

## **Penambangan data formulir pendaftaran untuk memprediksi kinerja mahasiswa Prodi Pendidikan Matematika UNISMA menggunakan CHAID**

**Siti Nurul Hasana<sup>1,\*</sup>**

<sup>1</sup>*Pendidikan Matematika, Universitas Islam Malang, Jl. Mayjen Haryono No. 193 Malang*

*\*Corresponding author: s.nurulhasana@unisma.ac.id*

---

### **Abstrak**

Tujuan penelitian ini adalah menggali lebih dalam data formulir pendaftaran mahasiswa baru untuk memprediksi prestasi belajar mahasiswa yang rendah. Perguruan Tinggi sebenarnya telah memiliki basis data yang cukup sejak penerimaan mahasiswa baru untuk dapat dimanfaatkan sebagai sumber informasi maupun tambahan pertimbangan dalam upaya menjaga kualitas dan profesionalitas pelayanan pendidikan di masa yang kompetitif. Mengetahui lebih dini faktor-faktor yang berhubungan dengan masalah prestasi belajar mahasiswa merupakan bentuk upaya untuk menjaga kualitas kinerja mahasiswa yang merupakan salah satu variabel yang ikut menentukan tingkat kualitas pelayanan pendidikan yang disediakan Perguruan Tinggi. Penelitian ini dilaksanakan di Program Studi Pendidikan Matematika, FKIP, Universitas Islam Malang (UNISMA). Data yang dipergunakan adalah data formulir pendaftaran mahasiswa baru Prodi Pendidikan Matematika tahun 2016, 2017, dan 2018. Analisa data dari formulir pendaftaran dilakukan dengan menerapkan Chi-square Automatic Interaction Detector (CHAID) dengan pohon keputusan sebagai hasil akhir pemodelan prediksinya sehingga akan lebih mudah dibaca dan dipahami. Avilla (dalam Hasana, 2013) menjelaskan bahwa CHAID merupakan suatu teknik untuk mendeteksi adanya interaksi dalam sebuah model prediksi. Hasil analisa CHAID dalam penelitian ini menunjukkan bahwa mahasiswa yang berpotensi memiliki kinerja rendah adalah mahasiswa dengan karakteristik yang melibatkan variabel Pendapatan Ayah, Jenis Kelamin, Pilihan Prodi, Jenis Asal Sekolah, dan Hobi. Hasil analisa ini diharapkan dapat menjadi salah satu pertimbangan untuk pengambilan keputusan dan kebijakan di Prodi Pendidikan Matematika, sehingga permasalahan kinerja mahasiswa yang rendah dapat segera teridentifikasi dan teratasi sedini mungkin.

**Kata-kata kunci:** *CHAID; formulir pendaftaran mahasiswa baru; kinerja mahasiswa*

### **Abstract**

*The purpose of this study is to dig deeper or to mine data from student registration forms to predict the low student learning achievements. The University actually has had a sufficient database since accepting the prospective new students which can be utilized as resources or additional considerations in efforts to maintain the quality and professionalism of educational services in this competitive era. Knowing earlier about the factors related to the problem of student learning achievement is a form of effort to maintain the quality of student performance which is one of the variables determining the level of educational service quality provided by universities. This study was conducted at Mathematics Education Study Program, Teacher Training and Education Faculty, University of Islam Malang (UNISMA). Data used in this research is extracted from registration forms of Mathematics Education students in 2016, 2017, and 2018. The analysis of registration form data was done by applying the Chi-square Automatic Interaction Detector (CHAID) with a decision tree as the final result of prediction modeling so that it will be easier to read and understand. Avilla (in Hasana, 2013) explains that CHAID is a technique for detecting interactions in a prediction model. The results of CHAID analysis show that students who have the opportunity to gain low*

*performance are students with characteristics which are involving the variables of Father's Salary, Gender, Study Program Choice, Types of Senior High School, and Hobby. Those five independent variables are predicted to have associations with low student performance. The results of this analysis are expected to be one of the considerations for Mathematics Education Study Program to make decisions and policies, so that the problems of low student performance can be immediately identified and resolved as early as possible.*

***Keywords:*** CHAID; student registration form; student's performance

---

## **PENDAHULUAN**

Semakin pesatnya perkembangan dunia pendidikan di Indonesia menunjukkan peningkatan dalam persaingan jasa pendidikan. Hal ini menuntut para pengelola perguruan tinggi untuk semakin kompetitif dan mampu memberikan pelayanan dan pengelolaan yang lebih profesional dan berkualitas. Tuntutan ini dirasakan pula oleh Program Studi Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan (FKIP), Universitas Islam Malang (UNISMA). Informasi dan pengetahuan berbasis data banyak sekali dibutuhkan sebagai sumber informasi dan bahan pertimbangan dalam proses mempertahankan dan meningkatkan pelayanan pendidikan. Analisis data akademik yang dimiliki dapat menjadi salah satu bentuk upaya yang dapat dilakukan oleh Program Studi Pendidikan Matematika demi menjaga kualitasnya. Hal ini dikarenakan hasil analisa tersebut akan dapat memberikan pengetahuan yang selanjutnya dapat digunakan sebagai bahan pertimbangan dalam menentukan kebijakan yang tepat oleh pengelola, menjaga dan memperbaiki kinerja mahasiswa, mengevaluasi proses pembelajaran, serta melakukan langkah-langkah lainnya sebagai upaya mengontrol kualitas.

