LAPORAN PENELITIAN
KEBIJAKAN DAN KELEMBAGAAN
DANA PNBP TAHUN ANGGARAN 2012

PREDIKSI DAN PEMETAAN DATA MAHASISWA FAKULTAS TEKNIK UNIVERSITAS NEGERI GORONTALO MENGGUNAKAN PENDEKATAN DATA MINING

Lillyan Hadjaratie, S.Kom, M.Si

JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS NEGERI GORONTALO
OKTOBER 2012

ABSTRAK

Penelitian "Prediksi dan Pemetaan Data Mahasiswa Fakultas Teknik Menggunakan Pendekatan Data Mining" bertujuan untuk membangkitkan informasi dan pengetahuan dari data akademik kemahasiswaan dengan: (1) Mengklasifikasi data mahasiswa aktif dan lulusan untuk memprediksi kategori IPK dan kategori Lama Studi; (2) Mengklaster data mahasiswa aktif dan lulusan berdasarkan kategori IPK dan Lama_Studi.

Metode klasifikasi yang digunakan untuk mengklasifikasi data mahasiswa aktif dan lulusan dalam penelitian ini adalah metode *Decision Tree,* Artificial Neural Network dan K-Nearest Neighbour, sedangkan metode yang digunakan untuk mengklaster data mahasiswa aktif dan lulusan adalah metode Hirarkis (*Average Between Lingkage*) dan Non-Hirarkis (K-*Means*).

Penelitian ini menghasilkan informasi akademik berupa hasil klasifikasi data mahasiswa aktif dan lulusan untuk memprediksi kategori IPK dan Lama_Studi serta pemetaan data mahasiswa aktif dan lulusan berdasarkan sasaran mutu IPK, sebagai sebuah sistem peringatan dini (*early morning*) dan bahan pertimbangan dalam proses pengambilan kebijakan dan keputusan.

Kata Kunci : Data Mining, Klasifikasi, Klasterisasi, Data Mahasiswa, IPK Lama Studi

LEMBAR IDENTITAS DAN PENGESAHAN

1. Judul	:	Prediksi dan Pemetaan Data Mahasiswa Fakultas Teknik Universitas Negeri Gorontalo Menggunakan Pendekatan Data Mining
2. Ketua Tim Pengusul		
a. Nama Lengkap	:	Lillyan Hadjaratie, S.Kom, M.Si
b. Jenis Kelamin	:	P
c. NIP	:	19800414 200212 2002
d. Jabatan struktural	:	-
e. Jabatan Fungsional	:	Lektor
f. Fakultas / Jurusan	:	Teknik / Teknik Informatika
g. Pusat Penelitian	:	Lembaga Penelitian Universitas Negeri Gorontalo
h. Alamat	:	Jl. Jend.Sudirman No.6 Kel. Dulalowo Kota Gorontalo Propinsi Gorontalo.
i. Telpon/fax	:	081356139486
j. Alamat rumah	:	Jln Gelatik No. 68, Kel. Heledulaa Utara Kota Gorontalo
k. Telpon/fax/email	:	081356139486 / - / lillyan.hadjaratie@gmail.com
3. Jangka waktu penelitian	:	6 bulan
4. Pembiayaan		
Jumlah biaya yang diajukan	:	Rp. 9.000.000 (sembilan juta rupiah)

Gorontalo , Oktober 2012

Mengetahui,
Dekan Fakultas Teknik

Ketua Peneliti

Ir. Rawiyah Husnan, MT
NIP : 19640427 199403 2001

Lillyan Hadjaratie, S.Kom. M.Si
NIP : 19800417 200212 2002

Menyetujui,
Ketua Lembaga Penelitian
Universita Negeri Gorontalo

Dr. Fitriyane Lihawa, M.Si
NIP. 196912091993032001

KATA PENGANTAR

Segala puja dan puji syukur hanya milik Allah S.W.T sang khalik pemilik alam raya, karena berkat rahmat dan ridhoNya penelitian dan laporan penelitian ini dapat diselesaikan.

Adapun laporan ini disusun melalui serangkaian aktivitas diantaranya pengumpulan data, pembersihan data, Transformasi data, Klasifikasi dan Klasterisasi data mahasiswa aktif dan lulusan Fakultas Teknik Universitas Negeri Gorontalo.

Terlepas dari kekurangan yang terdapat dalam penelitian ini, kami menyadari dalam penyusunan penelitian ini tidak lain juga karena bantuan dari berbagai pihak, oleh karenanya Kami mengucapkan terima kasih yang sebesar-besarnya.

Akhir kata semoga penelitian ini dapat berguna dan bermanfaat bagi yang membutuhkannya dan Kami sangat mengharapkan masukan dan saran yang dapat dan ada kelanjutan dari penelitian ini sehingga penelitian ini menjadi lebih baik dan berguna.

Gorontalo, Oktober 2012

Tim Peneliti

DAFTAR ISI

ABSTRAK i
LEMBAR IDENTITAS DAN PENGESAHAN ii
KATA PENGANTAR iii
DAFTAR ISI iv
DAFTAR TABEL vi
DAFTAR GAMBAR viii
DAFTAR LAMPIRAN ix

BAB I. PENDAHULUAN 1
A. LATAR BELAKANG 1
B. RUMUSAN MASALAH 2
C. TUJUAN PENELITIAN 2
D. MANFAAT DAN URGENSI PENELITIAN 2

BAB II. KERANGKA TEORI 2
A. Knowledge Discovery in Database (KDD) 3
B. Data Kategorikal 6
C. Teknik Klasifikasi 6
D. Teknik Klasterisasi 7

BAB III. METODE PENELITIAN 10
A. Lokasi dan Waktu Penelitian 10
B. Materi Penelitian 10
C. Alat Penelitian 10
D. Alur Penelitian 11
E. Jadwal Pelaksanaan Penelitian 13

BAB IV. HASIL PENELITIAN DAN PEMBAHASAN 14
A. Praproses Data 14
B. Klasifikasi 16
C. Klasterisasi 39
D. Perbandingan Hasil Prediksi 46
E. Tingkat *Sensitivity* dan *Importance* Variabel Bebas 47
F. Perbadingan Rata-Rata 50

BAB V. KESIMPULAN DAN SARAN 55
A. Kesimpulan 55
B. Saran 56

DAFTAR PUSTAKA 57
LAMPIRAN 59

DAFTAR TABEL

Hal

Tabel 1. Jadwal Pelaksanaan Penelitian 13
Tabel 2. Kategori Data Mahasiswa Aktif 15
Tabel 3. Kategori Data Mahasiswa Lulusan 16
Tabel 4. Nilai signifikansi variabel bebas terhadap variabel IPK mahasiswa aktif 19
Tabel 5. Klasifikasi dan Karakteristik Data Mahasiswa Aktif berdasarkan variabel IPK 19
Tabel 6. Jumlah Mahasiswa Aktif setiap Klasifikasi berdasarkan Kategori IPK 20
Tabel 7. Nilai signifikansi variabel bebas terhadap variabel Alamat_Asal mahasiswa aktif 23
Tabel 8. Klasifikasi dan Karakteristik Data Mahasiswa Aktif berdasarkan variabel Alamat_Asal 24
Tabel 9. Jumlah Mahasiswa Aktif setiap Klasifikasi berdasarkan Kategori Alamat_Asal 24
Tabel 10. Nilai signifikansi variabel bebas terhadap variabel IPK mahasiswa lulusan 27
Tabel 11. Klasifikasi dan Karakteristik Data Mahasiswa Lulusan berdasarkan variabel IPK 28
Tabel 12. Jumlah Mahasiswa Lulusan setiap Klasifikasi berdasarkan Kategori IPK 28
Tabel 13. Nilai signifikansi variabel bebas terhadap variabel Lama_Studi mahasiswa lulusan 30
Tabel 14. Klasifikasi dan Karakteristik Data Mahasiswa Lulusan berdasarkan variabel Lama_Studi 31
Tabel 15. Jumlah Mahasiswa Lulusan setiap Klasifikasi berdasarkan Kategori Lama_Studi 31
Tabel 16. *Final Cluster Centers* Mahasiswa Aktif 40
Tabel 17. Final Cluster Centers Mahasiswa Lulusan 41
Tabel 18. Anggota Cluster Data Mahasiswa Aktif 43
Tabel 19. Anggota Cluster Data Mahasiswa Lulusan 45
Tabel 20. Ketepatan Uji Prediksi IPK dan Alamat_Asal dari data Mahasiswa Aktif 46
Tabel 21. Ketepatan Uji Prediksi IPK dan Alamat_Asal dari data Mahasiswa Lulusan 47
Tabel 22 Daftar Independent Variabel Importance berdasarkan variabel terikat IPK 47
Tabel 23. Perbandingan rata-rata variabel Jenis_Kelamin terhadap variabel IPK dari data mahasiswa aktif 50
Tabel 24. Perbandingan rata-rata variabel Jurusan terhadap variabel IPK dari data mahasiswa aktif 50
Tabel 25. Perbandingan rata-rata variabel Jalur masuk terhadap variabel IPK dari data mahasiswa aktif 50

Tabel 26. Perbandingan rata-rata variabel Pekerjaan_Ortu terhadap variabel IPK dari data mahasiswa aktif 51
Tabel 27. Perbandingan rata-rata variabel Penghasilan_Ortu terhadap variabel IPK dari data mahasiswa aktif 51
Tabel 28. Perbandingan rata-rata variabel Asal_Sekolah terhadap variabel IPK dari data mahasiswa aktif 51
Tabel 29. Perbandingan rata-rata variabel Alamat_Asal terhadap variabel IPK dari data mahasiswa aktif 52
Tabel 30. Perbandingan rata-rata variabel Pendidikan_Ortu terhadap variabel IPK dari data mahasiswa aktif 52
Tabel 31. Perbandingan rata-rata variabel Jenis_Kelamin terhadap variabel IPK dari data mahasiswa lulusan 52
Tabel 32. Perbandingan rata-rata variabel Tempat_Lahir terhadap variabel IPK dari data mahasiswa lulusan 53
Tabel 33. Perbandingan rata-rata variabel Jurusan terhadap variabel IPK dari data mahasiswa lulusan 53
Tabel 34. Perbandingan rata-rata variabel Jenjang terhadap variabel IPK dari data mahasiswa lulusan 53
Tabel 35. Perbandingan rata-rata variabel Status terhadap variabel IPK dari data mahasiswa lulusan 54
Tabel 36. Perbandingan rata-rata variabel Predikat terhadap variabel IPK dari data mahasiswa lulusan 54
Tabel 37. Perbandingan rata-rata variabel Lama_Studi terhadap variabel IPK dari data mahasiswa lulusan 54

DAFTAR GAMBAR

Hal

Gambar 1. Tahapan dalam KDD (Han & Kamber 2006) 3
Gambar 2. Bagan Alir Penelitian 11
Gambar 3. Diagram Classification Tree berdasarkan IPK Mahasiswa Aktif 17
Gambar 4. Diagram Classification Tree berdasarkan Alamat_Asal Mahasiswa Aktif 21
Gambar 5. Diagram *Classification Tree* berdasarkan IPK Mahasiswa Lulusan 26
Gambar 6. Diagram Classification Tree berdasarkan Lama Studi Mahasiswa Lulusan 29
Gambar 7. Persentase Independent Variabel Importance berdasarkan IPK Mahasiswa Aktif 32
Gambar 8. Persentase Independent Variabel Importance berdasarkan Alamat_Asal 33
Gambar 9. Persentase Independent Variabel Importance berdasarkan IPK Mahasiswa 34
Gambar 10. Persentase Independent Variabel Importance berdasarkan Lama_Studi 35
Gambar 11. Klasifikasi KNN Data Mahasiswa Aktif berdasarkan variabel IPK 36
Gambar 12. Klasifikasi KNN Data Mahasiswa Aktif berdasarkan variabel Alamat_Asal 37
Gambar 13. Klasifikasi KNN Data Mahasiswa Lulusan berdasarkan variabel IPK 38
Gambar 14. Klasifikasi KNN Data Mahasiswa Lulusan berdasarkan variabel Lama_Studi 39
Gambar 15. Dendogram klasterisasi data mahasiswa aktif 42
Gambar 16. Dendogram klasterisasi data mahasiswa lulusan 44
Gambar 17. Kurva ROC variabel bebas data mahasiswa aktif berdasarkan variabel terikat IPK 48
Gambar 18. Kurva ROC variabel bebas data mahasiswa lulusan berdasarkan variabel terikat IPK 49

DAFTAR LAMPIRAN

Hal

Lampiran 1. Klasifikasi Data Mahasiswa Aktif dengan menggunakan metode *Artificial Neural Network* berdasarkan IPK 59

Lampiran 2. Klasifikasi Data Mahasiswa Aktif dengan menggunakan metode *Artificial Neural Network* berdasarkan Alamat_Asal 60

Lampiran 3. Klasifikasi Data Mahasiswa Lulusan dengan menggunakan metode *Neural Network* berdasarkan IPK..... 61

Lampiran 4. Klasifikasi Data Mahasiswa Lulusan dengan menggunakan metode *Neural Network* berdasarkan Lama_Studi 62

BAB I
PENDAHULUAN

A. Latar Belakang

Program Studi merupakan garda terdepan dalam penyelenggaraan pendidikan dari sebuah Perguruan Tinggi, sehingga sudah seyogyanya setiap penyelenggara Program Studi harus mengetahui kondisi aktual program studinya, salah satunya dengan melakukan evaluasi guna meningkatkan mutu dan efisiensi perguruan tinggi termasuk peningkatan kualitas lulusan.

Fakultas Teknik dengan 10 Program Studi merupakan salah satu fakultas di lingkungan Universitas Negeri Gorontalo yang memiliki jumlah mahasiswa yang cukup banyak. Hal ini terlihat dari peningkatan jumlah calon mahasiswa baru pada setiap tahun ajaran. Permasalahan yang sering terjadi adalah masih banyaknya jumlah mahasiswa yang lulus dengan lama studi melampaui waktu yang telah ditetapkan dengan perolehan Indeks Prestasi Kumulatif (IPK) yang relatif rendah. Faktor-Faktor yang dapat mempengaruhi lama studi dan perolehan IPK mahasiswa dapat berasal dari prestasi akademik mahasiswa maupun non-akademik. Hal ini bisa mempengaruhi mutu lulusan Perguruan Tinggi, karena pada umumnya IPK dan Lama Studi merupakan bagian dari sasaran mutu sebuah Perguruan Tinggi.

Seiring dengan terus bertambahnya jumlah mahasiswa di Fakultas Teknik UNG maka jumlah data kemahasiswaan terus meningkat sehingga terjadi penumpukan data yang belum diolah dengan optimal untuk menggali informasi dan pengetahuan baru melalui pola-pola yang terbentuk dari penumpukan data tersebut. Jumlah data yang terus meningkat ini memerlukan beberapa teknik ataupun metode untuk mengolahnya menjadi sebuah informasi dan pengetahuan yang dapat digunakan sebagai bahan pertimbangan pimpinan dalam proses pengambilan kebijakan dan keputusanjuga sebagai peringatan dini (*early warning*) bagi mahasiswa tertentu yang berdasarkan hasil prediksi dinyatakan berpotensi lulus dengan melampaui ketentuan lama studi ataupun berpotensi lulus dengan perolehan IPK yang rendah.

Salah satu teknik yang dapat digunakan untuk menemukan pola guna membangkitkan informasi dan pengetahuan dari kumpulan data yang tersedia adalah dengan menggunakan teknik Data Mining. Dalam Data Mining terdapat banyak metode yang dapat diterapkan, diantaranya adalah metode yang menjalankan fungsi klasifikasi dan klasterisasi. Fungsi klasifikasi digunakan untuk melakukan prediksi dan fungsi klasterisasi digunakan untuk memetakan data secara otomatis untuk dianalisa lebih lanjut pola keterkaitan antara satu variabel dengan variabel penelitian lainnya dan dipresentasikan dalam bentuk visualisasi.

B. Rumusan Masalah

Dari latar belakang diatas maka dapat dirumuskan suatu permasalahan yaitu "Bagaimana prediksi dan pemetaan data mahasiswa Fakultas Teknik UNG dengan menggunakan pendekatan data mining?".

C. Tujuan Penelitian

Penelitian ini bertujuan untuk membangkitkan informasi dari data akademik kemahasiswaan dengan: (1) Mengklasifikasi data mahasiswa aktif dan lulusan untuk memprediksi kategori IPK dan kategori Lama Studi mahasiswa aktif; (2) Mengklaster data mahasiswa aktif dan lulusan untuk memetakan data mahasiswa berdasarkan sasaran mutu IPK dan Lama Studi.

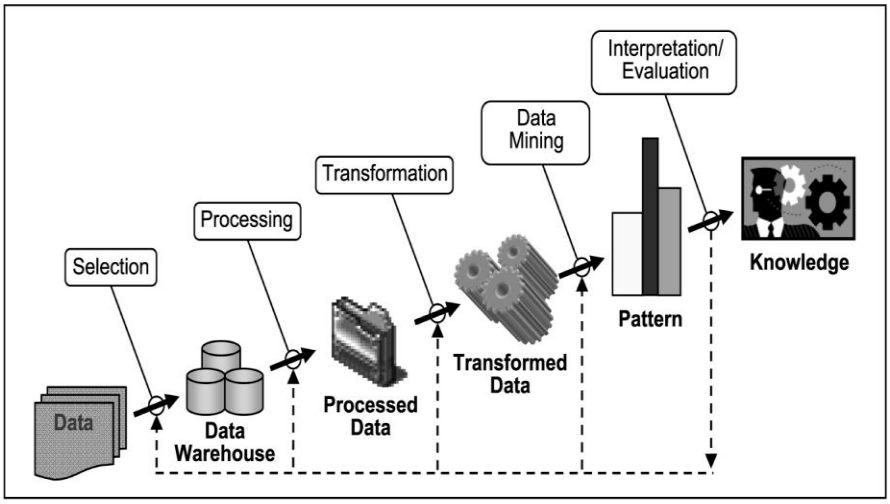
D. Manfaat dan Urgensi Penelitian

Manfaat yang dapat diperoleh melalui hasil penelitian ini yakni diharapkan dapat menghasilkan informasi akademik berupa hasil prediksi dan pemetaan data mahasiswa Fakultas Teknik UNG yang telah diklasfikasi dan diklaster berdasarkan sasaran mutu universitas yakni IPK dan Lama Studi. Urgensinya, informasi potensial yang dihasilkan melalui hasil prediksi dan pemetaan data mahasiswa ini bisa dijadikan sebagai sebuah sistem peringatan dini (*early morning*) bagi mahasiswa yang diprediksi berpotensi lulus dengan lama studi melampau waktu yang telah ditentukan dan dengan perolehan IPK yang rendah. Informasi yang dihasilan dapat menjadi rekomendasi bagi lembaga sebagai bahan pertimbangan bagi pimpinan dalam pengambilan kebijakan dan keputusan.

BAB II
KERANGKA TEORI

A. *Knowledge Discovery in Database* (KDD)

` *Knowledge Discovery in Database* (KDD) adalah proses menentukan informasi yang berguna serta pola-pola yang ada dalam data. Informasi ini terkandung dalam basis data yang berukuran besar yang sebelumnya tidak diketahui dan potensial bermanfaat (Han & Kamber, 2006). Data Mining merupakan salah satu langkah dari serangkaian proses *iterative* KDD. Tahapan proses KDD dapat dilihat pada Gambar 1.



Gambar 1. Tahapan dalam KDD (Han & Kamber 2006)

Tahapan proses KDD terdiri dari :

1. *Data Selection*

   Pada proses ini dilakukan pemilihah himpunan data, menciptakan himpunan data target, atau memfokuskan pada subset variabel (sampel data) dimana penemuan (*discovery*) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas yang terpisah dari basis data operasional.

2. *Pre-Processing* dan*Cleaning Data*

   Pre-Processing dan *Cleaning Data* dilakukan membuang data yang tidak konsisten dan *noise*, duplikasi data, memperbaiki kesalahan data, dan bisa diperkaya dengan data eksternal yang relevan.

3. *Tranformation*

   Proses ini mentransformasikan atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses *mining* dengan cara melakukan peringkasan (agregasi),

4. *Data Mining*

   Proses data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik, metode atau algoritma tertentu sesuai dengan tujuan dari proses KDD secara keseluruhan.

5. *Interpretation / Evaluasi*

   Proses untuk menerjamahkan pola-pola yang dihasilkan dari data mining, Mengevaluasi (menguji) apakah pola atau informasi yang ditemukan bersesuaian atau bertentangan dengan fakta atau hipotesa sebelumnya. Pengetahuan yang diperoleh dari pola-pola yang terbentuk dipresentasikan dalam bentuk visualisasi.

Fungsi-fungsi dalam data mining mengacu pada Larose (2005) yang membaginya ke dalam enam fungsi yaitu (Susanto & Suryadi, 2010) :

1. Fungsi Deksripsi (*description*)

   Fungsi deskripsi adalah cara yang digunakan untuk menggambarkan sekumpulan data secara ringkas. Banyak cara yang digunakan dalam memberikan gambaran secara ringkas bagi sekumpulan data yang besar jumlahnya dan banyak macamnya yaitu deskripsi grafis, deskripsi lokasi dan deskripsi keragaman.

2. Fungsi Estimasi (*estimation*)

   Fungsi estimasi adalah fungsi untuk memperkirakan suatu hal yang sudah ada datanya. Fungsi estimasi terdiri dari dua cara yaitu estimasi titik dan estimasi selang kepercayaan.

