

## Desain dan Kajian Awal *Profiling* Kelulusan Mahasiswa

Marta Hayu Raras S.R.S<sup>\*1)</sup>, Andi Rahadiyan Wijaya<sup>\*2)</sup>

<sup>1,2)</sup>Departemen Teknik Mesin dan Industri, Fakultas Teknik, Universitas Gadjah Mada  
Jl. Grafika No.2, Yogyakarta, Indonesia

\*Email: m.hayu@mail.ugm.ac.id

### Abstrak

Tingginya proporsi mahasiswa terlambat lulus, munculnya mahasiswa yang mengalami Drop Out serta adanya fakta bahwa tidak semua alumni menggunakan ijazah S2 dalam mencari pekerjaan, menimbulkan keingintahuan prodi terhadap pola kelulusan mahasiswa terkait masa studi, performansi dan pekerjaan setelah lulus dari prodi S2 Teknik Industri UGM. Oleh karena itu, perlu dilakukan penelitian mengenai bagaimana cara mengklasifikasikan data yang dimiliki oleh prodi dan hasil tracer study untuk mendapatkan pola kelulusan mahasiswa sebagai bahan evaluasi dalam penyelenggaraan prodi. Penelitian ini merupakan kajian awal dimana atribut dan data yang digunakan masih bersifat sementara, sehingga aturan transformasi yang dihasilkan pada penelitian ini kedepannya masih dapat berubah. Selain aturan transformasi, penelitian ini juga menghasilkan desain untuk penelitian selanjutnya yaitu menetapkan tujuan Educational Data Mining kasus ini sebagai student modelling yang akan dicapai dengan cara profiling menggunakan metode Decision Tree dengan algoritma ID3 dan C4.5. Selanjutnya kedua algoritma tersebut akan dibandingkan untuk menentukan pola terbaik. Penelitian lanjutan sangat diperlukan untuk dapat merealisasikan tujuan penelitian kasus ini.

**Kata Kunci:** C4.5 algorithm, decision tree, ID3 algorithm, profiling.

### 1. Pendahuluan

Salah satu aspek penting untuk mengevaluasi keberhasilan penyelenggaraan program studi di perguruan tinggi adalah mahasiswa (Ridwan *et al.*, 2013). Mahasiswa merupakan peserta didik pada jenjang pendidikan perguruan tinggi yang haknya diatur dalam Undang-Undang no. 20 tahun 2003 Bab V tentang Peserta Didik.

Demi kelancaran proses akademik, perguruan tinggi pasti memiliki data mahasiswa yang tersimpan dalam sistem informasi. Data tersebut terdiri dari data registrasi mahasiswa, data akademis mahasiswa setiap semesternya hingga data kelulusan mahasiswa. Setelah mahasiswa lulus data-data tersebut cenderung tidak digunakan secara maksimal. Oleh karena itu, data-data ini perlu dimanfaatkan untuk dapat menggali informasi sedalam-dalamnya. Namun, tidak mudah untuk membuat prediksi dengan cara memanfaatkan berbagai data mentah yang dimiliki oleh institusi, sehingga diperlukan teknik *Educational Data Mining* untuk membantu mengubah data mentah dari sistem menjadi sebuah informasi yang berpotensi memiliki dampak positif terhadap dunia pendidikan (Kaur *et al.*, 2015).

Setiap institusi pendidikan pasti ingin berkontribusi untuk memajukan dunia pendidikan. Salah satunya adalah Program Studi S2 Teknik Industri, Departemen Teknik Mesin & Industri Universitas Gadjah Mada. Di awal berdirinya prodi tahun 2008, mahasiswa cenderung mengalami keterlambatan waktu kelulusan. Waktu kelulusan untuk jenjang S2 ini normalnya adalah 4 semester, namun tidak sedikit mahasiswa menempuh pendidikan dengan keterlambatan lulus lebih dari 4 semester. Demi meningkatkan kualitasnya, prodi mengeluarkan kebijakan bahwa batas maksimum kelulusan adalah 8 semester. Sejak itu tingkat keterlambatan kelulusan dapat berkurang, namun mulai terdapat mahasiswa yang harus *Drop Out* karena tidak dapat menyelesaikan masa studinya sesuai ketentuan yang berlaku. Sekitar 64% mahasiswa yang masuk pada tahun 2008-2015 mengalami keterlambatan lulus dengan proporsi 52% terlambat dibawah 4 semester sedangkan 12% mengalami keterlambatan mulai dari 4 semester ke atas. Sisanya sebanyak masing-masing 5% untuk mahasiswa *Drop Out* akibat tidak dapat

