahlusunah wal Jamaah.
3. Memperluas jaringan kerja sama secara kontinyu dan selektif untuk mempercepat dan menjaga serta menjamin pencapaian tujuan universitas.

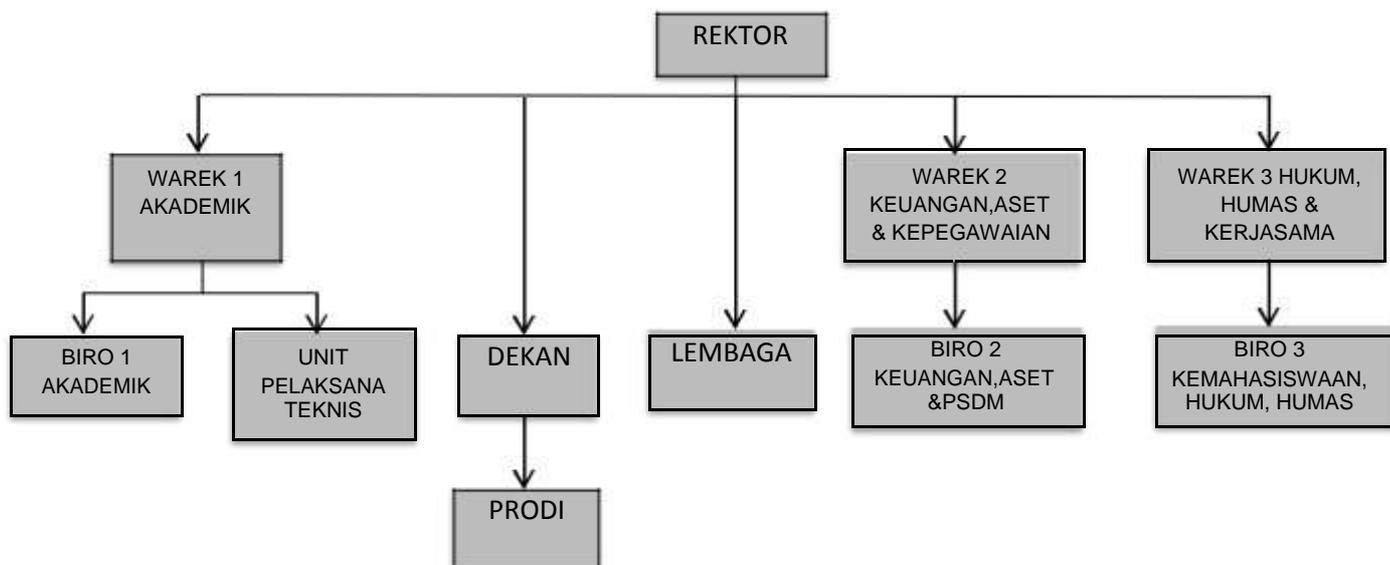
Adapun yang menjadi tujuan dalam mewujudkan Visi dan Misi universitas yaitu :

1. Membangun *University Branding* melalui *global networking for global partnership* sesuai dengan ajaran Ahlusunah wal Jamaah.
2. Menjadi pusat penyelenggara akademik yang unggul (*Academic Excellence*) berstandar nasional dan internasional.
3. Menjadi pusat penelitian IPTEK yang unggul (*Research Excellence*).

### **3.1.2 Struktur Organisasi**

Adapun struktur organisasi yang dijalankan di Universitas Islam Nahdlatul Ulama Jepara terdiri dari pimpinan universitas, pimpinan fakultas dan program studi, ketua program studi, kepala biro, kepala lembaga, kepala UPT, kepala bagian dan kepala tata usaha fakultas dan program pascasarjana.

Pimpinan universitas adalah Rektor dan Wakil Rektor, yang memiliki tugas antara lain menjalankan fungsi pengelolaan universitas secara keseluruhan, melaksanakan penyelenggaraan pendidikan, penelitian dan pengabdian kepada kepada masyarakat, membina *civitas* akademika (dosen & mahasiswa) serta membina hubungan dengan alumni, baik di lingkungan universitas dan masyarakat. Adapun struktur organisasi Universitas Islam Nahdlatul Ulama (UNISNU) jepara bisa dilihat dari gambar 3.1.



Gambar 3.1. Struktur Organisasi

### 3.1.3 Strategi Promosi UNISNU Jepara

Pada saat ini Universitas Islam Nahdlatul Ulama Jepara tidak memiliki strategi promosi secara tertulis. Proses promosi yang selama ini dilakukan oleh panitia atau Tim penerimaan mahasiswa baru hanya sebatas pencetakan spanduk, pamflet dan leaflet.

Untuk jalur penerimaan mahasiswa baru Universitas ini memiliki dua jalur yaitu jalur umum dan jalur beasiswa. Penerimaan mahasiswa baru jalur beasiswa memiliki persyaratan nilai akademis yang baik, mahasiswa yang mendaftar pada jalur ini tidak perlu mengikuti tes ujian masuk.

#### 1. Jalur Khusus

Jalur khusus merupakan jalur yang disiapkan bagi siswa yatim piatu dengan prestasi akademik unggul atau hafidz/hafidzah peserta pada jalur ini diterima sebagai mahasiswa baru tanpa tes dan pembebasan biaya sumbangan pengembangan institusi (SPI), bagi hafidz/hafidzah memperoleh pembebasan biaya SPP tetap selama masa studi maksimal 8 semester dengan persyaratan sebagai berikut :

- a. Lulusan baru yang memperoleh ranking 1,2, atau 3 pada kelas XII semester gasal parallel satu sekolah dan memiliki nilai rata-rata rapor minimal 8,5 untuk kelas X dan XI semester gasal dan genap dan pada kelas XII semester gasal.
- b. Khusus program studi ekonomi dan bisnis harus memiliki nilai rata-rata 9.0 pada kelas X dan XI semester gasal dan genap dan pada kelas XII semester gasal.
- c. Hafidz/hafidzah al-quran 30 juz, dibuktikan dengan syahadah/ijasah penghafal al quran serta lulus tes hapalan al-quran.
- d. Bagi siswa yatim piatu dibuktikan dengan surat keterangan dari desa/kelurahan.

## 2. Jalur Prestasi

Jalur prestasi ini adalah jalur pilihan bagi siswa yang berprestasi juara pada lomba/kejuaraan akademik ataupun non akademik saat SMA/MA/SMK/Paket C atau sederajat. Pendaftar pada jalur ini akan memperoleh keringanan berupa masuk tanpa tes serta pembebasan membayar SPI untuk juara 1 tingkat propinsi/nasional/internasional dan potongan SPI sebesar 50% bagi juara 1 di tingkat kabupaten. Adapun persyaratannya yaitu :

- a. Lulusan baru atau lulusan tahun sebelumnya untuk program studi hukum keluarga, komunikasi dan penyiaran islam serta budidaya perairan.
- b. Bagi prestasi akademik nilai rata-rata rapor kelas X dan XI pada semester gasal dan genap adalah 8,5
- c. Bagi prestasi non akademik memiliki kewajiban untuk aktif dalam kegiatan organisasi kemahasiswaan sesuai dengan keahliannya dan bersedia dikirim mengikuti lomba.kejuaraan untuk mewakili UNISNU.
- d. Pernah juara 1 lomba akademik atau non akademik yang diselenggarakan atau diakui oleh UNISNU, instansi/lembaga/badan pemerintahan/KONI/Dewan kesenian baik perorangan atau beregu.

### 3. Jalur Umum

Jalur umum terbuka bagi lulusan SMA/MA/SMK/Paket C atau sederajat tanpa batasan tahun lulus. Jalur ini terbagi menjadi regular I dan regular II, jalur regular II adalah untuk pendaftar yang sudah bekerja dibuktikan dengan surat keterangan bekerja. Pendaftara jalur regular I untuk jurusan program studi hukum keluarga, komunikasi dan penyiaran islam serta budidaya perairan memperoleh potongan 50% biaya SPI .

Informasi mengenai universitas ini dapat juga diperoleh melalui jalur digital

diantaranya yaitu melalui website <https://unisnu.ac.id> dan media social seperti instagram <https://www.instagram.com/unisnujepara> dan twitter <https://twitter.com/unisnujepara>.

### 3.2 Objek Penelitian

Data penerimaan calon mahasiswa baru pada Universitas Islam Nahdlatul Ulama Jepara periode 2014-2018 yang akan menjadi objek pada penelitian ini. Adapun sumber data ini berasal dari *database* PMB UNISNU Jepara tahun 2014-2018 yang didapat langsung dari bagian Unit Pelaksana Teknis (UPT) Universitas Islam Nahdlatul Ulama Jepara. Adapun yang menjadi atribut dari data tersebut dapat dilihat pada tabel 3.1 berikut :

Tabel 3.1 Daftar Atribut Data

No	Nama Atribut	Type Atribut	Nilai	MetaData
1	Nama	Polynominal	Sani Nurmalasari (8332)	Missing : 0 Least :- Most :-
2	Jenis Kelamin	Polynominal	P/L (2)	Missing : 0 Least : L (4589) Most : P (4752)
3	Alamat	Polynominal	Jepara (386)	Missing : 1 Least : Yogyakarta (1) Most : Jepara (7655)
4	Tahun Lahir	Integer	1990, 2018	Missing : 1 Min : 0 Max : 2018
5	Jenis Sekolah	Polynominal	SMA, MA, SMK, Paket C (4)	Missing : 593 Least : Paket C (225) Most : MA (3235)
6	Pekerjaan Orang tua	Polynominal	Wiraswasta, PNS, Petani, (13)	Missing : 2 Least : TNI/POLRI (43) Most : Wiraswasta (4647)

No	Nama Atribut	Type Atribut	Nilai	MetaData
7	Prodi	Polynominal	Manajemen, Akuntansi, (19)	<i>Missing</i> : 3 <i>Least</i> : Pendidikan Guru PAUD (129) <i>Most</i> : Manajemen (1553)
8	Angkatan	Integer	2014,2015,2016, 2017,2018 (5)	<i>Missing</i> : 0 <i>Min</i> : 2014 <i>Max</i> : 2018
9	Referensi	Polynominal	Radio, Teman (6)	<i>Missing</i> : 2 <i>Least</i> : Radio (52) <i>Most</i> : Teman (6193)
10	NIM	Real	141120001251	<i>Missing</i> : 0 <i>Min</i> : 0 <i>Max</i> : 182610000637
11	Nilai Ujian	Integer	540, 500	<i>Missing</i> : 0 <i>Min</i> : 0 <i>Max</i> : 560

Dari atribut diatas dapat dilihat hasil visualisasinya pada gambar 3.2. Berdasarkan hasil pengamatan data dari jumlah total 9.341 *record* terdapat 598 atribut yang hilang, sehingga hanya 8.743 data yang memiliki atribut lengkap. sementara jumlah mahasiswa yang berjenis kelamin laki-laki adalah 4.589 sedangkan yang berjenis kelamin perempuan adalah 4.752.

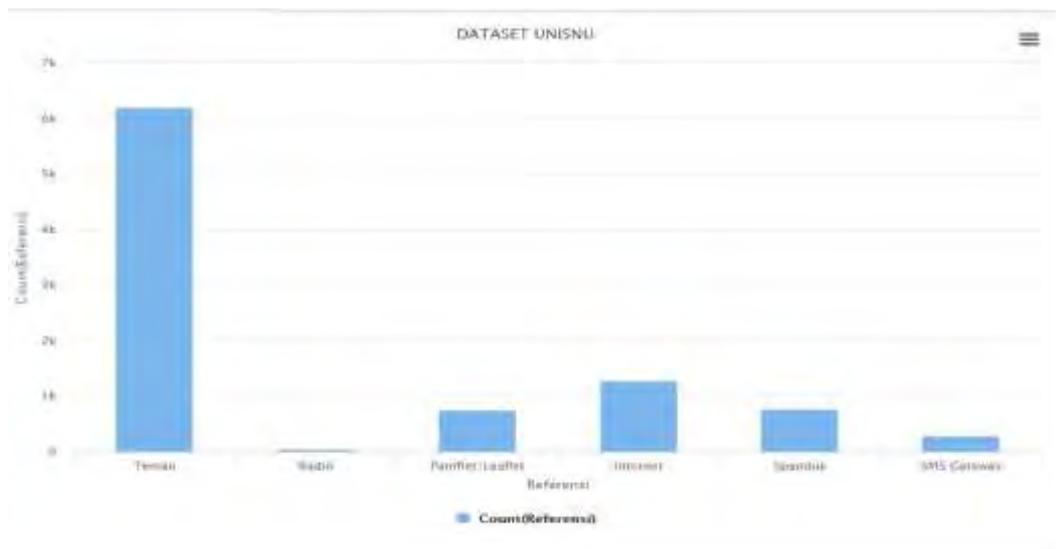
Untuk alamat atau asal kota kebanyakan berasal dari kota Jepara sejumlah 7.655 dengan total nama kota yang berbeda sebanyak 386 kota. Sementara untuk data tahun lahir masih memiliki atribut yang hilang dan juga atribut yang tidak valid sehingga belum bisa dilihat jelas perbandingannya. Begitu juga dengan atribut jenis asal sekolah yang masih memiliki atribut yang hilang sebanyak 593.

Row ...	Nama	JE...	ALAMAT	YEAR(ta.Tan...	JenisSekola...	Pekerjaan O...	NilaiUjian	NIM	Prodi	Angkatan	Referensi
1	WAHYU PRA...	P	JEPARA	1995	SMA	Tukang Kayu...	300	0	Akuntansi	2014	Teman
2	NURUL HIDA...	L	JEPARA	1991	SMK	Tukang Kayu...	400	141120001251	Akuntansi	2014	Teman
3	ELLA USWAT...	P	JEPARA	1996	SMK	Lain-lain	400	141120001215	Akuntansi	2014	Teman
4	RISA MARQIY...	P	JEPARA	1994	SMA	Sopir	410	141120001192	Akuntansi	2014	Teman
5	PRASETYO B...	L	JEPARA	1994	SMA	Pegawai Swa...	300	0	Akuntansi	2014	Teman
6	INDAH MUAR...	P	JEPARA	1994	SMK	Wiraswasta	270	141120001277	Akuntansi	2014	Teman
7	SETYANNING...	P	JEPARA	1992	SMA	Pegawai Neg...	0	0	Akuntansi	2014	Teman
8	DWI AYU BE...	P	JEPARA	1995	SMA	Pegawai Neg...	400	141120001360	Akuntansi	2014	Radio
9	AHMAD IBNU...	L	JEPARA	1994	SMA	Tukan BatuB...	400	141120001193	Akuntansi	2014	PandettLeaflet
10	KRISDAMAYA...	P	JEPARA	1995	SMK	Wiraswasta	220	141120001217	Akuntansi	2014	Teman
11	PUTRI SRI S...	P	JEPARA	1995	SMA	Wiraswasta	410	141120001252	Akuntansi	2014	Teman
12	TIGUR RAMA...	L	JEPARA	1996	SMA	Tukang Kayu...	400	141120001253	Akuntansi	2014	Teman
13	INTAN SUCL...	P	JEPARA	1996	SMK	Wiraswasta	410	141120001194	Akuntansi	2014	Internet
14	LILIK MUSFI...	P	JEPARA	1996	SMK	Wiraswasta	400	141120001361	Akuntansi	2014	Spanduk

Gambar 3.2. Visualisasi *Dataset*

Sementara nilai ujian juga memiliki banyak atribut kosong dikarenakan ketentuan pengisian atribut ini didasarkan pada nilai hasil ujian masuk mahasiswa baru sehingga jika nilai bernilai nol atau kosong maka calon mahasiswa tersebut tidak mengikuti ujian masuk penerimaan mahasiswa baru. Untuk pendaftar yang memilih jalur beasiswa prestasi maka tidak perlu mengikuti ujian masuk dan atribut nilai ujiannya pun akan menjadi kosong. Sedangkan untuk atribut NIM memiliki ketentuan bahwa seseorang tersebut apabila memiliki atribut NIM berarti telah diterima menjadi mahasiswa baru di Universitas.

Kemudian atribut referensi yang menunjukkan dari mana informasi mengenai perguruan tinggi ini didapat oleh calon mahasiswa akan menjadi atribut penting dasar penentuan strategi promosi. Adapun visualisasi dari atribut tersebut dapat dilihat pada gambar 3.3.



Gambar 3.3. Visualisasi Data Atribut Referensi

### 3.3 Strategi Promosi yang Digunakan

Berdasarkan hasil observasi yang telah dilakukan di Universitas Islam Nahdlatul Ulama Jepara didapatkan beberapa hasil strategi promosi yang digunakan di Universitas ini dan diterapkan ke dalam *dataset*. Strategi promosi tersebut didapat berdasarkan dengan data penerimaan mahasiswa baru atau *dataset* yang telah diberikan sebagai bahan pertimbangannya. Yang menjadi bahan pertimbangan pembuatan strategi promosi antara lain :

1. Pengurangan jumlah penerima beasiswa 100% dari Universitas bagi pendaftar yang berada di dalam kota Jepara, sebab dalam kota Jepara, Universitas Islam Nahdlatul Ulama hanya memiliki sedikit pesaing yang lebih unggul sehingga tidak perlu promosi yang berlebihan di dalam kota Jepara.
2. Mengoptimalkan pendaftar pada program studi yang memiliki peminat sedikit, baik itu untuk pendaftar dalam kota maupun dari luar kota.
3. Pemberian potongan biaya pendidikan diutamakan untuk pendaftar yang baru lulus, pekerjaan orang tuanya hanya petani atau tidak bekerja dan memilih program studi yang jarang diminati.

Adapun beberapa strategi promosi tersebut adalah dapat dilihat pada tabel 3.2.

berikut :

Tabel 3.2 Strategi Promosi

No	Strategi Promosi	Potongan SPI	PERSYARATAN						
			Kota Asal	Program Studi	Status Graduate	Nilai Tes	Pekerjaan Orang tua	Referensi	Jalur masuk
1	SP1	20%	Jepara	Komunikasi dan Penyiaran Islam, Pendidikan Guru PAUD, Teknik Elektro	-	-	-	Teman	Umum
2	SP2	30%	luar Jepara	-	-	-	-	Internet	Umum
3	SP3	10%	Jepara	Selain Manajemen, Akuntansi, Pendidikan Agama Islam	-	-	-	Internet	Umum
4	SP4	50%	-	Komunikasi dan Penyiaran Islam, Pendidikan Guru PAUD, Teknik Elektro	Fresh Graduate	-	Wiraswasta	-	Tidak Diketahui
5	SP5	80%	-	Komunikasi dan Penyiaran Islam, Pendidikan Guru PAUD, Teknik Elektro	Fresh Graduate	Baik	Tidak Bekerja	-	Umum
6	SP6	40%	-	-	-	Sangat Baik	-	-	Umum

Berikut adalah penjelasan dari tabel 3.2 :

1. Bagi calon pendaftar yang berasal dari kota Jepara dengan referensi berasal dari teman serta memilih jalur umum dengan program studi yang peminatnya kurang seperti program studi komunikasi dan penyiaran islam, pendidikan guru PAUD, dan teknik elektro maka akan mendapatkan potongan biaya pendidikan sebesar 20% dan teman yang memberi referensi apabila sedang menempuh studi di UNISNU akan mendapatkan voucher biaya pendidikan. (SP1)
2. Bagi calon pendaftar yang berasal dari luar kota dengan referensi berasal dari internet serta memilih jalur umum maka akan mendapatkan potongan biaya pendidikan sebesar 30%.(SP2)
3. Bagi calon pendaftar yang berasal dari kota Jepara memilih jalur umum dan referensinya berasal dari internet serta bukan memilih program studi yang paling

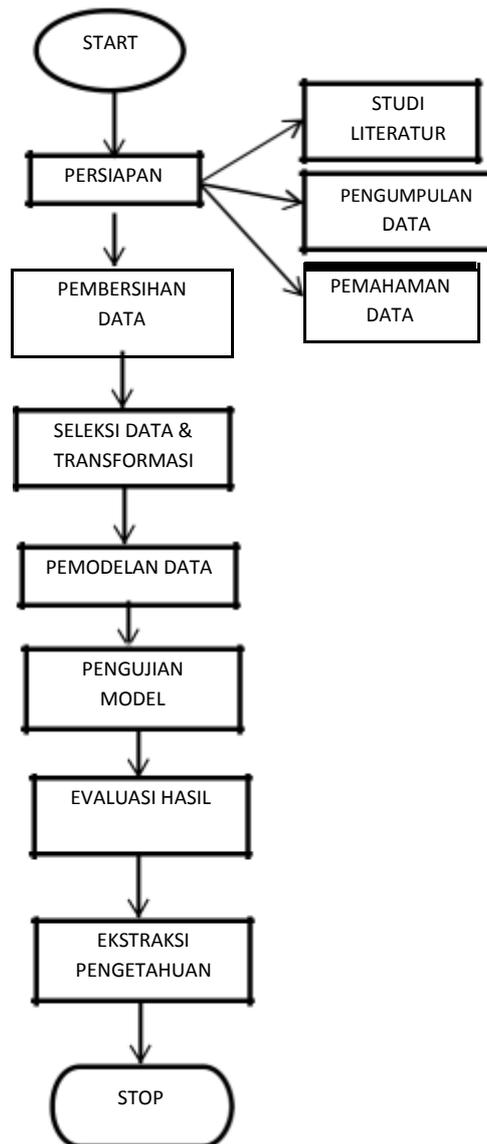
sering diminati maka akan mendapatkan potongan biaya pendidikan sebesar 10%.(SP3)

4. Bagi calon pendaftar yang atribut pekerjaan orang tuanya wiraswasta dan jalur masuknya tidak diketahui serta memilih program studi yang jarang peminatnya maka akan mendapatkan potongan biaya pendidikan sebesar 50%. (SP4)
5. Bagi calon pendaftar yang atribut pekerjaan orang tuanya sebagai petani atau tidak bekerja kemudian dia berasal dari SMA dan memilih program studi yang jarang diminati serta status graduate nya adalah Fresh Graduate dan kategori nilai tesnya baik maka akan mendapat potongan biaya pendidikan sebesar 80%. (SP5)
6. Bagi pendaftar jalur umum yang kategori nilai tesnya sangat baik maka akan mendapatkan potongan biaya SPI 40%. (SP6)
7. Diluar ketentuan dari strategi promosi yang telah disebutkan diatas maka selain itu tidak diterapkan strategi promosi apapun. (SP0)

### **3.4 Metodologi Penelitian**

Pada penelitian ini metode yang digunakan dimulai dari studi literatur untuk menemukan metode *data mining* yang terbaik untuk digunakan. Selanjutnya adalah pengumpulan data yang berhubungan dengan masalah yang akan diteliti. Setelah data terkumpul kemudian pra proses data yaitu antara lain proses pembersihan data, pemilihan atribut, pendiskritan dan pembineran serta transformasi atribut. Hal ini dimaksudkan agar data mudah diolah dan diterapkan dengan menggunakan algoritma *deep Learning* pada proses selanjutnya yaitu *data mining*.