3. Fungsi Prediksi (*prediction*)

   Fungsi prediksi adalah memperkirakan hasil dari informasi yang belum diketahui untuk mendapatkan informasi baru yang akan muncul selanjutnya. Cara memprediksi dalam fungsi ini adalah *Regresi Linier*.

4. Fungsi Klasifikasi (*clasification*)

   Fungsi klasifikasi atau menggolongkan suatu data. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma *Mean Vector*, algoritma K-*Nearest Neighbor* (KNN), algoritma ID3, algorimta C.45, algoritma C.50, algoritma *Naive Bayes*, *Neural Network*, dan algoritma lainnya. Fungsi klasifikasi ini banyak juga digunakan untuk melakukan prediksi.

5. Fungsi Klasterisasi (*clustering*)

   Fungsi klasterisasi merupakan fungsi untuk mengelompokkan data. Data yang dikelompokkan disebut objek yang memiliki kemiripan atribut kemudian dikelompokkan ke dalam kelompok yang berbeda. Teknik atau metode klasterisasi dapat diklompokkan menjadi dua kategori besar, yaitu : (1) Metode Hirarki (*Hierarchical Clustering*) dan (2) Metode Metode Partisi (*Partitional Hierarchical*). Metode Hirarki dibagi lagi menjadi dua kategori yaitu *Agglomerative* (*bottom-up*) dan *Divisive* (*top-down*). Algoritma yang digunakan antara lain adalah *Single Lingkage Hierarchical, Compelete Linkage Hierarchical, Average Linkage Hierarchical, Centroid Lingkage Hierarchical* dan algoritma pengembangan lainnya. Sedangkan Algoritma yang banyak digunakan pada Metode Partisi antara lain adalah K-*Means*, K-*Modes*, K-*Medoids*, *Expectation-Maximization* (EM) dan algoritma lainnya.

6. Fungsi Asosiasi (*association*)

   Fungsi asosiasi adalah untk menemukan aturan hubungan (*association rule*) yang mampu mengidentifikasi item-item yang menjadi objek. Algoritma yang dapat digunakan adalah algoritma *Generalizes Association Rules, Quantitative Association Rules, Assynchronus Parallel Mining* dan algoritma lainnya.

Fungsi atau pendekatan data mining yang digunakan dalam penelitian ini adalah klasifikasi dan klasterisasi. Klasifikasi digunakan untuk mengklasifikasi data mahasiswa lulusan dan menemukan pola yang akan digunakan untuk memprediksi Lama Studi IPK mahasiswa aktif. Klasterisasi digunakan untuk memetakan data akademik kemahasiswaan secara otomatis berdasarkan atribut kemiripannya, sehingga apabila ada data baru dapat dipetakan secara otomatis.

B. Data Kategorikal (*Categorical*)

Data kategorikal adalah data non-numerik yang bersifat simbolik, dimana variabelnya minimal memiliki dua relasi (Kantardzic, 2003). Data kategorikal biasanya merupakan data pengamatan sedangkan data numerik merupakan data hasil pengukuran. Data kategorikal diklasifikasikan menjadi dua, yaitu :

1. Data Nominal

   Data nominal yaitu data kategori yang tidak dapat dinyatakan bahwa kategori yang satu lebih baik dari kategori lainnya.

   Contoh : data jenis kelamin mahasiswa (pria,wanita), pekerjaan orang (PNS, Pengusaha, Karyawan, Tani), alamat tinggal (kota, kabupaten), SLTA asal (SMA, SMK), jalur seleksi (SNPT jalur undangan, SNPTN jalur ujian, Lokal, Jurusan), data lama studi (tepat waktu, melampaui waktu).

2. Data Ordinal

   Data ordinal yaitu data kategori yang mempunyai urutan tertentu namun jarak antar kategori sulit untuk dinyatakan sama.

   Contoh : data Indeks Prestasi (rendah, sedang, tinggi), Jumlah mata kuliah yang mengulang (sedikit, banyak), pendidikan orang tua (SD, SMP, SLTA, PT).

C. Teknik Klasifikasi (*Clasification*)

Teknik klasifikasi merupakan pendekatan untuk menjalankan fungsi klasifikasi dalam Data Mining yaitu untuk menggolongkan data. Teknik klasifikasi ini dapat pula digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Beberapa algoritma yang dapat digunakan antara

lain adalah algoritma Decission Tree C.45, algoritma C.50, *Artificial Neural Networks* (ANN), K-*Nearest Neighbor* (KNN), algoritma *Naive Bayes*, *Neural Network* serta algoritma lainnya.

Beberapa penelitian yang menggunakan teknik data mining pada data set Akademik dan Kemahasiswaan telah banyak dilakukan, antara lain adalah penelitian yang dilakukan oleh Al-Radaideh, dkk (2006) yang menganalisa dan mengevaluasi data akademik dengan menggunakan metode pohon keputusan (*decision tree*) untuk mendapatkan kinerja dari siswa yang selanjutnya dapat digunakan untuk mengetahui kualitas perguruan tinggi; Pramudyo (2008) melakukan penelitian tentang klasifikasi mahasiswa baru berdasarkan prediksi Indeks Prestasi Semester (Studi kasus Program Studi Teknik Informatika Universitas Bina Darma Palembang) dengan menggunakan metode *Case Base Reasoning* (CBR); Jananto (2010) melakukan penelitian untuk mengklasifikasi kinerja akademik mahasiswa dengan menggunakan algoritma *Supervised Learning In Quest* (*SLIQ*); Sunjana (2010) dalam penelitian Aplikasi Mining Data Mahasiswa dengan menggunakan metode klasifikasi *Decision Tree*; serta penelitian relevan lainnya yang telah membandingkan beberapa algoritma klasifikasi data mining, seperti penelitian Leidiyana (2011) yang mengkomparasi algoritma C.45, algoritma *Naïve Bayes* dan *Neural Network*. Hasil penelitian yang diperoleh dari hasil pengujian dengan mengukur kinerja ketiga algoritma tersebut diketahui bahwa algoritma C.45 memiliki nilai *accuracy* paling tinggi, diikuti oleh *Neural Network* dan yang terendah adalah *Naïve Bayes*. Hasil penelitian tersebut menunjukkan bahwa metode klasifikasi dengan menggunakan algoritma *decision tree* C.45 dapat melakukan klasifikasi data dengan baik dan dengan tingkat akurasi yang cukup tinggi.

D. Teknik Klasterisasi (*Clustering*)

Klasterisasi merupakan suatu teknik atau metode untuk mengelompokkan sejumlah besar data menjadi suatu bagian-bagian kecil data yang mempunyai atribut kemiripan dalam sifat, letak, ciri atau filter lain yang telah ditentukan Dalam mengelompokkan suatu data menjadi suatu bagian-bagian kecil diperlukan

suatu *centroid of a group data sets* atau sebuah titik yang akan dijadikan nilai parameter utama dari keseluruhan titik data yang teradapat pada kelompok.

Teknik atau metode klasterisasi dapat diklompokkan menjadi dua kategori besar, yaitu : (1) Metode Hirarki (*Hierarchical Clustering*) dan (2) Metode Non-Hirarki/Partisi (*Partitional Hierarchical*). Metode Hirarki dibagi lagi menjadi dua kategori yaitu *Agglomerative* (*bottom-up*) dan *Divisive* (*top-down*). Algoritma yang digunakan antara lain adalah *Single Lingkage Hierarchical, Compelete Linkage Hierarchical, Average Linkage Hierarchical, Centroid Lingkage Hierarchical* dan algoritma pengembangan lainnya. Sedangkan Algoritma yang banyak digunakan pada Metode Partisi antara lain adalah K-*Means*, K-Modes, K-*Medoids*, *Expectation-Maximization* (EM) dan lain sebagainya.

Data yang digunakan dalam penelitian ini adalah data kategorikal. Dalam penelitiannya, Fahim dkk (2008), Gunawan (2009) dan Karlita (2011) menyatakan bahwa Algoritma K-*Modes* merupakan metode pengembangan dari K-*Means* yang mampu mengelompokkan data kategorikal dan menghasilkan klaster yang lebih stabil dengan waktu komputasi yang lebih singkat daripada metode K-*Means.* Martiana dkk (2009) dalam penelitiannya pernah mengkomparasikan penggunaan metode *Aglomerative Hierarchical* (*Single Linkage Hierarchical dan Centroid Linkage Hierarchical*) dengan metode *Partitional Clustering* yakni K-*Means,* untuk mengklaster bidang tugas akhir mahasiswa Teknik Informatika PENS berdasarkan nilai. Hasil percobaan dalam penelitian tersebut menunjukkan bahwa metode *Centroid Linkage Hierarchical* memiliki nilai varians yang paling kecil, yang menandakan bahwa metode tersebut menghasilkan *cluster* paling baik dibandingkan dengan metode *Single Linkage Hierarchical,* dan K-*Means* dalam kasus tersebut.

Expectation-Maximization (EM) merupakan algoritma *clustering* yang masuk dalam kategori *Partitional Clustering*, menggunakan perhitungan probabilitas bukan perhitungan jarak seperti umumnya algritma *clustering*. Budiarti dkk (2006) dalam penelitian yang berjudul Studi Karakteristik Kelulusan Peserta Didik dengan Menggunakan Teknik *Clustering* (algoritma EM), mengemukakan bahwa algoritma EM merupakan metode klasterisasi data

numerik yang juga mampu mengelompokkan data kategorikal (data numerik yang telah didiskretisasi ke dalam bentuk data kategorikal), meskipun kinerja algoritma EM lebih stabil pada data numerik, terutama data distribusi eksponensial, seperti yang dikemukakan oleh Balakrishnan & Ling (2012).

BAB III
METODE PENELITIAN

A. Lokasi dan Waktu Penelitian

1. Lokasi Penelitian

Penelitian ini dilaksanakan di laboratorium *Rekayasa Engineering* Jurusan Tekni Informatika, Fakultas Teknik Universitas Negeri Gorontalo. Penetapan lokasi penelitian ini didasarkan pada pertimbangan akses data lebih mudah karena lokasi penelitian berada pada lingkup kerja peneliti sehingga proses pengumpulan data, pengolahan, analisa hingga implementasi relatif lebih mudah dilakukan.

2. Waktu Pelaksanaan Penelitian

Penelitian diselesaikan selama 6 (enam) bulan yaitu mulai bulan April sampai dengan Oktober 2012.

B. Materi Penelitian

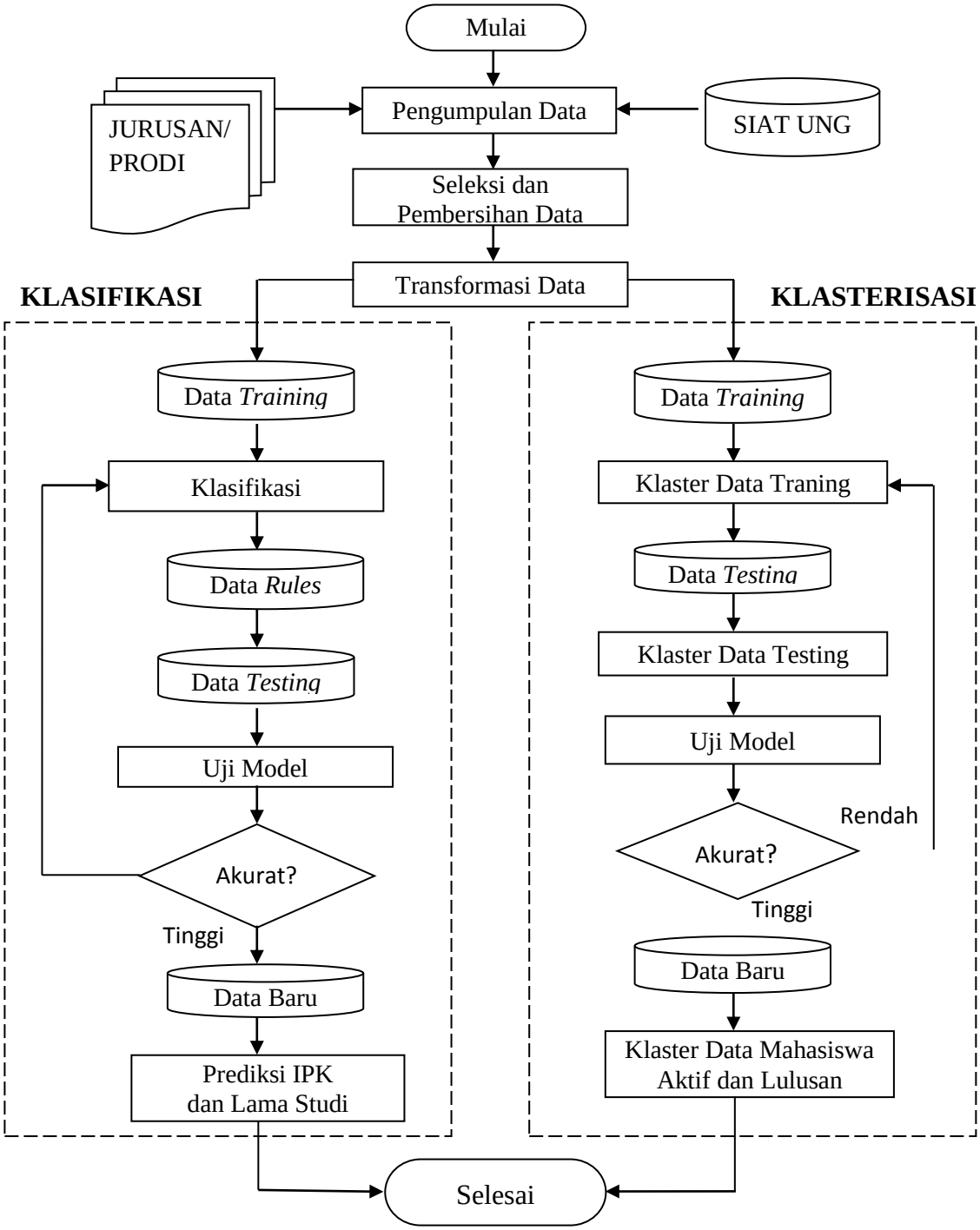
Bahan penelitian utama adalah data-data yang dikumpulkan dari setiap Jurusan/Program Studi yang ada di Fakultas TeknikUNG dan juga berasal dari database Sistem Informasi Akademik Terpadu (SIAT) UNG. Data-data tersebut meliputi data kemahasiswaan yang bersifat akademik dan non akademik.

C. Alat Penelitian

Alat yang digunakan dalam penelitian ini adalah Perconal Computer (PC) dengan spesifikasi prosesor Intel (R) Core ™2, Mainboard *pc CHIPS*, Hardisk 160 GB, Memori 1024 MB RAM, keyboard, optical mouse, monitor *Acer* 17', Printer Canon IP 2720, DVD-RQ. Sedangkan untuk softwarenya adalah Sistem Operasi Windows XP Professional, Pengolah Kata Microsoft Office Word 2007, Microsoft Office Excel 2007, *tools* SPSS ver 20, *tools* Matlab ver 7 dan Rapid Miner ver 5.0

D. Alur Penelitian

Tahapan proses penelitian ini digambarkan dalam suatu bagan alir seperti pada Gambar 2.



Gambar 2. Bagan Alir Penelitian

Secara detail tahapan penelitian diuraikan sebagai berikut :

1. Mengumpulkan Data

Proses ini ditujukan untuk mengumpulkan data mahasiswa di semua jurusan pada Fakultas Teknik, berupa data akademik dan data non-akademik, baik data mahasiswa aktif maupun data mahasiswa lulusan. Data mahasiswa lulusan dimaksudkan untuk mencari dan membentuk pola perolehan IPK dan Lama Studi yang akan digunakan untuk memprediksi perolehan IPK dan Lama Studi dari mahasiswa aktif, sebagai peringatan dini (*early warning*). Data lulusan juga dibutuhkan dalam proses pengklasteran guna mendapatkan pemetaan profil lulusan dan mahasiswa aktif berdasarkan sasaran mutu IPK dan Lama Studi.

Data akademik dan non_akademik yang dikumpulkan adalah dipisahkan menjadi data mahasiswa aktif dan data mahasiswa lulusan. Data mahasiswa aktif terdiri dari jurusan, jalur masuk, jenis kelamin, alamat asal, asal sekolah, IPK, pekerjaan orang tua, pendidikan terakhir orang tua dan penghasilan orang tua. Data mahasiswa lulusan terdiri dari data jurusan, jenjang, status, jenis kelamin, predikat, lama studi dan IPK. Data mahasiswa aktif yang digunakan dalam penelitian ini dari tahun angkatan 2008 sampai dengan tahun angkatan 2011, sedangkan data lulusan dari tahun lulus 2003 sampai dengan tahun lulus 2012.

2. Seleksi dan Pembersihan Data

Pada tahapan ini dilakukan pemilihan himpunan data, menciptakan himpunan data target dan atau memfokuskan pada subset variabel (sampel data) yang telah dikumpulkan pada tahap pengumpulan data. Pada tahap ini pula dilakukan pemusnahan data yang tidak konsisten dan mengandung *noise*, duplikasi data dan memperbaiki kesalahan dan ketidaklengkapan data, juga bisa diperkaya dengan data eksternal yang relevan.

3. Transformasi Data

Tahapan ini mentransformasikan atau menggabungkan data ke dalam format yang lebih tepat untuk melakukan proses *mining* dengan cara melakukan peringkasan (*agregation*) disertai dengan proses pendeskripsian data. Proses transformasi data dilakukan untuk data yang bersifat kategorikal.

4. Klasifikasi dan Klasterisasi

Ada kesamaan tahapan pada proses klasifikasi dan klasterisasi. Pada kedua proses ini data mahasiswa yang sudah ditransformasi dan siap untuk di *mining*, dipisahkan menjadi data *training*, data *testing* dan data baru. Hanya saja untuk proses klasifikasi, data *training* adalah data mahasiswa lulusan yang memiliki data *target*, yakni IPK dan Lama Studi. Data *testing* adalah data mahasiswa lulusan tapi tidak mengikutsertakan data *target*. Hal ini dimaksudkan untuk menguji hasil prediksi apakah dapat melakukan prediksi dengan tingkat generalisasi dan akurasi yang tinggi, sebelum diterapkan pada data baru, yakni data mahasiswa aktif yang belum memiliki informasi IPK akhir dan Lama Studi. Sedangkan pada proses klasterisasi, data *training* adalah data mahasiswa aktif dan lulusan, adapun data *testing* merupakan sebagian data mahasiswa yang tidak diproses dalam pengklasteran data *training*, untuk menguji performansi dan akurasi, sebelum diterapkan pada data baru. Selain itu perbedaan mendasar dari kedua proses ini adalah algoritma yang digunakan pada proses klasifikasi dan klasterisasi.

E. Jadwal Pelaksanaa Penelitian

Tabel 1. Jadwal Pelaksanaan Penelitian

Kegiatan	Bulan ke					
	1	2	3	4	5	6
Pegumpulan Data	■	■				
Seleksi dan Pembersihan Data		■	■			
Transformasi Data		■	■			
Klasifikasi dan Klasterisasi Data			■	■	■	
Pembuatan Laporan				■	■	■

BAB IV
HASIL PENELITIAN DAN PEMBAHASAN

A. PRAPROSES DATA

Setelah data dikumpulkan maka dilanjutkan dengan praproses data dengan melakukan seleksi dan pembersihan terhadap data mahasiswa serta proses tranformasi data agar data siap untuk di *mining*. Data hasil seleksi berupa data dengan atribut yang diperlukan serta membuang data yang tidak dibutuhkan, seperti pada data mahasiswa aktif, atribut yang dibuang adalah tanggal lahir, usia, angkatan, semester, nama ayah, nama ibu, jumlah tanggungan orang tua, tahun lulus SMA, dan kode pos. Adapun atribut yang terpilih untuk data mahasiswa aktif adalah NIM, nama, jurusan, jenis kelamin, jalur masuk, alamat asal, asal sekolah, IPK, pendidikan terakhir orang tua, pekerjaan orang tua dan penghasilan orang tua. Untuk data mahasiswa lulusan, atribut yang terpilih adalah NIM, nama, jurusan, jenis kelamin, jenjang, status, predikat, lama studi dan IPK.

Data hasil seleksi kemudian dibersihkan dari data yang mengandung *noise*, duplikasi data, data yang tidak konsisten, memperbaiki kesalahan dan ketidaklengkapan data. Untuk data yang tidak lengkap atau tidak tersedia (*missing value*) maka dilakukan proses pemusnahan pada sejumlah *record*, akan tetapi jika persentase data yang tidak tersedia (*missing value*) cukup besar maka dilakukan proses transformasi data dengan memasukkan nilai rata-rata (*mean*) dari masing-masing variabel tersebut pada *cell* yang mengandung *missing value*. Data mahasiswa yang mengandung *missing value* adalah data mahasiswa aktif pada atribut atau variabel asal sekolah, penghasilan orang tua, pekerjaan orang tua dan pendidikan terakhir orang tua.

Transformasi data dilakukan pula pada pengkategorian data untuk setiap atribut atau variabel yang digunakan. Hasil pengkategorian serta frekuensi data sebagai bagian dari proses pendeskripsian data ditunjukkan oleh Tabel 2 dan Tabel 3.