menyelesaikan studi sesuai waktu yang ditetapkan dan mahasiswa yang keluar sebelum masa studi berakhir.

Tingginya proporsi mahasiswa yang mengalami keterlambatan lulus, hingga munculnya mahasiswa yang mengalami *Drop Out* ini, mendorong penelitian untuk mengetahui pola kelulusan mahasiswa terkait masa studi dan performansi mahasiswa sebagai bahan evaluasi. Selain itu adanya fakta bahwa tidak semua alumni menggunakan ijazah S2 dalam mencari pekerjaan juga menimbulkan keingintahuan prodi terhadap pola kelulusan mahasiswa terkait pekerjaan setelah mahasiswa lulus dari prodi S2 Teknik Industri UGM.

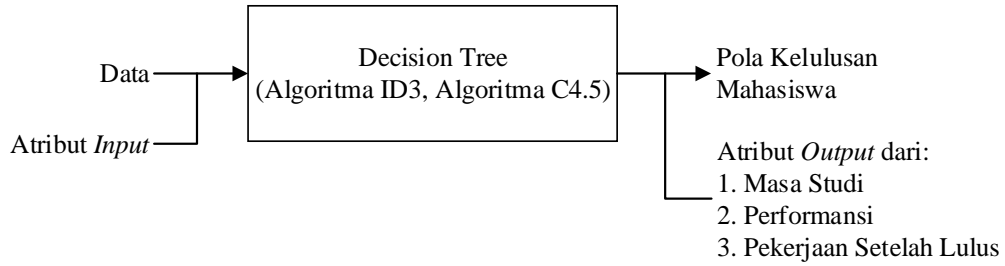
Penggunaan *Educational Data Mining* memiliki berbagai macam tujuan. Beberapa penelitian menggunakan EDM untuk memprediksi pola akademis dengan meninjau ketepatan masa studi peserta didik dalam menempuh pendidikannya seperti penelitian yang dilakukan oleh Abu-oda & El-halees (2015), Kusumawati *et al* (2015), Saefulloh & Moedjiono (2013), Jaman (2013), Ameri *et al* (2016), Ojha *et al* (2017), Aulck *et al* (2016). Selain itu terdapat penelitian dengan tujuan memprediksi pola akademis dengan meninjau performansi peserta didik seperti penelitian yang dilakukan oleh Ahmed & Elaraby (2014), Asif *et al* (2017), Ridwan *et al* (2013), Wanli *et al* (2014), Thakur *et al* (2014). Dan ada pula penelitian yang bertujuan untuk memprediksi pola akademis peserta didik dengan meninjau performansi peserta didik terhadap eksperimen atau skenario yang diberikan seperti penelitian yang dilakukan Márquez-Vera *et al* (2016) yaitu penelitian untuk memprediksi pola akademis dengan cara melakukan 3 kali eksperimen. Serta penelitian yang dilakukan oleh Zacharis (2016) untuk memprediksi performansi peserta didik menggunakan skenario *blended learning*.

Berdasarkan uraian di atas, kedepannya penelitian ini bertujuan untuk memprofilkan data yang dimiliki oleh prodi dan data hasil *tracer study* untuk mendapatkan pola kelulusan mahasiswa sebagai bahan evaluasi dalam penyelenggaraan Program Studi S2 Teknik Industri Universitas Gadjah Mada.