Penambangan data (data mining) merupakan teknologi yang sesuai untuk menjadi alat bantu otomatis dalam proses pengambilan keputusan dan penentuan kebijakan. Penambangan data yang memanfaatkan informasi statistik dan atau algoritma machine learning untuk menemukan pola yang menarik dari data pendidikan biasa disebut sebagai Educational Data Mining (EDM) (Elakia et.al., 2014). Proses penambangan data akan mampu mengekstraksi data yang selanjutnya dapat diperoleh informasi yang sedalam-dalamnya dan sebanyak-banyaknya serta menyajikan informasi tersebut secara lebih sederhana untuk dapat dengan mudah dipelajari (Marthasari, 2017).

Kinerja mahasiswa merupakan salah satu indikator yang biasa digunakan Perguruan Tinggi untuk menjaga kualitas pelayanan pendidikan dan pembelajaran yang diberikan. Prediksi kinerja mahasiswa yang rutin dilakukan akan bermanfaat untuk mengetahui permasalahan prestasi akademik mahasiswa sejak dini. Mahasiswa yang diprediksi memiliki

kinerja rendah akan dapat cepat teridentifikasi dan mendapatkan bantuan sehingga dapat diperbaiki untuk kedepannya. Permasalahan bisa dapat teridentifikasi dan solusi dapat segera diambil.

Di perguruan tinggi, keberhasilan belajar mahasiswa ditunjukkan dengan prestasi belajar yang dicapainya berdasarkan evaluasi hasil belajar di kelas. Menurut Suryanto (dalam Slameto, 2010), prestasi belajar adalah keseluruhan kompetensi dan hasil yang dicapai dari proses belajar di suatu institusi atau lembaga pendidikan yang dinyatakan dalam nilai-nilai. Prestasi belajar mahasiswa merupakan salah satu unsur yang digunakan untuk mengukur kinerja mahasiswa. Ukuran yang digunakan dalam menghitung prestasi hasil belajar mahasiswa adalah melalui Indeks Prestasi Kumulatif (IPK). IPK sebagai tolak ukur prestasi belajar mahasiswa menjadi salah satu faktor penentu kualitas dari perguruan tinggi. Terbukti dalam borang akreditasi program studi yang ditetapkan dalam Peraturan Badan Akreditasi Nasional PT No 4/2017, IPK menjadi salah satu faktor pada bagian data Mahasiswa dan Lulusan. Semakin tinggi rata-rata IPK mahasiswa maka penilaian pada isian borang pun akan mendapat nilai yang lebih maksimal. Hal ini menunjukkan bahwa kualitas dari suatu program studi pun akan ditetapkan semakin baik.

Prediksi prestasi belajar mahasiswa yang rutin dilakukan akan bermanfaat untuk mengetahui permasalahan kinerja akademik mahasiswa sejak dini. Baradwaj dan Pal (dalam Ahmed dan Elaraby, 2013) menyampaikan bahwa tujuan utama perguruan tinggi adalah menyediakan pendidikan yang berkualitas bagi mahasiswa, dan salah satu yang dapat menjadi upaya untuk mencapai tingkat kualitas terbaik dalam pendidikan tinggi adalah dengan memprediksi kinerja mahasiswa. Banyak sekali pendekatan yang dapat dilakukan dalam melakukan prediksi ini, salah satunya adalah menggunakan metode klasifikasi dengan pohon keputusan. Dalam penelitiannya, Ahmed dan Elaraby (2013) menggunakan klasifikasi melalui algoritma ID3 untuk memprediksi kinerja mahasiswa. Data dalam penelitian Ahmed dan Elaraby (2013) diperoleh dari basis data mahasiswa yang dimiliki suatu institusi pendidikan tinggi, yang selanjutnya dipergunakan untuk memprediksi nilai akhir mahasiswa.

Penelitian lain yang bertujuan untuk memprediksi kinerja peserta didik dilakukan oleh Ramaswami dan Bhaskaran (2010) di suatu sekolah menengah atas di India. Ramaswami dan Bhaskaran (2010) menjelaskan bahwa kinerja akademik peserta didik dipengaruhi oleh banyak faktor, sehingga diperlukan model penambangan data yang bersifat prediktif untuk hal ini agar kinerja peserta didik yang rendah beserta faktor yang paling

dominan berpengaruh dapat segera teridentifikasi. Penelitian Ramaswami dan Bhaskaran (2010) menggunakan model prediksi CHAID dengan data diperoleh dari siswa dan basis data di sekolah. Hasil penelitian Ramaswami dan Bhaskaran (2010) menunjukkan bahwa CHAID dapat bekerja secara efektif dengan akurasi prediksi yang lebih baik sehingga sangat berguna untuk dipergunakan menganalisa hubungan antar variabel yang berpengaruh terhadap kinerja siswa sekolah menengah atas.