Setelah proses *data mining* selesai maka hasil yang didapat akan dipresentasikan dengan cara menyajikan pola atau informasi yang ditemukan melalui proses presentasi pengetahuan. Untuk lebih jelasnya dapat dilihat pada gambar 3.2



Gambar 3.2 Metode Penelitian

1. Persiapan meliputi proses studi literature, pengumpulan data, dan pemahaman data.
  - a. Pada proses studi literature dilakukan dengan membaca hasil penelitian sebelumnya yang terkait, mencari informasi dan membaca artikel yang membahas subjek penelitian dan objek penelitian.
  - b. Pada proses pengumpulan data dilakukan dengan cara mendatangi kepala UPT Universitas Islam Nahdlatul Ulama Jepara untuk meminta data-data yang dibutuhkan.

- c. Sedangkan pada proses pemahaman data yaitu dilakukan analisis terhadap data yang tersedia untuk dapat ditentukan penerapan algoritma yang tepat.
2. Dengan proses pembersihan data menghilangkan *noise* dan *inconsistency* pada data seperti menghilangkan data yang memiliki *missing value*.
3. Pemilihan data yaitu memilih atribut data yang relevan untuk proses analisis selanjutnya. Kemudian proses transformasi data yaitu mengubah bentuk data agar sesuai dengan kebutuhan data pada proses data mining.
4. Pemodelan Data yaitu proses menerapkan metode data mining *deep learning* untuk mendapatkan informasi yang diinginkan yaitu target calon mahasiswa baru
5. Pengujian Model yaitu proses untuk menguji keakuratan model yang telah dibuat.
6. Evaluasi Hasil yaitu proses mengevaluasi hasil dari pemodelan data dan pengujian model untuk dapat disimpulkan hasilnya.

## BAB IV

### IMPLEMENTASI DAN PEMBAHASAN HASIL

#### 4.1 Tahap Persiapan

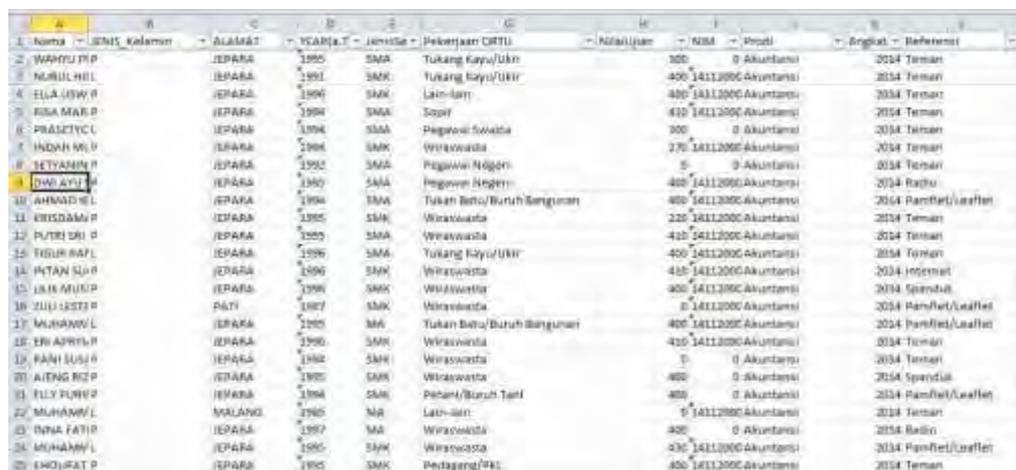
Pada tahap ini terdapat 3 proses yaitu studi literatur, pengumpulan data dan pemahaman data.

##### 1. Studi Literatur

Pada tahap studi literature dilakukan dengan mencari informasi dan artikel terkait dengan subjek dan objek dari penelitian yang akan dilakukan. Yang menjadi sumber dari informasi dalam proses ini antara lain jurnal-jurnal resmi, buku, serta artikel.

##### 2. Pengumpulan Data

Selanjutnya dilakukan pengumpulan data dengan cara melakukan observasi di Universitas Islam Nahdlatul Ulama Jepara dan mengajukan permintaan kebutuhan data kepada kepala UPT Universitas Islam Nahdlatul Ulama Jepara. Data yang diperlukan didapat dari *Database* Penerimaan Mahasiswa Baru (PMB) dan *Database* Mahasiswa pada Universitas Islam Nahdlatul Ulama Jepara. Jumlah *record* / *tuple* data yang didapat secara keseluruhan adalah 9.341 *record*. Adapun tampilan data yang merupakan gabungan dari data PMB dan data Mahasiswa yang didapat dari Universitas Islam Nahdlatul Ulama Jepara dapat dilihat pada gambar 4.1



No	Nama	JENIS Kelamin	ALAMAT	YEAR/B.T	JenisSex	Pekerjaan ORTU	AlamatJum	NIM	Prodi	Engkat	Referensi
1	WAHYU P P		JEPARA	1995	SMA	Tukang Kayu/Ukii		300	0 Akuntansi		2014 Teman
2	MURULHEL		JEPARA	1991	SMK	Tukang Kayu/Ukii		400	14112000 Akuntansi		2014 Teman
3	ELLA JHW P		JEPARA	1996	SMK	Lain-lain		400	14112000 Akuntansi		2014 Teman
4	RISA MAR P		JEPARA	1994	SMA	Soepi		410	14112000 Akuntansi		2014 Teman
5	PRASYCYL		JEPARA	1994	SMA	Pegawai Swasta		300	0 Akuntansi		2014 Teman
6	INDAH M D		JEPARA	1994	SMK	Wiraswasta		270	14112000 Akuntansi		2014 Teman
7	SETYANR P		JEPARA	1992	SMA	Pegawai Negeri		5	0 Akuntansi		2014 Teman
8	DWIAYU P		JEPARA	1990	SMA	Pegawai Negeri		400	14112000 Akuntansi		2014 Ratu
9	AHMAD HEL		JEPARA	1994	SMA	Tukang Batu/Buruh Bangunan		400	14112000 Akuntansi		2014 Ramfiet/Liaffet
10	KHISDAM P		JEPARA	1995	SMK	Wiraswasta		220	14112000 Akuntansi		2014 Teman
11	PUTRI SRI P		JEPARA	1995	SMA	Wiraswasta		410	14112000 Akuntansi		2014 Teman
12	TIGUR RAFI		JEPARA	1996	SMA	Tukang Kayu/Ukii		400	14112000 Akuntansi		2014 Teman
13	HYAN SLO P		JEPARA	1996	SMK	Wiraswasta		410	14112000 Akuntansi		2014 Injemat
14	IBN ALI P		JEPARA	1998	SMK	Wiraswasta		400	14112000 Akuntansi		2014 Spandak
15	ZULI LESTI P		PATI	1997	SMK	Wiraswasta		0	14112000 Akuntansi		2014 Ramfiet/Liaffet
16	MUHAMM L		JEPARA	1995	SMK	Tukang Batu/Buruh Bangunan		400	14112000 Akuntansi		2014 Ramfiet/Liaffet
17	EW APRIL P		JEPARA	1996	SMK	Wiraswasta		410	14112000 Akuntansi		2014 Teman
18	PANI SUSI S		JEPARA	1992	SMK	Wiraswasta		0	0 Akuntansi		2014 Teman
19	AIRANG RCP		JEPARA	1995	SMK	Wiraswasta		400	0 Akuntansi		2014 Spandak
20	ELLY FURN P		JEPARA	1994	SMK	Pebani/Buruh Tani		400	0 Akuntansi		2014 Ramfiet/Liaffet
21	MUHAMM L		MALANG	1995	MA	Lain-lain		0	14112000 Akuntansi		2014 Teman
22	DINA FATI P		JEPARA	1997	MA	Wiraswasta		400	0 Akuntansi		2014 Balin
23	MUHAMM L		JEPARA	1995	SMK	Wiraswasta		230	14112000 Akuntansi		2014 Ramfiet/Liaffet
24	ENOKRAT P		JEPARA	1995	SMK	Pedagang/IKI		400	14112000 Akuntansi		2014 Teman

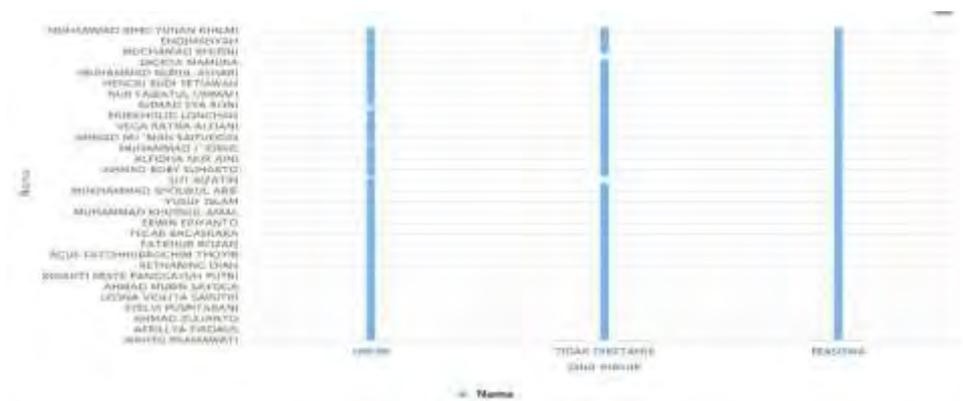
Gambar 4.1. Data PMB (Sumber : UPT UNISNU Jepara )

3. Pemahaman Data

Data yang diperoleh dari UPT UNISNU Jepara berjumlah 9.341 *record* dan memiliki 11 atribut yaitu :

a. Nama

Atribut nama merupakan ID yang nilai tiap *tuple* nya berbeda, meskipun ada kesamaan beberapa nama depan tapi tidak ada duplikat data. Perbandingan antara atribut Nama dengan atribut jalur masuk dapat dilihat pada gambar 4.2 berikut :



Gambar 4.2 Perbandingan Atribut Nama Terhadap Atribut Jalur Masuk  
(Sumber : Rapidminer 9.3 Data Unisnu Jepara)

b. Alamat

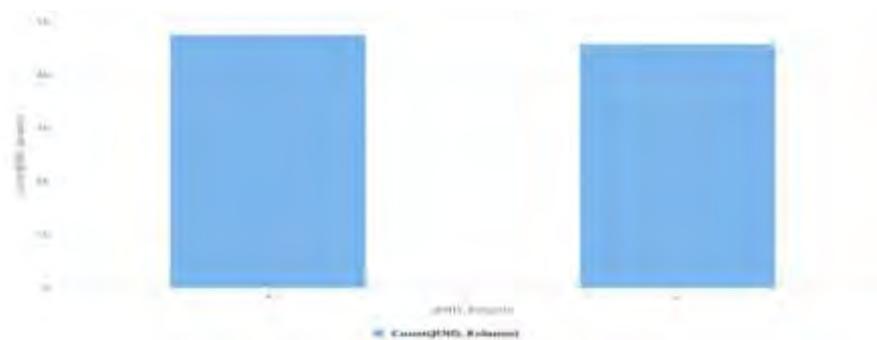
Alamat adalah atribut yang memiliki beberapa format bertipe polynominal yang berbeda sehingga perlu dilakukan penyamaan format agar nilai yang di dapat konsisten. Sebagai visualisasi data alamat dapat dilihat pada gambar 4.3



Gambar 4.3 Visualisasi Atribut Alamat  
(Sumber : Rapidminer 9.3 Data Unisnu Jepara)

c. Jenis Kelamin

Atribut jenis kelamin memiliki tipe polynominal. Atribut ini hanya memiliki dua nilai yaitu L untuk laki-laki dan P untuk perempuan. Visualisasi atribut jenis kelamin dapat dilihat pada gambar 4.4.

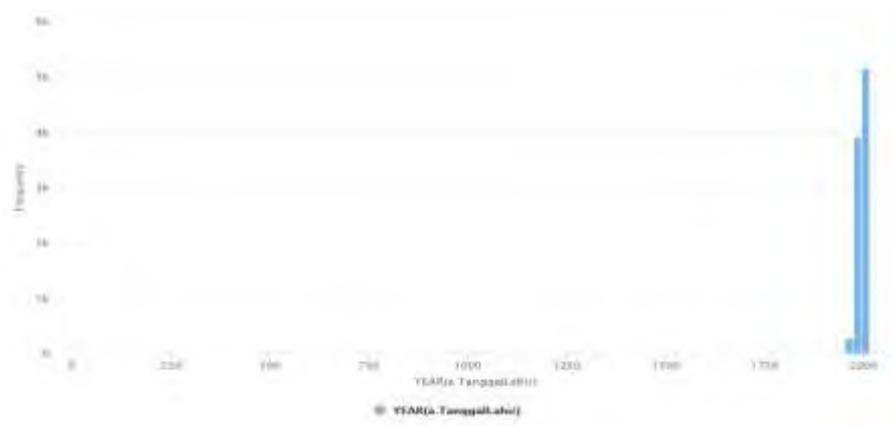


Gambar 4.4. Visualisasi Atribut Jenis Kelamin

(Sumber : Rapidminer 9.3 Data Unisnu Jepara)

d. Tahun Lahir

Atribut tahun lahir memiliki tipe integer yang memiliki nilai antara 0-2018 sehingga ada kemungkinan terjadi kesalahan dalam proses *input* sehingga atribut ini harus melalui proses pembersihan data menghilangkan kesalahan tersebut. Gambar 4.5 berikut adalah visualisasi data pada atribut tahun lahir.

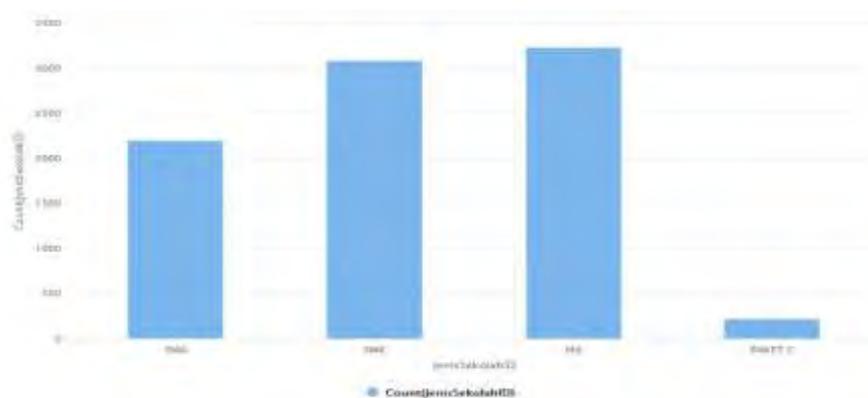


Gambar 4.5. Visualisasi Atribut Tahun Lahir

(Sumber : Rapidminer 9.3 Data Unisnu Jepara)

e. Jenis Asal Sekolah

Atribut jenis sekolah bertipe polynominal. Nilai dari atribut ini antara lain SMA, MA, SMK, dan Paket C meskipun terdapat beberapa *record* yang kosong tetapi tidak begitu banyak dan nilai dari atribut ini cukup stabil dan konsisten. Pada gambar 4.6 adalah tampilan visualisasi dari atribut jenis asal sekolah tersebut.



Gambar 4.6. Visualisasi Atribut jenis Asal Sekolah

(Sumber : Rapidminer 9.3 Data Unisnu Jepara)

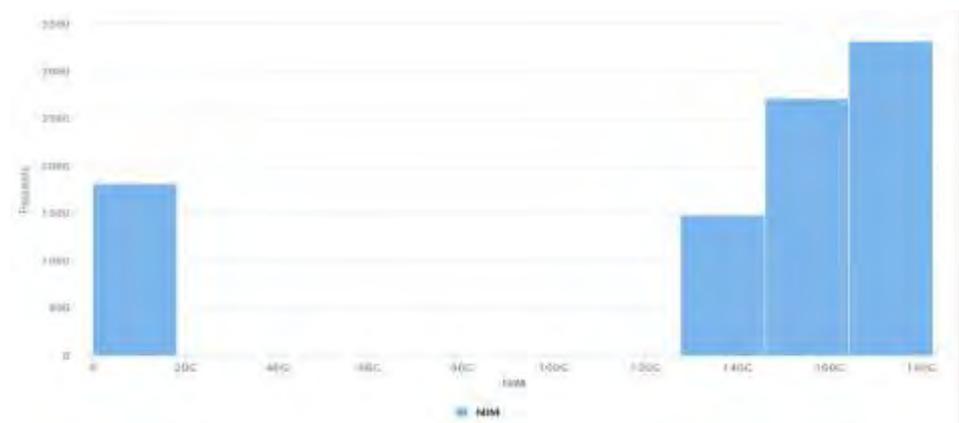
f. Pekerjaan Orang Tua

Atribut pekerjaan orang tua ini bertipe kategorikal yang memiliki jumlah kategori 13 jenis pekerjaan orang tua antara lain yaitu : nelayan, petani, wiraswasta, PNS, TNI/POLRI, Tidak Bekerja dan lain sebagainya. Nilai dari atribut ini sangat stabil dan konsisten meskipun memiliki 1 nilai yang hilang, tetapi secara keseluruhan atribut ini tidak memerlukan penanganan pada pra proses data yang terlalu rumit. Pada gambar 4.7 merupakan visualisasi data atribut pekerjaan orang tua.



#### h. NIM

Atribut ini juga memiliki tipe integer. Atribut ini bukanlah merupakan ID, keberadaan atribut ini adalah sebagai indikator bahwa pendaftar telah resmi diterima sebagai mahasiswa di UNISNU Jepara apabila telah memiliki atribut ini. Oleh karena itu banyak juga nilai dari atribut ini yang kosong sehingga diperlukan praproses data sebelum dapat diolah dalam proses *data mining*. Pada Gambar 4.9 dapat dilihat visualisasi datanya.

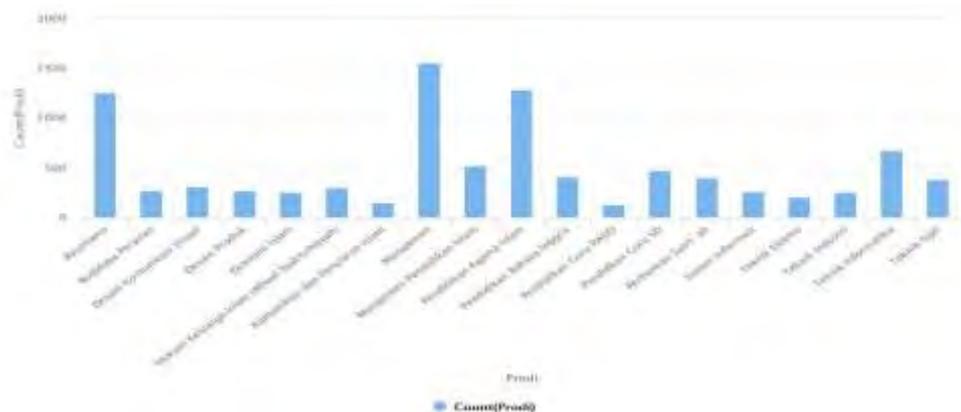


Gambar 4.9 Visualisasi Atribut NIM

(Sumber : Rapidminer 9.3 Data Unisnu Jepara)

#### i. Program Studi

Atribut ini adalah atribut kategorikal yang terdiri dari 19 kategori. Nilai yang paling sering muncul antara lain adalah : Akuntansi, Manajemen dan Pendidikan Agama Islam. Semetara yang paling sedikit adalah : Komunikasi dan Penyiaran Islam, Pendidikan Guru PAUD, dan Teknik Elektro. Pada Gambar 4.10 dapat dilihat visualisasi datanya

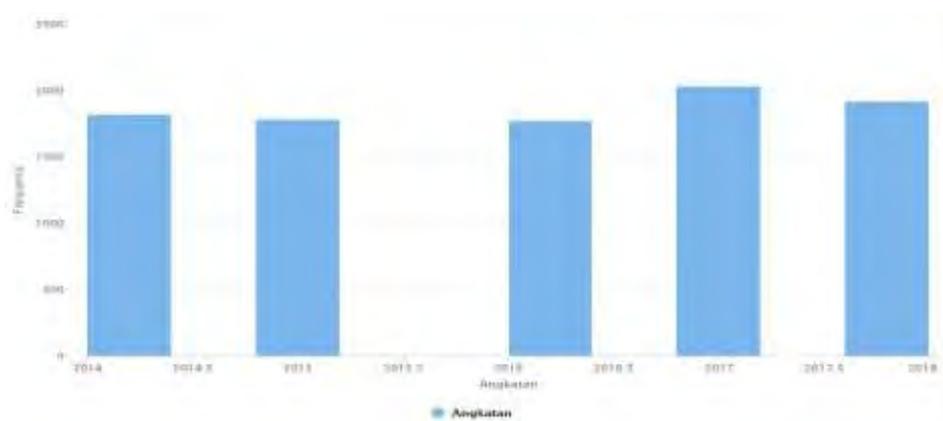


Gambar 4.10 Visualisasi Atribut Program Studi

(Sumber : Rapidminer 9.3 Data Unisnu Jepara)

j. Angkatan

Karena pada proses pengumpulan data, data yang dipilih merupakan data pada tahun 2014-2018 maka nilai dari atribut ini pun memiliki sebaran 2014-2018. Pada lima tahun tersebut tidak terjadi perbedaan jumlah pendaftar yang signifikan seperti terlihat pada gambar 4.11.

