



APLIKASI DATA MINING UNTUK MENENTUKAN LULUSAN MAHASISWA TEPAT WAKTU

Alwendi^{1,*}, Andi Saputa Mandopa², Lela Budiarti³

¹Dosen ilmu Komputer, Fakultas Teknik Universitas Graha Nusantara Padangsidimpuan, Indonesia

²Dosen Pend. Matematika, FKIP Universitas Graha Nusantara Padangsidimpuan, Indonesia

³Dosen Program Studi Teknologi Informasi Universitas Muhammadiyah Tapanuli Selatan, Indonesia

EMAIL: alwendi60@gmail.com andimandopa100@gmail.com lelabudiarti1@gmail.com

Diterima : 22 Oktober 2024. Disetujui : 10 November 2024. Dipublikasikan : 05 Desember 2024.

ABSTRACT - *Untimely graduation rates have become an important problem in the world of education that can affect the quality of education and cause additional costs for students and educational institutions. This research aims to build a prediction model to predict the acceptance of prospective new students at Graha Nusantara University, especially in the Computer Science study program. In achieving the graduation process, students must meet the requirements including GPA, Gender, Region of Origin, School of Origin. The research results show that the prediction model built using the rule association algorithm in this study had good accuracy, namely 38.3 percent. This research examines the application of data mining methods with the association rule algorithm, to predict student graduation on time at Graha Nusantara University. This research produces findings that this method makes it easier to determine student graduation patterns. Association rule is a data mining technique that finds patterns, dependencies, and connections between items in data. These algorithms can be used to explain patterns in seemingly independent data, such as relational databases and transactional databases.*

Keywords : Data Mining, Association Rules, Timely Graduation.

ABSTRAK - Tingkat kelulusan yang tidak tepat waktu telah menjadi masalah penting dalam dunia pendidikan yang dapat mempengaruhi kualitas pendidikan dan menimbulkan biaya tambahan bagi mahasiswa dan lembaga pendidikan. Penelitian ini bertujuan untuk membangun model prediksi untuk memprediksi penerimaan calon mahasiswa baru di Universitas Graha Nusantara, khususnya di program studi Ilmu Komputer. Dalam mencapai proses kelulusan, mahasiswa harus memenuhi persyaratan antara lain IPK, Jenis Kelamin, Daerah Asal, Asal Sekolah. Hasil penelitian menunjukkan bahwa model prediksi yang dibangun menggunakan algoritma asosiasi rule, pada penelitian ini memiliki akurasi yang baik, yaitu sebesar 38,3 persen. Penelitian ini mengkaji penerapan metode data mining dengan algoritma Algoritma asosiasi rule, untuk memprediksi kelulusan mahasiswa tepat waktu di Universitas Graha Nusantara. Penelitian ini menghasilkan temuan bahwa metode ini memudahkan dalam menentukan pola kelulusan

mahasiswa . Asosiasi rule merupakan teknik data mining yang menemukan pola, ketergantungan, dan koneksi antara item dalam data. Algoritma ini dapat digunakan untuk menjelaskan pola dalam data yang tampaknya independen, seperti basis data relasional dan basis data transaksional.

Kata kunci : Data Mining, Rule Asosiasi, Kelulusan Tepat Waktu.