Ramaswami dan Bhaskaran (2010) menyebutkan bahwa kinerja peserta didik bergantung pada beragam faktor seperti kepribadian, kondisi sosial ekonomi, kondisi psikologis serta berbagai variabel-variabel lingkungan yang lain. Perguruan tinggi, termasuk UNISMA, sebenarnya selalu rutin memperoleh data-data tersebut sejak membuka registrasi untuk calon mahasiswa baru. Proses pendaftaran masuk ke perguruan tinggi selalu melalui tahapan mengisi formulir pendaftaran yang sebagian besar berisi data demografis yang berisi data-data yang dapat berpengaruh terhadap kinerja mahasiswa nantinya. Data-data ini setiap tahunnya dihimpun dan menjadi database Perguruan Tinggi. Diperlukan penambahan data yang lebih mendalam untuk dapat mengoptimalkan database ini agar tidak hanya menjadi arsip tetapi juga memberikan pengetahuan lebih yang bisa dimanfaatkan untuk kepentingan kontrol kualitas. Oleh karena itu dilakukanlah suatu penelitian untuk memprediksi kinerja mahasiswa Program Studi Pendidikan Matematika UNISMA melalui data-data demografis mahasiswa yang terhimpun dari formulir pendaftaran yang diisi ketika mahasiswa melakukan proses registrasi awal dengan menggunakan CHAID.

Menurut Gallagher (dalam Kunto dan Hasana, 2006), Chi-square Automatic Interaction Detector (CHAID) merupakan teknik pendugaan variabel dependen melalui proses iteratif yang menguji satu persatu beberapa variabel independen yang terlibat dalam klasifikasi dan selanjutnya variabel-variabel independen tersebut akan diurutkan berdasarkan pada tingkat signifikansi statistik chi-square terhadap variabel dependennya. Proses ini selanjutnya akan berhenti setelah tidak ditemukan lagi variabel independen yang signifikan secara statistik. Hasil dari proses CHAID ini disusun dalam bentuk diagram pohon yang tersusun hirarkis mewakili suatu model prediksi dimana bagian akar pohon ini adalah populasi yang diteliti dan batang-batangnya menghubungkan segmen-segmen yang merupakan variasi dari variabel respon yang terlibat (Ramaswami dan Bhaskaran, 2010). Analisis CHAID digunakan ketika data yang dipakai adalah data dengan variabel-variabel kategorik. Metode CHAID hanya efektif apabila diterapkan pada data dengan pengamatan

yang banyak (du Toit et al, 1986). Avilla (dalam Hasana, 2013) menyebutkan bahwa target utama dari teknik CHAID adalah untuk mendeteksi adanya interaksi dalam suatu model prediksi.

Langkah paling mendasar dalam proses CHAID adalah memilih karakteristik khusus yang relevan pada proses klasifikasi dalam pembentukan model prediksi (Witten dan Frank, 2005). Sharp et al. (dalam Kunto dan Hasana, 2006) menyebutkan 3 elemen utama CHAID sebagai berikut:

1) Uji Chi-square

Menurut Liu dan Setiono (1995), Chi-square merupakan uji statistika yang umum untuk mengukur perbedaan antar variabel.

2) Koreksi Bonferroni

Ritschard (dalam Hasana, 2013) menyatakan bahwa salah satu karakteristik utama dari CHAID adalah penggunaan koreksi Bonferroni untuk nilai p-value yang didapatkan. Koreksi Bonferroni adalah suatu proses koreksi yang digunakan ketika beberapa uji statistik untuk kebebasan atau ketidakbebasan dilakukan secara bersamaan (Sharp et.al dalam Hasana, 2013)

3) Algoritma penggabungan kategori-kategori variabel.

Kass (dalam Hasana, 2013) menjelaskan bahwa ada 3 jenis variabel prediktor dalam CHAID, yaitu:

1) Prediktor Monotonik (*Monotonic Predictor*)

Prediktor monotonik ini adalah prediktor berbentuk kategori yang berskala ordinal. CHAID dapat menggabungkan atau mengkombinasikan kategori-kategori pada variabel jenis ini hanya jika keduanya berdekatan satu sama lain. Dengan kata lain, variabel-variabel ini memiliki kategori yang mengikuti urutan aslinya, contohnya usia atau pendapatan.

2) Prediktor Bebas (*Free Predictor*)

Prediktor bebas ini adalah prediktor berbentuk kategori yang berskala nominal. CHAID dapat menggabungkan atau mengkombinasikan kategori-kategori pada variabel jenis ini walaupun keduanya berdekatan atau tidak satu sama lain. Jadi kombinasi apapun memungkinkan untuk terbentuk, contohnya pekerjaan, kelompok etnik, dan area geografis.

3) Prediktor Mengambang (*Floating Predictor*)

Pada prakteknya di banyak kasus, sering ditemukan kategori-kategori prediktor yang berskala ordinal tetapi di dalamnya memiliki satu kategori yang tidak termasuk ordinal, atau dengan kata lain posisinya dalam skala ordinal itu tidak diketahui. Prediktor jenis ini disebut dengan istilah “mengambang”. CHAID akan memperlakukan kategori-kategori pada variabel ini seperti monotonik kecuali untuk kategori yang mengambang, yang dapat berkombinasi dengan kategori manapun. Contoh: data usia dengan kategori 25-34 tahun, 35-44 tahun, 45-54 tahun, 55-64 tahun, > 64 tahun, tidak diketahui.

Penelitian ini akan menambang data formulir pendaftaran mahasiswa Prodi Pendidikan Matematika dan mengekstraknya menjadi variabel-variabel prediktor. Selanjutnya variabel-variabel prediktor ini akan disusun dalam kategori-kategori yang dapat digunakan dalam teknik CHAID untuk memprediksi kinerja mahasiswa melalui nilai IPK-nya. Hipotesis penelitian ini menyebutkan bahwa terdapat beberapa variabel prediktor yang dominan berpengaruh terhadap nilai IPK mahasiswa Prodi Pendidikan Matematika.