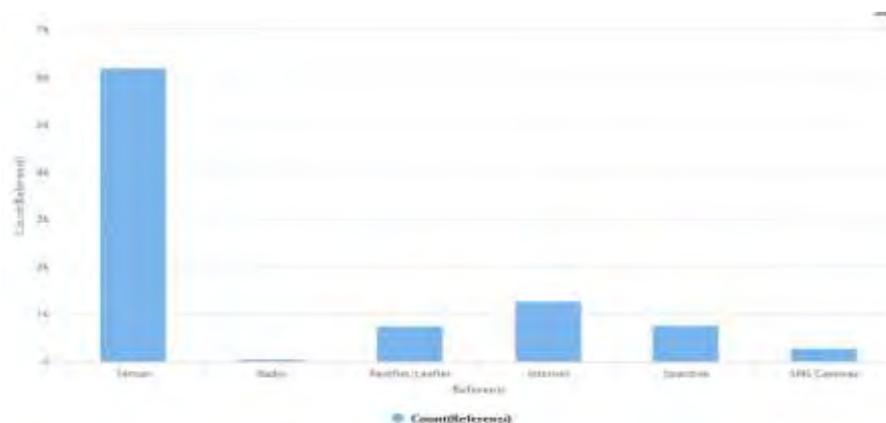


Gambar 4.11 Visualisasi Atribut Angkatan

(Sumber : Rapidminer 9.3 Data Unisnu Jepara)

#### k. Referensi

Atribut ini merupakan kategorikal yang memiliki nilai antara lain : Internet, Pamflet/Leaflet, Radio, SMS Gateway, Spanduk, dan Teman. Dengan nilai terbesar yaitu Teman dan yang terkecil adalah Radio seperti yang terlihat pada gambar 4.12.

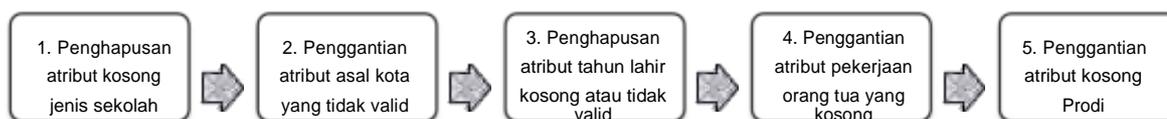


Gambar 4.12 Visualisasi Atribut Referensi

(Sumber : Rapidminer 9.3 Data Unisnu Jepara)

## 4.2 Pembersihan Data

Proses pembersihan data dilakukan untuk memperbaiki kesalahan pada data yang akan mempengaruhi kualitas data dan menyebabkan proses analisis *data mining* menjadi lebih sulit dilakukan. Tujuan dari proses ini adalah untuk membuat data lebih konsisten. Pada proses pembersihan data dilakukan langkah-langkah sebagaimana yang terlihat pada gambar 4.13.



Gambar 4.13. Langkah-langkah Pembersihan Data

### 1. Penghapusan Atribut Kosong Jenis Sekolah.

Dari jumlah record 9.341 data setelah dilakukan proses penghapusan terhadap atribut jenis sekolah yang kosong maka jumlah data berkurang sebanyak 594 *record*

sehingga jumlah data secara keseluruhan menjadi 8.747 *record* data. Proses penghapusan *record* ini dipilih karena jumlah data yang kosong berjumlah hamper 5% dari total keseluruhan data sehingga bila dilakukan pengisian atribut dengan metode nilai terbesar atau modus dikhawatirkan akan berpengaruh terhadap hasil akhir dalam proses *data mining*.

No	Nama	JENIS_Kota	ALAMAT	YEAR	JenisSekolah	Pekerjaan	NilaiUjian	NIM	Prodi	Angkatan	Referensi
1	PUTRI AYU P	JERAMBA	1997	MA	Wiraswast	0	1511	Akuntansi	2015	Spanduk	
2	MUHAMMAD SIDI	JERAMBA	1998	MA	Wiraswast	400	1511	Akuntansi	2015	Internet	
3	JULIANSARI S	JERAMBA	1998	MA	Wiraswast	400	0	Akuntansi	2015	Teman	
4	TAMARA SETI P	JERAMBA	1999	MA	Petani/Bu	0	1511	Akuntansi	2015	Pamflet/Leaflet	
5	NINA A P	JERAMBA	1997	MA	Wiraswast	300	1511	Akuntansi	2015	Teman	
6	ERHAPAM P	JERAMBA	1998	MA	Ayuhaywan	0	0	Akuntansi	2015	Teman	
7	ARINI DEVA P	JERAMBA	1997	MA	Petani/Bu	0	1511	Akuntansi	2015	Spanduk	
8	ARI MUHAMMAD	JARANG	1997	MA	Wiraswast	0	1511	Akuntansi	2015	Teman	
9	MIFTAH D	JERAMBA	1996	MA	Petani/Bu	0	1511	Akuntansi	2015	Teman	
10	PARISA IRI D	JERAMBA	1999	MA	Wiraswast	200	1711	Akuntansi	2017	Teman	
11	MURRI RIZKA D	KUALA LUMPUR	2000	MA	Wiraswast	300	1711	Akuntansi	2017	Teman	
12	SITI ARIYATI P	JERAMBA	1998	MA	Wiraswast	100	1811	Akuntansi	2018	Spanduk	
13	SHAFWA RIZKA D	JERAMBA	2000	MA	Pedagang	0	0	Akuntansi	2018	Teman	
14	PUTRI LINDA P	JERAMBA	2000	MA	Lain-lain	0	0	Akuntansi	2018	Teman	
15	MUSLIMAH D	JERAMBA	1998	MA	Pegawai S	0	1811	Akuntansi	2018	Pamflet/Leaflet	
16	MIFTYANNI D	JERAMBA	2000	MA	Wiraswast	0	1811	Akuntansi	2018	Teman	
17	RITA RIZKA D P	JERAMBA	2005	MA	Tidak Berk	0	0	Akuntansi	2018	Spanduk	
18	RIZKI QADIR D	JERAMBA	2000	MA	Pegawai S	0	1811	Akuntansi	2018	Teman	
19	ELINDA KHALI	SEMARANG	1995	MA	Petani/Bu	0	0	Budidaya	2018	Teman	
20	ACHMAD ALI	JERAMBA	1997	MA	Petani/Bu	0	1711	Budidaya	2017	Internet	
21	ANGGAS SARI	SURABAYA	1997	MA	Wiraswast	0	1511	Desain Gra	2015	Teman	
22	ABBA SUTERA P	SURABAYA	1988	MA	Petani/Bu	0	1711	Desain Gra	2017	Teman	
23	ADITYA SARI	SEMARANG	1997	MA	Pegawai S	0	1711	Desain Gra	2017	Teman	
24	INDRA PUTRA	JERAMBA	1999	MA	Wiraswast	0	0	Desain Gra	2018	Teman	

Gambar 4.14. Data Atribut Jenis Sekolah yang Dihapus

2. Penggantian Atribut Asal Kota yang tidak Valid

Pada atribut asal kota terdapat 10 *record* yang isinya tidak valid atau bukan merupakan nama kota oleh sebab itu dilakukan teknik pembersihan data dengan cara penggantian atribut. Nilai yang digunakan untuk mengisi atribut yang tidak valid adalah nilai modus atau yang nilai terbanyak dari keseluruhan *record* yang ada. Gambar 4.15. merupakan tampilan *record* yang mengalami penggantian atribut.

No	Nama	JENIS_Kota	ASAL KOTA	YEAR	JenisSekolah	Usia saat i	Pekerjaan	NilaiUjian	NIM	Prodi	Angkatan	Referensi
1	FAISAL L	31 AGUSTUS	1999	SMK	18	Wiraswast	0	171270	Desain Gra	2017	Teman	
2	MICHAEL	1997	1997	MA	20	Wiraswast	0	171290	Desain Gra	2017	Spanduk	
3	LISA NIP	01 MEI 1999	1999	MA	18	Wiraswast	0	171510	Komuni	2017	Teman	
4	LAILA SP	27-08-1997	1997	SMK	18	Wiraswast	0	151110	Manajer	2015	Teman	
5	AHMAD L	17 JANUARI 1	2016	SMK	0	Wiraswast	200	161110	Manajer	2016	Pamflet/Leaflet	
6	DWI RIZKI P	23 JUNI 1998	1998	SMK	15	Wiraswast	200	171110	Manajer	2017	Teman	
7	NINA A P	8 JUNI 1999	1999	MA	18	Pedagang	300	171110	Manajer	2017	Teman	
8	AMIN FL	09 JUNI 1988	1988	SMA	31	Wiraswast	0	171310	Pendidik	2017	Teman	
9	SRI MULI P	31-01-2000	2000	MA	18	Pegawai S	0	181130	Pendidik	2018	Teman	
10	NGATDI L	10071994	1994	SMA	21	Petani/Bu	0	151220	Teknik E	2015	Internet	

Gambar 4.15. Atribut Asal Kota yang tidak Valid

### 3. Penghapusan Atribut tahun lahir yang kosong atau tidak valid

Pada proses penghapusan atribut tahun lahir yang kosong atau tidak valid sebanyak 111 *record* data dihapus. Data yang dihapus tersebut meliputi 1 *record* data atribut yang kosong, dan 110 *record* data yang tidak valid. Sehingga jumlah data keseluruhan setelah melalui proses ini menjadi 8.629 *record* data. Pada gambar 4.16 bisa dilihat tampilan atribut yang tidak valid tersebut.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
3	Nama	JENIS_Kel	ASAL_KOT	YEAR	JenisSekol	Pekerjaan	NilaiUjian	NIM	Prodi	Angkatan	Referensi								
2	MELIK P	JEPARA	2014	SMK	Wiraswast	410	1411	Akuntansi	2014	Teman									
3	ULFA P	JEPARA	2014	MA	Wiraswast	430	1411	Akuntansi	2014	Teman									
4	INNA P	JEPARA	2014	SMK	Wiraswast	0	0	Akuntansi	2014	Internet									
5	LINDA P	JEPARA	2014	SMA	Wiraswast	420	1411	Akuntansi	2014	Spanduk									
6	NOR R P	JEPARA	2014	SMA	Tukang Bat	430	1411	Akuntansi	2014	Teman									
7	TIARA B	JEPARA	2014	SMA	Wiraswast	410	1411	Akuntansi	2014	Teman									
8	PITRI P	JEPARA	2013	SMA	Pegawai P	400	1511	Akuntansi	2013	Pamflet/Leaflet									
9	LUTHI P	JEPARA	2015	SMA	Sopir	240	1511	Akuntansi	2015	Pamflet/Leaflet									
10	CHAFI P	JEPARA	2015	SMA	Wiraswast	240	1511	Akuntansi	2015	Spanduk									
11	MIHA P	JEPARA	2015	SMA	Lain-lain	0	1511	Akuntansi	2015	Teman									
12	ANIK P	JEPARA	2015	SMA	Wiraswast	260	1511	Akuntansi	2015	Teman									
13	UMI L P	JEPARA	2015	MA	Wiraswast	0	1511	Akuntansi	2015	Teman									
14	ARVIN L	JEPARA	2015	SMA	Pegawai P	260	1511	Akuntansi	2015	Teman									
15	MYA P P	JEPARA	2015	SMK	Pegawai S	0	1511	Akuntansi	2015	Internet									
16	MEGA R	JEPARA	2015	SMA	Wiraswast	0	0	Akuntansi	2015	Pamflet/Leaflet									
17	RING A P	JEPARA	2015	SMK	Wiraswast	400	1511	Akuntansi	2015	Teman									
18	ALYA R P	JEPARA	2016	MA	Wiraswast	200	1611	Akuntansi	2016	Teman									
19	SILVIA P	JEPARA	2016	SMA	Wiraswast	420	1611	Akuntansi	2016	Teman									
20	AGUS L	JEPARA	2014	SMK	Smpir	430	0	Budidaya	2014	Teman									
21	AGNOR R	JEPARA	2014	SMK	Tukang Ka	400	0	Budidaya	2014	Teman									
22	MIRO P	JEPARA	2015	MA	Petani/Bu	0	0	Budidaya	2015	Teman									
23	LAILA P	JEPARA	2015	SMK	Wiraswast	0	0	Desain Ko	2015	Teman									
24	NISFU L	JEPARA	2016	MA	Wiraswast	0	1611	Desain Ko	2016	Spanduk									
25	MUHAL	JEPARA	2014	SMK	Wiraswast	410	1412	Desain Pro	2014	Teman									

Gambar 4.16 Data Atribut Tahun Lahir yang tidak Valid

### 4. Penggantian atribut pekerjaan orang tua yang kosong

Sebanyak 1 *record* data atribut pekerjaan orang tua yang kosong dan diganti nilainya sesuai dengan nilai modus atau jenis pekerjaan terbanyak pada *dataset* yaitu : Wiraswasta. sebagai ilustrasi prosesnya dapat dilihat pada gambar 4.17.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Nama	JENIS_Kel	ASAL_KOT	YE	JenisSekol	Pekerjaan	ORTU	NilaiUjian	NIM	Prodi	Angka	Referensi				
2	AL AMIP	JEPARA	199	MA				400	0	Manag	2015	Teman				
3																
4	Setelah Penggantian Atribut															
5																
6	AL AMIP	JEPARA	199	MA		Wiraswasta		400	0	Manag	2015	Teman				
7																
8																
9																
10																
11																
12																
13																
14																
15																

Gambar 4.17. Ilustrasi Penggantian Atribut Pekerjaan Orang Tua

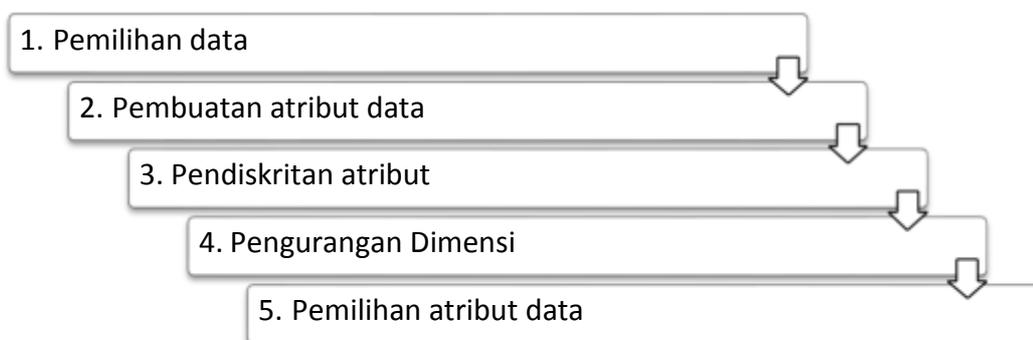
#### 5. Penggantian atribut kosong Prodi

Pada proses ini terdapat 3 *record* data yang diganti karena atribut Prodi nya bernilai kosong. Penggantian ini dilakukan dengan mengisi nilai modus atau nilai terbanyak yang terdapat pada atribut program studi yaitu : Akuntansi, Manajemen dan Pendidikan Agama Islam. Contoh ilustrasi proses ini dapat dilihat pada gambar 4.18 sehingga hasil data setelah melalui proses pembersihan berjumlah 8.629 *record*.

	A	B	C	D	E	G	H	I	J	K	L	M	N
1	Nama/ENIS_Kel	ASAL_KOT	YEAR	Ta	JenisSeko	Pekerjaan	OR	Nilai	Nilai	Prodi	Angka Referensi		
2	HAFID L	JEPARA	1998	SMK	Wiraswasta		0	0			2014 Teman		
3	MUHFA L	JEPARA	1992	SMK	Tukang Kayu/		0	0			2014 Teman		
4	NUR Z L	JEPARA	1987	MA	Petani/Buruh	400	0				2014 Internet		
5													
6	MENJADI												
7	HAFID L	JEPARA	1998	SMK	Wiraswasta		0	0	Akuntansi		2014 Teman		
8	MUHFA L	JEPARA	1992	SMK	Tukang Kayu/		0	0	Manajemen		2014 Teman		
9	NUR Z L	JEPARA	1987	MA	Petani/Buruh	400	0		Pendidikan Agama Islam		2014 Internet		
10													
11													
12													
13													
14													

Gambar 4.18. Ilustrasi Penggantian Nilai Atribut Prodi

### 4.3 Seleksi Data dan Transformasi



Gambar 4.19. Langkah-langkah seleksi dan transformasi data

Berdasarkan gambar 4.19 yaitu langkah-langkah seleksi dan transformasi data, maka proses ini pun terbagi menjadi 5 bagian yaitu :

#### 1. Pemilihan Data

Data yang dipilih merupakan yang yang berasal dari *database* PMB dan *database* Mahasiswa Universitas Islam Nahdlatul Ulama Jepara yang berjumlah 9.341 *record*

yang setelah melalui proses pembersihan mengalami pengurangan menjadi 8.629 *record* data. Semua data tersebut akan dipilih untuk melalui pra proses selanjutnya sampai dengan proses *data mining*. Untuk hasil lebih jelasnya dapat dilihat pada gambar 4.20 berikut

Nama	ENIS	ASAL	FAKULTAS	Jenjang	Pekerjaan	Status	NIM	Tes	Angkasa	Penerima
KARIMIA P	EPARA	1901	SMA	Tulang 64	000	0	Akuntansi	2014	Terima	
WIKASIA L	EPARA	1901	SMA	Tulang 64	400	141120001	Akuntansi	2014	Terima	
DUAUWIP	EPARA	1904	SMA	Luhur	400	141120001	Akuntansi	2014	Terima	
ESANARA P	EPARA	1904	SMA	Siam	410	141120001	Akuntansi	2014	Terima	
FRACENY	EPARA	1904	SMA	Pegawai 7	300	0	Akuntansi	2014	Terima	
WARIKAP	EPARA	1904	SMA	Witasari	210	141120001	Akuntansi	2014	Terima	
ISTANARA P	EPARA	1907	SMA	Pegawai 7	0	0	Akuntansi	2014	Terima	
WYANIV P	EPARA	1909	SMA	Pegawai 7	400	141120001	Akuntansi	2014	Revisi	
SHAMARA S	EPARA	1904	SMA	Tukan 801	200	141120001	Akuntansi	2014	Pemeriksaan	
RIHANA P	EPARA	1905	SMA	Witasari	310	141120001	Akuntansi	2014	Terima	
FUFIARA P	EPARA	1905	SMA	Witasari	410	141120001	Akuntansi	2014	Terima	
WIKASIA L	EPARA	1906	SMA	Tulang 62	400	141120001	Akuntansi	2014	Terima	
WYANIV P	EPARA	1906	SMA	Witasari	410	141120001	Akuntansi	2014	Revisi	
LEKASIA P	EPARA	1906	SMA	Witasari	400	141120001	Akuntansi	2014	Revisi	
ZULIESTA P	FATI	1907	SMA	Witasari	0	141120001	Akuntansi	2014	Pemeriksaan	
WYANARA P	EPARA	1908	MA	Tukan 801	400	141120001	Akuntansi	2014	Pemeriksaan	
DIANARA P	EPARA	1908	SMA	Witasari	410	141120001	Akuntansi	2014	Terima	
RAMDIA P	EPARA	1904	SMA	Witasari	0	0	Akuntansi	2014	Terima	
AFTARIA P	EPARA	1907	SMA	Witasari	000	0	Akuntansi	2014	Revisi	
LYXARA P	EPARA	1904	SMA	Pegawai 7	400	0	Akuntansi	2014	Pemeriksaan	
MILIANA S	MALANG	1905	MA	Luhur	0	141120001	Akuntansi	2014	Terima	
WYANARA P	EPARA	1907	MA	Witasari	400	0	Akuntansi	2014	Revisi	
MULIANA S	EPARA	1905	SMA	Witasari	410	141120001	Akuntansi	2014	Pemeriksaan	
KARIMIA P	EPARA	1905	SMA	Pegawai	400	141120001	Akuntansi	2014	Terima	

Gambar 4.20 Hasil Pemilihan Data

## 2. Pembuatan Atribut Data.

Pada gambar 4.20. yang memperlihatkan daftar atribut data yang diterima dari kepala UPT Universitas Islam Nahdlatul Ulama Jepara dengan jumlah atribut sebanyak 11. Dari ke-11 atribut ini terdapat atribut yang memiliki informasi yang kurang lengkap sehingga perlu dibuat atribut baru yang berdasarkan atribut lama yang memiliki informasi yang lebih akurat. Atribut yang dimaksud antara lain adalah :

### a. Pembuatan Atribut Re-Registrasi dan Jalur Masuk

Nilai Atribut NIM didapat jika calon mahasiswa yang mendaftar dapat melanjutkan ke tahap selanjutnya sampai diterima menjadi mahasiswa baru. Sementara Nilai Tes merupakan nilai hasil ujian masuk. Nilai tes yang berisi *value null* menunjukkan bahwa mahasiswa tersebut tidak mengikuti tes ujian masuk. Dikarenakan atribut NIM dan Nilai Tes kurang memberikan informasi yang cukup mendalam, oleh karena itu dilakukan proses pembuatan fitur/ atribut baru berdasarkan data dari atribut NIM dan Nilai Tes sehingga menghasilkan informasi yang lebih bermanfaat. Atribut yang dibuat dari data NIM dan Nilai Tes

adalah Re-Registrasi dan Jalur Masuk. Kedua atribut ini dibuat dengan ketentuan sebagai berikut :

- 1) Jika NIM bernilai kosong atau 0 maka mahasiswa tersebut tidak melakukan Re-Registrasi atau Re-Registrasi bernilai TIDAK. selain itu mahasiswa melakukan Re-Registrasi atau Re-Registrasi bernilai YA.
- 2) Jika Nilai Tes bernilai kosong atau 0 dan Re-Registrasi bernilai YA maka mahasiswa tersebut mengikuti jalur masuk BEASISWA.
- 3) Jika Nilai Tes bernilai Kosong atau 0 dan Re-Registrasi bernilai TIDAK maka jalur masuk TIDAK DIKETAHUI.
- 4) Jika Nilai Tes berisi *value* dan Re-Registrasi bernilai YA maka Jalur masuk adalah UMUM.