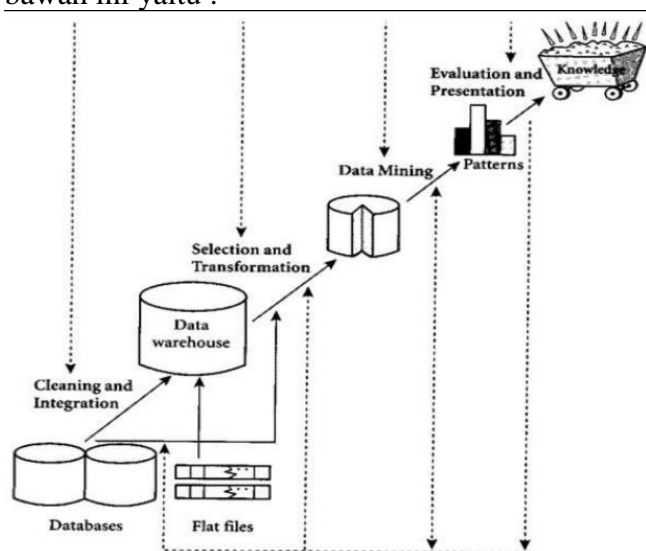
I. PENDAHULUAN

Ketepatan waktu lulus mahasiswa menjadi salah satu indikator penilaian kelayakan program studi sebagai unit pelaksana pendidikan pada perguruan tinggi. Mengetahui faktor yang mempengaruhi waktu lulus mahasiswa akan membantu program studi dan dosen dalam mengambil keputusan untuk meningkatkan kuantitas mahasiswa lulus tepat waktu. Tujuan dari penelitian ini adalah untuk mendapatkan gambaran tentang karakteristik mahasiswa yang mempengaruhi ketepatan waktu lulus mahasiswa program studi S1 Ilmu Komputer dengan menggunakan metode Data mining menggunakan rule asosiasi. tidak dapat dipisahkan dalam kegiatan suatu perguruan tinggi, sehari-harinya, baik di universitas atau tingkat. Namun terkadang banyak perguruan tinggi yang belum memaksimalkan untuk memanfaatkan data-data yang ada dan tak jarang pula ada yang mengabaikannya, terlebih jika data tersebut sudah usang dan tidak diperlukan lagi. Pada penelitian ini, penulis mencoba menggali data tersebut agar dapat lebih bermanfaat sehingga dapat memberikan informasi dan pengetahuan bagi perguruan tinggi. Data yang akan digali adalah Ketepatan waktu lulus mahasiswa yang dikategorikan menjadi dua yakni “Tepat” dan “Tidak Tepat”, IPK, Jenis Kelamin, Daerah Asal, Asal Sekolah. Dan data lainnya yang berkaitan dengan profil mahasiswa, yaitu bagaimana dengan memanfaatkan data-data alumni atau mahasiswa yang ada pada sebuah perguruan tinggi, khususnya di Universitas Graha Nusantara Perkembangan teknologi informasi untuk mengumpulkan dan menyimpan berbagai tipe data telah jauh meninggalkan kemampuan untuk menganalisis, meringkas dan mengekstraksi “pengetahuan” dari data. Para peneliti melihat

peluang itu untuk melahirkan sebuah teknologi baru yang menjawab kebutuhan ini, yaitu Data Mining. Teknik data mining dengan model Association rules (aturan asosiasi) atau affinity analysis (analisis afinitas) yang digunakan dalam penelitian ini menerapkan algoritma apriori sebagai pemrosesan data. Ide dasar dari algoritma ini adalah dengan mengembangkan frequent itemset. Dengan menggunakan satu item dan secara rekursif mengembangkan frequent itemset dengan dua item, tiga item dan seterusnya hingga frequent itemset dengan semua ukuran. Untuk mengembangkan frequent set dengan dua item, dapat menggunakan frequent set item. Alasannya adalah bila set satu item tidak melebihi isupport minimum, maka sembarang ukuran item set yang lebih besar tidak akan melebihi support minimum tersebut. Secara umum, mengembangkan set dengan fc-item menggunakan frequent set dengan $k - 1$ item yang dikembangkan dalam langkah sebelumnya. Setiap langkah memerlukan sekali pemeriksaan keseluruhan isi database (Santoso, 2017). Data mining pada dasarnya berhubungan dengan analisa data dan penggunaan teknik-teknik perangkat lunak dalam mencari pola dan keteraturan dalam himpunan data yang sifatnya tersembunyi. Dengan diperolehnya informasi-informasi yang berguna dari data-data yang ada, hubungan antar item dalam transaksi, maupun informasi-informasi yang potensial kemudian dapat diekstrak, dianalisa, dan diteliti lebih lanjut dari berbagai sudut pandang. Menurut Gartner Group, data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Analisis asosiasi di definisikan, merupakan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk support (minimum support) dan syarat minimum untuk confidence (minimum confidence), yang nantinya berguna dalam menentukan interesting association rules, yaitu untuk dibandingkan dengan batasan (threshold) yang ditentukan oleh user. Bila memenuhi kedua batasan, maka sebuah rule dapat disebut interesting rule.

II. METODE PENELITIAN

Pada penelitian Sebagai rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap ini bersifat interaktif, dengan secara langsung atau melalui basic pengetahuan. Ada 7 tahapan data mining seperti yang terlihat pada gambar 2.1 di bawah ini yaitu :



Gambar 2.1 Tahap-tahap Data mining

1. Pembersihan Data (Data Cleaning)
2. Integrasi Data (Data Integration)
3. Pemilihan Data (Data Selection)
4. Transformasi Data (Data Transformation)
5. Penambangan Data (Data Mining)
6. Evaluasi Pola (Pattern Evaluation)
7. Presentasi Pengetahuan (Knowledge Presentation)