## **METODE**

Penelitian ini akan menggunakan data demografis dan data akademik dari keseluruhan 3 angkatan mahasiswa Program Studi Pendidikan Matematika, FKIP, UNISMA. Tiga angkatan mahasiswa Program Studi Pendidikan Matematika yang dimaksudkan dalam penelitian ini adalah mahasiswa semester 3, 5, dan 7 pada periode Gasal 2019/2020. Data demografis, yang meliputi data diri, tempat tinggal, kondisi ekonomi mahasiswa dan lain sebagainya, diambil dari basis data UNISMA yang dihimpun melalui formulir pendaftaran calon mahasiswa baru. Sedangkan untuk data akademik berupa nilai IPK mahasiswa diambil dari basis data Program Studi Pendidikan Matematika.

Variabel-variabel dalam formulir pendaftaran yang digunakan untuk memprediksi kinerja mahasiswa disusun dalam bentuk kategori seperti dalam tabel 1 berikut.

**Tabel 1. Data Demografis dalam Penelitian**

<b>Variabel</b>	<b>Keterangan</b>	<b>Kategori</b>
<b>Pilihan Program Studi (PILPRODI)</b>	Program Studi yang dipilih pada Pilihan 1 dan Pilihan 2 ketika mendaftar ke Universitas Islam Malang	1=Pilihan 1; 2=Pilihan 2; 3=Pilihan 1 dan 2
<b>Jalur Beasiswa (BEASISWA)</b>	Data tentang apakah masuk ke Universitas Islam Malang melalui jalur beasiswa	1=Beasiswa; 2=Tidak Beasiswa

Variabel	Keterangan	Kategori
<b>Jenis Kelamin (GENDER)</b>	Data tentang jenis kelamin mahasiswa apakah laki-laki atau perempuan	1=Laki-laki; 2=Perempuan
<b>Status Pernikahan (MARSTAT)</b>	Data tentang status pernikahan calon mahasiswa saat mendaftar ke Universitas Islam Malang	1=Belum Menikah; 2=Menikah
<b>Jenis Asal Sekolah (ASALSEK)</b>	Jenis sekolah pada jenjang SMA/MA/SMK apakah Negeri, Swasta, atau Ma'arif	1=Negeri; 2=Swasta; 3=Ma'arif
<b>Jeda Waktu Masuk Kuliah (FRESHGRAD)</b>	Calon mahasiswa mendaftar ke Universitas Islam Malang langsung setelah lulus SMA/MA/SMK atau tidak	1=Tidak Langsung Kuliah; 2=Langsung Kuliah
<b>Hobi (HOBBY)</b>	Calon mahasiswa memiliki hobi olahraga atau seni, atau hobi dalam keduanya atau tidak hobi dalam keduanya	1=Olahraga; 2=Seni; 3=Olahraga dan Seni; 4=Tidak Keduanya
<b>Pendidikan Terakhir Ayah (PTAYAH)</b>	Jenjang pendidikan terakhir Ayah	1=SD/MI; 2=SMP/MTs; 3=SMA/MA/SMK; 4=Diploma; 5=Sarjana; 6=Lainnya
<b>Pendidikan Terakhir Ibu (PTIBU)</b>	Jenjang pendidikan terakhir Ibu	1=SD/MI; 2=SMP/MTs; 3=SMA/MA/SMK; 4=Diploma; 5=Sarjana; 6=Lainnya
<b>Pekerjaan Ayah (JOBAYAH)</b>	Jenis pekerjaan Ayah	1=PNS/TNI/POLRI; 2=Guru/Dosen; 3=Petani/Pekebun/Nelayan; 4=Swasta; 5=Wiraswasta; 6=Buruh; 7=Lainnya
<b>Pekerjaan Ibu (JOBIBU)</b>	Jenis pekerjaan Ibu	1=PNS/TNI/POLRI; 2=Guru/Dosen; 3=Petani/Pekebun/Nelayan; 4=Swasta; 5=Wiraswasta; 6=Buruh; 7=Ibu Rumah Tangga; 8=Lainnya
<b>Pendapatan Ayah (SALAYAH)</b>	Besar pendapatan Ayah setiap bulan	1=Sangat Tinggi; 2-Tinggi; 3=Sedang; 4=Rendah; 5=Tidak Menjawab
<b>Tempat Tinggal Saat ini (STAYNOW)</b>	Tempat tinggal mahasiswa apakah dengan orang tua, kos, pondok pesantren, atau yang lainnya.	1=Dengan Orang Tua; 2=Kos; 3=Pondok Pesantren; 4=Lainnya

Keseluruhan variabel-variabel tersebut akan digunakan untuk memprediksi mahasiswa dengan kinerja akademik rendah. Karena kinerja mahasiswa yang rendah merupakan salah satu bentuk masalah yang harus diselesaikan untuk menjaga kualitas Program Studi, maka dalam penelitian ini, kinerja mahasiswa yang rendah dijadikan target analisis.

Data yang telah dihimpun selanjutnya dianalisa menggunakan teknik analisa CHAID. Diepen dan Franses (dalam Hasana, 2013) menyebutkan bahwa CHAID adalah suatu algoritma non-parametrik dimana tidak ada asumsi distribusi yang harus dipenuhi untuk datanya. Gallagher (dalam Kunto dan Hasana, 2006) menjelaskan 4 langkah iteratif CHAID sebagai berikut.