Tabel 2. Kategori Data Mahasiswa Aktif

VARIABEL			FREKUENSI	
Nama Variabel	**Kategori**	**Nama Kategori**	**Jumlah**	**%**
Jurusan	1	Elektro	190	13,6
	2	Sipil	332	23,8
	3	Arsitektur	124	8,9
	4	Industri	71	5,1
	5	Informatika	596	42,7
	6	Kriya	82	5,9
Jalur	1	Undangan	204	14,6
	2	Ujian	219	15,7
	3	Mandiri	694	49,7
	4	Jurusan	278	19,9
Jenis_Kelamin	1	Laki-Laki	979	70,2
	2	Perempuan	416	29,8
Alamat_Asal	1	Kota Gorontalo	513	36,8
	2	Kabupaten Gorontalo	290	20,8
	3	Bone Bolango	175	12,5
	4	Boalemo	32	2,3
	5	Pohuwato	46	3,3
	6	Gorut	32	2,3
	7	Luar Daerah Gorontalo	307	22,0
Asal_Sekolah	1	SMA	580	41,6
	2	SMK	666	47,7
	3	Madrasah	46	3,3
	4	Penjenjangan	50	3,6
		Missing Value	53	3,8
IPK	1	Rendah	857	61,4
	2	Sedang	504	36,1
	3	Tinggi	34	2,4
Pendidikan_Ortu	1	Tidak Sekolah	110	7,88
	2	Pendidikan Dasar dan Menengah	779	55,85
	3	Pendidikan Tinggi	282	20,21
		Missing Value	224	16,06
Pekerjaan_Ortu	1	PNS	112	8,0
	2	Wiraswasta	370	26,5
	3	TNI/POLRI	23	1,6
	4	Petani/Nelayan	239	17,1
	5	Tidak Bekerja	14	1,0
	6	Lainnya	278	19,9
		Missing Value	359	25,7
Penghasilan_Ortu	1	Dibawah 1 juta	654	46,9
	2	1 – 3 juta	300	21,5
	3	3 – 5 juta	73	5,2
	4	5 – 10 juta	10	7
		Missing Value	358	25,7

Tabel 3. Kategori Data Mahasiswa Lulusan

VARIABEL			FREKUENSI	
Nama Variabel	**Kategori**	**Nama Kategori**	**Jumlah**	**%**
Jurusan	1	Sipil	225	16,3
	2	Informatika	733	53,2
	3	Elektro	135	9,8
	4	Kriya	118	8,6
	5	Arsitektur	100	7,3
	6	Industri	67	4,9
Jenjang	1	D3	1314	95,4
	2	S1	64	4,6
Status	3	Non Kependidikan	1314	95,4
	4	Kependidikan	64	4,6
Jenis_Kelamin	1	Laki-Laki	651	47,2
	2	Perempuan	727	52,8
Tempat_Lahir	1	Kota Gorontalo	605	43,9
	2	Kabupaten Gorontalo	271	19,7
	3	Bone Bolango	174	12,6
	4	Boalemo	21	1,5
	5	Pohuwato	26	1,9
	6	Gorut	24	1,7
	7	Luar Daerah Gorontalo	257	18,7
Predikat	1	Memuaskan	276	20,0
	2	Sangat Memuaskan	1071	77,7
	3	Terpuji	31	2,2
Lama_Studi	1	Tepat Waktu	189	13,7
	2	Melampaui	1189	86,3
IPK	1	Rendah	259	18,8
	2	Sedang	1066	77,4
	3	Tinggi	53	3,8

B. KLASIFIKASI

Data yang diklasifikasi adalah data mahasiswa aktif dan lulusan. Metode-metode yang digunakan dalam proses klasifikasi adalah metode *Decission*/*Classification Tree, Artificial Neural Network* dan K-*Neirest Neighbour*.

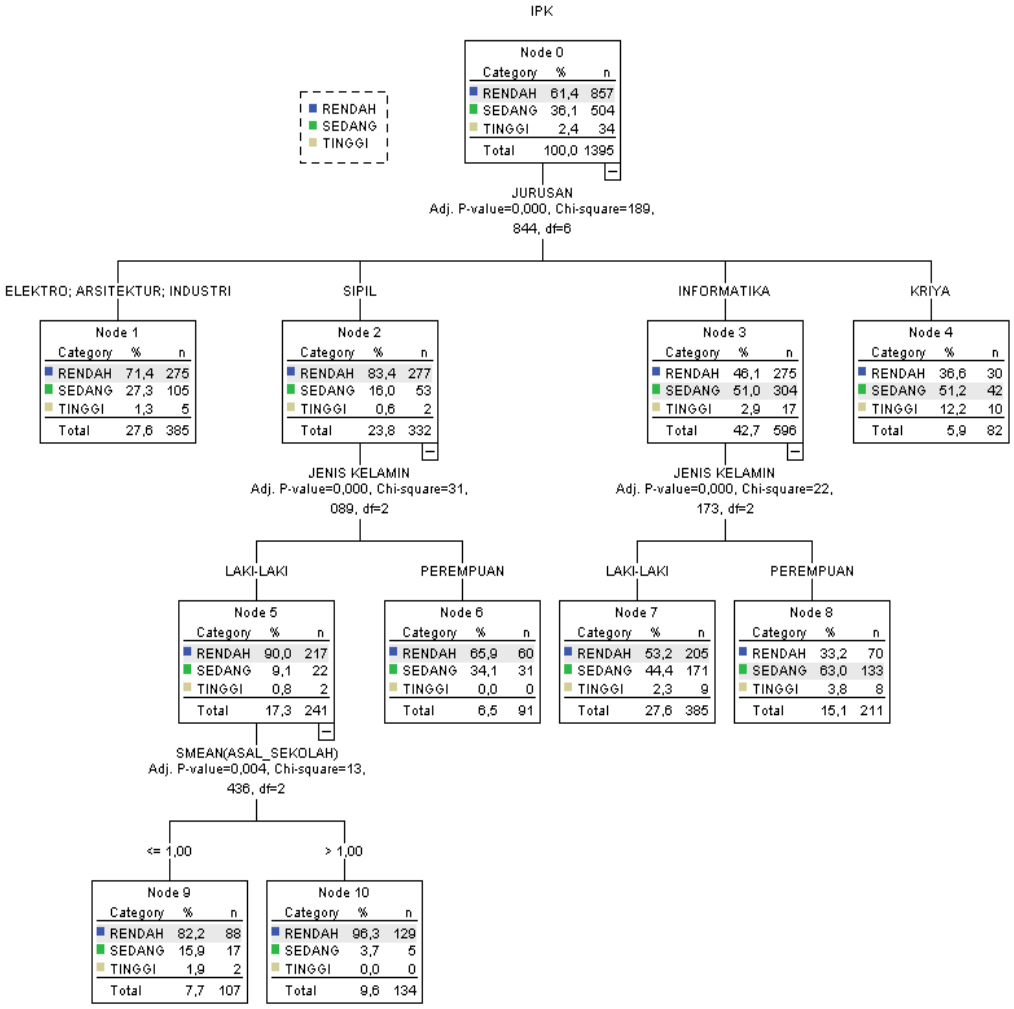
1. Decission Tree

Decission Tree (*Classification Tree*) merupakan salah satu metode yang dapat digunakan untuk mengklasifikasi data dengan pendekatan dependensi, dimana klasifikasi-klasifikasi dihasilkan karena adanya hubungan antara variabel terikat (*dependent variable*) dengan sejumlah variabel bebas (*independent variable*). Salah satu metode dependensi yang digunakan untuk membuat klasifikasi adalah analisis CHAID (CHi-squared Automatic Interaction Detection

analysis). Penelitan ini mengklasifikasikan data Mahasiswa aktif dan lulus dari Fakultas Teknik Universitas Negeri Gorontalo berdasarkan variabel terikat tertentu.

a. Klasifikasi data mahasiswa aktif berdasarkan variabel IPK dan Alamat_Asal

Klasifikasi data mahasiswa aktif dengan menggunakan metode pohon keputusan (*Decision Tree*) yang dilakukan berdasarkan variabel IPK terlihat pada Gambar 3.



Gambar 3. Diagram Classification Tree berdasarkan IPK Mahasiswa Aktif

Diagram pohon hasil analisis *classification tree* pada Gambar 3 menerangkan bahwa pada node teratas diketahui jumlah total mahasiswa aktif adalah 1395, terdiri dari 857 mahasiswa (61,4%) dengan IPK kategori **rendah**, 504 mahasiswa (36,1%) IPK kategori **sedang** dan 34 mahasiswa (2,4%) dengan IPK kategori **tinggi**. Tahap pertama pada analisis ini adalah tahap penggabungan, dimana dalam penelitian ini variabel Jurusan dibagi menjadi 6 kategori, yaitu : (1) Elektro; (2) Sipil; (3) Arsitektur; (4) Industri; (5) Informatika dan (6) Kriya. Setelah melalui analisis *classification tree*, variabel ini kemudian diringkas menjadi 4 kategori, seperti yang terlihat pada diagram pohon kedalaman yang ke-1, yaitu kategori Jurusan : (1) Elektro, Arsitektur dan Industri; (2) Sipil; (3) Informatika; dan (4) Kriya. Hal ini berarti kategori Elektro, Arsitektur dan Industri memenuhi syarat kesignifikanan *chi-square* untuk digabung menjadi satu kategori tunggal.

Penggabungan kategori juga terjadi pada variabel Asal_Sekolah, dimana dalam penelitian ini variabel Asal_Sekolah dibagi menjadi 4 kategori, yaitu (1) SMA; (2) SMK; (3) Madrasah; dan (4) Penjenjangan, yang selanjutnya diringkas menjadi 2 kategori, seperti yang terlihat pada diagram pohon kedalaman yang ke-3, yaitu kategori (1) SMA dan (2) SMK, Madrasah dan Penjenjangan. Hal ini berarti kategori SMK, Madrasah dan Penjenjangan memenuhi syarat kesignifikanan *chi-square* untuk digabung menjadi satu kategori tunggal berdasarkan analisis *classification tree* yang ada.

Dari *model summary* pada Tabel 5 dapat diketahui bahwa dalam penelitian ini sebenarnya terdapat 8 variabel bebas (Jenis_Kelamin, Jurusan, Jalur, Pendidikan_Ortu, Pekerjaan_Ortu, Penghasilan_Ortu, Asal_Sekolah dan Alamat_Asal), kemudian hasil analisis menunjukkan bahwa hanya ada 3 variabel bebas yang signifikan terhadap variabel terikatnya (IPK), yaitu variabel Jurusan, Jenis_Kelamin dan Asal_Sekolah, sehingga ada 5 variabel bebas yang tersisa dan tidak dianggap mempunyai hubungan dengan variabel terikat, yaitu variabel Jalur, Pendidikan_Ortu, Pekerjaan_Ortu, Penghasilan_Ortu, dan Alamat_Asal. Nilai *p-value* dan nilai uji *chi-square* dari masing-masing variabel bebas yang dianggap

mempunyai hubungan dengan variabel terikatnya dapat diringkas dalam Tabel 4 berikut.

Tabel 4. Nilai signifikansi variabel bebas terhadap variabel IPK mahasiswa aktif

Variabel Bebas	Nilai *p-value*	Nilai *chi-square*
Jurusan	0,000	189
Jenis_Kelamin	0,000	31 ; 22
Asal_Sekolah	0,004	12

Dari Tabel 5 dapat diketahui bahwa apabila dilakukan pengambilan keputusan berdasarkan nilai *p-value,* dimana ketiga nilai tersebut kurang dari α=0,05, yaitu 0,000 dan 0,004. Maka dapat ditarik kesimpulan bahwa keputusan uji *chi-square* adalah Tolak H_0, yang berarti bahwa terdapat hubungan antara ketiga variabel bebas tersebut (Jurusan, Jenis_Kelamin dan Asal_Sekolah) dengan variabel terikatnya, yaitu IPK.

Pohon klasifikasi pada Gambar 3 menunjukkan bahwa Mahasiswa Fakultas Teknik Universitas Negeri Gorontalo berdasarkan IPK dibagi menjadi 7 klasifikasi, yaitu :

Tabel 5. Klasifikasi dan Karakteristik Data Mahasiswa Aktif berdasarkan variabel IPK

Klasifikasi	Karakteristik
Ke-1	Mahasiswa dengan Jurusan Sipil, Jenis Kelamin Laki-Laki, dan Asal Sekolah SMA
Ke-2	Mahasiswa dengan Jurusan Sipil, Jenis Kelamin Laki-Laki, dan Asal Sekolah SMK, Madrasah dan Penjenjangan
Ke-3	Mahasiswa dengan Jurusan Sipil, Jenis Kelamin Perempuan
Ke-4	Mahasiswa dengan Jurusan Informatika, Jenis Kelamin Laki-Laki
Ke-5	Mahasiswa dengan Jurusan Informatika, Jenis Kelamin Perempuan
Ke-6	Mahasiswa dengan Jurusan Elektro, Arsitektur, Industri
Ke-7	Mahasiswa dengan Jurusan Kriya

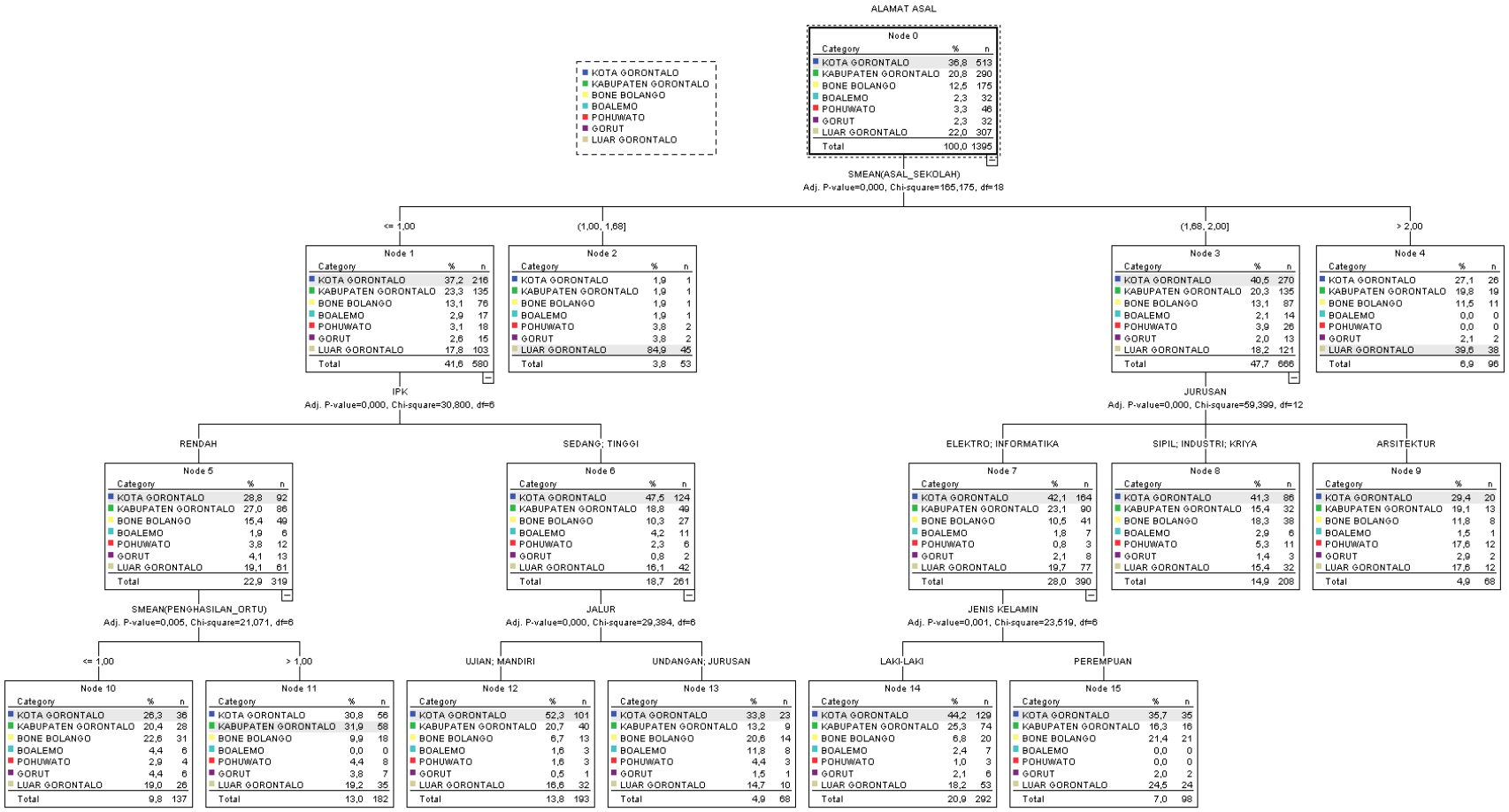
Dari ketujuh klasifikasi yang terbentuk, dapat ditabulasi jumlah masing-masing mahasiswa yang memiliki IPK dengan kategori rendah, sedang dan tinggi dalam Tabel 6 berikut.

Tabel 6. Jumlah Mahasiswa Aktif setiap Klasifikasi berdasarkan Kategori IPK

Klasifikasi	IPK		
	Rendah	Sedang	Tinggi
Ke-1	88	17	2
Ke-2	129	5	0
Ke-3	60	31	0
Ke-4	205	**171**	9
Ke-5	70	133	8
Ke-6	**275**	105	5
Ke-7	30	42	**10**

Hasil tabulasi IPK pada Tabel 6 di atas menunjukkan bahwa perolehan IPK dengan kategori **rendah** yang paling banyak dihasilkan oleh klasifikasi mahasiswa ke-6, yaitu karakteristik Mahasiswa Jurusan Elektro, Arsitektur, Industri. IPK dengan kategori **sedang** yang paling banyak dihasilkan oleh klasifikasi mahasiswa ke-4, yaitu karakteristik Mahasiswa Jurusan Informatika, Jenis Kelamin Laki-Laki. IPK dengan kategori **tinggi** yang paling banyak dihasilkan oleh klasifikasi mahasiswa ke-7, yaitu karakteristik Mahasiswa Jurusan Kriya.

Gambar 4 di bawah ini merupakan hasil klasifikasi data mahasiswa aktif Fakultas Teknik Universitas Negeri Gorontalo dengan menggunakan metode *Decision Tree,* yang diklasifikasikan berdasarkan variabel Alamat_Asal.



Gambar 4. Diagram Classification Tree berdasarkan Alamat_Asal Mahasiswa Aktif

Diagram pohon hasil analisis *classification tree* pada Gambar 4 menerangkan bahwa pada node teratas diketahui jumlah total mahasiswa aktif adalah 1395, terdiri dari 513 mahasiswa (36,8%) beralamat asal dari Kota Gorontalo, 290 mahasiswa (20,8%) beralamat asal dari Kabupaten Gorontalo, 175 mahasiswa (12,5%) beralamat asal dari Kabupaten Bone Bolango, 32 mahasiswa (2,3%) beralamat asal dari Kabupaten Boalemo, 46 mahasiswa (3,3%) beralamat asal dari Kabupaten Pohuwato, 32 mahasiswa (2,3%) beralamat asal dari Kabupaten Gorontalo Utara (Gorut) dan 307 mahasiswa (22,0%) beralamat asal dari Luar daerah Provinsi Gorontalo.

Tahap pertama pada analisis ini adalah tahap penggabungan, dimana dalam penelitian ini variabel Asal_Sekolah dibagi menjadi 4 kategori, yaitu (1) SMA; (2) SMK; (3) Madrasah; dan (4) Penjenjangan, yang selanjutnya diringkas menjadi 3 kategori, seperti yang terlihat pada diagram pohon kedalaman yang ke-1, yaitu kategori (1) SMA; (2) SMK; dan (3) Madrasah dan Penjenjangan. Hal ini berarti kategori Madrasah dan Penjenjangan memenuhi syarat kesignifikanan *chi-square* untuk digabung menjadi satu kategori tunggal berdasarkan analisis *classification tree* yang ada.

Penggabungan kategori juga terjadi pada variabel Jurusan, IPK, Penghasilan_Ortu dan Jalur. Pada penelitian ini variabel Jurusan dibagi menjadi 6 kategori yaitu : (1) Elektro; (2) Sipil; (3) Arsitektur; (4) Industri; (5) Informatika dan (6) Kriya, setelah melalui analisis *classification tree*, variabel ini kemudian diringkas menjadi 3 kategori, seperti yang terlihat pada diagram pohon kedalaman yang ke-2, yaitu kategori Jurusan : (1) Elektro dan Informatika; (2) Sipil, Industri dan Kriya; dan (3) Arsitektur. Variabel IPK dibagi menjadi 3 kategori yaitu (1) Rendah; (2) Sedang; dan (3) Tinggi, setelah melalui analisis *classification tree*, variabel ini kemudian diringkas menjadi 2 kategori, seperti yang terlihat pada diagram pohon kedalaman yang ke-2, yaitu kategori (1) Rendah dan (2) Sedang, Tinggi. Variabel Penghasilan_Ortu dibagi menjadi 4 kategori yaitu (1) Dibawah satu juta; (2) satu hingga tiga juta; (3) tiga hingga lima juta; dan (4) lima hingga 10 juta, setelah melalui analisis *classification tree*, variabel ini kemudian diringkas menjadi 2 kategori, seperti yang terlihat pada diagram pohon kedalaman

yang ke-3, yaitu kategori (1) Dibawah satu juta dan (2) Diatas satu juta. Variabel Jalur masuk Perguruan Tinggi dibagi menjadi 4 kategori yaitu (1) Undangan; (2) Ujian; (3) Mandiri; dan (4) Jurusan. Setelah melalui analisis *classification tree*, variabel ini kemudian diringkas menjadi 2 kategori, seperti yang terlihat pada diagram pohon kedalaman yang ke-3, yaitu kategori (1) Ujian, Mandiri dan (2) Undangan, Jurusan.