## 2. Metode

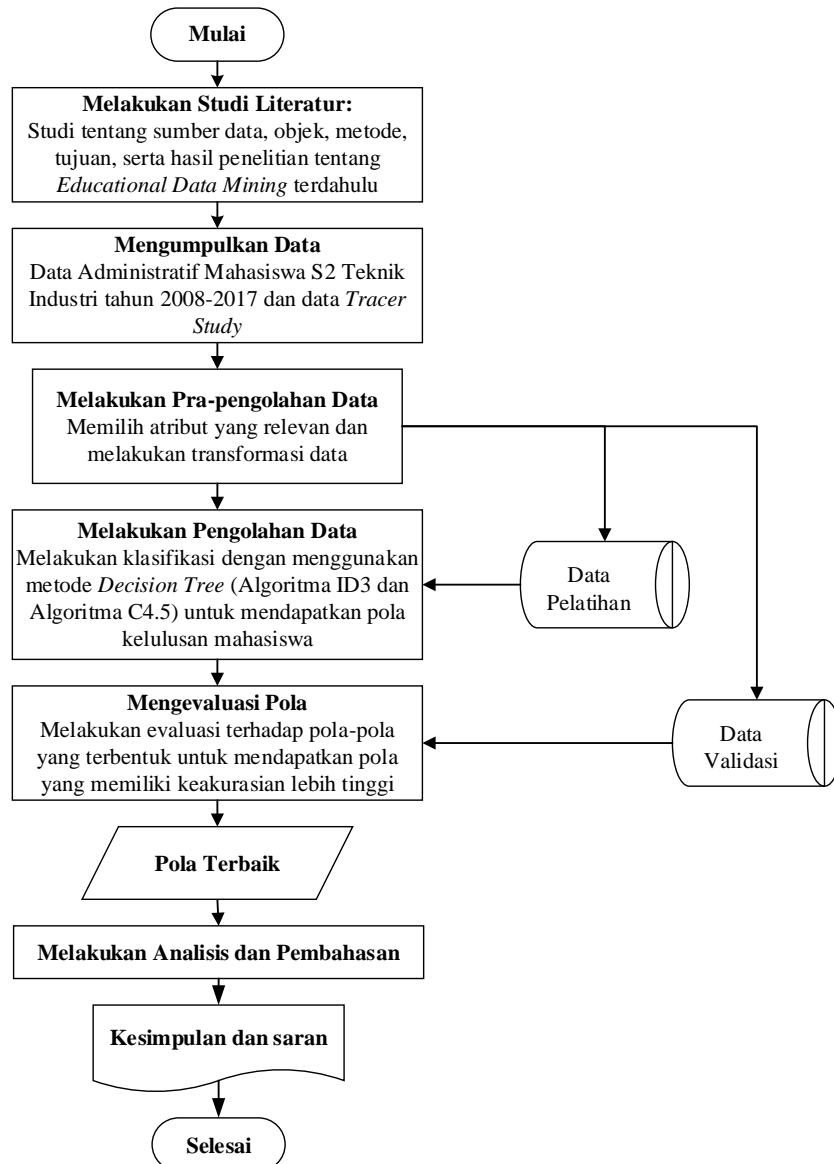
Metode yang digunakan dalam penelitian ini adalah *Decision Tree* dengan menggunakan 2 algoritma yaitu ID3 dan C4.5. Metode *Decision Tree* merupakan metode klasifikasi sederhana yang efektif dalam memprediksi dan menjelaskan hubungan antar atribut dan nilai target. *Decision Tree* berbeda dengan metode lain yang lebih fokus terhadap hasil akhir prediksi tanpa mengetahui hubungan antar atribut/ pola yang terjadi dalam prosesnya. Sehingga metode *Decision Tree* tepat digunakan untuk memprofilkan kelulusan mahasiswa. Seiring berjalannya waktu, metode *Decision Tree* memiliki berbagai macam algoritma. Algoritma ID3 dan C4.5 merupakan algoritma yang sering digunakan dalam kasus EDM. Menurut pengertiannya, algoritma C4.5 adalah pengembangan dari algoritma ID3 dimana algoritma C4.5 ini memiliki keunggulan yaitu lebih *robust* terhadap *noise*. Oleh karena itu algoritma C4.5 biasanya menghasilkan prediksi yang lebih akurat dibandingkan dengan algoritma ID3. Namun tidak semua penelitian terdahulu mendapatkan hasil yang serupa. Penelitian Pal (2013) menyimpulkan bahwa algoritma ID3 memiliki keakurasian lebih tinggi dibandingkan algoritma C4.5. Oleh karena itu, disamping tujuan utama penelitian yaitu untuk memprofilkan kelulusan mahasiswa, penelitian ini juga ingin membandingkan pola yang dihasilkan oleh kedua algoritma populer tersebut serta mencari tahu algoritma mana yang memiliki keakurasian lebih tinggi untuk kasus kelulusan mahasiswa S2 Teknik Industri UGM ini.