- 1) Setiap variabel independen yang didapa diperiksa satu per satu menggunakan uji Chi-square untuk menentukan kategori mana yang signifikan untuk menunjukkan perbedaan terhadap variabel dependen, dan sekaligus juga mengumpulkan semua kategori yang tidak signifikan.
- 2) Variabel independen mana yang paling signifikan ditentukan untuk menjadi yang terbaik dalam membedakan variabel dependen dan disusun berurutan berdasarkan nilai kesignifikanan hasil uji yang dilakukan.
- 3) Keseluruhan kategori variabel independen dibagi dan disusun berdasarkan peringkat yang paling signifikan.
- 4) Untuk setiap tingkatan yang terbentuk selanjutnya dilakukan langkah-langkah:
  - a) Pemeriksaan kategori variabel-variabel independen yang tersisa untuk menentukan peringkat yang paling signifikan dalam penentuan perbedaan variabel dependen selanjutnya, dan memisahkannya dengan yang tidak signifikan.
  - b) Penentuan variabel independen mana yang paling signifikan dan kemudian diteruskan lagi dengan pembagian datanya menggunakan variabel ini.
  - c) Pengulangan langkah ke-4 untuk semua subgrup sampai teridentifikasi semua pembagian yang telah signifikan secara statistik.

Magidson (dalam Hasana, 2013) menerangkan bahwa langkah-langkah analisis CHAID dapat dibagi menjadi tiga tahap, yaitu Penggabungan, Pemisahan dan Penghentian. Proses-proses yang terjadi dalam tahapan tersebut dijabarkan sebagai berikut.

### **1) Tahap Penggabungan**

Pada tahap ini, untuk setiap prediktor ( $X_1, X_2, \dots, X_k$ ) akan dilakukan proses seperti di bawah ini.

- a) Membentuk tabel kontingensi dua arah dengan variabel dependennya.
- b) Menghitung statistik chi-square untuk setiap pasang kategori yang dapat dipilih untuk digabung menjadi satu, untuk menguji kebebasannya dalam sebuah sub tabel



kontingensi  $2 \times j$  yang dibentuk oleh sepasang kategori tersebut dengan variabel dependennya yang mempunyai sebanyak  $j$  kategori.

- c) Untuk masing-masing nilai chi-square berpasangan, dilakukan penghitungan p-value berpasangan secara bersamaan. Di antara pasangan-pasangan yang tidak signifikan, dilakukan penggabungan sebuah pasangan kategori yang paling mirip (yaitu pasangan yang mempunyai nilai chi-square berpasangan terkecil) menjadi sebuah kategori tunggal, dan kemudian dilanjutkan ke langkah nomor d). Tetapi apabila semua pasangan kategori yang tersisa adalah signifikan, akan dilanjutkan ke langkah nomor e).
- d) Untuk suatu kategori gabungan yang terdiri dari 3 kategori atau lebih, akan dilakukan pengujian untuk melihat apakah suatu kategori prediktor seharusnya dipisah dengan menguji signifikansi antara kategori tersebut dengan kategori yang lain dalam satu kategori gabungan. Jika didapat nilai chi-square yang signifikan, akan dipisahkan kategori tersebut dengan yang lain. Jika lebih dari satu kategori yang bisa dipilih untuk dipisah, maka dipisahkan salah satu yang mempunyai nilai chi-square tertinggi. Kemudian kembali ke langkah nomor c).
- e) Dengan cara memilih, dilakukan penggabungan suatu kategori yang mempunyai sedikit pengamatan yang tidak sesuai dengan kategori lain yang paling mirip, seperti yang diukur oleh nilai chi-square berpasangan yang terkecil.
- f) Selanjutnya dilakukan penghitungan p-value terkoreksi Bonferroni didasarkan pada tabel yang telah digabung.

## **2) Tahap Pemisahan**

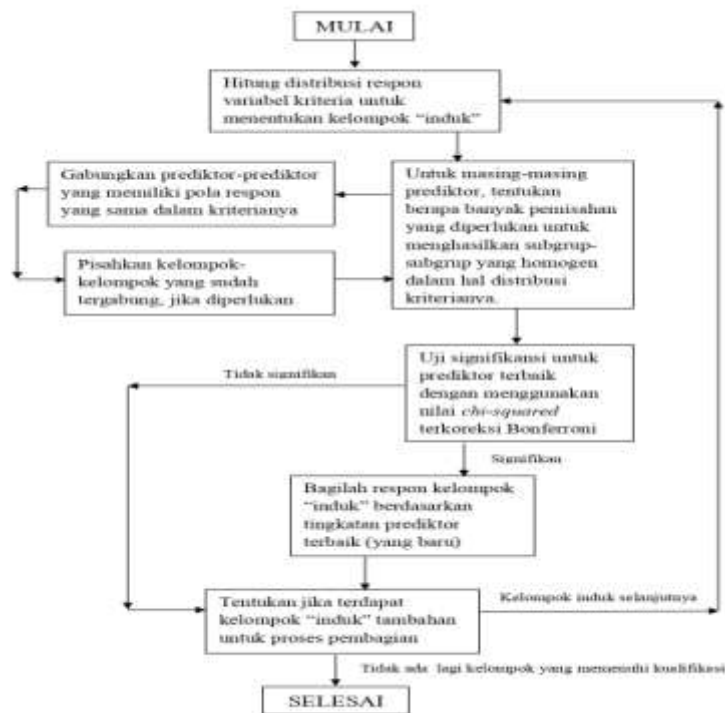
Pada tahap ini akan dipilih prediktor terbaik, yaitu prediktor dengan nilai p-value yang terendah, dan kemudian melakukan pembagian kelompok dengan prediktor ini (yaitu dengan menggunakan masing-masing kategori-kategori prediktor tersebut yang telah digabung secara optimal untuk menentukan sub pembagian dari kelompok induk menjadi sub kelompok yang baru). Jika tidak ada prediktor dengan nilai p-value yang signifikan, proses pembagian kelompok tidak akan dimulai.