Dari Model Summary pada Tabel 7 dapat diketahui bahwa dalam penelitian ini sebenarnya terdapat 8 variabel bebas (Jenis_Kelamin, Jurusan, Jalur, Pendidikan_Ortu, Pekerjaan_Ortu, Penghasilan_Ortu, Asal_Sekolah dan IPK), kemudian hasil analisis menunjukkan bahwa hanya ada 6 variabel bebas yang signifikan terhadap variabel terikatnya (Alamat_Asal), yaitu variabel Asal_Sekolah, IPK, Jurusan, Penghasilan_Ortu, Jalur dan Jenis_Kelamin. Sehingga ada 2 variabel bebas yang tersisa dan tidak dianggap mempunyai hubungan dengan variabel terikat, yaitu variabel Pendidikan_Ortu dan Pekerjaan_Ortu. Nilai *p-value* dan nilai uji *chi-square* dari masing-masing variabel bebas yang dianggap mempunyai hubungan dengan variabel terikatnya dapat diringkas dalam Tabel 8 berikut.

Tabel 7. Nilai signifikansi variabel bebas terhadap variabel Alamat_Asal mahasiswa aktif

Variabel Bebas	**Nilai *p-value***	**Nilai *chi-square***
Asal_Sekolah	0,000	165
IPK	0,000	30,800
Jurusan	0,000	59,399
Penghasilan_Ortu	0,005	21,071
Jalur	0,000	29,384
Jenis_Kelamin	0,001	23,519

Dari Tabel 7 dapat diketahui bahwa apabila dilakukan pengambilan keputusan berdasarkan nilai *p-value,* dimana ketiga nilai tersebut kurang dari α=0,05, yaitu 0,000, 0,001 dan 0,005. Maka dapat ditarik kesimpulan bahwa keputusan uji *chi-square* adalah Tolak H_0, yang berarti bahwa terdapat hubungan antara keenam variabel bebas tersebut (Asal_Sekolah, IPK, Jurusan,

Penghasilan_Ortu, Jalur dan Jenis_Kelamin) dengan variabel terikatnya, yaitu Alamat_Asal.

Pohon klasifikasi pada Gambar 4 menunjukkan bahwa Mahasiswa Fakultas Teknik Universitas Negeri Gorontalo berdasarkan Alamat Asal dibagi menjadi 10 klasifikasi, yaitu :

Tabel 8. Klasifikasi dan Karakteristik Data Mahasiswa Aktif berdasarkan variabel Alamat_Asal

Klasifikasi	Karakteristik
Ke-1	Mahasiswa dengan Asal Sekolah SMA, IPK Rendah, Penghasilan Orang Tua dibawah 1 Juta
Ke-2	Mahasiswa dengan Asal Sekolah SMA, IPK Rendah, Penghasilan Orang Tua di atas 1 Juta
Ke-3	Mahasiswa dengan Asal Sekolah SMA, IPK Sedang dan Tinggi, Jalur Masuk Ujian dan Mandiri
Ke-4	Mahasiswa dengan Asal Sekolah SMA, IPK Sedang dan Tinggi, Jalur Masuk Undangan dan Jurusan
Ke-5	Mahasiswa dengan Asal Sekolah SMK
Ke-6	Mahasiswa dengan Asal Sekolah SMK, Jurusan Elektro, Informatika, Jenis Kelamin Laki-Laki
Ke-7	Mahasiswa dengan Asal Sekolah SMK, Jurusan Elektro, Informatika, Jenis Kelamin Perempuan
Ke-8	Mahasiswa dengan Asal Sekolah SMK, Jurusan Sipil, Industri, Kriya
Ke-9	Mahasiswa dengan Asal Sekolah SMK, Jurusan Arsitektur
Ke-10	Mahasiswa dengan Asal Sekolah Madrasah dan Penjenjangan

Dari kesepuluh klasifikasi yang terbentuk, dapat ditabulasi jumlah masing-masing mahasiswa yang memiliki IPK dengan kategori rendah, sedang dan tinggi dalam Tabel 9 berikut.

Tabel 9. Jumlah Mahasiswa Aktif setiap Klasifikasi berdasarkan Kategori Alamat_Asal

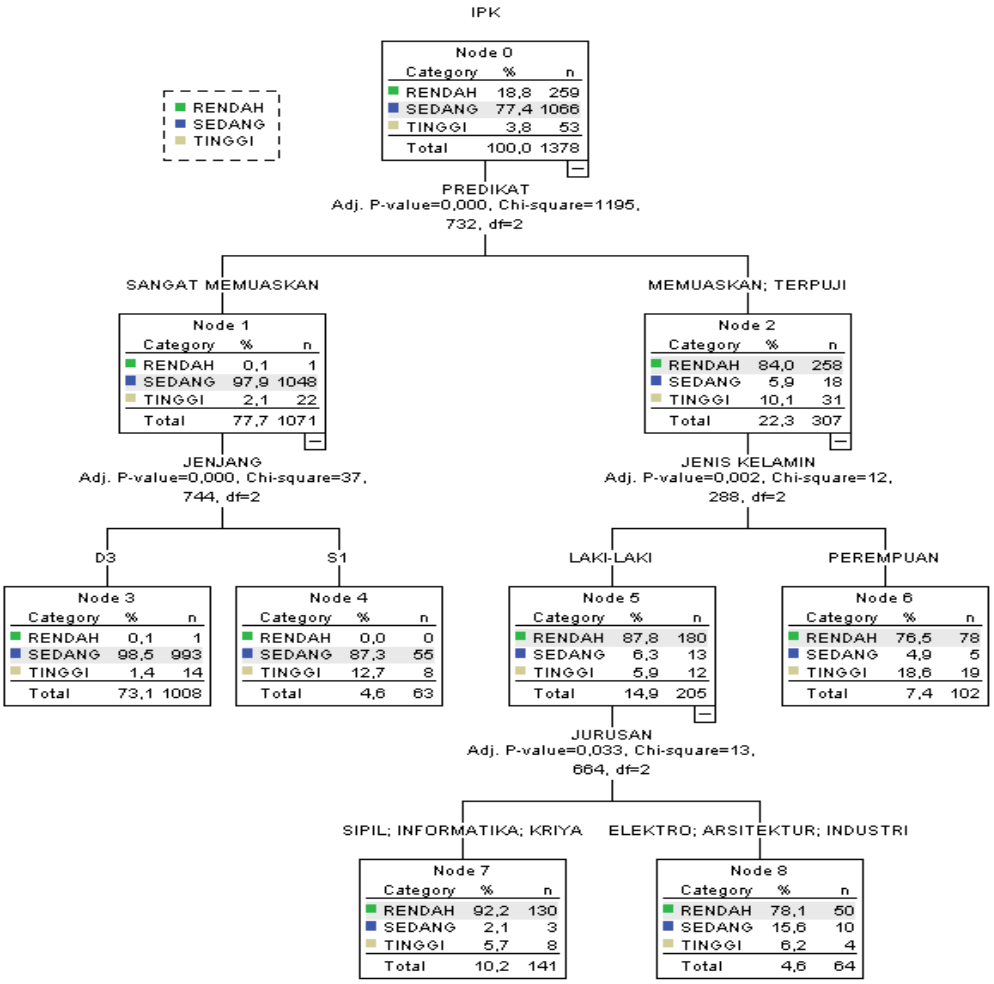
Klasifikasi	Alamat Asal						
	Kota Gorontalo	Kabupaten Gorontalo	Kabupaten Bone Bolango	Kabupaten Boalemo	Kabupaten Pohuwato	Kabupaten Gorut	Luar Daerah Gorontalo
Ke-1	36	28	31	6	4	6	26
Ke-2	56	58	18	0	8	**7**	35
Ke-3	101	40	13	3	3	1	32
Ke-4	23	9	14	**8**	3	1	10
Ke-5	1	1	1	1	2	2	45

Ke-6	**129**	**74**	20	7	3	6	**53**
Ke-7	35	16	21	0	0	2	24
Ke-8	86	32	**38**	6	11	3	32
Ke-9	20	13	8	1	**12**	2	12
Ke-10	26	19	11	0	0	2	38

Hasil tabulasi IPK pada Tabel 9 di atas menunjukkan bahwa jumlah mahasiswa Fakultas Teknik yang terbanyak berdasarkan alamat asal **Kota Gorontalo, Kabupaten Gorontalo** dan **Luar daerah Gorontalo**, berada pada klasifikasi ke-6 dengan karakteristik Mahasiswa Asal Sekolah SMK, Jurusan Elektro, Informatika, dan Jenis Kelamin Laki-Laki. Asal alamat dari **Kabupaten Bone Bolango** berada pada klasifikasi ke-8 dengan karakteristik Mahasiswa dengan Asal Sekolah SMK, Jurusan Sipil, Industri, Kriya. Asal alamat dari **Kabupaten Boalemo** berada pada klasifikasi ke-4 dengan karakteristik Mahasiswa Asal Sekolah SMA, IPK Sedang dan Tinggi, Jalur Masuk Undangan dan Jurusan. Asal alamat dari **Kabupaten Pohuwato** berada pada klasifikasi ke-9 dengan karakteristik Mahasiswa Asal Sekolah SMK dan Jurusan Arsitektur. Asal alamat dari **Kabupaten Gorut** berada pada klasifikasi ke-2 dengan karakteristik Mahasiswa Asal Sekolah SMA, IPK Rendah, Penghasilan Orang Tua di atas 1 Juta.

b. Hasil klasifikasi data lulusan berdasarkan variabel IPK dan Lama Studi

Klasifikasi data mahasiswa lulusan dengan menggunakan metode pohon keputusan (*Decision Tree*) berdasarkan IPK terlihat pada Gambar 5.



Gambar 5. Diagram *Classification Tree* berdasarkan IPK Mahasiswa Lulusan

Diagram pohon hasil analisis *classification tree* pada Gambar 5 menerangkan bahwa pada node teratas diketahui jumlah total mahasiswa lulusan adalah 1378, terdiri dari 259 mahasiswa (18,8%) dengan IPK kategori **rendah**, 1066 mahasiswa (77,4%) IPK kategori **sedang** dan 53 mahasiswa (3,8%) dengan IPK kategori **tinggi**. Tahap pertama pada analisis ini adalah tahap penggabungan, dimana dalam penelitian ini variabel Jurusan untuk mahasiswa lulusan dibagi

menjadi 6 kategori, yaitu : (1) Sipil; (2) Informatika; (3) Elektro; (4) Kriya; (5) Arsitektur dan (6) Industri. Setelah melalui analisis *classification tree*, variabel ini kemudian diringkas menjadi 2 kategori, seperti yang terlihat pada diagram pohon kedalaman yang ke-3, yaitu kategori Jurusan : (1) Sipil, Informatika, Kriya; (2) Elektro, Arsitektur, Industri. Hal ini berarti kategori Sipil, Informatika, Kriya memenuhi syarat kesignifikanan *chi-square* untuk digabung menjadi satu kategori tunggal, demikian halnya dengan kategori Elektro, Arsitektur, Industri.

Penggabungan kategori juga terjadi pada variabel Predikat, dimana dalam penelitian ini variabel Predikat dibagi menjadi 3 kategori, yaitu (1) Memuaskan; (2) Sangat Memuaskan; (3) Terpuji, yang selanjutnya diringkas menjadi 2 kategori, seperti yang terlihat pada diagram pohon kedalaman yang ke-1, yaitu kategori (1) Sangat Memuaskan dan (2) Memuaskan, Terpuji. Hal ini berarti kategori Memuaskan dan Terpuji memenuhi syarat kesignifikanan *chi-square* untuk digabung menjadi satu kategori tunggal berdasarkan analisis *classification tree* yang ada.

Dari *model summary* pada Tabel 10 dapat diketahui bahwa dalam penelitian ini sebenarnya terdapat 7 variabel bebas (Jenis_Kelamin, Tempat_Lahir, Jurusan, Jenjang, Lama_Studi, Status dan Predikat), kemudian hasil analisis menunjukkan bahwa hanya ada 4 variabel bebas yang signifikan terhadap variabel terikatnya (IPK), yaitu variabel Predikat, Jenjang, Jenis_Kelamin dan Jurusan, sehingga ada 3 variabel bebas yang tersisa dan tidak dianggap mempunyai hubungan dengan variabel terikat, yaitu variabel Tempat_Lahir, Lama_Studi, dan Status. Nilai *p-value* dan nilai uji *chi-square* dari masing-masing variabel bebas yang dianggap mempunyai hubungan dengan variabel terikatnya dapat diringkas dalam Tabel 10 berikut.

Tabel 10. Nilai signifikansi variabel bebas terhadap variabel IPK mahasiswa lulusan

Variabel Bebas	**Nilai *p-value***	**Nilai *chi-square***
Predikat	0,000	1195
Jenjang	0,000	37
Jenis_Kelamin	0,002	12
Jurusan	0,033	13

Dari Tabel 10 dapat diketahui bahwa apabila dilakukan pengambilan keputusan berdasarkan nilai *p-value*, dimana ketiga nilai tersebut kurang dari α=0,05, yaitu 0,000; 0,002 dan 0,033. Maka dapat ditarik kesimpulan bahwa keputusan uji *chi-square* adalah Tolak H_0, yang berarti bahwa terdapat hubungan antara ketiga variabel bebas tersebut (Predikat, Jenjang, Jenis_Kelamin dan Jurusan) dengan variabel terikatnya, yaitu IPK.

Pohon klasifikasi pada Gambar 5 menunjukkan bahwa Mahasiswa Lulusan Fakultas Teknik Universitas Negeri Gorontalo berdasarkan IPK dibagi menjadi 5 klasifikasi, yaitu :

Tabel 11. Klasifikasi dan Karakteristik Data Mahasiswa Lulusan berdasarkan variabel IPK

Klasifikasi	Karakteristik
Ke-1	Mahasiswa dengan Predikat Sangat Memuaskan, Jenjang D3
Ke-2	Mahasiswa dengan Predikat Sangat Memuaskan, Jenjang S1
Ke-3	Mahasiswa dengan Predikat Memuaskan dan Terpuji, Jenis Kelamin Laki-Laki, Jurusan Sipil, Informatika, Kriya
Ke-4	Mahasiswa dengan Predikat Memuaskan dan Terpuji, Jenis Kelamin Laki-Laki, Jurusan Elektro, Arsitektur, Industri
Ke-5	Mahasiswa dengan Predikat Memuaskan dan Terpuji, Jenis Kelamin Perempuan

Dari kelima klasifikasi yang terbentuk, dapat ditabulasi jumlah masing-masing mahasiswa yang memiliki IPK dengan kategori rendah, sedang dan tinggi dalam Tabel 12 berikut.

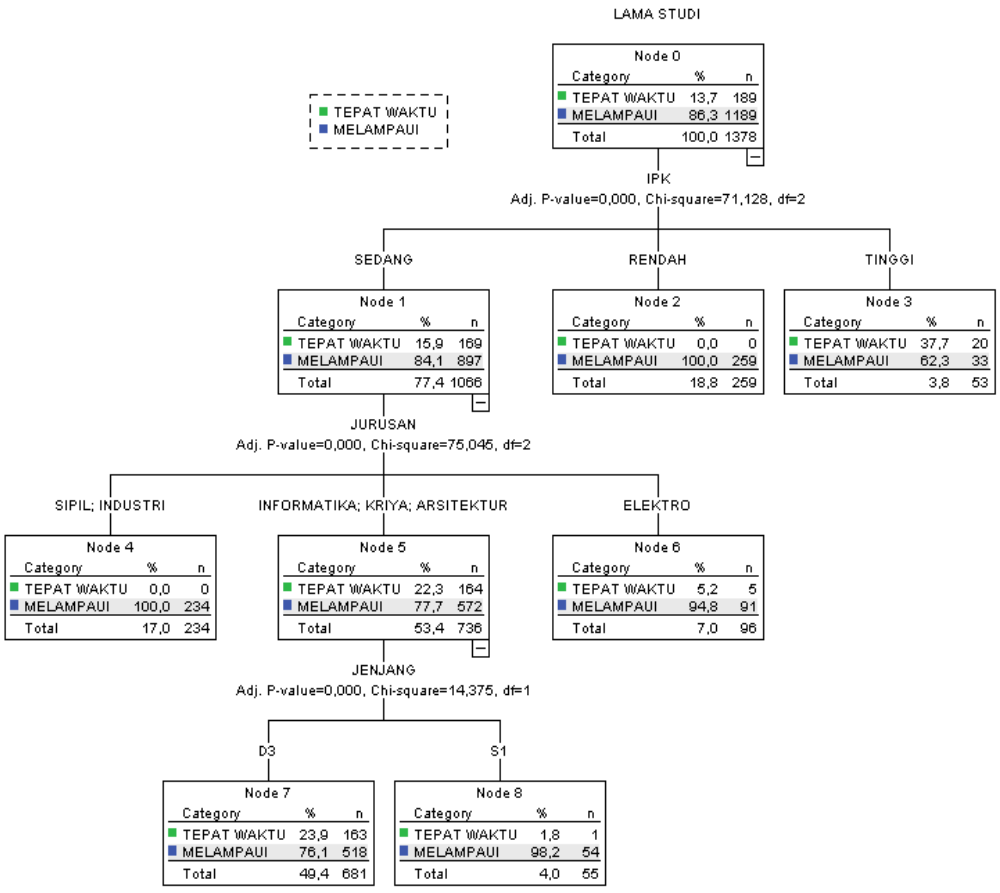
Tabel 12. Jumlah Mahasiswa Lulusan setiap Klasifikasi berdasarkan Kategori IPK

Klasifikasi	IPK		
	Rendah	Sedang	Tinggi
Ke-1	1	**993**	14
Ke-2	0	55	8
Ke-3	**130**	3	8
Ke-4	50	10	4
Ke-5	78	5	**19**

Hasil tabulasi IPK pada Tabel 12 di atas menunjukkan bahwa perolehan IPK dengan kategori **rendah** yang paling banyak dihasilkan oleh klasifikasi mahasiswa ke-3, yaitu karakteristik Mahasiswa dengan Predikat Memuaskan dan

Terpuji, Jenis Kelamin Laki-Laki, Jurusan Sipil, Informatika, Kriya. IPK dengan kategori **sedang** yang paling banyak dihasilkan oleh klasifikasi mahasiswa ke-1, yaitu karakteristik Mahasiswa dengan Predikat Sangat Memuaskan, Jenjang D3. IPK dengan kategori **tinggi** yang paling banyak dihasilkan oleh klasifikasi mahasiswa ke-5, yaitu karakteristik Mahasiswa dengan Predikat Memuaskan, Terpuji dan Jenis Kelamin Perempuan.

Gambar 6 di bawah ini merupakan hasil klasifikasi data mahasiswa lulusan Fakultas Teknik Universitas Negeri Gorontalo dengan menggunakan metode *Decision Tree*, yang diklasifikasikan berdasarkan variabel Lama_Studi.



Gambar 6. Diagram Classification Tree berdasarkan Lama Studi Mahasiswa Lulusan

Diagram pohon hasil analisis *classification tree* pada Gambar 6 menerangkan bahwa pada node teratas diketahui jumlah total mahasiswa lulusan adalah 1378, terdiri dari 189 mahasiswa (13,7%) dengan Lama_Studi kategori **tepat waktu**, 1189 mahasiswa (86,3%) Lama_Studi kategori **melampaui** batas masa studi. Tahap pertama pada analisis ini adalah tahap penggabungan, dimana dalam penelitian ini variabel Jurusan untuk mahasiswa lulusan dibagi menjadi 6 kategori, yaitu : (1) Sipil; (2) Informatika; (3) Elektro; (4) Kriya; (5) Arsitektur dan (6) Industri. Setelah melalui analisis *classification tree*, variabel ini kemudian diringkas menjadi 3 kategori, seperti yang terlihat pada diagram pohon kedalaman yang ke-2, yaitu kategori Jurusan : (1) Sipil, Industri; (2) Informatika, Kriya, Arsitektur dan (3) Elektro. Hal ini berarti kategori Sipil dan Industri memenuhi syarat kesignifikanan *chi-square* untuk digabung menjadi satu kategori tunggal, demikian halnya dengan kategori Informatika, Kriya dan Arsitektur.

Dari Model Summary pada Tabel 13 dapat diketahui bahwa dalam penelitian ini sebenarnya terdapat 7 variabel bebas (Jenis_Kelamin, Tempat_Lahir, Jurusan, Jenjang, Lama_Studi, Status dan Predikat), kemudian hasil analisis menunjukkan bahwa hanya ada 3 variabel bebas yang signifikan terhadap variabel terikatnya (Lama_Studi), yaitu variabel IPK, Jurusan dan Jenjang, sehingga ada 4 variabel bebas yang tersisa dan tidak dianggap mempunyai hubungan dengan variabel terikat, yaitu variabel Tempat_Lahir, Jenis_Kelamin, Status dan Predikat. Nilai *p-value* dan nilai uji *chi-square* dari masing-masing variabel bebas yang dianggap mempunyai hubungan dengan variabel terikatnya dapat diringkas dalam Tabel 13 berikut.