Secara garis besar, hubungan antara *input*, proses dan *output* penelitian ini ditunjukkan pada Gambar 1. berikut:



Gambar 1. Garis Besar Penelitian

Sedangkan untuk detail tahap penelitian ini digambarkan dengan jelas menggunakan diagram alir pada Gambar 2. berikut:



Gambar 2. Diagram Alir Tahap Penelitian

### 3. Hasil dan Pembahasan

Hal yang pertama kali dilakukan adalah melakukan pengambilan data, yaitu data administratif mahasiswa S2 Teknik Industri tahun 2008-2017. Namun jika hanya menggunakan data-data yang tersedia tidak akan ada bedanya dengan penelitian-penelitian yang sudah banyak dilakukan sebelumnya. Penelitian ini memiliki perbedaan karena bersifat Deep Mining, dimana tidak hanya menggunakan data-data yang sudah tersedia namun dituntut sekreatif mungkin dalam mendapatkan data-data diluar database, misalnya dengan Tracer study yaitu melakukan pelacakan jejak alumni baik menggunakan kuesioner, social media, search engine, ristek dikti dan lain sebagainya. Selain itu output dari penelitian ini adalah multi output, tidak seperti penelitian sebelumnya yang biasanya hanya 1 output saja (ketepatan masa studi atau performansi).

Setelah data-data didapatkan perlu dilakukan seleksi dan juga transformasi data. Seleksi data dilakukan agar lebih efisien pada saat melakukan proses klasifikasi, sedangkan transformasi data dilakukan untuk mengubah bentuk data menjadi lebih sesuai dan dapat diolah. Namun pada kajian awal ini belum semua data didapatkan oleh karena itu tahap seleksi data belum dapat dilakukan. Jika diasumsikan semua data perlu dimasukkan ke dalam proses mining, maka data tersebut perlu ditransformasikan. Tabel 1 berikut menjelaskan aturan transformasi dari data-data sementara yang sudah didapatkan.

**Tabel 1.** Aturan Transformasi

Atribut	Keterangan
Jenis Kelamin	1= Perempuan 2= Laki-Laki
Tahun Masuk	1= Ganjil 2= Genap
Semester	1= Ganjil 2= Genap
Tempat Lahir	1= Yogyakarta 2= Jawa Tengah 3= Jawa Barat 4= Jawa Timur 5= Kalimantan 6= Sulawesi 7= Sumatra 8= Riau 9= Maluku
Umur masuk S2	1= < 25 2= 25-30 3= > 30
Biaya	1= Biaya Sendiri 2= Beasiswa/ BPPS
Suku (dilihat dari nama)	1= Jawa (Nama yang tidak memiliki marga) 2= Batak (Nama yang memiliki marga Batak) 3= Timur (Nama yang memiliki marga Timur)
Minat Studi	1= Teknik Produksi 2= OR 3= Ergonomi 4= Manufaktur
SKS	1= < 46 SKS