## **3) Tahap Penghentian**

Keseluruhan proses akan berulang kembali ke langkah nomor a pada tahap Penggabungan untuk menganalisis sub kelompok berikutnya. Proses akan mengalami

penghentian ketika semua sub kelompok telah dianalisis dan juga telah berisi pengamatan-pengamatan dengan jumlah yang terlalu sedikit.

Perreault dan Barksdale (dalam Hasana, 2013) menyusun algoritma CHAID dalam bentuk diagram alir sebagai di bawah ini.



Gambar 1. Diagram Alir Algoritma CHAID

Keseluruhan proses CHAID dalam penelitian ini dilakukan dengan bantuan software IBM SPSS versi 25. Penggunaan Teknik CHAID lebih mudah diaplikasikan dengan sudah tersedianya berbagai perangkat lunak yang menyediakan algoritma CHAID sehingga proses kontrol rutin terhadap kualitas kinerja mahasiswa dapat dikerjakan dengan efisien dan efektif.

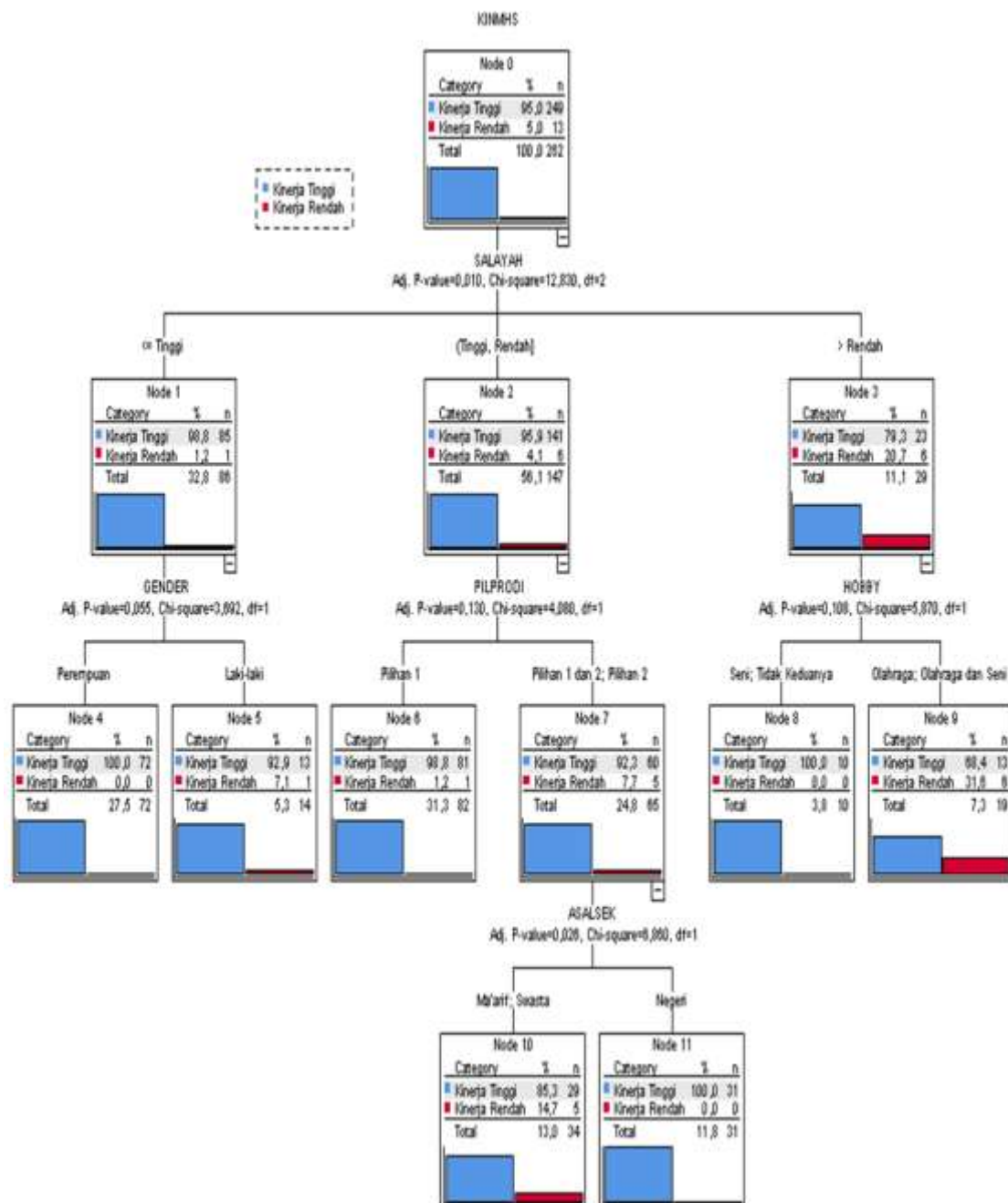
## HASIL DAN PEMBAHASAN

Hasil analisa CHAID menunjukkan bahwa dari 13 variabel demografis yang dipilih dari formulir pendaftaran mahasiswa baru, terdapat 5 variabel yang berasosiasi dengan kinerja mahasiswa yang kurang memuaskan. Kelima variabel tersebut adalah variabel Pendapatan Ayah, Jenis Kelamin, Pilihan Program Studi, Asal Sekolah, dan Hobi. Hal ini secara lengkap disajikan dalam tabel 2 berikut.

**Tabel 2. Ringkasan Model Analisa CHAID**

Spesifikasi	Metode Penyusunan Model	CHAID
	Variabel Dependen	Kinerja Mahasiswa
	Variabel Independen	Pilihan Prodi, Jalur Beasiswa, Jenis Kelamin, Status Pernikahan, Jenis Asal Sekolah, Jeda Waktu Masuk Kuliah, Hobi, Pendidikan Terakhir Ayah, Pendidikan Terakhir Ibu, Pekerjaan Ayah, Pekerjaan Ibu, Pendapatan Ayah, Tempat Tinggal Saat Ini
Results	Variabel Independen yang Terlibat	Pendapatan Ayah, Jenis Kelamin, Pilihan Prodi, Jenis Asal Sekolah, Hobi
	Hasil Jumlah Node dalam Diagram Pohon	12
	Jumlah Node Terminal dalam Diagram Pohon	7
	Kedalaman Diagram Pohon	3

Asosiasi kelima variabel disusun oleh CHAID dalam bentuk diagram pohon yang disajikan pada gambar berikut.



**Gambar 2. Diagram Pohon CHAID Prediksi Kinerja Mahasiswa**

Dari diagram tersebut tampak bahwa terdapat 12 node yang terbentuk. Node induk (parent node) merupakan node Kinerja Mahasiswa dan memiliki 11 node anakan (child node). Akan tetapi, yang paling menjadi fokus dalam analisa adalah 7 node akhir (terminal node). Ketujuh node tersebut merupakan kumpulan kemungkinan klasifikasi untuk dapat