NIM	Nama	Jurusan	Prodi	Angkatan	Re-Registrasi	Jalur Masuk
0001 L	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	TIDAK DITAKHUI
0001 L	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	BEASISWA
0001 L	SPARA	MA	Widawati	Teknik Sipil	2018 internet	BEASISWA
0001 L	SPARA	MA	Widawati	Teknik Sipil	2018 internet	BEASISWA
0001 L	SPARA	MA	Pegawai negeri	Teknik Sipil	2018 internet	BEASISWA
0001 P	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	BEASISWA
0001 P	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	BEASISWA
0001 L	SPARA	MA	Tulang sayu/ultra	Teknik Sipil	2018 Pamflet/Lauffer	BEASISWA
0001 L	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	BEASISWA
0001 P	SPARA	MA	Widawati	Teknik Sipil	2018 Pamflet/Lauffer	BEASISWA
0001 P	SPARA	MA	Tidak Benar	Teknik Sipil	2018 Terlanir	BEASISWA
0001 L	SPARA	MA	Pesawat	Teknik Sipil	2018 internet	BEASISWA
0001 L	SPARA	MA	Widawati	Teknik Sipil	2018 SMS online	BEASISWA
0001 L	SPARA	MA	Tulang sayu/ultra	Teknik Sipil	2018 Terlanir	BEASISWA
0001 L	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	BEASISWA
0001 L	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	TIDAK DITAKHUI
0001 P	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	BEASISWA
0001 L	SPARA	MA	Pegawai swasta	Teknik Sipil	2018 Terlanir	TIDAK DITAKHUI
0001 P	SPARA	MA	Pegawai swasta	Teknik Sipil	2018 Terlanir	BEASISWA
0001 L	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	BEASISWA
0001 P	SPARA	MA	Widawati	Teknik Sipil	2018 Terlanir	BEASISWA
0001 L	SPARA	MA	Pesawat	Teknik Sipil	2018 internet	BEASISWA

Gambar 4.21 Pembuatan Atribut Re-Registrasi dan Jalur Masuk.

#### b. Pembuatan Atribut Usia Saat Mendaftar

Pada gambar 4.21. menunjukkan hasil yang didapat dari pembuatan atribut Re-Registrasi dan Jalur Masuk. Selain dua atribut diatas, akan dibuatkan juga atribut Usia saat mendaftar yang didapat dari nilai atribut tahun lahir yang dikurangi tahun mendaftar atau Angkatan. Hasil yang didapat dapat dilihat pada gambar 4.22 berikut ini.

No	Nama	ASAL	JK	Usia	Pendidikan	Pekerjaan	Tahun Pendaftaran	Referensi	Strategi Promosi
8981	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	TIDAK	TIDAK DITAKHLASI
8982	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8983	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Internet	YA	BAKOSWA
8984	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Internet	YA	BAKOSWA
8985	SPANA	18 SMA	Pegawai Reguler	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8986	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8987	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8988	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8989	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8990	SPANA	18 SMA	Tidak Berkarya	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8991	SPANA	18 SMA	Teman	2018	Teknik Sipil	2018	Internet	YA	BAKOSWA
8992	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	SMS Gateway	YA	BAKOSWA
8993	SPANA	18 SMA	Tukang Kayu/Laki	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8994	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8995	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	TIDAK	TIDAK DITAKHLASI
8996	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Internet	YA	BAKOSWA
8997	SPANA	18 SMA	Pegawai Swasta	2018	Teknik Sipil	2018	Teman	TIDAK	TIDAK DITAKHLASI
8998	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
8999	SPANA	18 SMA	Widawati	2018	Teknik Sipil	2018	Teman	YA	BAKOSWA
9000	SPANA	18 SMA	Teman	2018	Teknik Sipil	2018	Internet	YA	BAKOSWA

Gambar 4.22 Pembuatan Atribut Usia Saat Mendaftar

c. Pembuatan atau Penerapan Atribut Strategi Promosi

Sebagaimana aturan-aturan yang telah ditetapkan sebelumnya yaitu strategi promosi yang akan diterapkan maka aturan-aturan tersebut diterapkan ke dalam *dataset* untuk dijadikan atribut target atau label. Hasil dari penerapan aturan-aturan strategi promosi dapat dilihat pada gambar 4.23 berikut ini.

No	Nama	ASAL	JK	Usia	Pekerjaan	Tahun Pendaftaran	Referensi	Strategi Promosi				
1	WIDAWATI	SPANA	1801	SMA	19	Tukang Ka	400 14112 Akuntansi	2018	Teman	TIDAK	UNLUM	SPU
2	MARULI	SPANA	1801	SMA	21	Tukang Ka	400 14112 Akuntansi	2018	Teman	YA	UNLUM	SPU
3	SILA	SPANA	1801	SMA	18	Lan-lan	400 14112 Akuntansi	2018	Teman	YA	UNLUM	SPU
4	WIDAWATI	SPANA	1801	SMA	20	Sopir	400 14112 Akuntansi	2018	Teman	YA	UNLUM	SPU
5	WIDAWATI	SPANA	1801	SMA	20	Pegawai B	400 14112 Akuntansi	2018	Teman	TIDAK	UNLUM	SPU
6	WIDAWATI	SPANA	1801	SMA	20	Widawati	170 14112 Akuntansi	2018	Teman	YA	UNLUM	SPU
7	WIDAWATI	SPANA	1801	SMA	22	Anggota P	0 0 Akuntansi	2018	Teman	TIDAK	TIDAK DITAKHLASI	SPU
8	WIDAWATI	SPANA	1801	SMA	23	Pegawai P	400 14112 Akuntansi	2018	Radio	YA	UNLUM	SPU
9	WIDAWATI	SPANA	1801	SMA	20	Tukang Rot	400 14112 Akuntansi	2018	Parafit	YA	UNLUM	SPU
10	WIDAWATI	SPANA	1801	SMA	18	Widawati	220 14112 Akuntansi	2018	Teman	YA	UNLUM	SPU
11	WIDAWATI	SPANA	1801	SMA	19	Widawati	400 14112 Akuntansi	2018	Teman	YA	UNLUM	SPU
12	WIDAWATI	SPANA	1801	SMA	19	Widawati	400 14112 Akuntansi	2018	Teman	YA	UNLUM	SPU
13	WIDAWATI	SPANA	1801	SMA	18	Tukang Ka	400 14112 Akuntansi	2018	Teman	YA	UNLUM	SPU
14	WIDAWATI	SPANA	1801	SMA	18	Widawati	400 14112 Akuntansi	2018	Internet	YA	UNLUM	SPU
15	WIDAWATI	SPANA	1801	SMA	18	Widawati	400 14112 Akuntansi	2018	Sportif	YA	UNLUM	SPU
16	WIDAWATI	SPANA	1801	SMA	27	Widawati	0 14112 Akuntansi	2018	Parafit	YA	BAKOSWA	SPU
17	WIDAWATI	SPANA	1801	SMA	13	Tukang Bat	400 14112 Akuntansi	2018	Parafit	UNLUM	SPU	SPU
18	WIDAWATI	SPANA	1801	SMA	18	Widawati	400 14112 Akuntansi	2018	Teman	UNLUM	SPU	SPU
19	WIDAWATI	SPANA	1801	SMA	20	Widawati	0 0 Akuntansi	2018	Teman	TIDAK	TIDAK DITAKHLASI	SPU
20	WIDAWATI	SPANA	1801	SMA	19	Widawati	400 0 Akuntansi	2018	Internet	TIDAK	UNLUM	SPU
21	WIDAWATI	SPANA	1801	SMA	20	Parafit	400 0 Akuntansi	2018	Parafit/TIDAK	UNLUM	SPU	SPU
22	WIDAWATI	MALANG	1801	MA	29	Lan-lan	0 14112 Akuntansi	2018	Teman	YA	BAKOSWA	SPU
23	WIDAWATI	SPANA	1801	MA	27	Widawati	400 0 Akuntansi	2018	Radio	TIDAK	UNLUM	SPU
24	WIDAWATI	SPANA	1801	MA	19	Widawati	400 14112 Akuntansi	2018	Parafit	YA	UNLUM	SPU
25	WIDAWATI	SPANA	1801	MA	18	Pedagang	400 14112 Akuntansi	2018	Teman	YA	UNLUM	SPU

Gambar 4.23. Penerapan Label Strategi Promosi

### 3. Pendiskritan Atribut

Untuk bisa memproses data dengan menggunakan metode *deep learning*, data yang bertipe integer atau nominal harus diubah tipe datanya menjadi kategorikal atau polynominal. Untuk itulah dilakukan proses transformasi pendiskritan data pada atribut usia saat mendaftar dan atribut nilai ujian.

Proses pendiskritan data pada atribut usia saat mendaftar menjadi atribut Status Graduate terdiri dari beberapa langkah yaitu :

- Sebaran data usia  $\leq 19$  tahun maka dilakukan pendiskritan menjadi Fresh Graduate.
- Sebaran data usia  $\leq 30$  tahun maka didiskritkan menjadi Graduate.
- Sebaran data usia  $> 30$  tahun maka didiskritkan menjadi Long Graduate.

Sedangkan untuk proses pendiskritan pada atribut nilai ujian menjadi Hasil Tes langkah-langkahnya adalah sebagai berikut :

- Sebaran data nilai  $\geq 500$  didiskritkan menjadi Sangat Baik
- Sebaran data nilai  $< 500 - 450$  didiskritkan menjadi Baik
- Sebaran data nilai  $< 450 - 300$  didiskritkan menjadi Agak Baik
- Sebaran data nilai  $< 300 - 200$  didiskritkan menjadi Sedang
- Sebaran data nilai  $< 200 - 100$  didiskritkan menjadi Kurang
- Sebaran data nilai  $< 100 - 10$  didiskritkan menjadi Sangat Kurang
- Sebaran data nilai 0 didiskritkan menjadi Tidak Tes

Untuk lebih jelasnya proses pendiskritan data dapat dilihat pada gambar 4.24

	A	B	C	D	E	F	G	H
1	Status Graduate	usia saat mendaftar			hasil tes			
2	Fresh Graduate	15			300 Agak Tinggi			
3	Graduate	25			400 Tinggi			
4	Fresh Graduate	30			400 Tinggi			
5	Graduate	30			410 Tinggi			
6	Graduate	30			300 Agak Tinggi			
7	Graduate	30			200 Sedang			
8	Graduate	32			0 Tanpa Tes			
9	Fresh Graduate	19			400 Tinggi			
10	Graduate	30			400 Tinggi			
11	Fresh Graduate	19			320 Sedang			
12	Fresh Graduate	19			410 Tinggi			
13	Fresh Graduate	18			400 Tinggi			
14	Fresh Graduate	18			410 Tinggi			
15	Fresh Graduate	18			400 Tinggi			
16	Graduate	37			0 Tanpa Tes			
17	Fresh Graduate	19			400 Tinggi			
18	Fresh Graduate	18			410 Tinggi			
19	Graduate	30			0 Tanpa Tes			
20	Fresh Graduate	19			400 Tinggi			
21	Graduate	30			400 Tinggi			
22	Graduate	39			0 Tanpa Tes			
23	Fresh Graduate	17			400 Tinggi			
24	Fresh Graduate	19			400 Tinggi			
25	Fresh Graduate	19			400 Tinggi			

Gambar 4.24 Proses Pendiskritan

#### 4. Pengurangan Dimensi

Pada proses pengurangan dimensi, atribut yang melalui proses ini adalah Asal Kota. Hal ini dikarenakan terdapat perbedaan format dalam mengisi *record* asal kota, untuk itu dilakukan proses pengurangan dimensi. Terdapat 190 *record* yang mengalami perubahan setelah melalui proses ini sebagaimana yang terlihat pada gambar 4.25 berikut ini.

ID	Name	Address	City	Gender	Education	Age	Year
106	RAU, RHAER	SEPARA	1994	MA	33 Wiraswasta	0	2012 Ternan
108	VIVI FADH P	SEPARA	1990	SME	18 Tokang Kayu/Kayu	400	2014 Ternan
120	LAILA FAT P	SEPARA	1997	MA	18 Wiraswasta	340	2015 Ternan
171	RAU, RHAER	SEPARA	1996	MA	18 Wiraswasta	0	2014 Ternan
172	RAU, RHAER	SEPARA	1994	MA	18 Wiraswasta	0	2017 Ternan
173	RAU, RHAER	SEPARA	1997	MA	18 Wiraswasta	0	2015 Ternan
174	AHMAD, S L	SEPARA	1997	MA	18 Petani/Buruh Tani	300	2015 Ternan
175	ADI SAPU I	SEPARA	1999	SMA	18 Tukang Kayu/Kayu	0	2014 Pemfint/Laallai
176	SHASAH L L	SEPARA	1988	SMA	30 Wiraswasta	0	2018 Spanduk
177	NIA RANAF P	DIBOLOGA	1988	SMA	18 Wiraswasta	220	2016 Spanduk
178	TONGI HAJI	DILI BAI TIRI	1987	SME	18 Pegawai Negeri	300	2015 Ternan
179	YAMRONI L	KOTA BANGSRI	1999	MA	18 Petani/Buruh Tani	0	2017 Ternan
180	MUHAMMIL	BANGSRI	1996	SMA	13 Wiraswasta	300	2015 Ternan
181	DIDIK PRAL	KOTA BANGSRI	1997	SMA	20 Wiraswasta	240	2017 Ternan
182	MOCH KAL	DIBOLOGA	1998	PAKET C	18 Wiraswasta	300	2014 Ternan
183	RAU, RHAER P	TROSO	1998	MA	18 Petani/Buruh Tani	0	2018 Ternan
184	HIFOLITU P	TROSO	1994	MA	18 Pegawai Swasta	0	2018 Pemfint/Laallai
185	RAU, RHAER P	SEMURUNG	1991	SMA	25 Petani/Buruh Tani	0	2018 Ternan
186	RAU, RHAER P	BINANGSRI	1993	SME	18 Wiraswasta	0	2017 Ternan
187	MUHAMMIL	SEPARA	1998	SME	13 Wiraswasta	0	2017 Ternan
188	MUHAMMIL	SEPARA	1998	SME	13 Wiraswasta	0	2017 Ternan
189	MUHAMMIL	KEDIRI	1974	SME	30 Wiraswasta	0	2014 Ternan
190	MUHAMMIL P	JOGJAKARTA	1990	MA	18 Wiraswasta	410	2014 Spanduk

Gambar 4.25. Pengurangan Dimensi

#### 5. Pemilihan Atribut

Setelah melalui berbagai proses sebelumnya maka selanjutnya dilakukan proses pemilihan atribut. Proses ini dilakukan karena setelah melewati tahapan sebelumnya terdapat beberapa jenis atribut yang sudah tidak relevan atau informasinya tidak diperlukan lagi karena telah ada atribut lain yang lebih dapat mempresentasikan informasinya lebih lengkap dan jelas. Beberapa jenis atribut tersebut antara lain :

##### a. Nama

Atribut nama tidak dipilih karena atribut ini tidak memiliki informasi yang relevan yang dibutuhkan untuk proses *data mining* penentuan strategi promosi.

##### b. Jenis kelamin

Atribut jenis kelamin tidak dipilih karena atribut ini memiliki korelasi yang rendah terhadap atribut-atribut yang lain.

c. Re-Registrasi

Atribut ini tidak dipilih karena informasi yang terkandung didalam atribut ini telah tersampaikan pada atribut jalur masuk.

d. Angkatan

Atribut ini tidak dipilih karena nilai informasi yang terkandung didalamnya tidak cukup banyak dan memiliki korelasi yang banyak terhadap atribut-atribut lainnya.

e. Tahun Lahir

Atribut ini tidak dipilih karena informasi yang terkandung didalamnya telah tersampaikan pada atribut status graduate yang telah dibuat pada proses pembuatan atribut.

sehingga jumlah atribut yang terpilih dapat dilihat pada gambar 4.26 berikut

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ASAL ROKTA	jenisSekolah	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir	namaLahir
2	JEPARA	SMA					10 Tukang Kayu/Ular		100 Akuntansi	Teman	UNJUM					SPD
3	JEPARA	SMK					23 Tukang Kayu/Ular		400 Akuntansi	Teman	UNJUM					SPD
4	JEPARA	SMK					12 Lompah		400 Akuntansi	Teman	UNJUM					SPD
5	JEPARA	SMA					20 Sopir		410 Akuntansi	Teman	UNJUM					SPD
6	JEPARA	SMA					20 Pegawai Swasta		300 Akuntansi	Teman	UNJUM					SPD
7	JEPARA	SMK					20 Wirawasta		120 Akuntansi	Teman	UNJUM					SPD
8	JEPARA	SMA					22 Pegawai Negeri		0 Akuntansi	Teman	TIDAK DITETAPKAN					SPD
9	JEPARA	SMA					19 Pegawai Negeri		400 Akuntansi	Radio	UNJUM					SPD
10	JEPARA	SMA					20 Tukang Batu/Kurati Ba		400 Akuntansi	Pamflet/Laaff	UNJUM					SPD
11	JEPARA	SMK					18 Wirawasta		220 Akuntansi	Teman	UNJUM					SPD
12	JEPARA	SMA					19 Wirawasta		410 Akuntansi	Teman	UNJUM					SPD
13	JEPARA	SMA					18 Tukang Kayu/Ular		400 Akuntansi	Teman	UNJUM					SPD
14	JEPARA	SMK					18 Wirawasta		410 Akuntansi	internet	UNJUM					SPD
15	JEPARA	SMK					19 Wirawasta		400 Akuntansi	Sendiuk	UNJUM					SPD
16	MAN	SMK					27 Wirawasta		0 Akuntansi	Pamflet/Laaff	MANORAMA					SPD
17	JEPARA	MA					19 Tukang Batu/Kurati Ba		400 Akuntansi	Pamflet/Laaff	UNJUM					SPD
18	JEPARA	SMK					18 Wirawasta		410 Akuntansi	Teman	UNJUM					SPD
19	JEPARA	SMK					20 Wirawasta		0 Akuntansi	Teman	TIDAK DITETAPKAN					SPD
20	JEPARA	SMK					22 Wirawasta		400 Akuntansi	Sendiuk	UNJUM					SPD
21	JEPARA	SMK					20 Pelani/Buruh Tani		400 Akuntansi	Pamflet/Laaff	UNJUM					SPD
22	MALANG	MA					29 Lompah		0 Akuntansi	Teman	MANORAMA					SPD
23	JEPARA	MA					17 Wirawasta		400 Akuntansi	Radio	UNJUM					SPD
24	JEPARA	SMK					19 Wirawasta		430 Akuntansi	Pamflet/Laaff	UNJUM					SPD
25	JEPARA	SMK					19 Pedagang/PJL		400 Akuntansi	Teman	UNJUM					SPD

Gambar 4.26. Pemilihan Atribut

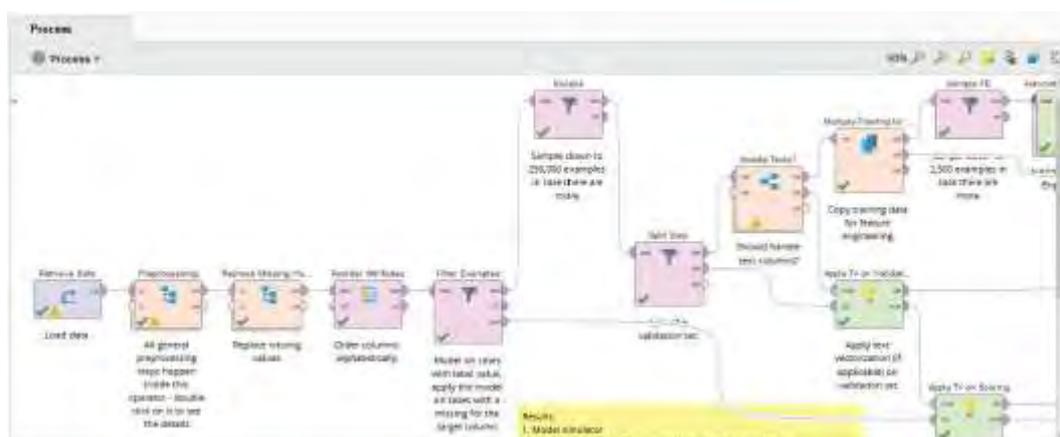
#### 4.4 Proses Pemodelan Data

Setelah semua tahapan pra proses data selesai, kini dimulailah proses pemodelan data. Proses ini dilakukan dengan menggunakan *tools* Rapidminer 9.3 sehingga proses pemodelan data dapat dilakukan secara otomatis. Proses ini terdiri dari beberapa langkah yaitu:

*Retrieve Data* atau pengambilan data yaitu data diambil dari *local repository* Rapidminer dalam bentuk file Excel kemudian dilanjutkan *Praproses* pada proses ini

dilakukan pengecekan ulang terhadap data diantaranya yaitu pengecekan target atribut, kolom yang tidak dipakai, pendiskritan dan penyatuan semua nilai atribut. Kemudian proses berikutnya adalah *Replace Missing Value* pada proses ini dilakukan pengecekan jika ada atribut yang hilang yang kemudian akan diganti dengan nilai rata-rata dari atribut yang hilang tersebut atau dengan nilai modus.

Proses selanjutnya adalah *Reorder Attribute* pada proses ini pengaturan atribut ditata ulang dengan aturan alfabet secara *ascending*, kemudian melalui *Filter Example* untukantisipasi bila model pada kasus tidak memiliki nilai label, maka selanjutnya akan diterapkan model pada kasus yang tidak memiliki nilai pada kolom target tersebut. Selanjutnya proses *Sample* pada tahap ini di buat pembatasan sample data pada angka 250.000 untukantisipasi jika data latih melebihi angka tersebut sebagaimana yang terlihat pada gambar 4.27.