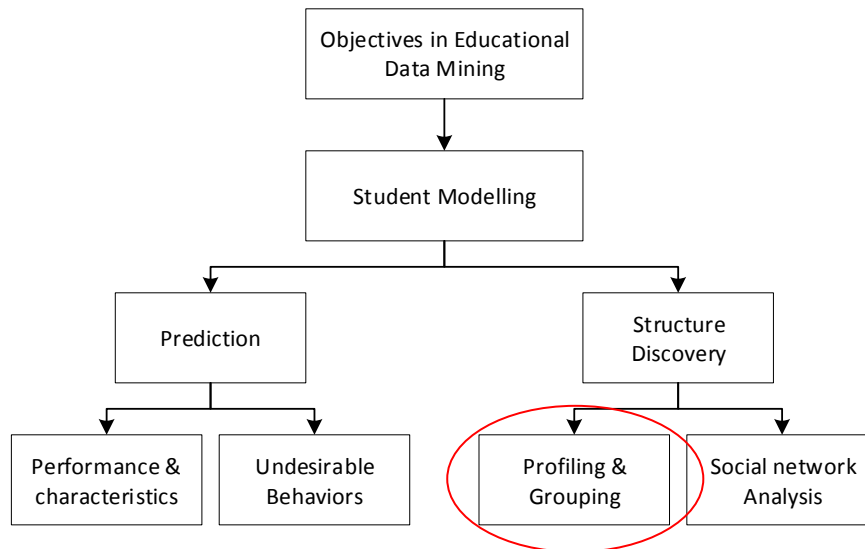
Atribut	Keterangan
	2= $\geq 46$ SKS
Dosen Pembimbing	Masing-masing Dosen S2 TI UGM yang sudah pernah membimbing tesis.
Pernah mengalami perubahan tesis	1= ya 2= tidak
Minat Tesis	1= Teknik Produksi 2= OR 3= Ergonomi 4= Manufaktur
Topik Tesis	1 = Desain Intrumen 2 = Aplikasi Ergonomi 3 = Pengambilan Keputusan 4 = Penjadwalan 5 = Desai Proses Manufaktur 6 = Analisis Pengaruh 7 = Analisis Kesuksesan produk 8 = Keandalan Manusia 9 = Pengembangan Produk 10= Supply Chain 11= Desain Alat Bantu Manufaktur 12= Manajemen Proyek
Keterbaruan Tesis	1= Metode 2= Kasus 3= Metode dan Kasus
IPK	1= $> 3.75$ 2= $3.5 - 3.75$ 3= $3.25 - 3.5$ 4= $< 3.25$
Nilai Tesis	1= A 2= B 3= C
Lama Pengerjaan Tesis	1= 6 bulan 2= 7 - 12 bulan 3= $> 12$ bulan
Lama Masa Studi	1= $< 800$ hari 2= $\geq 800$ hari

Keterangan pada tabel di atas dapat berubah karena pada saat pembuatan aturan transformasi tersebut belum melibatkan seluruh data yang dimiliki, dan dilakukan berdasarkan intensitas dari history data yang dimiliki.

Setelah transformasi data, hal yang dilakukan kemudian adalah mengolahnya menggunakan Decision Tree dengan algoritma ID3 dan algoritma C4.5. Kemudian di validasi manakah algoritma yang memiliki keakurasian yang lebih tinggi untuk kasus profiling kelulusan mahasiswa S2 Teknik Industri UGM ini.

#### 4. Simpulan

Menurut Backshinategh et al (2017), salah satu yang menjadi tujuan dari penelitian Educational Data Mining adalah Student Modelling, dimana student Modelling ini terbagi menjadi 2 yaitu prediksi dan penemuan struktur. Jika didasarkan pada penelitian yang dilakukan oleh Backshinategh et al (2017), kedepannya penelitian ini akan masuk ke dalam penelitian yang bertujuan untuk menemukan struktur/ pola dengan cara Profiling. Uraian tersebut dapat dengan jelas dilihat pada Gambar 3 berikut ini:



Gambar 3. Bagan Survey Penelitian EDM (Backshinategh et al, 2017)

Namun untuk sementara, paper ini terbatas hanya pada desain dan kajian awal dari profiling kelulusan mahasiswa S2 Teknik Industri UGM. Sehingga untuk selanjutnya perlu dilakukan penelitian lanjutan untuk benar-benar merealisasikan tujuan penggunaan Educational Data Mining dalam penemuan pola kelulusan mahasiswa.