memprediksi kinerja mahasiswa yang rendah. Dalam penelitian ini, kinerja mahasiswa yang rendah yang menjadi target pengamatan karena dianggap merupakan fokus yang harus dicermati untuk dapat diketahui dan segera dicarikan solusinya. Dalam hal ini maka ketujuh node yang terbentuk pada diagram merupakan karakteristik kandidat mahasiswa dengan potensi kinerja rendah.

Dari ketujuh kandidat tersebut dianalisa oleh CHAID dan didapatkan 3 karakteristik yang paling berasosiasi dengan kinerja mahasiswa yang rendah. Ketiga karakteristik tersebut secara berurutan adalah sebagai berikut, Tabel 3.

**Tabel 3. Karakteristik Mahasiswa yang Diprediksi Berkinerja Rendah**

<b>Node</b>	<b>Karakteristik</b>	<b>Persentase Kemungkinan</b>
<b>Node 9</b>	Mahasiswa dengan pendapatan Ayah sedang sampai sangat tinggi (pendapatan Ayah > Rp. 1.500.000,-) dan memiliki hobi olahraga saja dan atau olahraga dan seni	46,2%
<b>Node 10</b>	Mahasiswa dengan pendapatan Ayah rendah dan tinggi (Pendapatan Ayah $\leq$ Rp. 1.500.000,- dan Pendapatan Ayah > Rp. 2.500.000,- dan $\leq$ Rp. 3.500.000,-), memilih Prodi Pendidikan Matematika di pilihan kedua dan atau pilihan pertama dan kedua, serta berasal dari sekolah Ma'arif dan Swasta	38,5%
<b>Node 5</b>	Mahasiswa dengan pendapatan Ayah tinggi sampai rendah (Pendapatan Ayah $\leq$ Rp. 3.500.000,-) dan berjenis kelamin laki-laki	7,7%

Tabel 3 menunjukkan bahwa variabel independen pertama yang berasosiasi dengan kinerja mahasiswa yang rendah adalah variabel tingkat pendapatan Ayah. Variabel-variabel independen berikutnya yang berasosiasi dengan kinerja mahasiswa yang rendah adalah Jenis Kelamin, Pilihan Prodi, Jenis Asal Sekolah, dan Hobi. Dari tabel tersebut juga diketahui bahwa 46,2% mahasiswa dengan karakteristik seperti di node 9 akan berpotensi berkinerja rendah. Begitu juga pada node 10, 38,5% mahasiswa dengan karakteristik seperti pada node 10 akan berpotensi berkinerja rendah. Yang terakhir, 7,7% mahasiswa dengan karakteristik seperti pada node 5 akan berpotensi juga untuk berkinerja rendah. Dari hasil analisis CHAID menggunakan software IBM SPSS versi 25 tersebut diketahui bahwa analisa yang dilakukan

memiliki standar eror yang sangat kecil yaitu 0,013 dengan resiko kesalahan yang diperkirakan adalah 5%.