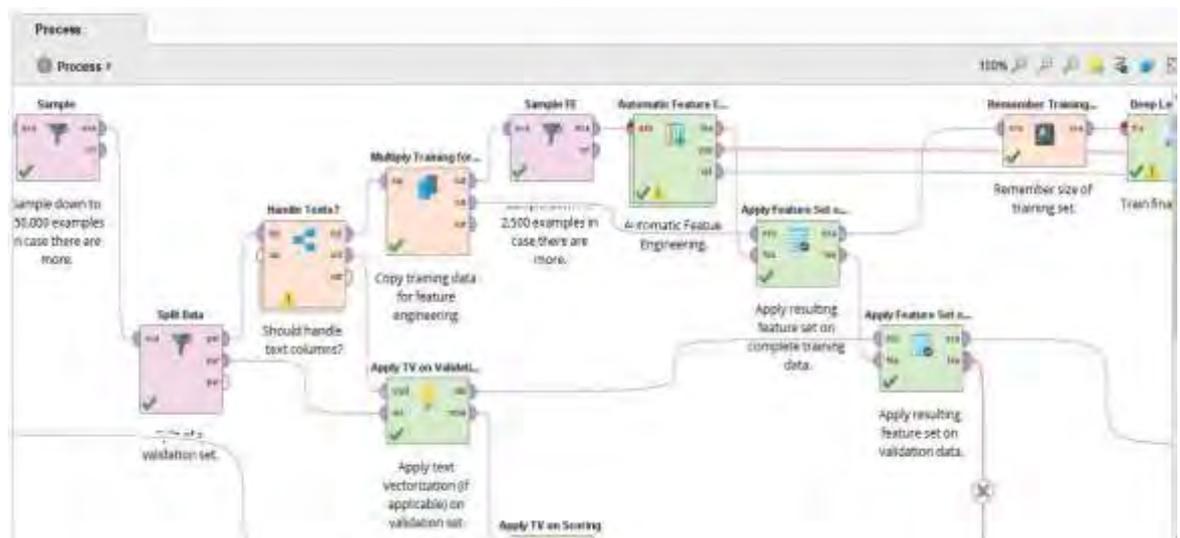


Gambar 4.27 Automodel PraProses Data (Sumber :Rapidminer 9.3)

Kemudian dilakukan *Split data* yaitu proses pemisahan data untuk dibagi menjadi data latih dan data uji dengan persentase yang dapat diatur sedemikian rupa misalnya sebesar 80% untuk data latih dan 20% untuk data uji. Kemudian proses *Handle Text* dilakukan untuk dasar pembuatan *feature extraction* yang berasal dari kolom text. Selanjutnya *Apply TV on Validation* proses ini digunakan pada data yang akan menjadi data latih yang akan dibuat menjadi sebuah model. Yang pada proses *Apply TV Scoring* model tersebut diterapkan terhadap data uji.

Langkah berikutnya adalah Multiple Training For Feature Engineering. Pada proses ini data latih digandakan agar bisa dilakukan proses *feature engineering*. Lalu *Sample FE* untuk membuat data sample bertingkat yang berasal dari data latih. Proses ini tidak bisa diterapkan bila label bernilai numerical. Untuk ukuran sample dibatasi sebanyak 2500 data. Kemudian pada proses *Automatic FE* dilakukan alidasi model terhadap kumpulan atribut data untuk memastikan bahwa proses *split data* dilakukan sama. proses ini meliputi seleksi fitur dan pembuatan fitur.

Selanjutnya *Apply Feature Set on Training data*. Proses ini bekerja menerapkan hasil fitur set yang telah dihasilkan pada proses sebelumnya terhadap keseluruhan data latih. Lalu *Apply Feature Set on Validation data*. Pada proses ini diterapkan hasil fitur set terhadap semua data uji. Untuk lebih jelasnya proses no 7 sampai no 15 dapat dilihat pada gambar 4.28.



Gambar 4.28 Automodel *Splitting Data* (Sumber :Rapidminer 9.3)

Proses Selanjutnya adalah *Remember Training* proses ini dilakukan untuk mengingat jumlah dan ukuran dari hasil data latih. Lalu proses utama yaitu *Deep Learning Final Model*. Pada proses ini dijalankan algoritma *deep learning multy-layer perceptron* dengan menggunakan H2O 3.8.2.6. Adapun jenis layer yang digunakan pada proses ini adalah :

1. *Input Layer* yaitu data latih berlabel
2. *Recifier Layer* yaitu *hidden layer* atau layer yang berisi sejumlah fungsi aktivasi *Recifier* atau *ReLU (Rectified Linear Unit)* untuk membentuk model. Pada pemodelan ini *hidden layer* yang digunakan berjumlah 2 layer.
3. *Softmax layer* juga merupakan salah satu fungsi aktivasi yang berkerja dengan menyamaratakan batasan output sehingga mudah diinterpretasikan sebagai suatu kemungkinan. Fungsi ini membatasi output dalam rentang 0-1. Untuk itu jumlah node pada lapisan ini harus sama dengan dengan output klasifikasinya.

Parameter yang terdapat dalam proses ini yaitu :

1. Epoch adalah perulangan atau iterasi yang dilakukan terhadap kumpulan data yang terjadi apabila seluruh kumpulan data telah melalui proses latih terhadap neural network yang kemudian dikembalikan lagi ke awal untuk melalui proses itu lagi.
2. *Dropout* yaitu metode yang berfungsi mencegah terjadinya *overfitting* dengan cara mengacuhkan sebagian dari unit objek.
3. L1 yaitu metode regulasi yang membatasi nilai absolut dan menurunkan beberapa nilai sampai ke angka 0 untuk mengurangi kerumitan dan mencegah terjadinya *overfitting*
4. L2 yaitu metode regulasi yang membatasi bobot dari jumlah yang dikuadratkan.
5. *Mean Rate* yaitu perhitungan yang berguna dalam perhitungan *mean square error (MSE)*.
6. *Rate RMS* yaitu yaitu nilai perhitungan akar dari *means square*.
7. *Momentum* yaitu parameter control saat permulaan proses dilakukan.
8. *Mean weight* yaitu nilai bobot rata-rata
9. *RMS weight* yaitu bobot nilai dari akar kuadrat rata-rata.
10. *Mean Bias* yaitu nilai bias dari rata-rata
11. *RMS Bias* yaitu nilai bias dari akar kuadrat rata-rata.
12. *Duration Training* yaitu lama waktu yang dibutuhkan untuk proses melatih data.

Selain itu pada metode ini diterapkan juga beberapa jenis validasi model untuk menilai keakuratan dari model yang dihasilkan. Beberapa metode validasi yang digunakan yaitu :

1. *MSE (Means Square Error)*

Adalah salah satu metode untuk mengukur tingkat keakuratan suatu model peramalan. MSE dapat dianalogikan sebagai varian ditambah kuadrat bias model. Semakin rendah nilai MSE maka semakin akurat model peramalan tersebut.

2. R kuadrat

Uji R kuadrat atau uji determinasi ini merupakan salah satu metode uji regresi untuk menginformasikan baik atau tidaknya suatu model peramalan. Nilai R yang kecil berarti variable dependen sangat terbatas sementara nilai R yang mendekati 1 berarti bahwa variable dependen sudah dapat memberikan informasi yang dibutuhkan untuk proses prediksi.

3. *Logloss*

Dalam melakukan proses klasifikasi terdapat *matriks performance*. *Logloss* ini adalah fungsi untuk menghitung *performance* data, dan kualitas data. Semakin mendekati angka 0 maka semakin baik.

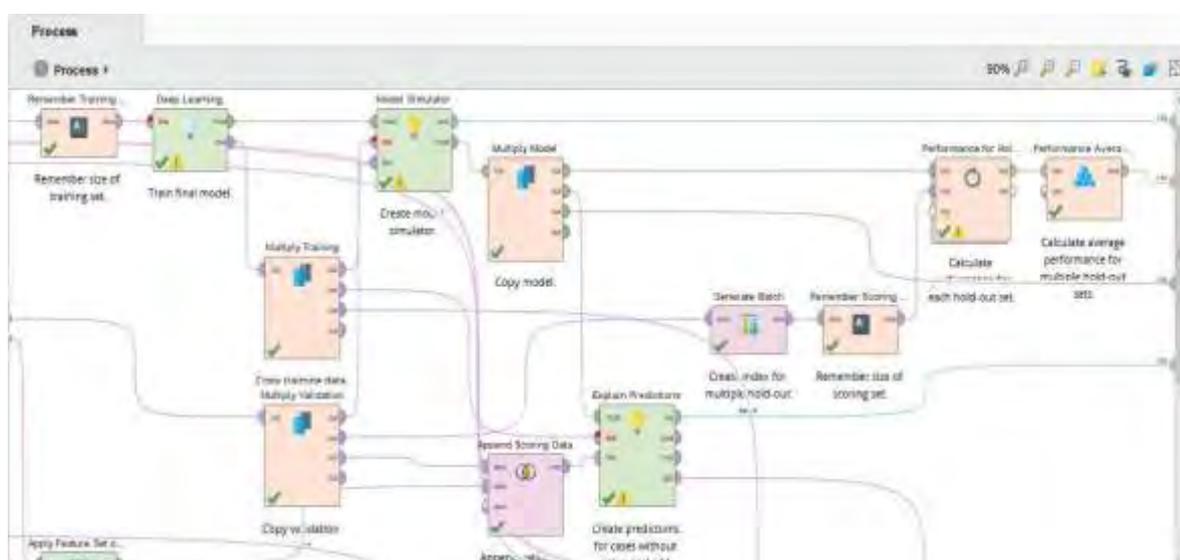
4. *Clasification Error*

Error klasifikasi adalah proporsi jumlah pola yang diklasifikasikan secara salah oleh model yang diterapkan. Sehingga semakin kecil nilai error klasifikasi maka semakin baik hasil dari pemodelan data tersebut.

Selanjutnya dilanjutkan dengan proses *Multiply Training* untuk membuat *copy* dari data latih dan *Multiply Validation* untuk membuat *copy* dari data uji. Kemudian dibuat *Model Simulato* untuk membuat simulasi model menjadi lebih mudah bila akan dilakukan perubahan terhadap *input* atau *output* data. Selain itu juga dapat menampilkan proses prediksi yang dilakukan. Operator *Multiply Model* lalu digunakan untuk mengcopy model. *Append Scoring data* dilakukan untuk menggabungkan hasil data uji yang berasal dari dua atau lebih data uji dan kemudian menggabungkannya.

Proses Selanjutnya adalah *Explain Prediction* yaitu proses untuk mengidentifikasi atribut mana yang memiliki peran paling besar dalam membentuk model. *Generating Batch* untuk menambahkan atribut baru ke dalam model untuk uji validasi model data. Metode yang digunakan adalah Hold-Out. Lalu proses *Remember Scoring* digunakan untuk menyimpan hasil validasi data.

Proses *Performance to Hold Out* dijalankan proses validasi terhadap model dengan metode Hold-Out sebanyak 7 *iterations*. Kemudian proses *Performance Average* proses ini menghitung jumlah rata-rata dari vector kinerja yang dikirimkan sebagai *output*. Untuk lebih jelasnya proses no 16 sampai dengan 27 dapat dilihat pada gambar 4.29.



Gambar 4.29 Automodel *Deep Learning* (Sumber :Rapidminer 9.3)

#### 4.5 Proses Pengujian Model

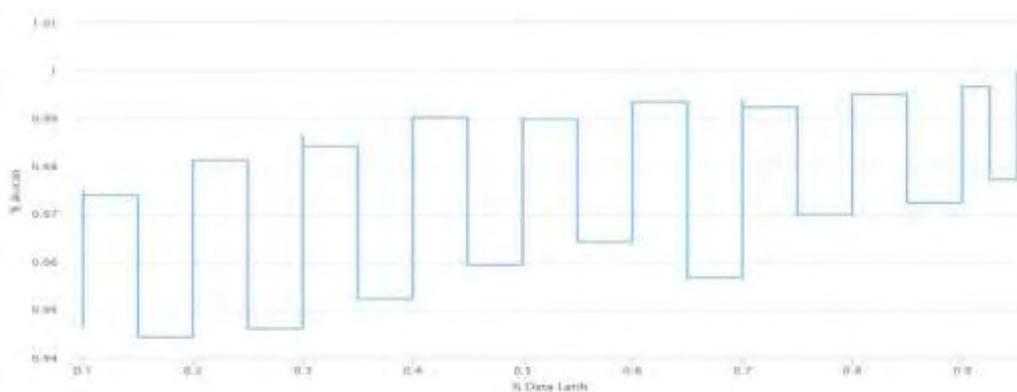
Setelah pada proses sebelumnya dilakukan pemodelan data maka pada proses ini dilakukan evaluasi hasil pemodelan tersebut dalam proses pengujian model. Pengujian model dilakukan dengan melakukan serangkaian proses untuk pembuatan model dengan mengubah beberapa variable yang mungkin dapat mempengaruhi hasil akhir. Adalah variable tersebut yaitu persentase data latih dan jumlah *epoch*. Sebagaimana yang terlihat

pada tabel.4.1 yang menunjukkan matrix hasil pengujian dengan metode *deep learning* pada Rapidminer 9.3.

Tabel 4.1. Matrixs Pengujian

NO	% Data Latih	Jumlah Data Latih (Record)	% Akurasi per Epoch														
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	10%	863	94,6 3%	94,7 7%	94,8 1%	95,3 9%	95,26 %	94,7 4%	96,1 6%	96,0 5%	96,0 5%	96,8 6%	96,9 4%	97,42 %	97,51 %	97,04 %	97,40 %
2	20%	1724	94,4 4%	94,6 3%	95,8 8%	96,2 1%	97,24 %	97,3 6%	97,2 6%	97,5 7%	97,8 5%	97,8 5%	97,5 5%	97,59 %	97,79 %	98,05 %	98,13 %
3	30%	2590	94,6 2%	96,0 6%	97,5 0%	97,1 3%	97,54 %	97,5 2%	98,0 5%	98,1 0%	98,2 4%	98,3 5%	98,3 3%	98,68 %	98,68 %	98,42 %	98,42 %
4	40%	3452	95,2 4%	96,9 4%	97,1 6%	97,3 2%	98,32 %	98,3 8%	98,4 9%	98,4 6%	98,6 5%	98,8 1%	98,5 4%	98,84 %	98,89 %	99,03 %	99,03 %
5	50%	4315	95,9 4%	97,3 7%	97,8 6%	98,0 2%	98,09 %	98,4 7%	98,5 7%	98,7 3%	98,5 7%	98,8 6%	98,8 6%	98,90 %	98,96 %	99,03 %	98,99 %
6	60%	5177	96,4 3%	97,2 4%	98,0 5%	97,8 9%	98,54 %	98,6 2%	98,5 4%	98,9 0%	98,8 6%	98,9 0%	99,1 1%	99,11 %	99,15 %	99,15 %	99,35 %
7	70%	6039	95,6 8%	97,0 8%	98,2 7%	98,1 1%	98,11 %	98,7 0%	98,5 9%	99,1 4%	99,1 4%	99,3 0%	99,4 1%	99,41 %	99,41 %	99,41 %	99,24 %
8	80%	6905	97,0 0%	97,9 7%	98,6 2%	98,4 6%	98,86 %	98,8 6%	99,1 1%	99,1 9%	99,3 5%	99,3 5%	99,5 1%	99,51 %	99,43 %	99,43 %	99,51 %
9	90%	7766	97,2 5%	98,5 4%	98,3 8%	99,3 5%	99,19 %	99,3 5%	99,5 1%	99,5 1%	99,3 5%	99,5 1%	99,5 1%	99,51 %	99,51 %	99,51 %	99,67 %
10	95%	8198	97,7 3%	98,0 5%	98,7 0%	99,3 5%	99,68 %	99,3 5%	99,3 5%	99,6 8%	99,6 8%	99,6 8%	99,6 8%	100%	100%	100%	100%

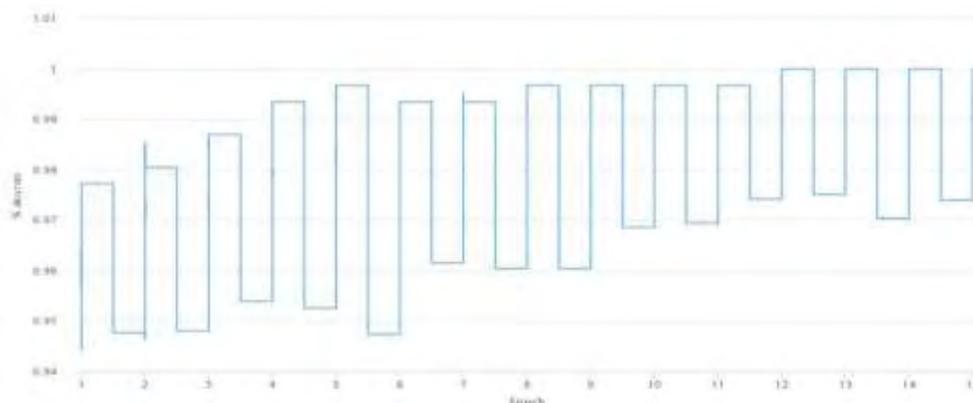
Berdasarkan tabel 4.1 dapat dilihat bahwa pada tiap pembagian data latih dengan persentase yang berbeda didapatkan persentase akurasi yang berbeda. Semakin banyak jumlah data latih maka semakin besar persen akurasinya hal ini menunjukkan bahwa persentase data latih dan persentase akurasi berbanding lurus. Untuk lebih jelasnya dapat dilihat pada gambar 4.30.



Gambar 4.30 Grafik Perbandingan %Akurasi dan %Data Latih.

(Sumber :Rapidminer 9.3)

Akan tetapi besar akurasi tidak selalu berbanding lurus dengan jumlah *epoch*. Seperti yang terlihat pada gambar 4.31.



Gambar 4.31 Grafik Perbandingan %Akurasi dan *Epoch* (Sumber :Rapidminer 9.3)

Maka dari hasil analisis tabel 4.1 serta gambar 4.30 dan gambar 4.31 dapat disimpulkan bahwa hasil pengujian yang optimal adalah pengujian dengan persentase akurasi 100% terdapat pada persentase data latih 95% dan jumlah *Epoch* 12,13,14 dan 15.

Untuk mengetahui hasil yang paling optimal dari keempat kandidat yang telah ditetapkan maka diperlukan parameter pembanding lain. Parameter yang dipilih untuk pembanding adalah parameter waktu (*Duration*). Hasil perbandingan tersebut dapat dilihat pada tabel 4.2.

Tabel 4.2 Matrix Perbandingan Persentase Data Latih dan Durasi

% Data Latih	Durasi Training Per Epoch (sec)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
10%	0.247	0.514	0.705	1.311	1.166	1.455	1.562	1.737	1.911	2.117	2.268	2.688	2.575	2.798	2.957
20%	0.379	0.790	2.357	1.668	2.283	2.766	2.866	3.163	3.381	4.380	4.452	4.489	5.015	5.277	6.088
30%	0.710	1.122	1.936	2.344	2.988	3.442	4.151	4.820	5.161	6.684	12.498	9.577	9.034	8.620	10.431
40%	0.861	1.641	2.280	3.220	4.255	5.169	5.864	7.078	7.404	8.203	9.782	10.576	9.624	10.422	11.220
50%	1.120	2.028	2.992	3.767	4.736	5.770	6.671	7.619	8.841	10.069	11.785	11.122	11.990	13.324	13.918
60%	1.251	2.500	3.438	4.435	5.620	7.087	8.521	14.151	11.023	11.251	12.141	13.693	14.233	15.480	16.494
70%	1.982	2.651	3.979	5.614	6.472	7.869	9.435	16.839	12.578	12.722	14.226	16.315	16.658	19.755	19.419

% Data Latih	Durasi Training Per Epoch (sec)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
80%	2.3 62	3.2 42	4.4 91	5.9 87	7.8 98	12.9 50	10.8 14	11.0 84	12.2 93	14.0 49	16.7 23	16.3 35	17.3 98	18.8 05	20.1 06
90%	1.6 13	3.1 80	4.7 99	6.2 46	7.7 98	9.37 7	10.7 00	12.1 77	13.7 14	16.8 80	16.5 98	17.8 91	19.5 21	20.9 74	22.7 16
95%	1.6 44	3.2 63	4.9 07	6.8 33	7.9 90	9.77 1	13.6 20	12.7 04	14.5 14	18.0 01	17.8 53	19.1 90	22.3 25	21.8 44	23.7 66

Berdasarkan Tabel 4.2 yaitu matriks perbandingan persentase data latih terhadap durasi per *epoch* maka dapat dilihat bahwa pada 4 matriks kandidat sebelumnya apabila dilihat dari parameter lain yaitu durasi maka durasi yang tercepat terdapat pada matriks 10.12 atau pada persentase data latih 95% dan *Epoch* yang berjumlah 12 dengan waktu 19.190. Matriks 10.13 yaitu pada persentase data latih 95% dengan 13 *epoch* durasi yang diperlukan adalah 22.325, pada matriks 10.14 yaitu pada persentase data latih 95% dengan 14 *epoch* dan durasi 199.190 dan pada matriks 10.15 yaitu pada persentase data latih 95% dengan 15 *epoch* yang berdurasi 23.766.

Maka kemudian dapat dilihat bahwa pada matriks 10.12 adalah yang memiliki durasi paling cepat diantara yang lainnya sehingga dapat disimpulkan bahwa hasil pengujian model yang paling optimal adalah dengan menggunakan data latih 95% dan 12 *epoch*.

#### 4.6 Evaluasi Hasil Pemodelan

Berdasarkan hasil pemodelan yang telah dilakukan sebagaimana yang terlihat pada gambar 4.32 hasil tersebut menunjukkan akurasi sebesar 100% dengan rata-rata error sebesar 0.00%. Hasil pemodelan yang didapat dengan menggunakan deep learning dapat dikatakan sangat baik dengan persentase keakuratan yang optimal.

Pada gambar 4.32 yaitu hasil pemodelan *deep learning* yang digunakan didapat nilai MSE (*Means Square Error*) sebesar 0.00199, semakin kecil nilai MSE (mendekati angka 0) maka semakin baik keakuratan model tersebut. Sementara itu nilai uji R kuadrat yang didapat sebesar 0.99720, semakin besar nilai uji R kuadrat (mendekati angka 1) maka semakin baik keakuratan sebuah model dan nilai *logloss* sebesar 0.00863 yang bila semakin kecil nilainya mendekati angka 0 semakin baik. Serta nilai *classification error* yang didapat sebesar

0.00171, yang mana semakin kecil nilai ini maka semakin baik tingkat keakuratan suatu model..