Tabel 13. Nilai signifikansi variabel bebas terhadap variabel Lama_Studi mahasiswa lulusan

Variabel Bebas	**Nilai *p-value***	**Nilai *chi-square***
IPK	0,000	71,128
Jurusan	0,000	75,045
Jenjang	0,000	14,375

Dari Tabel 13 dapat diketahui bahwa apabila dilakukan pengambilan keputusan berdasarkan nilai *p-value*, dimana ketiga nilai tersebut kurang dari α=0,05, yaitu 0,000. Maka dapat ditarik kesimpulan bahwa keputusan uji *chi-square* adalah Tolak H_0, yang berarti bahwa terdapat hubungan antara ketiga variabel bebas tersebut (IPK, Jurusan dan Jenjang) dengan variabel terikatnya, yaitu Lama_Studi.

Pohon klasifikasi pada Gambar 6 menunjukkan bahwa karakterisasi mahasiswa Lulusan Fakultas Teknik Universitas Negeri Gorontalo berdasarkan Lama Studi dibagi menjadi 6 klasifikasi, yaitu :

Tabel 14. Klasifikasi dan Karakteristik Data Mahasiswa Lulusan berdasarkan variabel Lama_Studi

Klasifikasi	Karakteristik
Ke-1	Mahasiswa dengan IPK kategori Sedang, Jurusan Sipil, Industri
Ke-2	Mahasiswa dengan IPK kategori Sedang, Jurusan Informatika, Kriya, Arsitektur, dan Jenjang D3
Ke-3	Mahasiswa dengan IPK kategori Sedang, Jurusan Informatika, Kriya, Arsitektur, dan Jenjang S1
Ke-4	Mahasiswa dengan IPK kategori Sedang, Jurusan Elektro
Ke-5	Mahasiswa dengan IPK kategori Rendah
Ke-6	Mahasiswa dengan IPK kategori Tinggi

Dari keenam klasifikasi yang terbentuk, dapat ditabulasi jumlah masing-masing mahasiswa yang memiliki IPK dengan kategori rendah, sedang dan tinggi dalam Tabel 15 berikut.

Tabel 15. Jumlah Mahasiswa Lulusan setiap Klasifikasi berdasarkan Kategori Lama_Studi

Klasifikasi	Lama Studi	
	Tepat Waktu	Melampaui
Ke-1	0	234
Ke-2	**163**	**518**
Ke-3	1	54
Ke-4	5	91
Ke-5	0	259
Ke-6	20	33

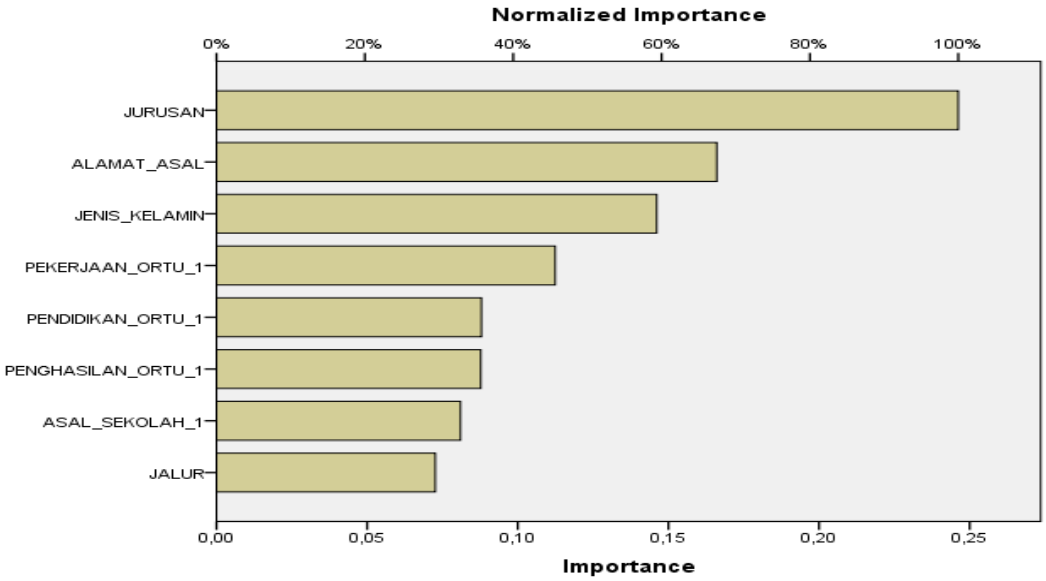
Hasil tabulasi IPK pada Tabel 15 di atas menunjukkan bahwa Lama Studi baik kategori **tepat waktu** maupun **melampaui** terbanyak dihasilkan oleh klasifikasi mahasiswa ke-2, yaitu karakteristik Mahasiswa dengan IPK kategori Sedang, Jurusan Informatika, Kriya, Arsitektur, dan Jenjang D3.

2. Artificial Neural Network

Penelitian ini menggunakan metode Jaringan saraf Tiruan (*Artificial Neural Network*) untuk mengetahui persentase dari variabel-variabel bebas yang dianggap penting (*independent variabel importance*) dan memiliki hubungan yang kuat dengan variabel terikatnya.

a. Klasifikasi data mahasiswa aktif berdasarkan variabel IPK dan Alamat_Asal

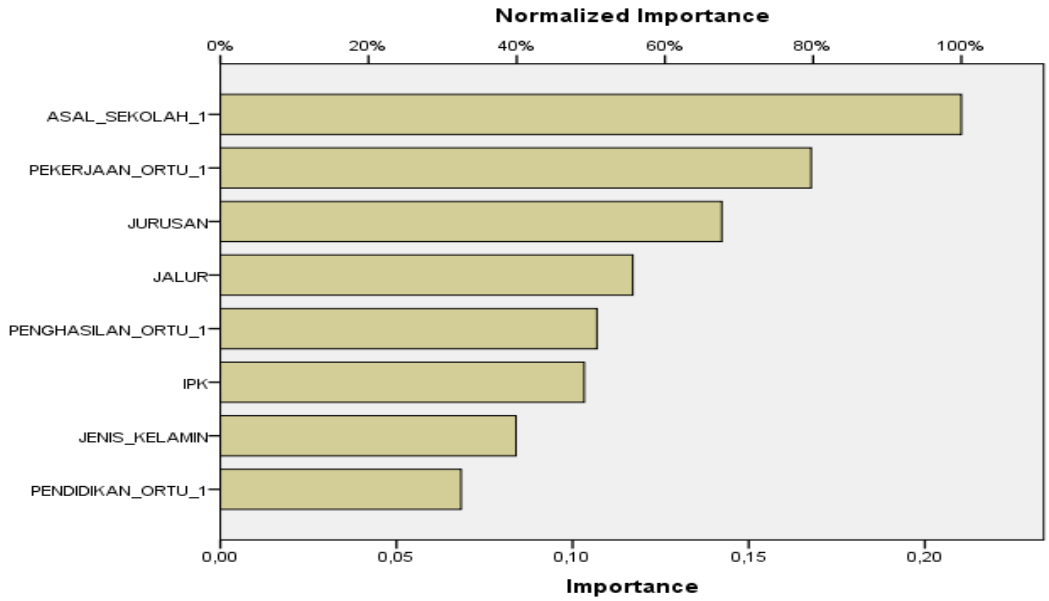
Klasifikasi data mahasiswa aktif dengan menggunakan metode Jaringan Saraf Tiruan (*Artificial Neural Network*) berdasarkan IPK terlihat pada Lampiran-1 dan daftar *Independent Variabel Importance* dapat dilihat pada Gambar 7.



Gambar 7. Persentase Independent Variabel Importance berdasarkan IPK Mahasiswa Aktif

Gambar 7 menunjukkan bahwa variabel bebas yang memiliki hubungan yang penting bagi variabel terikat IPK dengan persentase di atas 50% adalah variabel bebas Jurusan (100%), Alamat_Asal (67,4%) dan Jenis Kelamin (59,3%). Sekalipun persentasenya kurang dari 50%, dari ketiga variabel bebas yang berhubungan dengan data orang tua mahasiswa yaitu Pekerjaan_Ortu, Pendidikan_Ortu dan Penghasilan_Ortu, terlihat bahwa variabel Pekerjaan_Ortu memiliki hubungan yang lebih penting atau kuat dengan varibel terikatnya yaitu IPK. Variabel yang paling kecil nilai *importance* nyada adalah Jalur (29,5%).

Klasifikasi data mahasiswa aktif dengan menggunakan metode Jaringan Saraf Tiruan (*Artificial Neural Network*) berdasarkan Alamat_Asal terlihat pada Lampiran-2 dan daftar *Independent Variabel Importance* dapat dilihat pada Gambar 8.

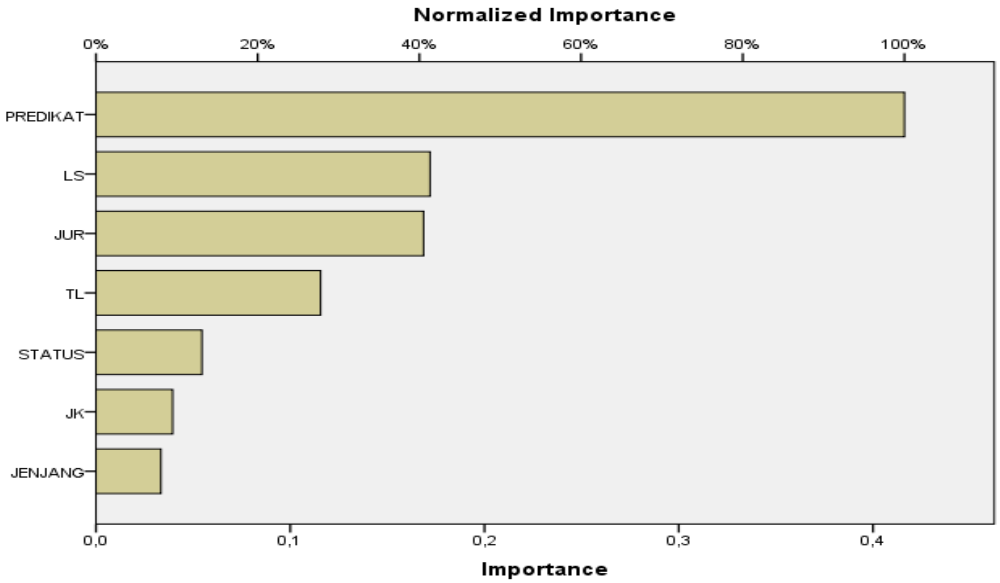


Gambar 8. Persentase Independent Variabel Importance berdasarkan Alamat_Asal

Gambar 8 menunjukkan bahwa variabel bebas yang memiliki hubungan yang penting bagi variabel terikat IPK dengan persentase di atas 50% adalah variabel bebas Asal_Sekolah (100%), Pekerjaan_Ortu (74,3%) dan Jurusan (52,4%). Variabel yang paling kecil nilai *importance* nyada adalah Pendidikan_Ortu (31,1%).

b. Klasifikasi data mahasiswa lulusan berdasarkan variabel IPK dan Lama_Studi

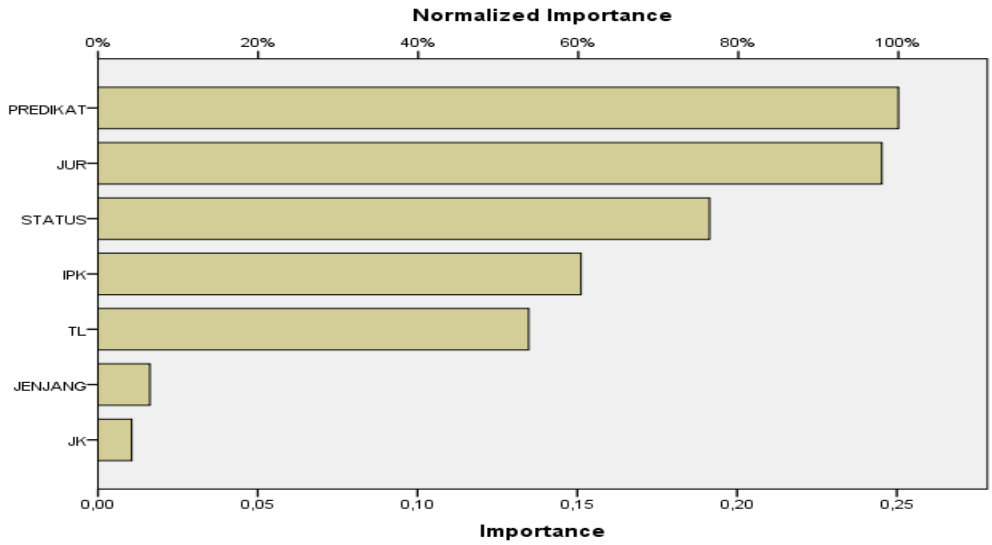
Klasifikasi data mahasiswa lulusan dengan menggunakan metode Jaringan Saraf Tiruan (*Artificial Neural Network*) berdasarkan IPK terlihat pada Lampiran-3 dan daftar *Independent Variabel Importance* dapat dilihat pada Gambar 9.



Gambar 9. Persentase Independent Variabel Importance berdasarkan IPK Mahasiswa

Gambar 9 menunjukkan bahwa variabel bebas yang memiliki hubungan yang penting bagi variabel terikat IPK dengan persentase di atas 50% adalah variabel Predikat (100%). Variabel yang paling kecil nilai *importance* nyada adalah Jenjang (8%).

Klasifikasi data mahasiswa lulusan dengan menggunakan metode Jaringan Saraf Tiruan (*Artificial Neural Network*) berdasarkan Lama_Studi terlihat pada Lampiran-4 dan daftar *Independent Variabel Importance* dapat dilihat pada Gambar 10.



Gambar 10. Persentase Independent Variabel Importance berdasarkan Lama_Studi

Gambar 10 menunjukkan bahwa variabel bebas yang memiliki hubungan yang penting bagi variabel terikat IPK dengan jumlah persentase terbesar adalah variabel Predikat (100%). Sedangkan yang memiliki nilai *importance* yang terkecil (di bawah 50%) adalah variabel Jenjang (6,5%) dan Jenis Kelamin (4,2%).

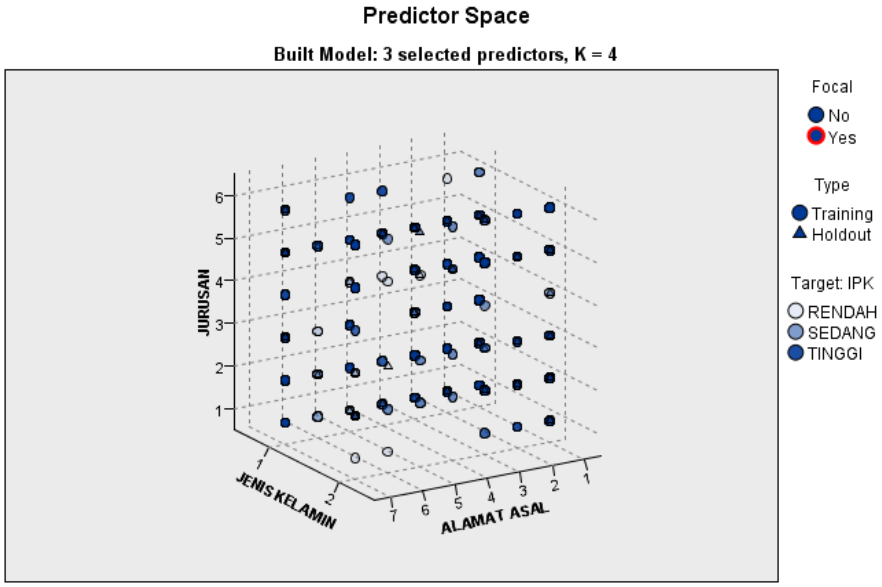
3. K-Neirest Neighbour

K-Neirest Neighbour (KNN) merupakan salah satu metode untuk mengklasifikasikan suatu data baru berdasarkan similaritas (kemiripan) dengan karakteristik pada data lama (*training set*), umumnya menggunakan metrik jarak (*euclidian*). Dalam penelitian ini, metode KNN digunakan untuk memprediksi kelas target (*variabel dependent*) IPK dan Alamat_Asal dari data mahasiswa aktif serta kelas target IPK dan Lama_Studi dari data lulusan mahasiswa Fakultas Teknik Universitas Negeri Gorontalo.

a. Klasifikasi Data Mahasiswa Aktif

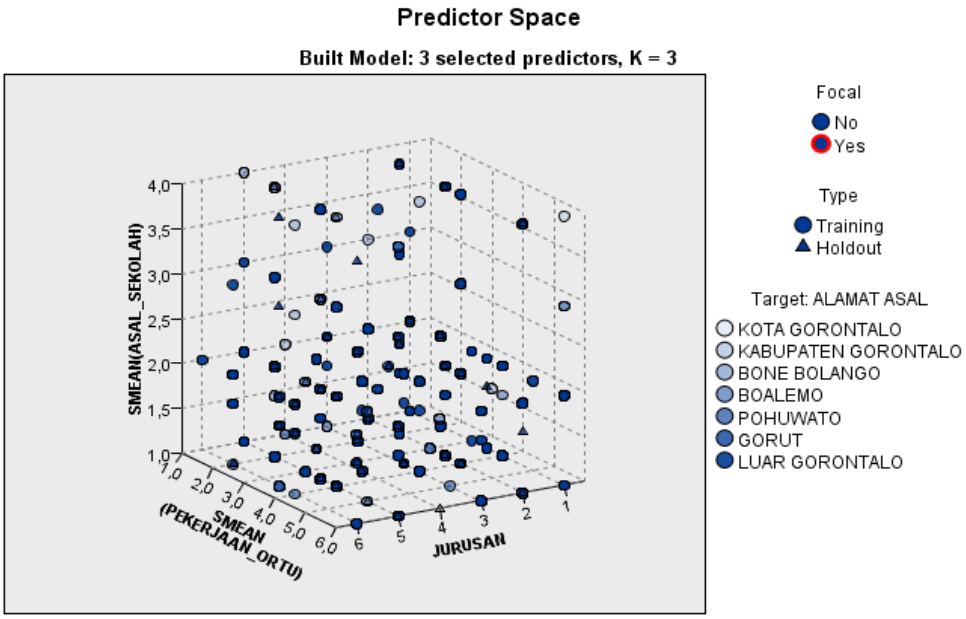
Klasifikasi data mahasiswa aktif dengan metode KNN dilakukan berdasarkan kelas target (*variabel dependent*) IPK dan Alamat_Asal, dengan komposisi data pelatihan (*training set*) 80% dan data uji (*test set*) 20%, serta pemilihan jumlah **k** secara otomatis (berkisar antara 3 hingga 5).

Klasifikasi data mahasiswa aktif berdasarkan kategori IPK ditunjukkan oleh Gambar 11. Hasil analisis menunjukkan bahwa model terbaik diperoleh pada proses kalasifikasi dengan jumlah **k** = 4 dan dari 8 variabel bebas yang digunakan untuk mengklasifikasi mahasiswa aktif berdasarkan variabel terikat IPK, terpilih 3 variabel bebas yang dianggap faktor penting dalam menentukan perolehan IPK, yaitu Jurusan, Jenis_Kelamin dan Alamat_Asal.



Gambar 11. Klasifikasi KNN Data Mahasiswa Aktif berdasarkan variabel IPK

Klasifikasi data mahasiswa aktif berdasarkan kategori Alamat_Asal ditunjukkan oleh Gambar 12. Hasil analisis menunjukkan bahwa model terbaik diperoleh pada proses kalasifikasi dengan jumlah **k** = 3 dan dari 8 variabel bebas yang digunakan untuk mengklasifikasi mahasiswa aktif berdasarkan variabel terikat Alamat_Asal, terpilih 3 variabel bebas yang dianggap faktor penting dalam menentukan perolehan Asal_Sekolah, Jurusan dan Pekerjaan_Ortu.