## **KESIMPULAN**

Hasil penelitian menunjukkan bahwa teknik-teknik Educational Data Mining (EDM) dapat dimanfaatkan untuk menggali pengetahuan yang sudah dimiliki dalam basis data perguruan tinggi. Hasil penambangan data akademik tersebut dapat secara efektif dan efisien digunakan untuk membantu upaya menjaga dan meningkatkan kualitas dan profesionalitas perguruan tinggi dalam memberikan layanan pendidikan dan pembelajaran di era yang kompetitif seperti sekarang ini. Dari hasil analisa CHAID terhadap data 3 angkatan di Program Studi Pendidikan Matematika, FKIP, UNISMA didapatkan bahwa terdapat 3 karakteristik mahasiswa yang diprediksi akan berkinerja rendah. Ketiga karakteristik tersebut secara berurutan melibatkan variabel independen Pendapatan Ayah, Jenis Kelamin, Pilihan Prodi, Jenis Asal Sekolah, dan Hobi. Dengan adanya hasil analisa CHAID tersebut diharapkan dapat menjadi salah satu pertimbangan untuk bisa memberikan solusi kepada mahasiswa-mahasiswa yang diprediksi akan berkinerja rendah. Sehingga kinerja mahasiswa kedepannya bisa terjaga dan bahkan ditingkatkan dengan adanya langkah-langkah solusi yang bisa diambil oleh Program Studi, baik melalui kebijakan maupun melalui metode-metode pembelajaran baru yang akan diterapkan berdasarkan pertimbangan hasil analisa.

## **UCAPAN TERIMA KASIH**

Peneliti menyampaikan banyak terimakasih kepada pihak-pihak yang membantu terselesaikannya artikel ini. Teruntuk Bapak Prof. Dr. H. Maskuri, M.Si. selaku Rektor UNISMA, Bapak Dr. Hasan Busri, M.Pd. selaku Dekan FKIP, Ibu Alifiani, S.Pd., M.Pd. selaku Kaprodi Pendidikan Matematika, serta LPPM UNISMA yang telah mengizinkan, memwadahi dan mendukung terlaksananya penelitian ini.

## **DAFTAR PUSTAKA**

- Ahmed, A.B.E.D. & Elaraby, I.S. (2014). Data Mining: A Prediction for Student's Performance Using Classification Method. *World Journal of Computer Application and Technology*, 2(2), 43-47.
- Alhammadi, D.A. & Aksoy, M.S. (2013). Data Mining-An Experimental Study. *International Journal of Computer Application*, 62(15), 31-34.

- Chandra, K., Nandhini, E., & Chandra, E. (2010). Knowledge Mining from Student Data. *European Journal of Scientific Research*, 47(1), 156–163.
- du Toit, S. H. C., Steyn, A.G.W., & Stumpf, R.H. (1986). *Graphical Exploratory Data Analysis*. New York: Springer.
- Elakia, Gayathri, Aarthi, & Naren J. (2014). Application of Data Mining in Educational Database for Predicting Behavioural Patterns of the Students. *International Journal of Computer Science and Information Technologies*, 5(3), 4649–4652.
- Hasana, S. N. (2013). Algoritma Forward dan Backward Berdasarkan Entropi Sebagai Alternatif Untuk Algoritma Segmentasi CHAID. *Tesis*. Universitas Gadjah Mada. Yogyakarta. Diakses dari <http://etd.repository.ugm.ac.id/penelitian/detail/66171>
- Kunto, Y. S. & Hasana, S. N. (2006). Analisis CHAID sebagai Alat Bantu Statistika untuk Segmentasi Pasar (Studi Kasus pada Koperasi Syari'ah Al-Hidayah). *Jurnal Manajemen Pemasaran*, 1(2), 88-97.
- Larose, D.T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken: Willey-Interscience, John Willey and Sons, Inc.
- Liu, H. & Setiono, R. (1995). Chi-square: Feature Selection and Discretization of Numeric Attributes. Proceedings of IEEE 7th International Conference on Tools with Artificial Intelligence. Diakses dari [https://www.researchgate.net/profile/Rudy\\_Setiono/publication/3619899\\_Chi2\\_Feature\\_Selection\\_and\\_Discretization\\_of\\_Numeric\\_Attributes/links/00b49517a0cf0d1069000000/Chi2-Feature-Selection-and-Discretization-of-Numeric-Attributes.pdf](https://www.researchgate.net/profile/Rudy_Setiono/publication/3619899_Chi2_Feature_Selection_and_Discretization_of_Numeric_Attributes/links/00b49517a0cf0d1069000000/Chi2-Feature-Selection-and-Discretization-of-Numeric-Attributes.pdf).
- Marthasari, G.I. (2017). Implementasi Teknik Data Mining untuk Evaluasi Kinerja Mahasiswa Berdasarkan Data Akademik. *Fountain of Informatics Journal*, 2(2), 20-27. doi: <http://dx.doi.org/10.21111/fij.v2i2.1216>
- Ramaswami, M. & Bhaskaran, R. (2010). A CHAID Based Performance Prediction Model in Educational Data Mining. *International Journal of Computer Science Issues*, 7(1)(1), 10-18.
- Slameto. (2010). *Belajar dan Faktor-Faktor yang Mempengaruhinya*. Jakarta Timur: Rineka Cipta.
- Witten, I.H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (2nd ed.). San Francisco: Morgan Kaufmann Publisher.