## DeepLearning

```

Model Metrics Type: Multinomial
Description: Metrics reported on full training frame
model id: rm-h2o-model-deep_learning-513233
frame id: rm-h2o-frame-deep_learning-546761
MSE: 0.001952269
R^2: 0.9972008
logloss: 0.00863423
hit ratios: [0.99829227, 0.99987805, 1.0, 1.0, 1.0, 1.0, 1.0]
CM: Confusion Matrix (vertical: actual; across: predicted):
      SP0  SP2  SP6  SP3  SP1  SP5  SP4  Error  Rate
SP0  7751  0    0    2    0    0    0  0.0003  = 2 / 7,753
SP2  0    40  0    0    0    0    0  0.0000  = 0 / 40
SP6  0    0  35  0    0    0    0  0.0000  = 0 / 35
SP3  0    0  0    91  0    0    0  0.0000  = 0 / 91
SP1  2    0  0    0  226  0    2  0.0174  = 4 / 230
SP5  4    0  0    1    1  22  0  0.2143  = 6 / 28
SP4  2    0  0    0    0    0  19  0.0952  = 2 / 21
Totals 7759  40  35  94  227  22  21  0.0017  = 14 / 8,198
    
```

Status of Neuron Layers (predicting Strategy: Promot, 7-class Classification, multinomial distribution, CrossEntropy loss, 14,407 weights/biases, 174.1 KB, 66,176 training samples, mini-batch size 1):

Layer	Units	Type	Scopout	L1	L1 Mean	Rate	Rate	RMS	Momentum	Mean Weight	Weight	RMS	Mean Bias	Bias	RMS
1	219	Input	(1.0)												
2	50	Rectifier	0.00	0.000010	0.000000	0.212950	0.338102	0.000000		1.500343	0.094200	0.502022	0.243721		
3	50	Rectifier	0.00	0.000010	0.000000	0.036423	0.067426	0.000000		0.019723	0.139025	1.023683	0.104558		
4	7	Softmax		0.000010	0.000000	0.164697	0.156633	0.000000		-1.207331	0.397732	-1.100965	0.067058		

Scoring History:

Timestamp	Duration	Training Speed	Epochs	Iterations	Samples Training	MSE Training	R^2 Training	LogLoss Training	Classification Error
2019-11-25 02:44:23	0.000 sec		0.00000	0	0.00000	NaN	NaN	NaN	NaN
2019-11-25 02:44:25	2.195 sec	5004 rows/sec	1.00000	1	8139.00000	0.02400	0.36619	0.29265	0.02231
2019-11-25 02:44:27	4.155 sec	5014 rows/sec	2.00000	2	16296.00000	0.01668	0.37657	0.25678	0.01830
2019-11-25 02:44:29	6.170 sec	5039 rows/sec	3.00000	3	24594.00000	0.02211	0.38259	0.24509	0.01269
2019-11-25 02:44:31	7.977 sec	5145 rows/sec	4.00000	4	32792.00000	0.01605	0.38169	0.22943	0.00974
2019-11-25 02:44:33	9.717 sec	5237 rows/sec	5.00000	5	40990.00000	0.01612	0.39140	0.21379	0.00659
2019-11-25 02:44:34	11.512 sec	5312 rows/sec	6.00000	6	49188.00000	0.01611	0.39142	0.20212	0.00756
2019-11-25 02:44:36	13.711 sec	5257 rows/sec	7.00000	7	57386.00000	0.01680	0.39184	0.21994	0.00671
2019-11-25 02:44:38	15.440 sec	5304 rows/sec	8.00000	8	65584.00000	0.01830	0.39536	0.21399	0.00293
2019-11-25 02:44:40	18.201 sec	5347 rows/sec	9.00000	9	73782.00000	0.01817	0.39554	0.21279	0.00384
2019-11-25 02:44:43	19.568 sec	5385 rows/sec	10.00000	10	81980.00000	0.01528	0.39256	0.21049	0.00720
2019-11-25 02:44:45	21.762 sec	5389 rows/sec	11.00000	11	90178.00000	0.00759	0.39636	0.21010	0.00293
2019-11-25 02:44:47	24.149 sec	5219 rows/sec	12.00000	12	98376.00000	0.01199	0.39720	0.20683	0.00171

H2O version: 3.8.2.6-m3.0.0

Gambar 4.32 Hasil Pemodelan Deep Learning

(Sumber :Rapidminer 9.3)

Dari hasil yang dapat terlihat pada gambar 4.33. yaitu *Performance Vector* dapat dilihat tingkat akurasi yang dihasilkan dari hasil pemodelan menggunakan data latih sebanyak 95% dari total data keseluruhan atau 8.198 *record* dan data uji sebanyak 5% atau 431 *record* dengan menggunakan metode *deep Learning* 4 layer dan 12 *epochs* didapat hasil prediksi sebesar 100% akurat dengan rincian sebagai berikut :

1. Strategi Promosi 0 (SP0) memiliki keakuratan sebesar 100%
2. Strategi Promosi 1 (SP1) memiliki keakuratan sebesar 100%
3. Strategi Promosi 2 (SP2) memiliki keakuratan sebesar 100%
4. Strategi Promosi 3 (SP3) memiliki keakuratan sebesar 100%
5. Strategi Promosi 4 (SP4) memiliki keakuratan sebesar 100%
6. Strategi Promosi 5 (SP5) memiliki keakuratan sebesar 100%
7. Strategi Promosi 6 (SP6) memiliki keakuratan sebesar 100%.

### PerformanceVector

```

PerformanceVector:
accuracy: 100.00% +/- 0.00% (micro average: 100.00%)
ConfusionMatrix:
True:  SP0  SP2  SP6  SP3  SP1  SP5  SP4
SP0:  292  0   0   0   0   0   0
SP2:  0   1   0   0   0   0   0
SP6:  0   0   2   0   0   0   0
SP3:  0   0   0   3   0   0   0
SP1:  0   0   0   0   7   0   0
SP5:  0   0   0   0   0   1   0
SP4:  0   0   0   0   0   0   1
Classification_error: 0.00% +/- 0.00% (micro average: 0.00%)
ConfusionMatrix:
True:  SP0  SP2  SP6  SP3  SP1  SP5  SP4
SP0:  292  0   0   0   0   0   0
SP2:  0   1   0   0   0   0   0
SP6:  0   0   2   0   0   0   0
SP3:  0   0   0   3   0   0   0
SP1:  0   0   0   0   7   0   0
SP5:  0   0   0   0   0   1   0
SP4:  0   0   0   0   0   0   1

```

Gambar 4.33 *Performance Vector*

Maka dari hasil tersebut dapat disimpulkan bahwa model yang dibangun dengan menggunakan metode *deep learning* ini dapat mengklasifikasikan semua jenis strategi promosi yang diterapkan dengan akurasi 100% akan tetapi persentase yang akurat ini baru bisa didapatkan dengan menggunakan data latih yang cukup banyak yaitu 95% dari total data atau sebanyak 8.198 *record*. Hal ini dikarenakan terdapat beberapa strategi promosi yang sulit untuk diklasifikasikan oleh model. Untuk lebih jelasnya dapat dilihat pada tabel 4.3 berikut ini :

Tabel 4.3 Perbandingan Hasil Akurasi Strategi Promosi

No	Strategi Promosi	Matrix Pengujian				
		6.11	7.11	8.11	9.15	10.7
1	SP0	99,45%	94,66%	99,74%	99,83%	100%
2	SP1	95,59%	93,88%	100%	100%	100%
3	SP2	100%	100%	87,50%	100%	100%
4	SP3	91,30%	100%	100%	100%	100%
5	SP4	80,00%	75%	75%	100%	0%
6	SP5	50,00%	100%	75%	50,00%	33,33%
7	SP6	100%	100%	100%	100%	100%

Dari tabel 4.3 dapat dilihat bahwa dari 5 contoh hasil pengujian yang telah dilakukan, strategi promosi yang selalu memiliki nilai akurasi tinggi yaitu strategi promosi SP2, SP3, dan SP6. Sedangkan strategi promosi yang sering memiliki nilai akurasi lebih rendah yaitu strategi promosi SP4 dan SP5. Untuk mengetahui alasannya maka dilakukan proses ekstraksi pengetahuan yang berasal dari *database* PMB Universitas Islam Nahdlatul Ulama Jepara.

#### 4.7 Ekstraksi Pengetahuan

Untuk memperoleh hasil pengetahuan yang bermanfaat dari kumpulan data yang telah digunakan dalam proses *data mining* maka dilakukan pengujian kembali terhadap hasil evaluasi pemodelan yang optimal untuk mengetahui seberapa besar suatu atribut berkorelasi terhadap hasil pemodelan dengan tidak mengikutsertakan atribut tertentu. Hasil dari pengujian tersebut dapat dilihat pada table 4.4.

Tabel.4.4 Pengujian Korelasi Atribut Terhadap Akurasi

No	Nama Atribut								% Akurasi
	Asal Kota	Jenis Sekolah	Jalur Masuk	Status Graduate	Pekerjaan Ortu	Prodi	Referensi	Hasil Tes	
1	-	√	√	√	√	√	√	√	99,03%
2	√	-	√	√	√	√	√	√	100%
3	√	√	-	√	√	√	√	√	100%
4	√	√	√	-	√	√	√	√	99,68%
5	√	√	√	√	-	√	√	√	99,68%
6	√	√	√	√	√	-	√	√	95,14%
7	√	√	√	√	√	√	-	√	96,75%
8	√	√	√	√	√	√	√	-	97,73%

Berdasarkan data pada table 4.4 maka urutan atribut yang paling berpengaruh terhadap akurasi sampai yang tidak berpengaruh adalah :

1. Program Studi (Prodi)
2. Referensi
3. Hasil Tes
4. Asal Kota
5. Status Graduate
6. Pekerjaan Orang tua
7. Jenis Sekolah
8. Jalur Masuk

Berdasarkan urutan dari atribut yang paling berpengaruh diatas, dapat dilihat bahwa pada strategi promosi SP4 dan SP5 memiliki persyaratan yang didasarkan pada atribut jalur masuk, pekerjaan orang tua dan status *graduate* yang memiliki korelasi rendah sehingga hal inilah yang menjadi penyebab strategi promosi SP4 dan SP5 menjadi sulit untuk di klasifikasikan.

Agar didapat data hasil yang lebih akurat dari strategi promosi SP4 dan SP5 maka dilakukan analisis terhadap data *Confusion Matrix* dari tiap jenis strategi promosi agar selanjutnya dapat dilakukan perbaikan terhadap strategi promosi SP4 dan SP5 agar menjadi lebih mudah diklasifikasikan. Lebih jelasnya dapat dilihat pada table 4.5 berikut.

Tabel 4.5 Tabel *Confusion Matrix* Strategi Promosi dengan Atribut yang dihilangkan.

No	Atribut yg dihilangkan	% Akurasi Strategi Promosi					
		SP1	SP2	SP3	SP4	SP5	SP6
1	Asal Kota	87,50%	0%	75%	50%	100%	100%
2	Jenis Sekolah	100%	100%	100%	100%	100%	100%
3	Jalur Masuk	100%	100%	100%	0%	100%	100%
4	Status Graduate	100%	100%	100%	50%	100%	100%
5	Pekerjaan Ortu	100%	0%	100%	50%	0%	0%
6	Prodi	0%	100%	0%	0%	0%	100%
7	Referensi	66,67%	0%	0%	100%	100%	100%
8	Hasil Tes	100%	50%	0%	50%	0%	0%

Pada percobaan pertama dilakukan pemodelan dengan menghilangkan atribut asal kota, pada percobaan ini yang nilai akurasinya menurun adalah strategi promosi SP2 dan SP4. Pada percobaan kedua atribut yang dihilangkan adalah jenis sekolah yang ternyata tidak berpengaruh apapun terhadap nilai akurasi.

Pada percobaan ketiga jalur masuk dihilangkan yang berakibat menurunnya tingkat akurasi pada strategi promosi SP4. Pada percobaan keempat status graduate dihilangkan hasilnya berkurangnya tingkat akurasi pada SP4. Pada percobaan kelima atribut pekerjaan orang tua dihilangkan yang menyebabkan berkurangnya tingkat akurasi pada strategi promosi SP2,SP4,SP5 dan SP6.

Pada percobaan keenam dihilangkan atribut prodi yang menyebabkan berkurangnya akurasi pada strategi promosi SP1,SP3,SP4, dan SP5. Pada percobaan ketujuh atribut referensi dihilangkan sehingga nilai akurasi pada strategi promosi SP1, SP2 dan SP3 berkurang. Pada percobaan terakhir yang dihilangkan adalah atribut hasil tes sehingga nilai akurasi pada SP2,SP3,SP4,SP5, dan SP6 berkurang.

Dari data pada table 4.5 bisa dilihat bahwa atribut yang mempengaruhi rendahnya nilai akurasi pada strategi promosi SP4 dan SP5 adalah atribut program studi (Prodi), Pekerjaan Orang Tua, dan Hasil Tes. Untuk itu dilakukan perbaikan terhadap persyaratan pada stretegi promosi SP4 dan SP5 dengan menghilangkan dan atau menambah *value* dari atribut yang mempengaruhi rendahnya tingkat akurasi. Untuk lebih jelasnya dapat dilihat pada tabel 4.6.

Tabel 4.6 Rekomendasi 1 Perbaikan Strategi Promosi

No	Strategi Promosi	Potongan SPI	PERSYARATAN						
			Kota Asal	Program Studi	Status Graduate	Nilai Tes	Pekerjaan Orang tua	Referensi	Jalur masuk
1	SP1	20%	Jepara	Komunikasi dan Penyiaran Islam, Pendidikan Guru PAUD, Teknik Elektro	-	-	-	Teman	Umum
2	SP2	30%	luar Jepara	-	-	-	-	Internet	Umum
3	SP3	10%	Jepara	Selain Manajemen, Akuntansi, Pendidikan Agama Islam	-	-	-	Internet	Umum
4	SP4	50%	-	Komunikasi dan Penyiaran Islam, Pendidikan Guru PAUD, Teknik Elektro	Fresh Graduate	-	-	-	Tidak Diketahui
5	SP5	80%	-	-	Fresh Graduate	Baik Sangat Baik	Tidak Bekerja Buruh Bangunan Nelayan	-	Umum
6	SP6	40%	-	-	-	Sangat Baik	-	-	Umum

Berdasarkan tabel 4.6 perubahan yang dilakukan terhadap strategi promosi SP4 adalah dengan menghilangkan persyaratan dari pekerjaan orang tua. Sementara pada strategi promosi SP5 dilakukan perubahan dengan cara menghilangkan persyaratan program studi dan menambah persyaratan pada hasil tes dan pekerjaan orang tua.

Setelah itu maka dilakukan pengujian terhadap *dataset* strategi promosi yang sudah mengalami perbaikan. Hasil dari pengujian tersebut dapat dilihat pada tabel 4.7 berikut.

Tabel 4.7. Matriks Pengujian Terhadap Rekomendasi 1 Strategi Promosi.

No	Strategi Promosi	Matrix Pengujian				
		6.11	7.11	8.11	9.15	10.12
1	SP0	99,78%	97,71%	99,91%	100,00%	100%
2	SP1	95,77%	98,18%	97%	94%	91%
3	SP2	95%	94%	93,75%	100%	100%
4	SP3	92,00%	89%	94%	100%	100%
5	SP4	100%	100%	100%	100%	100%
6	SP5	97,06%	100%	100%	90%	100%
7	SP6	100%	100%	100%	100%	100%

Dari hasil matriks pengujian terhadap rekomendasi 1 strategi promosi terdapat perubahan yaitu meningkatnya nilai akurasi dari strategi promosi SP4 dan SP5 tetapi terjadi penurunan nilai akurasi terhadap strategi promosi SP1. Untuk menghasilkan hasil klasifikasi strategi promosi yang akurat maka dilakukan analisis terhadap hasil strategi promosi SP1 dan dibuat rekomendasi perubahan strategi promosi yang kedua.

Dari hasil analisis diketahui bahwa ternyata persyaratan strategi promosi SP1 dan SP3 memiliki banyak kemiripan sehingga hal inilah yang membuat proses klasifikasi strategi promosi SP1 menurun keakuratannya. Maka dari itu sebagai rekomendasi strategi promosi SP1 dan SP3 digabungkan menjadi strategi promosi SP1&3. Untuk lebih jelasnya dapat dilihat pada tabel 4.8

Tabel 4.8 Rekomendasi 2 Strategi Promosi

No	Strategi Promosi	Potongan SPI	PERSYARATAN						
			Kota Asal	Program Studi	Status Graduate	Hasil Tes	Pekerjaan Orang tua	Referensi	Jalur masuk
1	SP1&3	20%	Jepara	Selain Manajemen	-	Baik	-	Teman	Umum
				Akuntansi dan				Internet	
				PAI					
2	SP2	30%	luar Jepara	-	-	-	-	Internet	Umum
3	SP4	50%	-	Komunikasi dan Penyiaran Islam,	Fresh Graduate	-	-	-	Tidak Diketahui
				Pendidikan Guru PAUD,					
				Teknik Elektro					
4	SP5	80%	-	-	Fresh Graduate	Baik	Tidak Bekerja	-	Umum
						Sangat Baik	Buruh Bangunan		
							Nelayan		
5	SP6	40%	-	-	-	Sangat Baik	-	-	Umum

Pada tabel 4.8 menunjukkan perubahan pada strategi promosi SP1 dan SP3 yang digabungkan menjadi SP1&3 dengan persyaratan asal kota jepara,program studi selain Manajemen, Akuntansi dan Pendidikan Agama Islam (PAI), Hasil Tes Baik, Referensi dari

Teman dan Internet serta Jalur Masuk Umum. Perubahan tersebut kemudian diterapkan kedalam *dataset* lalu dilakukan pengujian kembali untuk menilai keakuratannya.

Pengujian yang dilakukan yaitu dengan menerapkan aturan atau persyaratan strategi promosi baru yang direkomendasikan kedalam *dataset* untuk selanjutnya dilakukan kembali proses pemodelan untuk mencari nilai akurasi tertinggi yang didapat. Adapun proses pemodelan ini dilakukan sebanyak empat kali pengujian yaitu :

1. Pengujian dengan menggunakan 60% data latih dan 10 *epoch* (Matrik 6.10)
2. Pengujian dengan menggunakan 70% data latih dan 11 *epoch* (Matrik 7.11)
3. Pengujian dengan menggunakan 80% data latih dan 12 *epoch* (Matrik 8.12)
4. Pengujian dengan menggunakan 90% data lati dan 14 *epoch* (Matrik 9.14)

Hasil yang didapat berupa persentase akurasi dari tiap strategi promosi yang direkomendasikan seperti yang terlihat pada tabel 4.9.

Tabel 4.9 Matriks Pengujian Terhadap Rekomendasi 2 Strategi Promosi

No	Strategi Promosi	Matrix Pengujian			
		6.10	7.11	8.12	9.14
1	SP0	99,52%	99,42%	99,56%	100%
2	SP1&3	96,77%	99,48%	98%	100%
3	SP2	92%	92%	94,44%	100%
4	SP4	89%	100%	100%	100%
5	SP5	90,00%	91%	95%	100%
6	SP6	100%	100%	100%	100%

Dari hasil pengujian pada tabel 4.9, nilai akurasi strategi promosi secara keseluruhan meningkat dan tingkat akurasi tertinggi (100%) dapat diperoleh pada percobaan dengan jumlah data latih 90% dan jumlah *epoch* 14 (9.14). hasil tersebut lebih baik jika dibandingkan dengan rekomendasi strategi promosi yang pertama karena tingkat keakuratan tertinggi dapat diperoleh dengan menggunakan jumlah data latih yang lebih sedikit.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Setelah melalui proses yang panjang dan dihasilkan *output* dari penelitian, selanjutnya dapat disimpulkan bahwa :

1. Berdasarkan hasil pengujian, strategi promosi yang efektif pada Universitas Islam Nahdlatul Ulama Jepara adalah strategi promosi berupa potongan SPI 30% untuk yang berasal dari luar kota Jepara (SP2) dan strategi promosi berupa potongan SPI sebesar 40% bagi calon mahasiswa yang nilai tesnya sangat baik (SP6).
2. Sementara itu strategi promosi yang kurang efektif adalah strategi promosi SP1 dan SP3 karena memiliki kemiripan pada persyaratannya dan strategi promosi SP4 dan SP5 yang melibatkan atribut-atribut yang berkorelasi rendah terhadap strategi promosi seperti atribut pekerjaan orang tua dan jalur masuk.
3. Berdasarkan hasil penelitian maka dapat disimpulkan bahwa untuk penentuan strategi promosi yang efektif dan yang kurang efektif didasarkan pada nilai korelasi atribut-atribut yang diperoleh. Adapun atribut yang memiliki nilai korelasi tinggi yaitu program studi, referensi, hasil tes dan asal kota.
4. Sementara atribut yang kurang berpengaruh dalam menentukan strategi promosi adalah jenis sekolah dan jalur masuk dikarenakan nilai korelasinya rendah.
5. Proses pemilihan atribut sangat berpengaruh dalam menghasilkan pemodelan *deep learning* yang baik. Semakin tinggi korelasi antar atribut maka akan menghasilkan akurasi yang tinggi dalam pemodelan.
6. Untuk mencapai tingkat akurasi yang tinggi dalam pemodelan *deep learning* diperlukan jumlah data latih yang memadai sehingga model yang didapatkan berhasil dengan baik mengklasifikasikan data uji.
7. Untuk mengklasifikasikan calon mahasiswa dalam penerapan strategi promosi dengan metode *deep learning* diperlukan beberapa langkah yaitu pembersihan data,

seleksi data dan Transformasi. pembuatan model dengan menggunakan data latih, pengujian model dan ekstraksi pengetahuan.