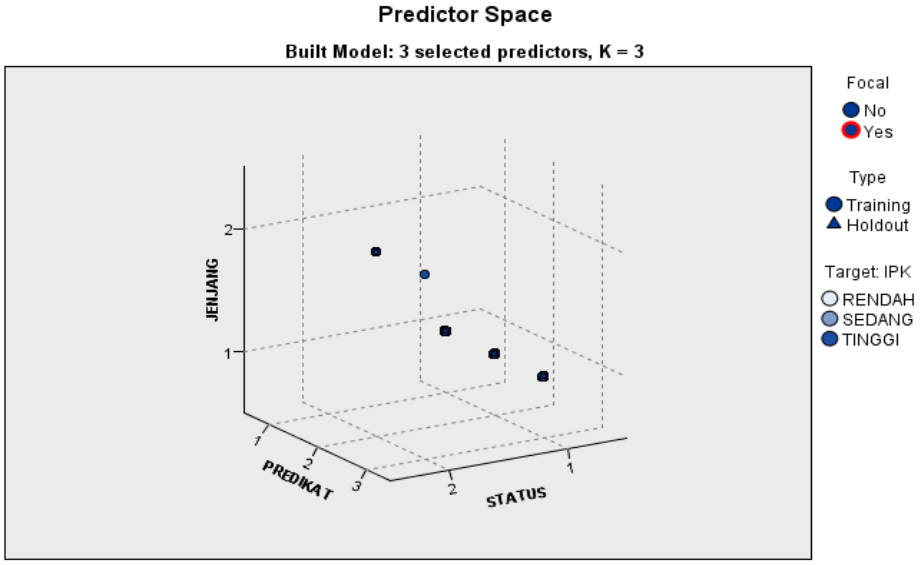


Gambar 12. Klasifikasi KNN Data Mahasiswa Aktif berdasarkan variabel Alamat_Asal

b. Klasifikasi Data Mahasiswa Lulusan

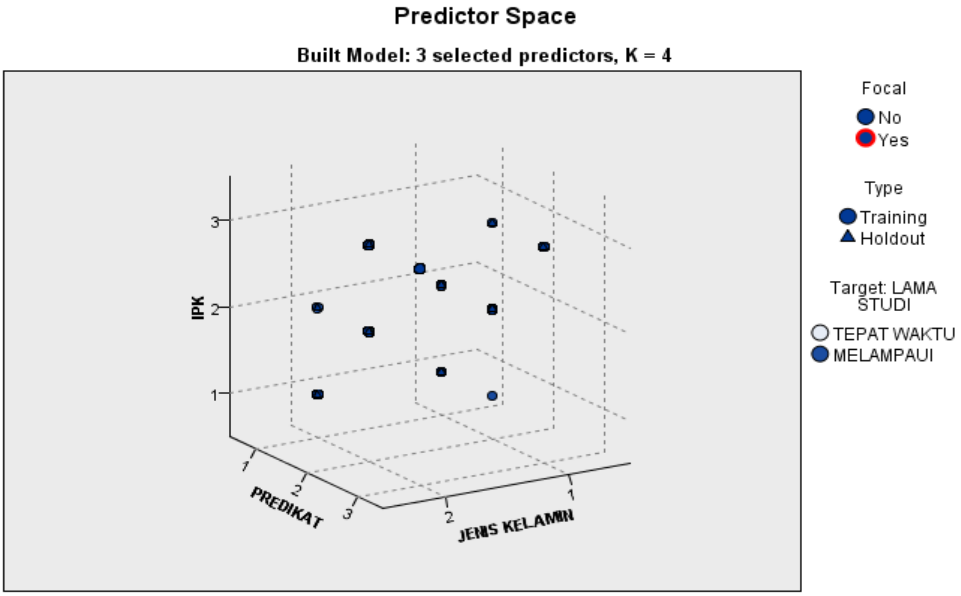
Klasifikasi data mahasiswa lulusan dengan metode KNN dilakukan berdasarkan kelas target (*variabel dependent*) IPK dan Lama_Studi, dengan komposisi data pelatihan (*training set*) 80% dan data uji (*test set*) 20%, serta pemilihan jumlah **k** secara otomatis (berkisar antara 3 hingga 5).

Klasifikasi data mahasiswa lulusan berdasarkan kategori IPK ditunjukkan oleh Gambar 13. Hasil analisis menunjukkan bahwa model terbaik diperoleh pada proses kalasifikasi dengan jumlah **k** = 3 dan dari 7 variabel bebas yang digunakan untuk mengklasifikasi mahasiswa aktif berdasarkan variabel terikat IPK, terpilih 3 variabel bebas yang dianggap faktor penting dalam menentukan perolehan IPK, yaitu Predikat, Jenjang dan Status.



Gambar 13. Klasifikasi KNN Data Mahasiswa Lulusan berdasarkan variabel IPK

Klasifikasi data mahasiswa lulusan berdasarkan kategori Lama_Studi ditunjukkan oleh Gambar 14. Hasil analisis menunjukkan bahwa model terbaik diperoleh pada proses kalasifikasi dengan jumlah **k** = 4 dan dari 7 variabel bebas yang digunakan untuk mengklasifikasi mahasiswa aktif berdasarkan variabel terikat Lama_Studi, terpilih 3 variabel bebas yang dianggap faktor penting dalam menentukan perolehan Predikat, IPK dan Jenis_Kelamin.



Gambar 14. Klasifikasi KNN Data Mahasiswa Lulusan berdasarkan variabel Lama_Studi

C. KLASTERISASI

Tujuan dari analisis klaster adalah mengelompokkan obyek berdasarkan kesamaan karakteristik di antara obyek-obyek tersebut, tanpa memiliki kelas target seperti pada teknik klasifikasi. Metode pengelompokkan yang umumnya digunakan pada analisis klaster adalah (1) Metode Hirarkis dan (2) Metode Non Hirarkis. Metode Hirarkis memulai pengelompokkan dengan dua atau lebih obyek yang memiliki kesamaan paling dekat, kemudian diteruskan pada obyek yang lain dan seterusnya hingga cluster akan membentuk semacam "pohon" dimana terdapat tingkatan (hirarki) yang jelas antar obyek, dari yang paling mirip hingga yang paling tidak mirip. Alat bantu untuk memperjelas proses hirarki ini disebut dengan **"dendogram".** Metode Non-Hirarkis dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua, tiga atau lainnya). Setelah jumlah cluster ditentukan, maka proses cluster dilakukan dengan tanpa mengikuti proses hirarki. Metode ini biasa disebut **"K-Means Cluster".**

1. Metode Non-Hirarkis (K-*Means Cluster*)

Data yang diklaster pada penelitian ini adalah data aktif dan lulusan dari Mahasiswa Fakultas Teknik Universitas Negeri Gorontalo.

a. Klasterisasi data mahasiswa aktif

Berdasarkan hasil *final cluster* data mahasiswa aktif dengan jumlah k = 3, maka diperoleh karakteristik dari masing-masing *cluster* adalah seperti yang terlihat pada Tabel 16 berikut.

Tabel 16. *Final Cluster Centers* Mahasiswa Aktif

VARIABEL	Cluster		
	1	**2**	**3**
IPK	1	2	2
Jalur	3	3	3
Alamat_Asal	7	2	1
Jenis_Kelamin	1	2	1
Smean(Pendidikan_Ortu)	2	2	3
Smean(Pekerjaan_Ortu)	4	3	1
Smean(Penghasilan_Ortu)	2	2	3
Smean(Asal_Sekolah)	2	2	2

Dari Tabel 16 di atas dapat didefiniskan bahwa :

- Cluster-1

  Berisikan mahasiswa-mahasiswa dengan perolehan IPK kategori rendah, jalur masuk mandiri, alamat asal dari luar daerah Provinsi Gorontalo, jenis kelamin laki-laki, pendidikan orang tua kategori menengah, pekerjaan orang tua Petani/Nelayan, penghasilan orang tua berkisar satu hingga tiga juta rupiah dan asal sekolah SMK.

- Cluster-2

  Berisikan mahasiswa-mahasiswa dengan perolehan IPK kategori sedang, jalur masuk mandiri, alamat asal dari Kabupaten Gorontalo, jenis kelamin perempuan, pendidikan orang tua kategori menengah, pekerjaan orang tua TNI/POLRI, penghasilan orang tua berkisar satu hingga tiga juta rupiah dan asal sekolah SMK.

- Cluster-3

  Berisikan mahasiswa-mahasiswa dengan perolehan IPK kategori rendah, jalur masuk mandiri, alamat asal dari Kota Gorontalo, jenis kelamin perempuan, pendidikan orang tua kategori tinggi, pekerjaan orang tua PNS, penghasilan orang tua berkisar tiga hingga lima juta rupiah dan asal sekolah SMK.

b. Klasterisasi data mahasiswa lulusan

Berdasarkan hasil *final cluster* data mahasiswa lulusan dengan jumlah k = 3, maka diperoleh karakteristik dari masing-masing *cluster* adalah seperti yang terlihat pada Tabel 17 berikut.

Tabel 17. Final Cluster Centers Mahasiswa Lulusan

VARIABEL	Cluster		
	1	2	3
Jenis_Kelamin	2	2	1
Tempat_Lahir	2	1	7
Jenjang	1	1	1
Status	1	1	1
Predikat	2	2	1
IPK	2	2	1
Lama_Studi	1	2	2

Dari Tabel 18 dapat didefiniskan bahwa :

- Cluster-1

  Berisikan mahasiswa-mahasiswa dengan jenis kelamin perempuan, tempat lahir Kabupaten Gorontalo, Jenjang D3, Status Non-Kependidikan, IPK kategori sedang dan lama studi tepat waktu.
- Cluster-2

  Berisikan mahasiswa-mahasiswa dengan jenis kelamin perempuan, tempat lahir Kota Gorontalo, Jenjang D3, Status Non-Kependidikan, IPK kategori sedang dan lama studi melampaui batas waktu.
- Cluster-3

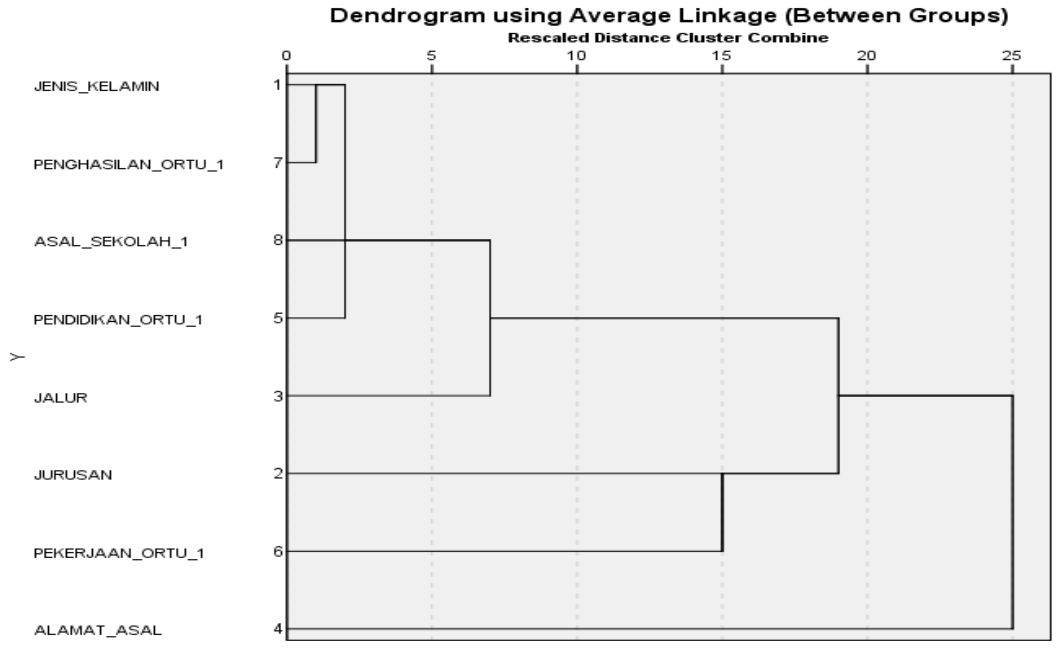
  Berisikan mahasiswa-mahasiswa dengan jenis kelamin laki-laki, tempat lahir luar daerah Provinsi Gorontalo, Jenjang D3, Status Non-Kependidikan, IPK kategori rendah dan lama studi melampaui batas waktu.

2. Metode Hirarkis

Data yang diklaster pada penelitian ini adalah data aktif dan lulusan dari Mahasiswa Fakultas Teknik Universitas Negeri Gorontalo

a. Klasterisasi data mahasiswa aktif

Gambar 15 menunjukkan hasil klasterisasi data mahasiswa aktif dengan menggunakan metode hirarkis, yang digambarkan dalam bentuk sebuah dendogram. Dendoram berguna untuk menunjukkan anggota cluster yang ada jika akan ditentukan berapa cluster seharusnya dibentuk.



Gambar 15. Dendogram klasterisasi data mahasiswa aktif

Tabel 18 merupakan tabel yang berisikan anggota cluster dari data Mahasiswa Aktif Fakultas Teknik Universitas Negeri Gorontalo.

Tabel 18. Anggota Cluster Data Mahasiswa Aktif

Variabel	4 Clusters	3 Clusters	2 Clusters
Jenis Kelamin	1	1	1
Jurusan	2	2	1
Jalur	1	1	1
Alamat Asal	3	3	2
Smean(Pendidikan_Ortu)	1	1	1
Smean(Pekerjaan_Ortu)	4	2	1
Smean(Penghasilan_Ortu)	1	1	1
Smean(Asal_Sekolah)	1	1	1

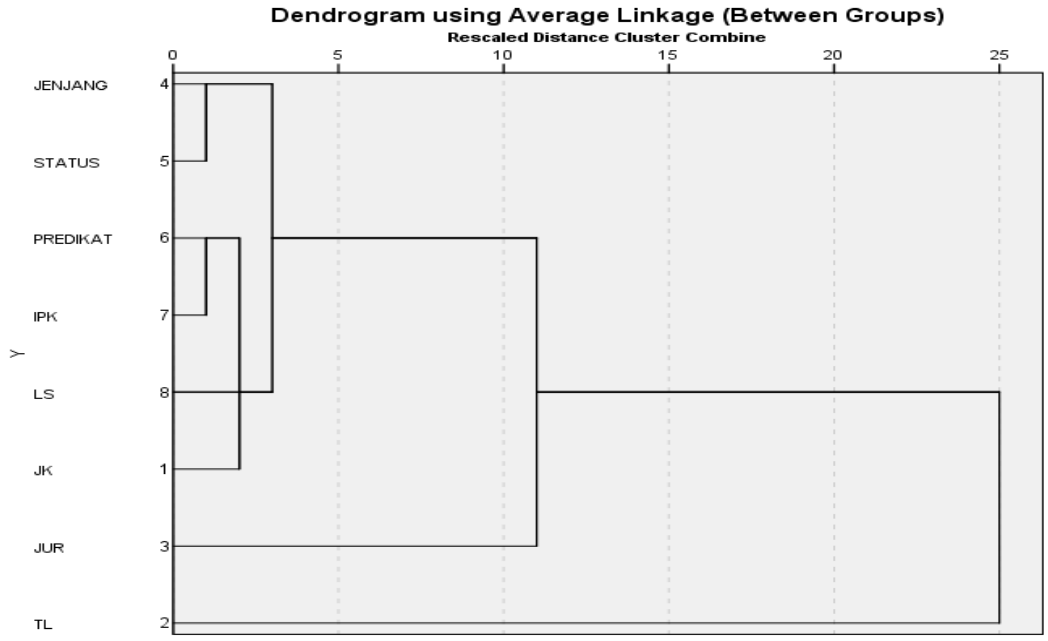
Dari Tabel 18 di atas dapat dijabarkan bahwa :

- Apabila diinginkan dibentuk 4 cluster, maka :
  - ✓ Variabel-variabel yang digunakan dalam menentukan anggota dari Cluster 1 adalah Jenis_Kelamin, Jalur_Masuk, Pendidikan_Ortu, Penghasilan_Ortu dan Asal_sekolah.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 2 adalah Jurusan.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 3 adalah Alamat_Asal.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 4 adalah Pekerjaan_Ortu.
- Apabila diinginkan dibentuk 3 cluster, maka :
  - ✓ Variabel-variabel yang digunakan dalam menentukan anggota dari Cluster 1 adalah Jenis_Kelamin, Jalur_Masuk, Pendidikan_Ortu, Penghasilan_Ortu dan Asal_sekolah.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 2 adalah Pekerjaan_Ortu dan Jurusan.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 3 adalah Alamat_Asal.

- Apabila diinginkan dibentuk 2 cluster, maka :
  - ✓ Variabel-variabel yang digunakan dalam menentukan anggota dari Cluster 1 adalah Jenis_Kelamin, Jurusan, Jalur_Masuk, Pendidikan_Ortu, Penghasilan_Ortu, Pekerjaan_Ortu dan Asal_sekolah.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 2 adalah Alamat_Asal.

b. Klasterisasi data mahasiswa lulusan

Gambar 16 menunjukkan hasil klasterisasi data mahasiswa aktif dengan menggunakan metode hirarkis, yang digambarkan dalam bentuk sebuah dendogram. Dendoram berguna untuk menunjukkan anggota cluster yang ada jika akan ditentukan berapa cluster seharusnya dibentuk.



Gambar 16. Dendogram klasterisasi data mahasiswa lulusan

Tabel 19 merupakan tabel yang berisikan anggota cluster dari data lulusan Mahasiswa Fakultas Teknik Universitas Negeri Gorontalo.

Tabel 19. Anggota Cluster Data Mahasiswa Lulusan

Variabel	4 Clusters	3 Clusters	2 Clusters
Jenis Kelamin	1	1	1
Tempat Lahir	2	2	2
Jurusan	3	3	1
Jenjang	4	1	1
Status	4	1	1
Predikat	1	1	1
IPK	1	1	1
Lama Studi	1	1	1

Dari Tabel 19 di atas dapat dijabarkan bahwa :

- Apabila diinginkan dibentuk 4 cluster, maka :
  - ✓ Variabel-variabel yang digunakan dalam menentukan anggota dari Cluster 1 adalah Jenis_Kelamin, Predikat, IPK dan Lama Studi.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 2 adalah Tempat_Lahir.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 3 adalah Jurusan.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 4 adalah Jenjang dan Status.
- Apabila diinginkan dibentuk 3 cluster, maka :
  - ✓ Variabel-variabel yang digunakan dalam menentukan anggota dari Cluster 1 adalah Jenis_Kelamin, Jenjang, Status, Predikat, IPK dan Lama_Studi.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 2 adalah Tempat_Lahir.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 3 adalah Jurusan.

- Apabila diinginkan dibentuk 2 cluster, maka :
  - ✓ Variabel-variabel yang digunakan dalam menentukan anggota dari Cluster 1 adalah Jenis_Kelamin, Jurusan, Jenjang, Status, Predikat, IPK dan Lama_Studi.
  - ✓ Variabel yang digunakan dalam menentukan anggota dari Cluster 2 adalah Tempat_Lahir.

D. Perbandingan Hasil Prediksi

a. Prediksi Kelas Target Mahasiswa Aktif

Tabel 20 menunjukkan perbedaan hasil ketepatan uji prediksi dari masing-masing kelas target dari data Mahasiswa Aktif, dengan menggunakan metode *Decission Tree* dan K-*Neirest Neigbour*.

Tabel 20. Ketepatan Uji Prediksi IPK dan Alamat_Asal dari data Mahasiswa Aktif

Kelas Target	Ketepatan Uji Prediksi	Metode Klasifikasi			
		Decission Tree		K-Neirest Neihbour	
		Jml	%	Jml	%
IPK	Tepat	932	**66,81**	913	**65,45**
	Tidak Tepat	463	33,19	482	33,55
Alamat_Asal	Tepat	571	40,93	496	35,56
	Tidak Tepat	824	**59,07**	899	**64,44**

Dari Tabel 20 di atas dapat dijabarkan bahwa ketepatan prediksi berdasarkan kelas target IPK memiliki presentase yang lebih besar daripada kelas target Alamat_Asal, yang menunjukkan bahwa IPK memiliki nilai lebih baik dari Alamat_Asal untuk dijadikan sebagai kelas target dalam proses prediksi, meskipun persentase ketepatan uji prediki yang dihasilkan cenderung rendah (kurang dari 70%).

b. Prediksi Kelas Target Mahasiswa Lulusan

Tabel 21 menunjukkan perbedaan hasil ketepatan uji prediksi dari masing-masing kelas target dari data Mahasiswa Lulusan, dengan menggunakan metode *Decission Tree* dan K-*Neirest Neigbour*.

Tabel 21. Ketepatan Uji Prediksi IPK dan Alamat_Asal dari data Mahasiswa Lulusan

Kelas Target	Ketepatan Uji Prediksi	Metode Klasifikasi			
		Decission Tree		K-Neirest Neihbour	
		Jml	%	Jml	%
IPK	Tepat	1306	**94,78**	1307	**94,85**
	Tidak Tepat	72	5,22	71	5,15
Lama_Studi	Tepat	1189	**86,28**	1191	**86,43**
	Tidak Tepat	189	13,72	187	13,57

Dari Tabel 21 di atas dapat dijabarkan bahwa ketepatan prediksi berdasarkan kelas target IPK memiliki presentase yang lebih besar daripada kelas target Lama_Studi, yang menunjukkan bahwa IPK memiliki nilai lebih baik dari Lama_Studi untuk dijadikan sebagai kelas target dalam proses prediksi. Apabila dibandingkan dengan persentase ketepatan uji prediksi yang diperoleh pada klasifikasi data aktif mahasiswa, maka hasil persentase ketepatan uji prediksi pada klasifikasi data lulusan lebih tinggi (lebih dari 85%). Sehingga dapat disimpulkan bahwa untuk memprediksi IPK dan Lama_Studi mahasiswa yang masih berstatus aktif dapat menggunakan karakteristik berupa variabel-variabel bebas yang digunakan pada data set lulusan mahasiswa Fakultas Teknik Universitas Negeri Gorontalo, yang terdiri dari Jurusan, Jenjang, Status, Predikat, dan Jenis_Kelamin.