Profiling ini diharapkan dapat memberi dampak positif bagi prodi sebagai bahan evaluasi penyelenggaraan Program Studi S2 Teknik Industri Universitas Gadjah Mada. Pola yang terbentuk dapat memberi gambaran terhadap prodi hal-hal apa saja yang dapat berdampak pada lama masa studi, performansi dan juga pekerjaan setelah lulus.

#### Daftar Pustaka

- Abu-oda, Ghadeer S., and Alaa M El-halees., 2015, Data Mining in Higher Education : University Student Dropout Case Study, *International Journal of Data Mining & Knowledge Management Process*, Vol 5, No 1: 15–27.
- Ahmed, Abeer Badr El Din., and Ibrahim Sayed Elaraby., 2014, Data Mining : A Prediction for Student's Performance Using Classification Method, *World Journal of Computer Application and Technology*, Vol 2 No 2: 43–47.
- Ameri, Sattar., Mahtab J. Fard., Ratna B. Chinnam., and Chandan K. Reddy., 2016, Survival Analysis Based Framework for Early Prediction of Student Dropouts, *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management - CIKM '16*, 903–12.
- Aulck, Lovenoor., Nishant Velagapudi., Joshua Blumenstock., and Jevin West., 2016, Predicting Student Dropout in Higher Education, *International Conference on Machine Learning and Application*, p 16-20

- Bakhshinategh, B., Zaiiane, O. R., ElAtia, S., & Ipperciel, D. (2017). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*.
- Jaman, Jajam Haerul., 2013, Prediksi Kelulusan Mahasiswa Dengan Metode Algoritma c4.5, *SYNTAX* Vol 2 No 2: 1–6.
- Kaur, Parneet., Manpreet Singh., and Gurpreet Singh Josan., 2015, Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector, *Procedia Computer Science* 57: 500–508.
- Kusumawati, Dewi., Wing Wahyu Winarno., and M Rudyanto Arief., 2015, Prediksi Kelulusan Mahasiswa Menggunakan Metode Neural Network Dan Particle Swarm Optimization, *Seminar Nasional Teknologi Informasi dan Multimedia*, hal 6–8.
- Márquez-vera, Carlos., Alberto Cano., Cristobal Romero., Amin Yousef., Mohammad Noaman., Habib Mousa Fardoun., and Sebastian Ventura, 2016, Early dropout prediction using data mining: a case study with high school students, Vol 33 No 1: 107–24.
- Ojha, Tushar., Gregory L Heileman., Manel Martinez-ramon., and Ahmad Slim., 2017, Prediction of Graduation Delay Based on Student Performance, *Institute of Electrical and Electronics Engineering*, 3454–3460.
- Ridwan, Mujib., Hadi Suyono., and M Sarosa., 2013, Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, *Eeccis* Vol 7 No 1: 59–64.
- Saefulloh, Asep., and Moedjiono., 2013, Penerapan Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Tepat Waktu, *InfoSys Journal* Vol 2 No 1: 41–54.
- Thakur, Gautam S., Mohammed M Olama., Allen W Mcnair., Sreenivas R Sukumar., and Scott Studham., 2014, Towards Adaptive Educational Assessments: Predicting Student Performance Using Temporal Stability and Data Analytics in Learning Management Systems, *Proceedings 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Wanli, Xing., Rui Guo., Eva Petakovic., and Sean Goggins., 2015, Participation-Based Student Final Performance Prediction Model through Interpretable Genetic Programming: Integrating Learning Analytics, Educational Data Mining and Theory, *Computers in Human Behavior* 47: 168–81.
- Zacharis, Nick Z., 2016, Predicting Student Academic Performance in Blended Learning Using Artificial Neural Networks, *International Journal of Artificial Intelligence & ApplicationsI*, Vol7 No 5: 17–29.