8. Hasil yang didapat dalam mengklasifikasikan data calon mahasiswa berdasarkan strategi promosi yaitu sebanyak 242 *record* diklasifikasikan kedalam strategi promosi tipe SP1, 42 diklasifikasikan kedalam strategi promosi tipe SP2, 89 *record* diklasifikasikan kedalam strategi promosi tipe SP3, 22 *record* untuk strategi promosi tipe SP4, 29 *record* untuk strategi SP5 dan 37 *record* diklasifikasikan kedalam strategi promosi tipe SP6.
9. Untuk mendapatkan penerapan strategi promosi yang efektif maka dilakukan percobaan perubahan terhadap beberapa jenis persyaratan atribut dalam strategi promosi tipe SP4 dan SP5 serta penggabungan strategi promosi SP1 dan SP3 yang ternyata berhasil mendapatkan tingkat keakuratan yang lebih baik dan membuat proses klasifikasi menjadi lebih efektif akan tetapi ini hanya merupakan rekomendasi strategi promosi.

## 5.2 Saran

Adapun saran yang penulis dapat diberikan untuk penelitian selanjutnya dan untuk institusi perguruan tinggi yaitu :

1. Saran untuk penelitian selanjutnya yaitu untuk mencoba menambah atribut lain yang lebih beragam yang bisa mempengaruhi strategi promosi pada perguruan tinggi yaitu atribut yang memiliki nilai korelasi yang tinggi dalam menentukan strategi promosi seperti atribut sertifikat atau piagam yang pernah didapat jumlah penghasilan orang tua, jumlah saudara, nilai atau ranking saat semester terakhir di SMA/MA/SMK dan lain sebagainya.
2. Selain itu saran untuk penelitian selanjutnya bila mengolah *dataset* untuk penentuan strategi promosi dengan menggunakan atribut antara lain yaitu asal sekolah, program studi, referensi, hasil tes, status graduate, asal sekolah dan jalur masuk maka untuk mendapatkan hasil akurasi yang tinggi diperlukan jumlah data latih yang cukup banyak. Semakin banyak data latih maka semakin besar akurasi yang didapatkan.

3. Untuk penelitian selanjutnya dapat mencoba menggunakan jenis fungsi aktifator lain dalam metode *deep learning* selain menggunakan *Rectifier* seperti *Tanh* , *Maxout* atau *Exponential* untuk bisa mendapatkan hasil yang lebih akurat atau untuk membandingkan penggunaannya mana yang lebih akurat.
4. Saran yang dapat diberikan untuk Institusi perguruan tinggi yaitu sebelum menentukan strategi promosi maka terlebih dahulu harus melihat kedalam data atribut calon mahasiswa dan membuat strategi promosi yang sesuai dengan data atribut tersebut.
5. Selain itu saran yang dapat diberikan untuk Institusi perguruan tinggi adalah dengan menambahkan beberapa atribut baru bagi calon mahasiswa yang wajib untuk diisi saat mendaftar. Atribut tersebut adalah yang sekiranya dapat membantu atau menjadi dasar untuk penentuan strategi promosi dimasa depan.

## DAFTAR PUSTAKA

- Alom, M.Z. et al. 2018.** "the History Began From AlexNet: A Comprehensive Survey on Deep Learning Approaches" .Cornell University Library.
- Candel, Alan and LeDell, Erin. **2019.** "Deep Learning With H2O". H2O.ai, Inc. Mountain View, CA 94043.
- Charniak, Eugene. 2018.** "Introduction to Deep Learning".The MIT Press Cambridge, Massachusetts London, England.
- Chiu, Susan dan Tavella Domingo.2008.** "Data Mining and Market Intelligence :For Optimal Marketing Return ". Elsevier. ISBN: 978-0-7506-8234-3
- Choelho, Orlando Bisacchi dan Silveria , Ismar Frango. **2017.** "Deep Learning applied to Learning Analytics and Educational Data Mining". VI Congresso Brasileiro de Informática na Educação (CBIE 2017).
- Du, Xiaoshan. 2006.** "Data Mining Analysis and Modeling for Marketing Based On Attribute of Customer Relations". Växjö University SE-351 95 Växjö, Sweden.
- Fox Wilton W.T. et al. **2018.** " Prediction Model for Students' Future Development by Deep Learning and Tensorflow Artificial Intelligence Engine". IEEE International Conference on Information Management.
- Gajic, Jelena . 2012.** "Importance Of Marketing Mix In Higher Education Institutions". Singidium Journal Of Applied Science. ISSN 2217-8090.
- Han, Jiawei et al. .2012.**"Data Mining Concepts and Techniques". Morgan Kaufman Series. USA.
- Hermawati, Fajar Astuti. 2013.**" Data Mining" Penerbit Andi. Yogyakarta.
- Kusrini, Luthfi Taufiq Emha. 2015. "Algoritma Data Mining " Penerbit Andi. Yogyakarta
- Land, Sebastian dan Fischer, Simon. 2012.** "RapidMiner 5 : RapidMiner in academic use". by Rapid-I GmbH.
- Layton, Robert. 2015."Learning Data Mining With Python" Pact Publishing. Birmingham.
- Larose, Daniel dan Larose Chantal.2015.**"Data Mining and Predictive Analysis". Wiley Series. ISBN 978-1-118-11619-7.
- Maheswari, Anil K. 2015. "Business Intelligence and data mining". Business Express Press.New York.
- Ohsawa, Yukio. 2009.**" Data Mining for Design and Marketing ". Chapman & Hall. ISBN: 978-1-4200-7019-4.
- Patil Akhilesh P, et al . 2017.** "Effective Deep Learning Model to Predict Student Grade Point Averages" IEEE International Conference on Computational Intelligence and Computing Research. 978-1-5090-6621-6/17.
- Prastyo, Pulung Hendro, Saharuna, Zawiyah dan Abduh, Ibrahim. 2017.**" Analisa data atribut mahasiswa untuk menentukan strategi promosi kampus menggunakan

**metode data mining”. Seminar Nasional Teknik Elektro dan Informatika (SNTEI) 2017 978-602-18168-2-0.**

**Primartha, Rifkie, 2018. “ Belajar Machine Learning “. Informatika Bandung ISBN: 978-602-6232-67-0.**

RapidMiner Studio Manual. 2014. [www.rapidminer.com](http://www.rapidminer.com).

Reutterer, Thomas, Hornik, Kurt March ,Nicolas, dan Gruber , **Kathrin. 2016.** “A data mining framework for targeted category promotions”. Springer. DOI 10.1007/s11573-016-0823-7.

Sadiq, Hussain. **2019.** “Prediction Model on Student Performance based on Internal Assessment using Deep Learning”. <https://doi.org/10.3991/ijet.v14i08.10001>.

**Siyoto, Sandu dan Ali Sodik. 2015. “ Dasar Metodologi Penelitian “. Literasi Media Publishing. Yogyakarta..**

**Skansi, Sandro. 2018. “Introduction to Deep Learning From Logical Calculus to Artificial Intelligence”. Springer. ISBN 978-3-319-73004-2 (e-book).**

Sultana Jabeen, Rani M. Usha, M.A.H. Farquad. **2019. “Student’s Performance Prediction using Deep Learning and Data Mining Methods”. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4.**

Suyanto1. 2019. **“Data mining untuk klasifikasi dan klasterisasi data”. Informatika. Bandung**

**Suyanto2, 2018. “Machine Learning”. Penerbit Informatika Bandung. ISBN :978-602-6232-78-6.**

**Tjiptono, Fandi dan Chandra, Gregorius. 2017.”Pemasaran Strategik Edisi 3”. Penerbit Andi. ISBN : 978-979-29-6720-3.**

**Whalley, Andrew. 2010. “Strategic Marketing “. Ventus Publishing Aps. ISBN :978-87-7681-643-8.**

Wang, Shu-**Ching et all .2014.” Systematic Approach for Digital Marketing Strategy through Data Mining Technology”. Journal of Computers Vol. 25, No. 3, October 2014.**

Witten, Ian. H. etall. 2016. **“Data Mining practical machine learning tools and techniques “. Morgan Kaufman Series. San Fransisco.**

Ying yang, Geoffrey I. **Webb and Xindong Wu. 2010. “Discretization Method in Data mining and Knowledge Discovery Handbook”. Springer New York. ISBN 978-0-387-09822-7.**

**Zaki, Mohammed J & Meira, Wagner. 2013. “Data Mining and Analysis”. Cambridge University Press.**

Zhang, Cha and Ma **Yunqian. 2012. “ Ensemble Machine Learning”. Springer New York Dordrecht Heidelberg London. ISBN 978-1-4419-9325-0**

<https://forlap.ristekdikti.go.id/perguruan tinggi/homegraphpt> diakses tanggal 3 juli pukul 13:40 WIB.

<https://www.universalclass.com/articles/business/marketing-strategies-promotion-advertising-and-public-relations.htm>. Diakses tanggal 2 september pukul 17:53.

<https://unisnu.ac.id/halaman/sejarah> diakses pada tanggal 9 september 2019 pukul 5:30.

## LAMPIRAN

### A. Dokumentasi Hasil Pengujian dengan Rapidminer

#### 9.3 1. Hasil pengujian 10% data latih dan 12 epoch

The screenshot shows the Rapid Miner interface with the following configuration:

- Model:** Feedforward Neural Network (FFNN) with 1 hidden layer and 1 output layer.
- Learning Rate:** 0.01
- Epochs:** 12
- Activation Function:** Sigmoid
- Loss Function:** Cross Entropy

**Training History:**

Iteration	Training Error	Validation Error	Classification Accuracy
1	0.999	0.999	0.000
2	0.999	0.999	0.000
3	0.999	0.999	0.000
4	0.999	0.999	0.000
5	0.999	0.999	0.000
6	0.999	0.999	0.000
7	0.999	0.999	0.000
8	0.999	0.999	0.000
9	0.999	0.999	0.000
10	0.999	0.999	0.000
11	0.999	0.999	0.000
12	0.999	0.999	0.000

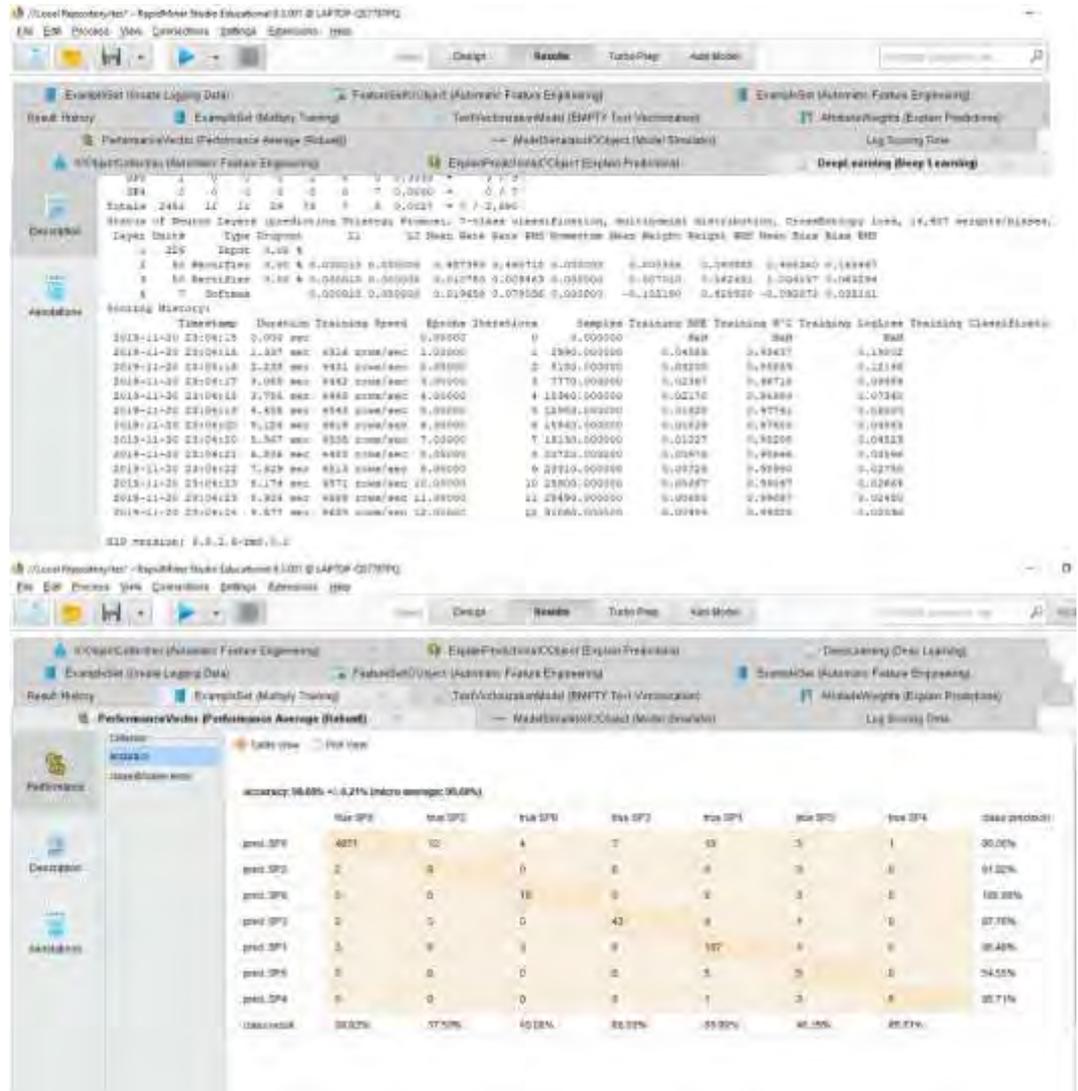
The screenshot shows the Performance View for the Feedforward Neural Network model. The overall accuracy is 91.42% (micro average 91.42%).

**Confusion Matrix:**

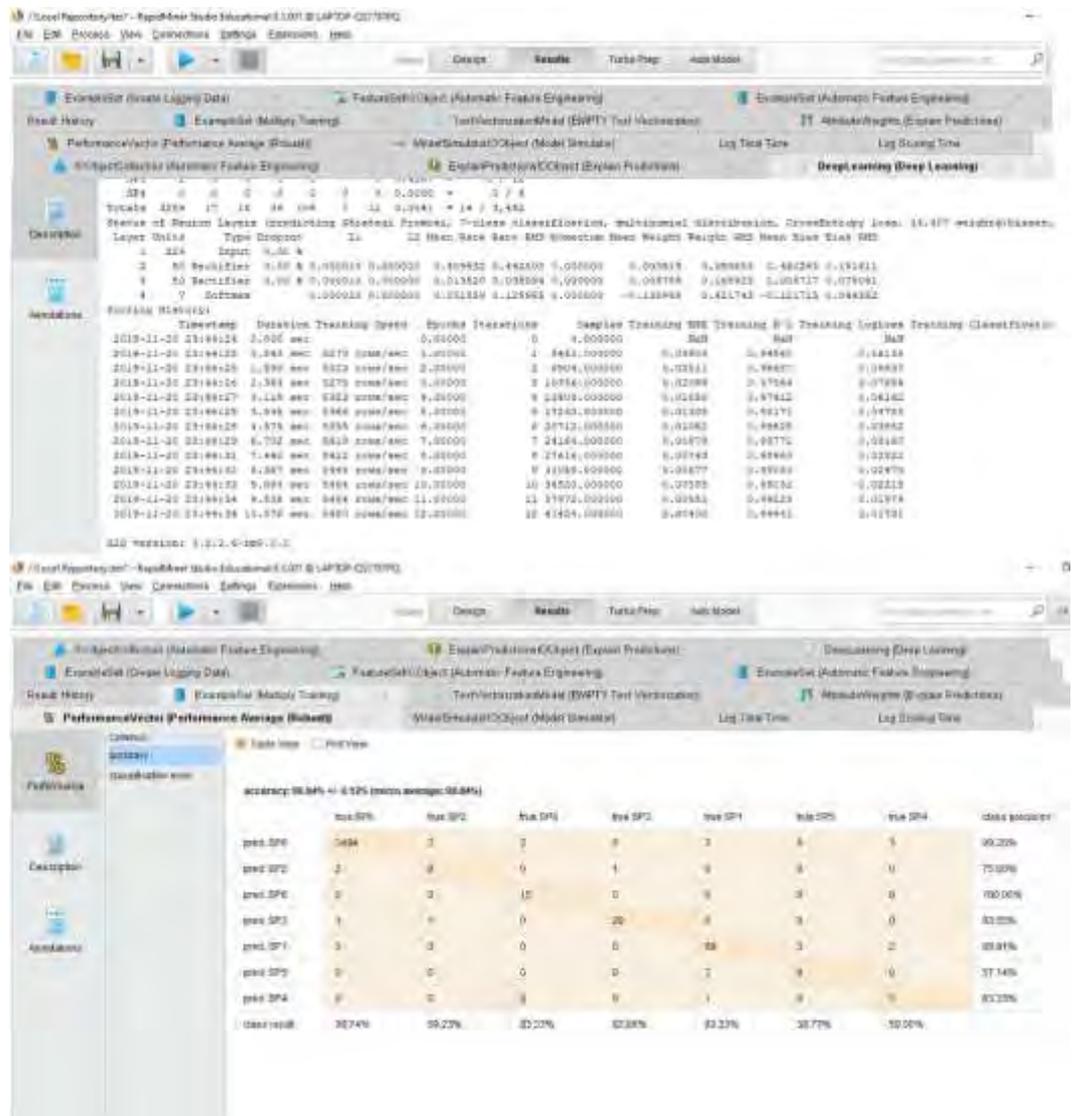
	Real SP1	Real SP2	Real SP3	Real SP4	Real SP5	Real SP6	Real SP7	Real SP8	Real SP9	Real SP10	Real SP11	Real SP12	Real SP13	Real SP14	Real SP15	Real SP16	Real SP17	Real SP18	Real SP19	Real SP20
Real SP1	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP2	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP4	1	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP5	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Real SP20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Overall	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%	91.42%



3. Hasil pengujian 30% data latih dan 12 epoch



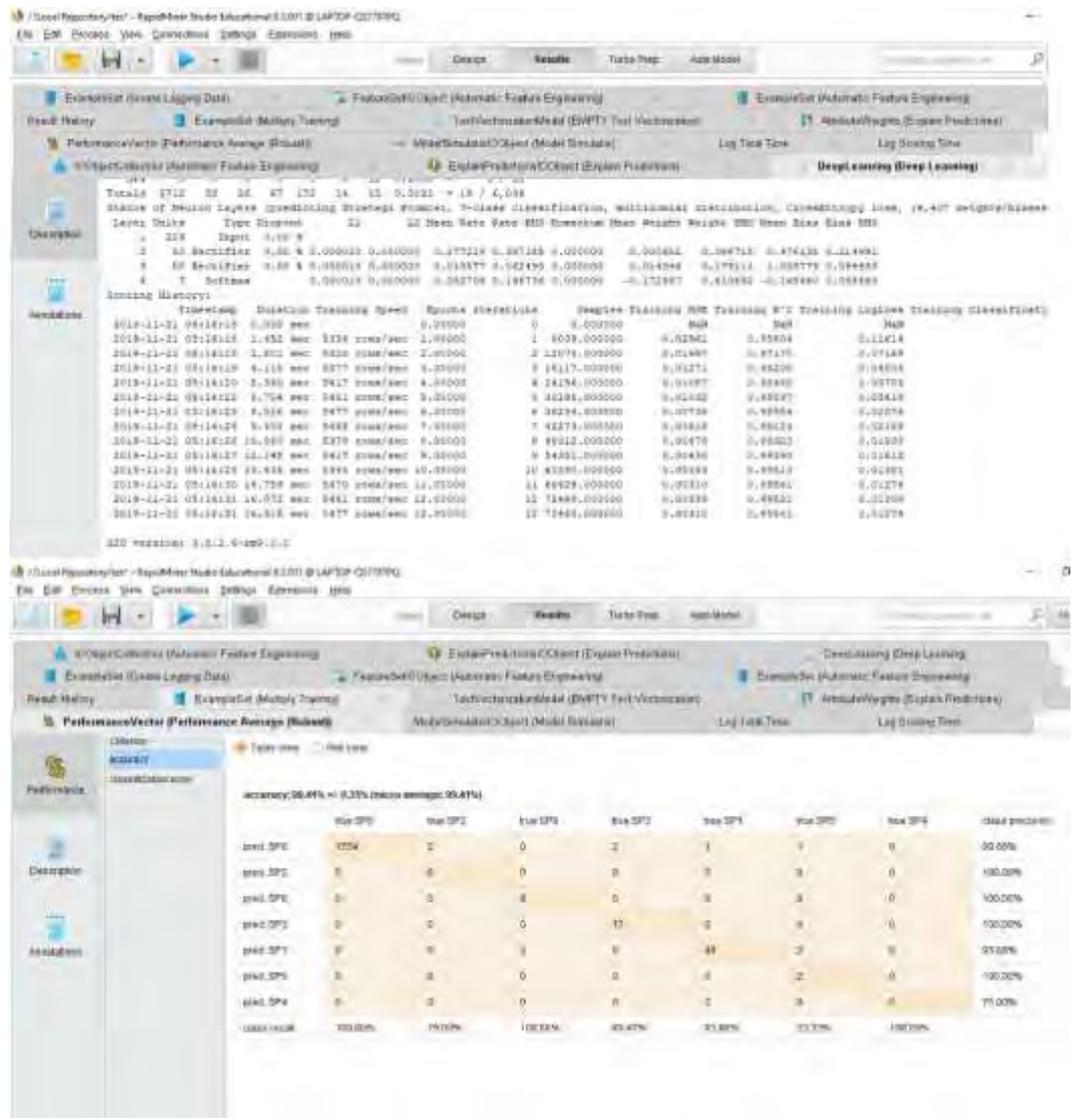
4. Hasil pengujian 40% data latih dan 12 epoch







7. Hasil pengujian 70% data latih dan 12 epoch



8. Hasil pengujian 80% data latih dan 12 epoch

The screenshot displays the MATLAB interface during a neural network training process. The top window shows the 'Deployment (Deep Learning)' tab with a 'Performance Vector Performance Average (ReLU)' table. The bottom window shows the 'Performance' tab with a 'Performance Vector Performance Average (ReLU)' table.