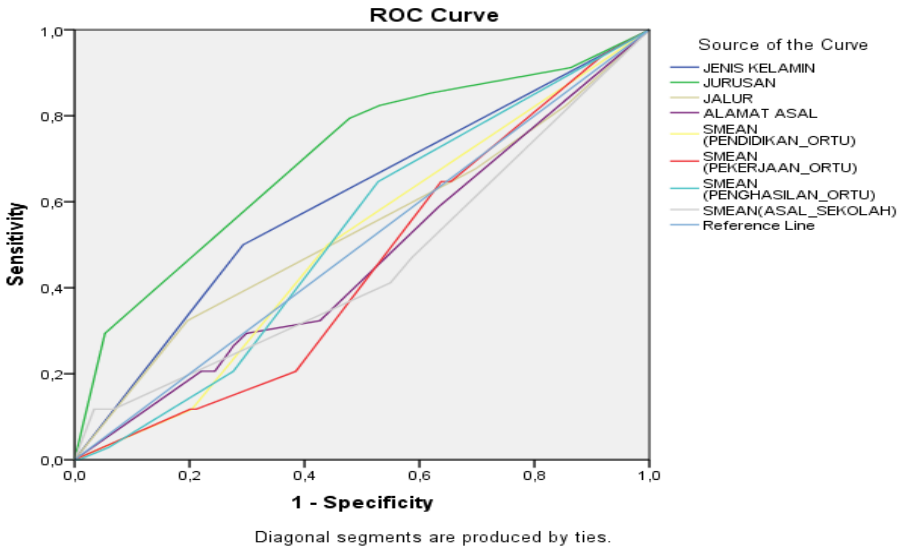
E. Tingkat *Sensitivity* dan *Importance* Variabel Bebas

Tabel 22 Daftar Independent Variabel Importance berdasarkan variabel terikat IPK

METODE KLASIFIKASI					
MAHASISWA AKTIF			**MAHASISWA LULUSAN**		
TREE	**ANN**	**K-NN**	***TREE***	**ANN**	**K-NN**
Jurusan	Jurusan	Jurusan	Predikat	Predikat	Predikat
Jenis_Kelamin	Alamat_Asal	Jenis_Kelamin	Jenjang	Lama_Studi	Jenjang
Asal_Sekolah	Jenis_Kelamin	Alamat_Asal	Jenis_Kelamin	Jurusan	Status

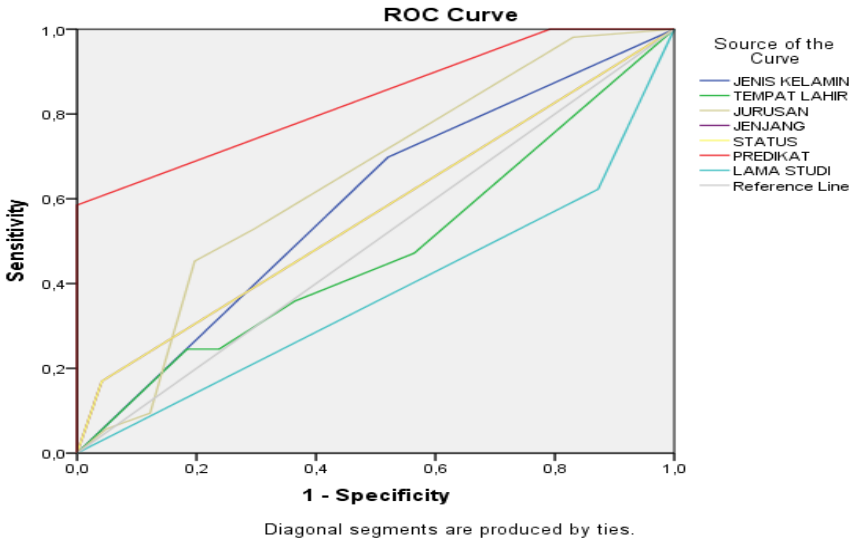
Tabel 22 menunjukkan bahwa dalam menentukan *Independent Variabel Importance* dari data mahasiswa aktif dan lulusan yang diklasifikasikan berdasarkan variabel terikat IPK, ketiga metode (*Decission Tree*, ANN dan KNN) menghasilkan variabel bebas yang sama pada level yang pertama, yaitu variabel bebas **Jurusan** untuk data mahasiswa aktif dan variabel bebas **Predikat** untuk data mahasiswa lulusan. Pada level kedua, variabel bebas yang dihasilkan oleh metode Tree dan K-NN menunjukkan hasil yang sama, yaitu variabel **Jenis_Kelamin** untuk data mahasiswa aktif dan variabel **Jenjang** untuk data mahasiswa lulusan, sedangkan Metode ANN menghasilkan variabel bebas yankg berbeda.

Gambar 17 berikut merupakan Kurva ROC (*Relative Operating Characteristics*) yang menunjukkan tingkat sensitivitas dari variabel-variabel bebas yang digunakan untuk menguji model prediksi data mahasiswa aktif berdasarkan variabel terikat IPK. Dua variabel bebas yang memiliki tingkat sensitivitas tertinggi sebagaimana yang tergambar pada Kurva ROC adalah variabel Jurusan dan Jenis_Kelamin. Hal ini menjelaskan bahwa tingkat sensitivitas variabel bebas yang dihasilkan oleh Kurva ROC memiliki kesamaan dengan *Independent Variabel Importance* yang dihasilkan oleh metode *Decission Tree* dan K-*Neirest Neighbour*.



Gambar 17. Kurva ROC variabel bebas data mahasiswa aktif berdasarkan variabel terikat IPK

Gambar 18 merupakan Kurva ROC (*Relative Operating Characteristics*) yang menunjukkan tingkat sensitivitas dari variabel-variabel bebas yang digunakan untuk menguji model prediksi data mahasiswa lulusan berdasarkan variabel terikat IPK. Variabel bebas yang memiliki tingkat sensitivitas tertinggi sebagaimana yang tergambar pada Kurva ROC adalah variabel Predikat. Hal ini menjelaskan bahwa tingkat sensitivitas variabel bebas yang dihasilkan oleh Kurva ROC memiliki kesamaan dengan *Independent Variabel Importance* yang dihasilkan oleh ketiga metode klasifikasi yang digunakan sebelumnya, yaitu metode *Decission Tree*, *Artificial Neural Network* (ANN) dan K-*Neirest Neighbour* (K-NN).



Gambar 18. Kurva ROC variabel bebas data mahasiswa lulusan berdasarkan variabel terikat IPK

F. Perbandingan Rata-Rata (*Compare Means*)

Pada penelitian ini, uji *Compare Means* perlu dilakukan untuk membandingkan rata-rata dari setiap variabel yang digunakan sekaligus menguji *linearity* nya.

Tabel 23. Perbandingan rata-rata variabel Jenis_Kelamin terhadap variabel IPK dari data mahasiswa aktif

JENIS KELAMIN	Mean	N	Minimum	Maximum	% of Total N
LAKI-LAKI	1,31	979	RENDAH	TINGGI	70,2%
PEREMPUAN	1,64	416	RENDAH	TINGGI	29,8%
Total	1,41	1395	RENDAH	TINGGI	100,0%

Tabel 24. Perbandingan rata-rata variabel Jurusan terhadap variabel IPK dari data mahasiswa aktif

JURUSAN	Mean	N	Minimum	Maximum	% of Total N
ELEKTRO	1,29	190	RENDAH	TINGGI	13,6%
SIPIL	1,17	332	RENDAH	TINGGI	23,8%
ARSITEKTUR	1,26	124	RENDAH	TINGGI	8,9%
INDUSTRI	1,38	71	RENDAH	TINGGI	5,1%
INFORMATIKA	1,57	596	RENDAH	TINGGI	42,7%
KRIYA	1,76	82	RENDAH	TINGGI	5,9%
Total	1,41	1395	RENDAH	TINGGI	100,0%

Tabel 25. Perbandingan rata-rata variabel Jalur masuk terhadap variabel IPK dari data mahasiswa aktif

JALUR	Mean	N	Minimum	Maximum	% of Total N
UNDANGAN	1,42	204	RENDAH	TINGGI	14,6%
UJIAN	1,36	219	RENDAH	TINGGI	15,7%
MANDIRI	1,43	694	RENDAH	TINGGI	49,7%
JURUSAN	1,39	278	RENDAH	TINGGI	19,9%
Total	1,41	1395	RENDAH	TINGGI	100,0%

Tabel 26. Perbandingan rata-rata variabel Pekerjaan_Ortu terhadap variabel IPK dari data mahasiswa aktif

SMEAN(PEKERJAAN_ORTU)	Mean	N	Minimum	Maximum	% of Total N
PNS	1,43	112	RENDAH	TINGGI	8,0%
WIRASWASTA	1,40	370	RENDAH	TINGGI	26,5%
TNI POLRI	1,30	23	**RENDAH**	**SEDANG**	1,6%
3,5	1,42	359	RENDAH	TINGGI	25,7%
PETANI NELAYAN	1,37	239	RENDAH	TINGGI	17,1%
TIDAK BEKERJA	1,36	14	**RENDAH**	**SEDANG**	1,0%
LAINNYA	1,45	278	RENDAH	TINGGI	19,9%
Total	1,41	1395	RENDAH	TINGGI	100,0%

Tabel 27. Perbandingan rata-rata variabel Penghasilan_Ortu terhadap variabel IPK dari data mahasiswa aktif

SMEAN(PENGHASILAN_ORTU)	Mean	N	Minimum	Maximum	% of Total N
DIBAWAH 1 JUTA	1,41	654	RENDAH	TINGGI	46,9%
1,5	1,42	358	RENDAH	TINGGI	25,7%
1 - 3 JUTA	1,40	300	RENDAH	TINGGI	21,5%
3 - 5 JUTA	1,34	73	RENDAH	TINGGI	5,2%
5 - 10 JUTA	1,60	10	**RENDAH**	**SEDANG**	0,7%
Total	1,41	1395	RENDAH	TINGGI	100,0%

Tabel 28. Perbandingan rata-rata variabel Asal_Sekolah terhadap variabel IPK dari data mahasiswa aktif

SMEAN(ASAL_SEKOLAH)	Mean	N	Minimum	Maximum	% of Total N
SMA	1,48	580	RENDAH	TINGGI	41,6%
1,7	1,38	53	RENDAH	TINGGI	3,8%
SMK	1,36	666	RENDAH	TINGGI	47,7%
MADRASAH	1,35	46	**RENDAH**	**SEDANG**	3,3%
PENJENJANGAN	1,34	50	RENDAH	TINGGI	3,6%
Total	1,41	1395	RENDAH	TINGGI	100,0%

Tabel 29. Perbandingan rata-rata variabel Alamat_Asal terhadap variabel IPK dari data mahasiswa aktif

ALAMAT ASAL	Mean	N	Minimum	Maximum	% of Total N
KOTA GORONTALO	1,49	513	RENDAH	TINGGI	36,8%
KABUPATEN GORONTALO	1,41	290	RENDAH	TINGGI	20,8%
BONE BOLANGO	1,39	175	RENDAH	TINGGI	12,5%
BOALEMO	1,53	32	RENDAH	TINGGI	2,3%
POHUWATO	1,26	46	RENDAH	TINGGI	3,3%
GORUT	1,09	32	**RENDAH**	**SEDANG**	2,3%
LUAR GORONTALO	1,34	307	RENDAH	TINGGI	22,0%
Total	1,41	1395	RENDAH	TINGGI	100,0%

Tabel 30. Perbandingan rata-rata variabel Pendidikan_Ortu terhadap variabel IPK dari data mahasiswa aktif

SMEAN(PENDIDIKAN_ORTU)	Mean	N	Minimum	Maximum	% of Total N
TIDAK SEKOLAH	1,41	324	RENDAH	TINGGI	52,94%
PENDIDIKAN MENENGAH	1,41	779	RENDAH	TINGGI	55,85%
PENDIDIKAN TINGGI	1,42	282	RENDAH	TINGGI	20,21%
Total	1,41	1395	RENDAH	TINGGI	100,0%

Berdasarkan beberapa tabel di atas dapat dijabarkan bahwa sebagian besar sebaran data dari nilai rata-rata setiap variabel bebas terhadap variabel terikat IPK, minimum berada kategori IPK rendah dan maksimum berada pada kategori IPK tinggi, kecuali pada variabel Pekerjaan_Ortu (TNI/POLRI dan Tidak Bekerja), Variabel Penghasilan_Ortu (5 – 10 juta), variabel Asal_Sekolah (Madrasah) dan varaibel Alamat_Asal (Gorut) yang maksimum hanya berada pada kategori IPK sedang.

Tabel 31. Perbandingan rata-rata variabel Jenis_Kelamin terhadap variabel IPK dari data mahasiswa lulusan

JENIS KELAMIN	Mean	N	Minimum	Maximum	% of Total N
LAKI-LAKI	1,75	651	RENDAH	TINGGI	47,2%
PEREMPUAN	1,94	727	RENDAH	TINGGI	52,8%
Total	1,85	1378	RENDAH	TINGGI	100,0%

Tabel 32. Perbandingan rata-rata variabel Tempat_Lahir terhadap variabel IPK dari data mahasiswa lulusan

TEMPAT LAHIR	Mean	N	Minimum	Maximum	% of Total N
KOTA GORONTALO	1,87	605	RENDAH	TINGGI	43,9%
KABUPATEN GORONTALO	1,85	271	RENDAH	TINGGI	19,7%
BONE BOLANGO	1,86	174	RENDAH	TINGGI	12,6%
BOALEMO	1,81	21	RENDAH	**SEDANG**	1,5%
POHUWATO	1,65	26	RENDAH	**SEDANG**	1,9%
GORUT	1,87	24	RENDAH	**SEDANG**	1,7%
LUAR DAERAH GORONTALO	1,82	257	RENDAH	TINGGI	18,7%
Total	1,85	1378	RENDAH	TINGGI	100,0%

Tabel 33. Perbandingan rata-rata variabel Jurusan terhadap variabel IPK dari data mahasiswa lulusan

JURUSAN	Mean	N	Minimum	Maximum	% of Total N
SIPIL	1,78	225	RENDAH	TINGGI	16,3%
INFORMATIKA	1,83	733	RENDAH	TINGGI	53,2%
ELEKTRO	1,77	135	RENDAH	TINGGI	9,8%
KRIYA	2,14	118	RENDAH	TINGGI	8,6%
ARSITEKTUR	1,84	100	RENDAH	TINGGI	7,3%
INDUSTRI	2,00	67	RENDAH	TINGGI	4,9%
Total	1,85	1378	RENDAH	TINGGI	100,0%

Tabel 34. Perbandingan rata-rata variabel Jenjang terhadap variabel IPK dari data mahasiswa lulusan

JENJANG	Mean	N	Minimum	Maximum	% of Total N
D3	1,84	1314	RENDAH	TINGGI	95,4%
S1	2,14	64	**SEDANG**	**TINGGI**	4,6%
Total	1,85	1378	RENDAH	TINGGI	100,0%

Tabel 35. Perbandingan rata-rata variabel Status terhadap variabel IPK dari data mahasiswa lulusan

STATUS	Mean	N	Minimum	Maximum	% of Total N
NON PENDIDIKAN	1,84	1314	RENDAH	TINGGI	95,4%
PENDIDIKAN	2,14	64	**SEDANG**	**TINGGI**	4,6%
Total	1,85	1378	RENDAH	TINGGI	100,0%

Tabel 36. Perbandingan rata-rata variabel Predikat terhadap variabel IPK dari data mahasiswa lulusan

PREDIKAT	Mean	N	Minimum	Maximum	% of Total N
MEMUASKAN	1,07	276	**RENDAH**	**SEDANG**	20,0%
SANGAT MEMUASKAN	2,02	1071	RENDAH	TINGGI	77,7%
TERPUJI	3,00	31	**TINGGI**	**TINGGI**	2,2%
Total	1,85	1378	RENDAH	TINGGI	100,0%

Tabel 37. Perbandingan rata-rata variabel Lama_Studi terhadap variabel IPK dari data mahasiswa lulusan

LAMA STUDI	Mean	N	Minimum	Maximum	% of Total N
TEPAT WAKTU	2,11	189	**SEDANG**	**TINGGI**	13,7%
MELAMPAUI	1,81	1189	RENDAH	TINGGI	86,3%
Total	1,85	1378	RENDAH	TINGGI	100,0%

Berdasarkan beberapa tabel di atas dapat dijabarkan bahwa sebagian besar sebaran data dari nilai rata-rata setiap variabel bebas terhadap variabel terikat IPK, minimum berada kategori IPK **rendah** dan maksimum berada pada kategori IPK **tinggi**, kecuali pada variabel Tempat_Lahir (Boalemo, Pohuwato dan Gorut) dan variabel Predikat (Memuaskan) yang minimum berada pada kategori IPK **rendah** dan maksimum hanya berada pada kategori IPK **sedang**. Variabel IPK (S1), variabel Status (Pendidikan) dan variabel Lama_Studi (Tepat Waktu) minimum berada pada kategori IPK **sedang** dan maksimum hanya berada pada kategori IPK **tinggi,** serta variabel Predikat (Terpuji) yang minimum dan maksimumnya berada pada IPK dengan kategori **tinggi**.

BAB V
KESIMPULAN DAN SARAN

A. Kesimpulan

1. Hasil klasifikasi dengan menggunakan metode *Decission Tree* untuk data mahasiswa aktif berdasarkan variabel IPK diperoleh 7 klasifikasi dan 10 klasifikasi berdasarkan variabel Alamat_Asal. Sedangkan hasil klasifikasi data mahasiswa lulusan berdasarkan variabel IPK diperoleh 5 klasifikasi dan 6 klasifikasi berdasarkan variabel Lama_Studi.
2. Ketepatan prediksi mahasiswa aktif berdasarkan kelas target IPK memiliki presentase yang lebih besar daripada kelas target Alamat_Asal, yang menunjukkan bahwa variabel IPK memiliki nilai lebih baik dari Alamat_Asal untuk dijadikan sebagai kelas target dalam proses prediksi. Persentase ketepatan uji prediki yang dihasilkan cenderung rendah (kurang dari 70%). Sedangkan ketepatan prediksi mahasiswa lulusan berdasarkan kelas target IPK memiliki presentase yang lebih besar daripada kelas target Lama_Studi, yang menunjukkan bahwa IPK memiliki nilai lebih baik dari Lama_Studi untuk dijadikan sebagai kelas target dalam proses prediksi. Apabila dibandingkan dengan persentase ketepatan uji prediksi yang diperoleh pada klasifikasi data aktif mahasiswa, maka hasil persentase ketepatan uji prediksi pada klasifikasi data lulusan lebih tinggi (lebih dari 85%).
3. Tingkat sensitivitas variabel bebas pada data mahasiswa aktif yang dihasilkan oleh Kurva ROC memiliki kesamaan dengan *Independent Variabel Importance* yang dihasilkan oleh metode *Decission Tree* dan K-*Neirest Neighbour*, sedangkan tingkat sensitivitas variabel bebas pada data mahasiswa lulusan yang dihasilkan oleh Kurva ROC memiliki kesamaan dengan *Independent Variabel Importance* yang dihasilkan oleh ketiga metode klasifikasi yang digunakan sebelumnya, yaitu metode *Decission Tree, Artificial Neural Network* (ANN) dan K-*Neirest Neighbour* (K-NN).
4. Sebagian besar sebaran data dari nilai rata-rata setiap variabel bebas terhadap variabel terikat IPK, minimum berada kategori IPK rendah dan maksimum

berada pada kategori IPK tinggi, kecuali pada variabel Pekerjaan_Ortu (TNI/POLRI dan Tidak Bekerja), Variabel Penghasilan_Ortu (5 – 10 juta), variabel Asal_Sekolah (Madrasah) dan varaibel Alamat_Asal (Gorut) yang maksimum hanya berada pada kategori IPK sedang. Begitu pula halnya dengan data mahasiswa lulusan, Sebagian besar sebaran data dari nilai rata-rata setiap variabel bebas terhadap variabel terikat IPK, minimum berada kategori IPK rendah dan maksimum berada pada kategori IPK tinggi, kecuali pada variabel Tempat_Lahir (Boalemo, Pohuwato dan Gorut) dan variabel Predikat (Memuaskan) yang minimum berada pada kategori IPK **rendah** dan maksimum hanya berada pada kategori IPK **sedang**. Variabel IPK (S1), variabel Status (Pendidikan) dan variabel Lama_Studi (Tepat Waktu) minimum berada pada kategori IPK **sedang** dan maksimum hanya berada pada kategori IPK **tinggi,** serta variabel Predikat (Terpuji) yang minimum dan maksimumnya berada pada IPK dengan kategori **tinggi.**

B. Saran

1. Data mahasiswa baik data mahasiswa aktif maupun data lulusan masih banyak yang tidak tersedia, sehingga disarankan kepada lembaga agar senantiasa memperbaharui data agar selalu lengkap dan tersedia serta mudah untuk diakses.
2. Hasil klasifikasi yang digunakan untuk memprediksi IPK dan Lama Studi dari mahasiswa aktif menunjukkan berbagai klasifikasi serta daftar variabel-variabel penting yang dapat menentukan variabel IPK dan Lama Studi, sehingga disarankan bagi lembaga agar dapat memperhatikan kedua hal tersebut untuk membantu meminimalisir potensi perolehan IPK dengan kategori rendah dan lama studi yang melampaui batas masa studi.

DAFTAR PUSTAKA

Al-Radaideh, Q.A., Al-Shawakfa, E.M dan Al-Najjar, M.I. 2006. *Mining Student Data Using Decission Tree. International Arab Conference on Informational Technology* (ACIT).

Balakrishnan & Ling. 2012. EM *Algorithm for One-Shot Device Testing Under The Exponential Distribution*. International Journal Computational Statistics & Data Analysis, Vol 56, Issue 3. Netherlands.

Budiarti, A. Sucahyo, G,Y. Ruldeviyani, Y. 2006. Studi Karakteristik Kelulusan Peserta Didik dengan Menggunakan Teknik *Clustering.* Proseding Seminar Nasional Sistem dan Informatika. Bali.