**Training Progress Summary:**

Epoch	Time	Duration	Training Speed	Epochs	Iterations	Sample Training RMSE	Training F1	Training Accuracy	Classification
2019-11-21 21:28:42	0:00	sec		0	0	0.00000	NaN	NaN	NaN
2019-11-21 21:28:43	0:01	sec	5647 rows/sec	1	0	0.9820,00000	0.1846	0.18724	0.1133
2019-11-21 21:28:44	0:02	sec	5436 rows/sec	2	0	0.9910,00000	0.9297	0.9726	0.9494
2019-11-21 21:28:44	0:03	sec	5785 rows/sec	3	0	0.9975,00000	0.9188	0.9588	0.9414
2019-11-21 21:28:47	0:06	sec	5544 rows/sec	4	0	0.9980,00000	0.9187	0.9576	0.9397
2019-11-21 21:28:48	0:08	sec	5321 rows/sec	5	0	0.9982,00000	0.9178	0.9567	0.9375
2019-11-21 21:28:50	0:10	sec	4949 rows/sec	6	0	0.9983,00000	0.9180	0.9569	0.9358
2019-11-21 21:28:52	0:12	sec	5381 rows/sec	7	0	0.9983,00000	0.9181	0.9568	0.9351
2019-11-21 21:28:53	0:13	sec	4937 rows/sec	8	0	0.9983,00000	0.9187	0.9568	0.9350
2019-11-21 21:28:54	0:14	sec	4934 rows/sec	9	0	0.9983,00000	0.9183	0.9568	0.9351
2019-11-21 21:28:58	0:18	sec	4984 rows/sec	10	0	0.9983,00000	0.9189	0.9567	0.9328
2019-11-21 21:28:59	0:19	sec	4911 rows/sec	11	0	0.9983,00000	0.9183	0.9567	0.9333
2019-11-21 21:28:57	0:18	sec	4936 rows/sec	12	0	0.9983,00000	0.9179	0.9562	0.9302
2019-11-21 21:29:00	0:21	sec	4921 rows/sec	13	0	0.9983,00000	0.9182	0.9565	0.9303

**Performance Vector Performance Average (ReLU) Summary:**

Category	Pre-SP1	Pre-SP2	Pre-SP3	Pre-SP4	Pre-SP5	Pre-SP6	Pre-SP7	Pre-SP8	Class Accuracy
Pre-SP1	100%	0	0	1	3	8	0	0	39.74%
Pre-SP2	1	7	0	0	0	0	0	0	67.03%
Pre-SP3	0	0	3	0	0	0	0	0	100.00%
Pre-SP4	2	3	0	14	0	0	0	0	100.00%
Pre-SP5	0	0	0	0	14	0	0	0	100.00%
Pre-SP6	0	0	0	0	0	1	3	0	75.00%
Pre-SP7	0	0	0	0	0	1	0	0	75.00%
Overall	88.07%	100.00%	100.00%	82.12%	85.47%	100.00%	100.00%		



10. Hasil pengujian 95% data latih dan 12 epoch

The image displays three sequential screenshots from the RapidMiner software interface, illustrating the training and testing process of a neural network model.

**Top Screenshot: Training Progress**  
 This window shows the 'DeepLearning (Deep Learning)' process. The 'Results' tab displays a table of training progress over 12 epochs. The table includes columns for Epoch, Training Time, and various performance metrics. The accuracy starts at 0% and reaches 100% by epoch 12.

Epoch	Training Time	Accuracy
1	0:00:00	0%
2	0:00:00	0%
3	0:00:00	0%
4	0:00:00	0%
5	0:00:00	0%
6	0:00:00	0%
7	0:00:00	0%
8	0:00:00	0%
9	0:00:00	0%
10	0:00:00	0%
11	0:00:00	0%
12	0:00:00	100%

**Middle Screenshot: Performance Metrics**  
 This window shows the 'Performance' tab for the 'DeepLearning (Deep Learning)' process. It displays a table of performance metrics for each epoch, including accuracy, precision, recall, and F1 score. All metrics are consistently 100% across all 12 epochs.

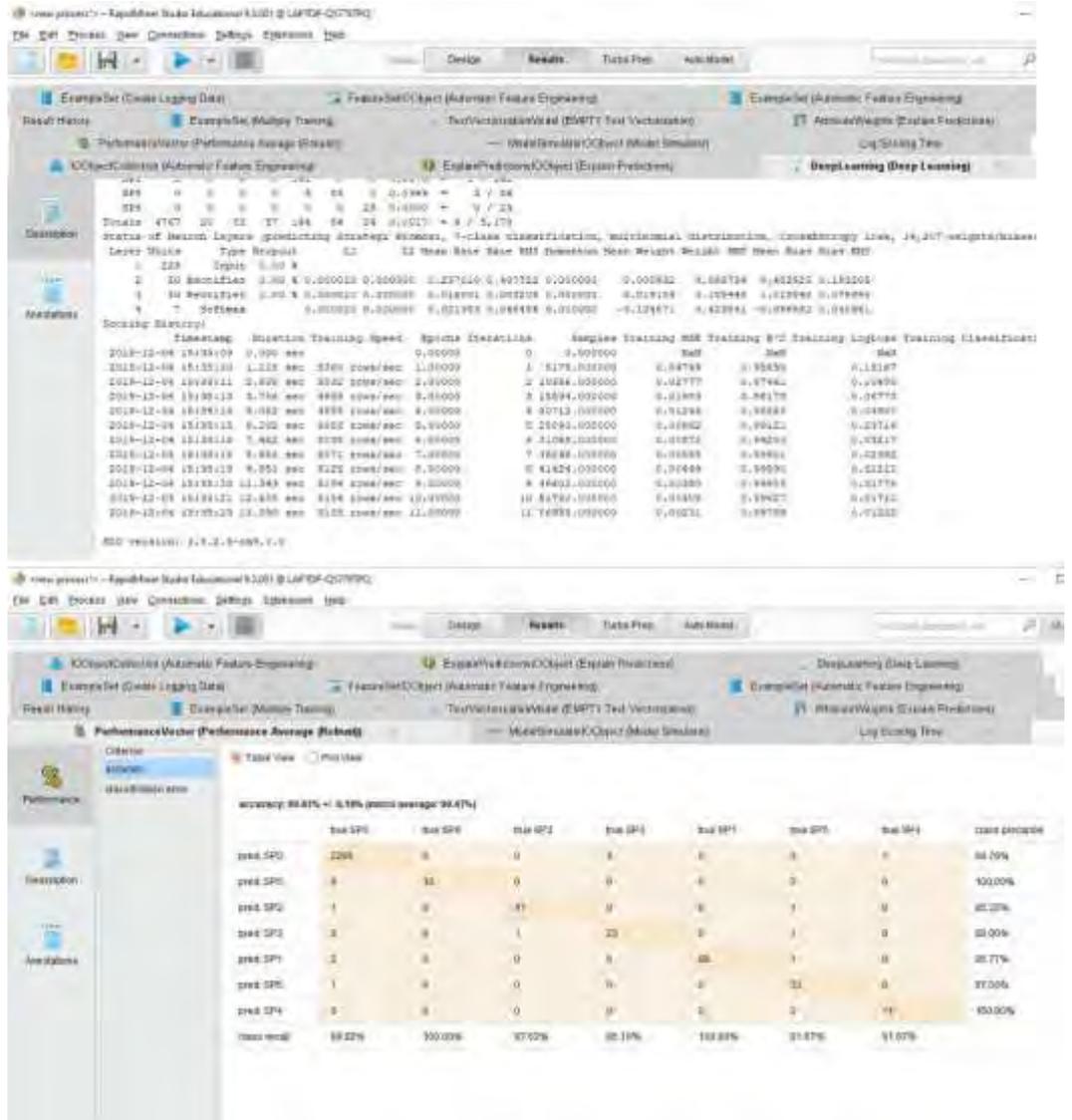
Epoch	Accuracy	Precision	Recall	F1 Score
1	100%	100%	100%	100%
2	100%	100%	100%	100%
3	100%	100%	100%	100%
4	100%	100%	100%	100%
5	100%	100%	100%	100%
6	100%	100%	100%	100%
7	100%	100%	100%	100%
8	100%	100%	100%	100%
9	100%	100%	100%	100%
10	100%	100%	100%	100%
11	100%	100%	100%	100%
12	100%	100%	100%	100%

**Bottom Screenshot: Classification Results**  
 This window shows the 'Explains' tab for the 'ExampleSet (Cross-Logging Data)' process. It displays a table of classification results for 14 instances. The table includes columns for Row No., Strategy, and various performance metrics. The results show a mix of 'Correct' and 'Wrong' classifications.

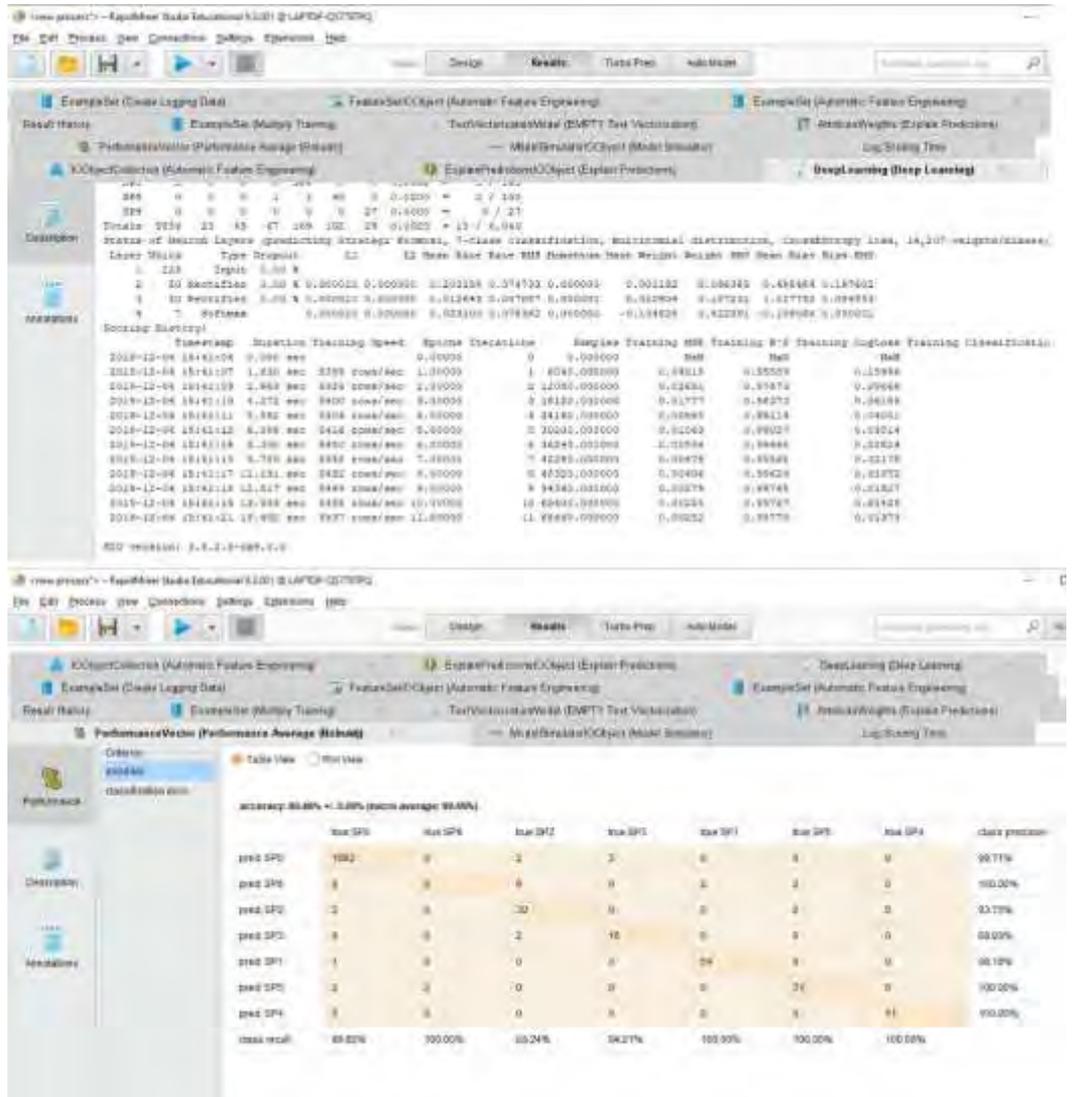
Row No.	Strategy	Classification	Actual Class	Result Type
1	SP1	SP1	SP1	Correct
2	SP1	SP1	SP1	Correct
3	SP1	SP1	SP1	Correct
4	SP1	SP1	SP1	Correct
5	SP1	SP1	SP1	Correct
6	SP1	SP1	SP1	Correct
7	SP1	SP1	SP1	Correct
8	SP1	SP1	SP1	Correct
9	SP1	SP1	SP1	Correct
10	SP1	SP1	SP1	Correct
11	SP1	SP1	SP1	Correct
12	SP1	SP1	SP1	Correct
13	SP1	SP1	SP1	Correct
14	SP1	SP1	SP1	Correct

**B. Hasil Pengujian Strategi Promosi Rekomendasi**

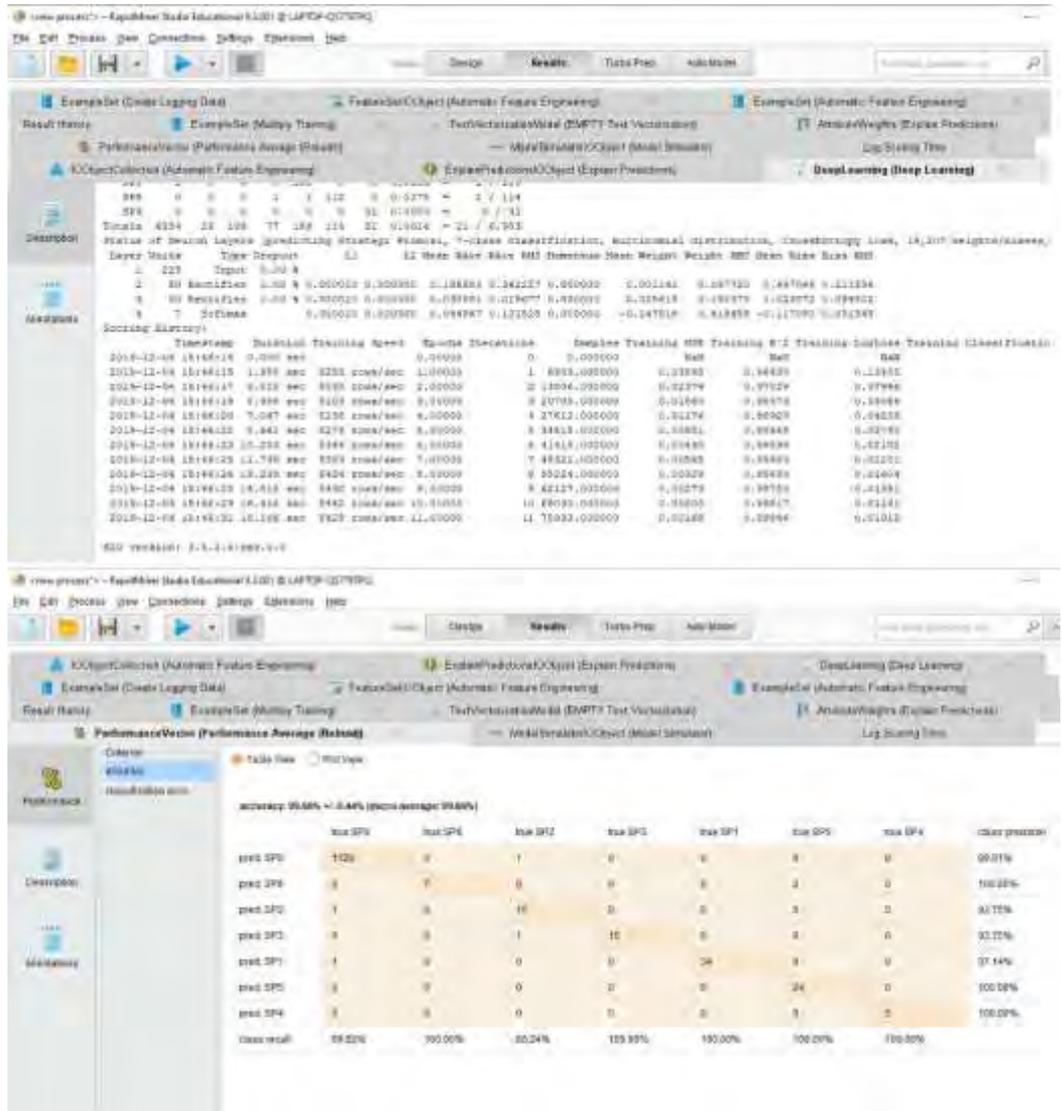
**1.1. Pengujian dengan data latih 60 % dan 11 epoch**



2. Pengujian dengan data latihan 70 % dan 11 epoch



3. Pengujian dengan data latih 80 % dan 11 epoch

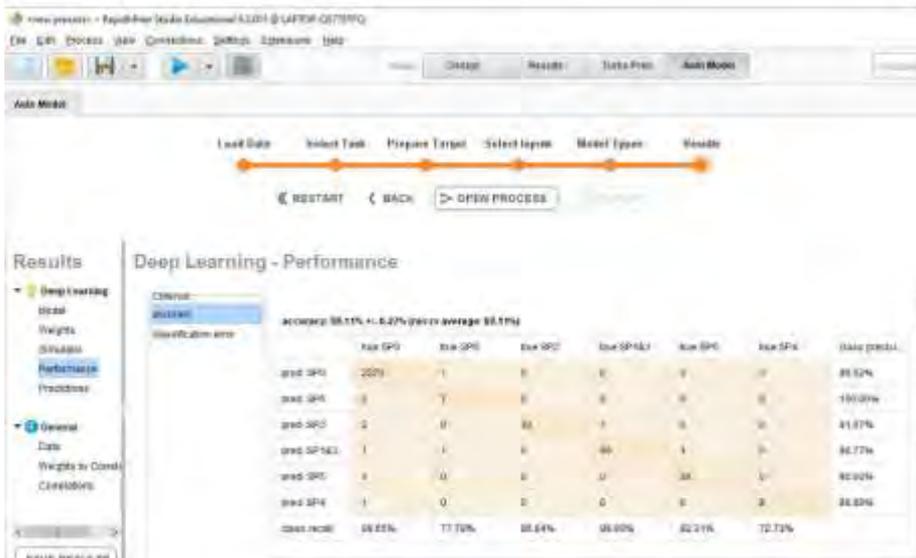
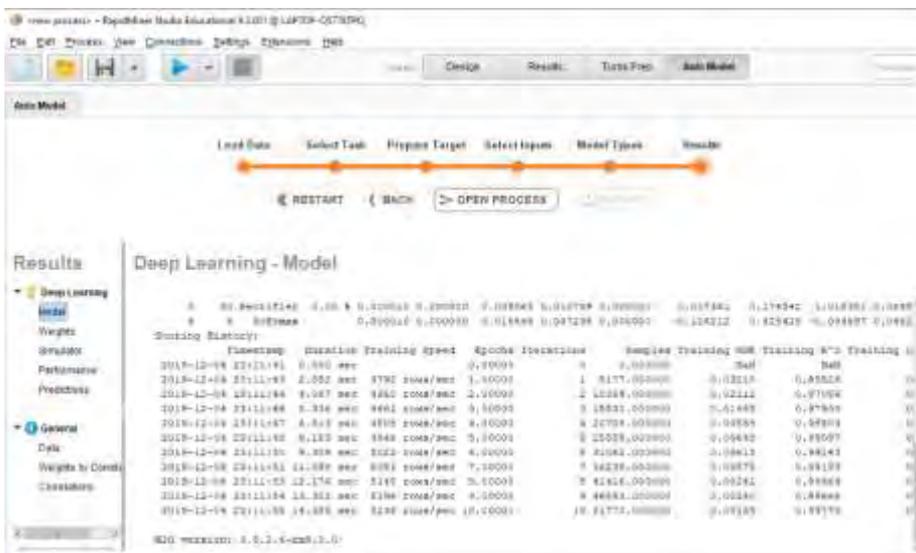






### C. Hasil Pengujian Strategi Promosi Rekomendasi

#### 2 1. Pengujian dengan data latih 60% dan 10 epoch









#### D. Surat Keterangan Riset

 UPT PUSAT DATA DAN PENGEMBANGAN IT  
UNIVERSITAS ISLAM NAHDLATUL ULAMA (UNISNU) JEPARA

Jl. Taman Siswa (Pekang) Tahunan Jepara | telp. 59427  
☎ 0291 595320 | ☎ 0291 592630 | ☎ voip 3004  
✉ [it@unisnu.ac.id](mailto:it@unisnu.ac.id) | <http://it.unisnu.ac.id>

UPT Pusat Data & Pengembangan IT 

**SURAT KETERANGAN**  
Nomor : 003/UPT PDIT.UNISNU/IX/2019

Yang bertanda tangan dibawah ini Kepala UPT Pusat Data dan Pengembangan IT Unisnu Jepara menerangkan bahwa:

Nama : Sani Nurmalasari  
NPM : 2018210068  
Jentang Pendidikan : PascaSarjana (S2)  
Bidang Minat : Sistem Informasi Bisnis

Benar telah melaksanakan Penelitian untuk tesis dengan judul “ Data Mining Untuk Menentukan Strategi Promosi” pada tanggal 14 September sampai dengan 30 September 2019 di Bagian Pusat Data dan IT Unisnu Jepara.

Demikian Surat Keterangan ini dibuat agar yang berkepentingan menjadikan periksa

Bagian Pusat Data dan Pengembangan IT  
UNISNU Jepara,

  
**Harmino Mulvo, S.Kom, M.Kom**  
NIY. 3 820204 13 096

## E. Dokumentasi Foto

### 1. Gedung Rektorat UNISNU Jepara



### 2. Gedung Dekanat UNISNU Jepara



3. Gedung Fakultas Sains dan Teknologi UNISNU Jepara



4. Gedung Perpustakaan



5. Masjid UNISNU Jepara



6. Pusat Pelayanan Kesehatan UNISNU Jepara