Fahim, A.M. Saake, G. Salem, A.M. Torkey, F.A & Ramadan, M.A. 2008. K-*Means for Spherical Cluter with Large Variance in Size*. Proceedings of International Conference on Mathematical, Computational and Statistical Sicence and Engineering.

Gunawan. 2009. Klasterisasi Data Kategorikal Menggunakan Algoritma K-Modes. Jurnal Muria Sains. Universitas Muria Kudus.

Han & Kamber. 2006. *Data Mining Concepts and Techniques* Edisi ke-2. San Fransisco : Morgan Kaufmann publisher.

Jenanto A. 2010. Penggunaan Algoritma SLIQ untuk Pengklasifikasian Kinerja Akademik Mahasiswa. Jurnal Teknologi Informasi DINAMIK Vol XV, No.1 : 66-72

Kantardzic M. 2003. *Data Mining : Concepts, Model, Methods, and Algorithm*. New Jersey : John Willey& Sons Inc.

Karlita, T. 2011. Klasterisasi Data Kategorikal dengan Menggunakan Algiritma *Modes Linkage.* Proseding Industrial Electronics Seminar. Surabaya.

Larose, D. T. 2005. *Discovering Knowledge in Data.*Canada : Wiley-Interscience.

Leidiyana H. 2011. Komparasi Algoritma Klasifikasi Data Mining dalam Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. Tesis. STMIK Nusa Mandiri. Yogyakarta.

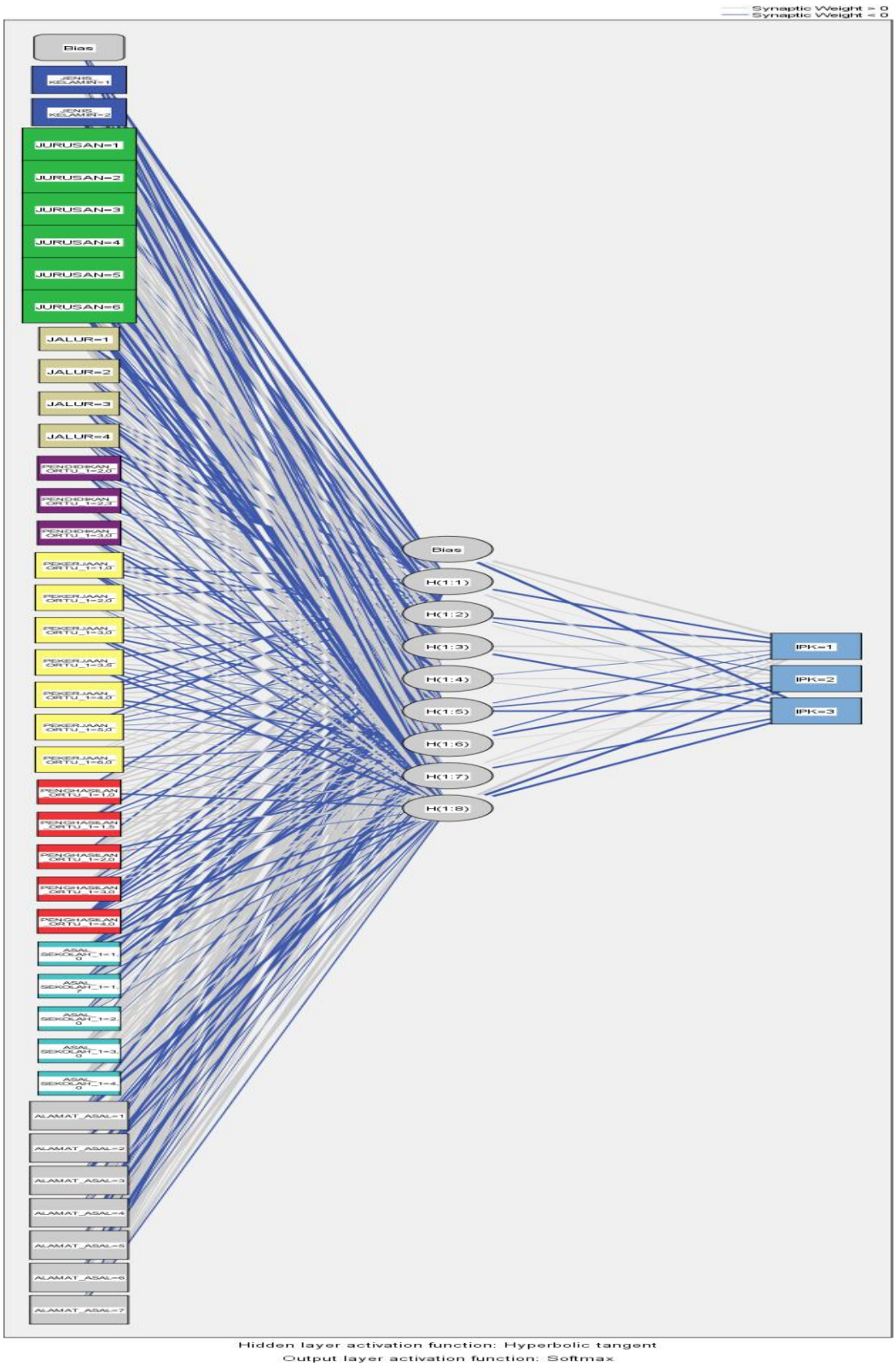
Martiana, E. Mutbada'i, N.R. Purnomo, E. 2009. Penggunaan Metode Pengklasteran Untuk Menentukan Bidang Tugas Akhir Mahasiswa Teknik Informatika PENS Berdasarkan Nilai. Proseding Industrial Electronics Seminar. Surabaya.

Pramudyo, A. S. 2008. *Case Base Reasoning* untuk Klasifikasi Mahasiswa Baru berdasarkan prediksi Indeks Prestasi Semester I (studi kasus Program Studi Teknik Informatika Universitas Bina Darma Palembang). Tesis. Universitas Gajah Mada. Yogyakarta.

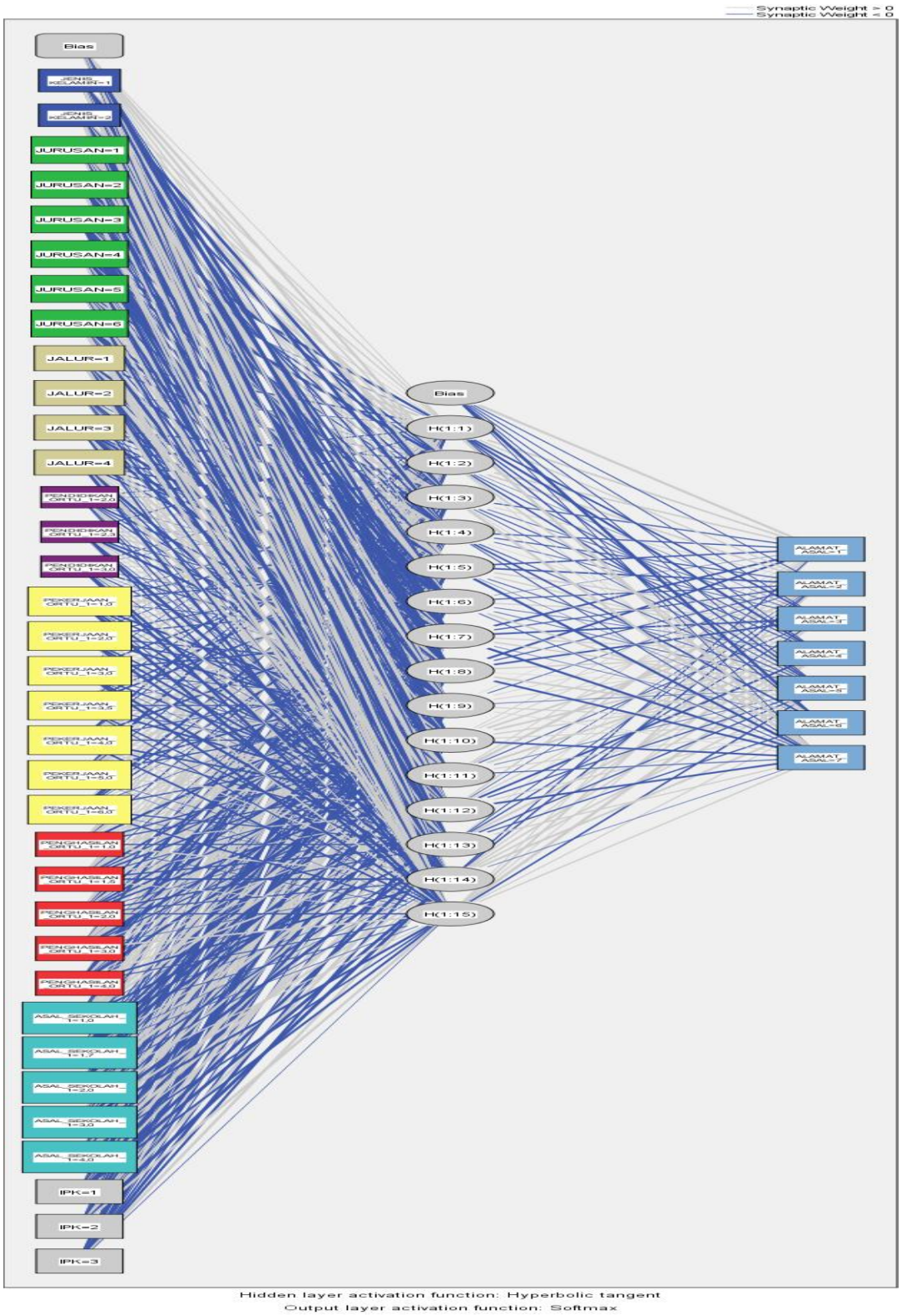
Sunjana. 2010. Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree. Proseding Seminar Nasional Aplikasi Teknologi Informasi. Yogyakarta.

Susanto, S., & Suryadi, D. 2010.Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data.CV Andi Offset.Yogyakarta.

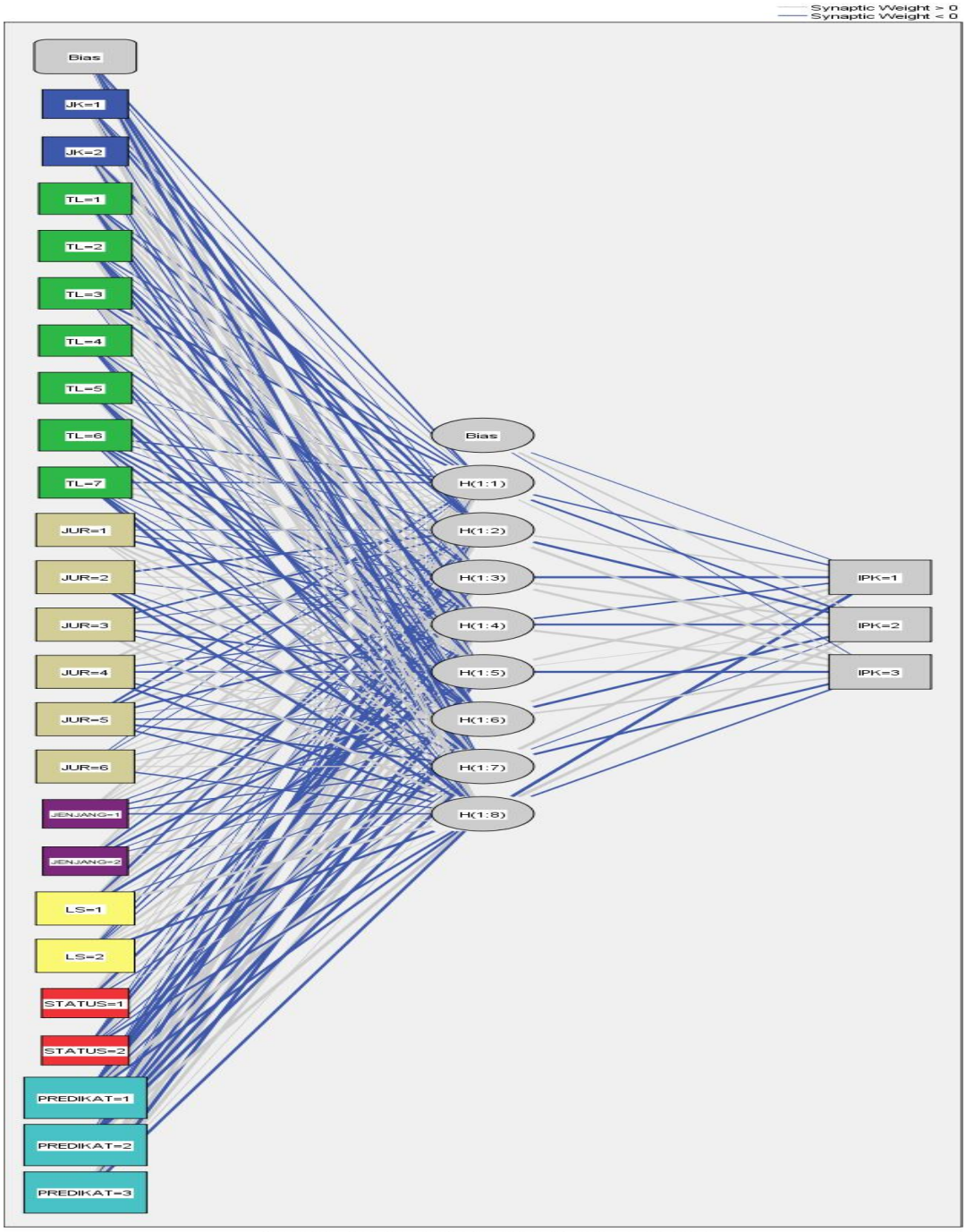
Lampiran 1. Klasifikasi Data Mahasiswa Aktif dengan menggunakan metode *Artificial Neural Network* berdasarkan IPK



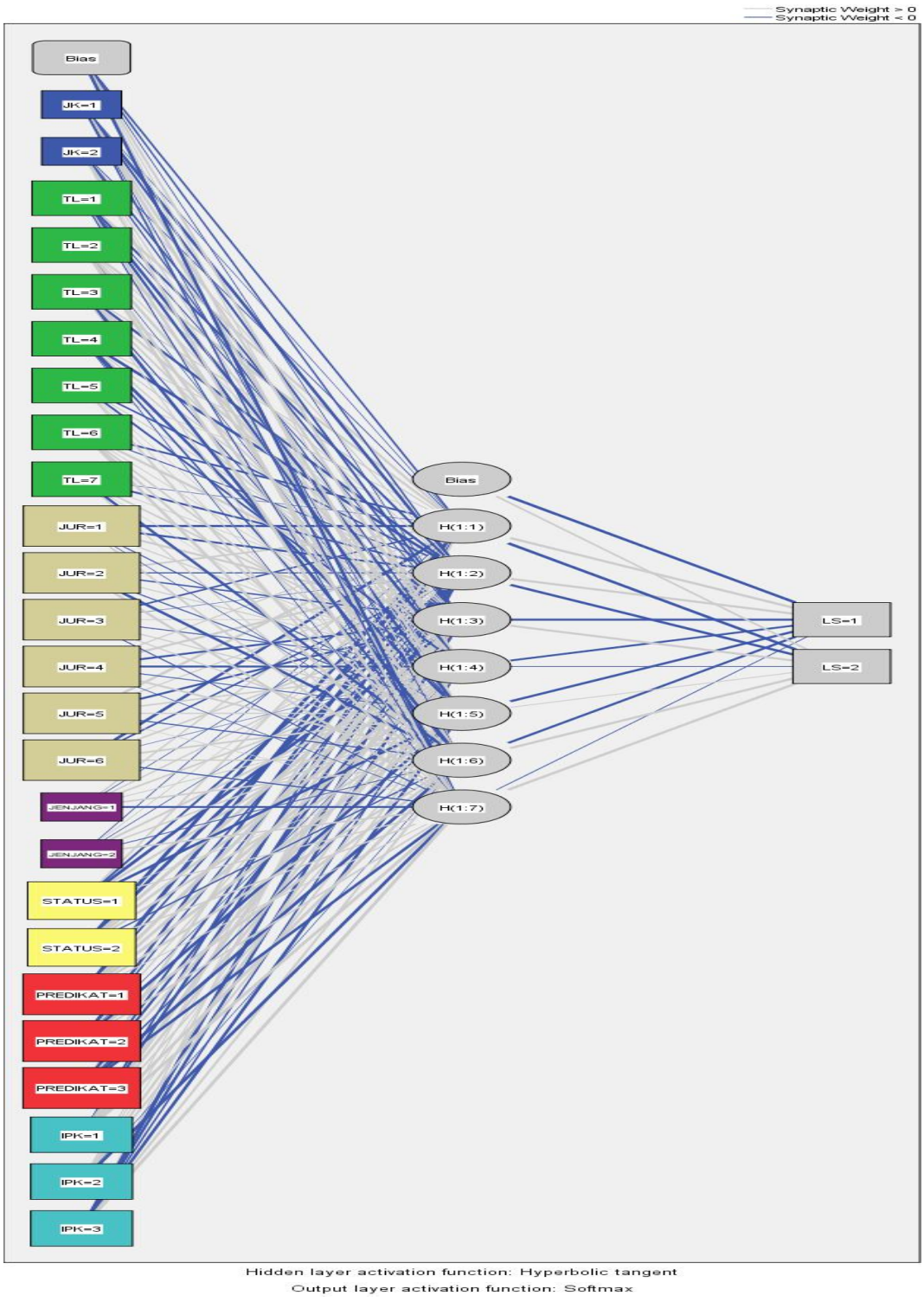
Lampiran 2. Klasifikasi Data Mahasiswa Aktif dengan menggunakan metode *Artificial Neural Network* berdasarkan Alamat_Asal



Lampiran 3. Klasifikasi Data Mahasiswa Lulusan dengan menggunakan metode *Neural Network* berdasarkan IPK



Lampiran 4. Klasifikasi Data Mahasiswa Lulusan dengan menggunakan metode *Neural Network* berdasarkan Lama_Studi



IDENTITAS PENELITI

1	Nama Lengkap (dengan gelar)	Lillyan Hadjaratie, S.Kom, M.Si
2	Jabatan Fungsional	Lektor
3	Jabatan Struktural	-
4	NIP	198004172002122002
5	NIDN	00187048001
6	Tempat dan Tanggal Lahir	Gorontalo, 17 April 1980
7	Alamat Rumah	Jln. Gelatik No. 17 Kota Gorontalo
8	Nomor Telepon/Faks/HP	081356139486
9	Alamat Kantor	Jln. Jenderal Sudirman No. 6 Kota Gorontalo
10	Nomor Telepon/Faks	0435-821125 / 0435-821752
11	Alamat email	lillyan.hadjaratie@gmail.com
12	Lulusan yang telah dihasilkan	± 100 orang
13	Mata kuliah yang diampu	1. Desain Sistem Basis Data 2. Sistem Data Warehouse 3. Pengantar Kecerdasan Buatan 4. Sistem Informasi Geografis

1. Riwayat Pendidikan

	S1	**S2**
Nama Perguruan Tinggi	STMIK Dipanegara Makassar	Institut Pertanian Bogor
Bidang Ilmu	Manajemen Informatika	Ilmu Komputer
Tahun Masuk-Lulus	1998-2002	2006-2011
Judul Skripsi/Thesis	Sistem Informasi Inventarisasi di Fakultas MIPA Universitas	Jaringan Saraf Tiruan untuk Prediksi Tingkat Kemajuan Studi

	Hasanuddin Makassar	Mahasiswa Diploma Manajemen Informatika UNG
Nama Pembimbing	Ir. Muh Hasbi, M.Kom	Prof. Dr. Ir. Kudang Boro Seminar, M.Sc

2. Pengalaman Penelitian

No	Tahun	Judul Penelitian	Pendanaan	
			Sumber	Jumlah (Juta Rp)
1	2006	Analisis Pengembangan e-Government di Provinsi Gorontalo	Penelitian Dosen Muda DIKTI	10 juta
2	2008	Analisis dan Desain Sistem Informasi Petani di Provinsi Gorontalo	-	-
3	2010	Studi Komparatif Sistem Pakar dan Sistem Pengambilan Keputusan di Bidang Pertanian	-	-

3. Pengalaman Pengabdian kepada Masyarakat

No	Tahun	Judul Pengabdian	Pendanaan	
			Sumber	Jumlah (Juta Rp)
1	2006	Pelatihan Dasar Komputer bagi Guru Sekolah Dasar	UNG	3 juta
2	2011	Pelatihan Pemanfaatan Teknologi Informasi di Lingkungan Kerja Dinas Pendidikan Kota Gorontalo	-	-
3	2011	Pelatihan Pemanfaatan Sistem Informasi Layanan Kelurahan di Kota Gorontalo	UNG	5 juta

4. Pengalaman Penulisan Artikel dalam Jurnal Ilmiah

No	Judul Artikel Ilmiah	Volume/ Nomor/ Tahun	Nama Jurnal
1	Analisis Pengembangan e-Government di Provinsi Gorontalo	2006	Jurnal Teknik UNG
2	Analisis dan Desain Sistem Informasi Petani di Provinsi Gorontalo	2008	Jurnal Ilmiah Agropolitan
3	Studi Komparatif Sistem Pakar dan Sistem Pengambilan Keputusan di Bidang Pertanian	2010	Jurnal Ilmiah Agropolitan

Gorontalo, Oktober 2012

Lillyan Hadjaratie, S.Kom, M.Si
NIP. 19804172002122002