Berdasarkan ini terlihat bahwa data mining hanya merupakan salah satu proses dari keseluruhan proses yang ada pada KDD, tetapi merupakan proses yang sangat penting dalam usaha menemukan pola-pola yang berguna bagi sejumlah data yang besar (data tersebut bisa disimpan dalam basis data, data warehouse, atau media penyimpanan informasi lainnya). Data Mining memperkirakan perilaku dan tren masa depan, memungkinkan bisnis untuk membuat keputusan yang proaktif dan berdasarkan pengetahuan. Data mining mampu menjawab permasalahan bisnis yang

secara tradisional terlalu lama untuk diselesaikan. Data mining menjelajah database untuk mencari pola tersembunyi, menemukan informasi yang prediktif yang mungkin dilewatkan para pakar karena berada di luar ekspektasi mereka. Business Objective Determination atau dalam hal ini adalah penentuan sasaran bisnis. Di dalam tahapan ini yang perlu dilakukan adalah mendefinisikan permasalahan atau tantangan bisnis yang jelas. Dalam hal ini adalah merupakan aspek yang sangat esensial dalam setiap proyek data mining. Data Preparation, pada tahapan KDD ini adalah bertujuan untuk mempersiapkan data yang diperlukan untuk proses mining sebuah data.

Tujuan yang lainnya adalah agar data yang digunakan benar-benar sesuai dengan permasalahan yang ada dan yang akan dipecahkan, dalam hal penyelesaiannya dengan penggunaan data preparation ini dapat dijamin kebenarannya dan dalam format yang sesuai atau tepat dengan yang diinginkan atau yang dituju. Pada tahapan data preparation ini merupakan tahapan yang paling banyak mengkonsumsi resources yang meliputi manusia, biaya, dan waktu yang tersedia. Di dalam tahapan ini terdapat data selection yang bertujuan sebagai pengidentifikasi semua sumber informasi internal dan eksternal dan memilih sebagian saja dari data yang akan diperlukan guna pengembangan dan penggunaan aplikasi data mining. Di samping data selection yang ada pada tahapan data preparation, data pre-processing juga berguna, dalam hal ini bertugas untuk meyakinkan kualitas dari data yang telah dipilih pada tahapan sebelumnya, tetapi dalam keadaan yang sebenarnya sering adanya permasalahan dalam penggunaannya yaitu meliputi Noisy dan Missing Values. Ada lagi metode yang digunakan untuk membantu dalam pelaksanaan data preparation yaitu Data Transformation yang bertugas dan berfungsi untuk mengubah data ke dalam model analitis serta memodelkan data agar sesuai dengan analisis yang diharapkan dan format data yang diperlukan oleh

algoritma data mining. Ada beberapa hal yang memang sangat perlu diperhatikan terlebih dahulu dalam menggunakan data transformation, antara lain adalah Categorical dan Quantitative Categorical adalah semua nilai yang mungkin ada, bersifat terbatas yang berdasarkan nominal dan ordinal. Quantitative adalah semua nilai yang mungkin dapat diukur perbedaannya. Dalam hal ini ada dua hal yang dapat dijadikan pertimbangan, antara lain yaitu continuous dan discrete. Continuous adalah nilai-nilai bilangan real seperti rata-rata transaksi dalam periode waktu. Sedangkan discrete adalah nilai-nilai bilangan bulat seperti jumlah transaksi dalam satu periode waktu. Tahapan selanjutnya adalah Data Mining, dalam tahapan ini adalah melakukan proses pencarian pengetahuan terhadap data yang ditransformasikan pada tahap sebelumnya. Dalam metode data mining telah dibagi menjadi 3 kategori yang meliputi Automated Exploration/Discovery, Prediction/Classification, Explanation/Description. Pada automated exploration/discovery merupakan sebuah clustering misalnya dapat dipakai untuk mendapatkan segmentasi pasar baru dari suatu produk. Kategori yang kedua adalah prediction/classification adalah belajar suatu model predictive, ANN, GA, regression dapat dipakai untuk peramalan nilai penjualan. Sedangkan kategori yang ketiga adalah explanation/description yang memiliki fungsi menjelaskan atau mendeskripsikan. Induksi Decision Tree dan Rule Extraction dapat digunakan untuk menjelaskan perilaku customer berdasarkan demographic dan sejarah pembeliannya. Pada tahapan Analysis of Result berfungsi untuk menginterpretasikan dan mengevaluasi output dari tahap mining: patterns. Pendekatan analisis yang digunakan akan bervariasi menurut operasi data mining yang digunakan, tetapi biasanya akan melibatkan teknik visualisasinya. Tahapan yang terakhir adalah Assimilation of Knowledge yang berfungsi untuk menggunakan hasil mining yang telah dievaluasi ke dalam perilaku organisasi dan

sistem informasi perusahaan. Tahapan proses KDD (menurut Daniel L. Silver) meliputi Data Consolidation (and Cleaning), Data Selection and Pre-processing, Data Mining, Interpretation and Evaluation. Pada tahapan Data Mining terdapat Evaluasi: dilakukan secara validasi statistic dan pengujian signifikannya; dilakukan review secara kualitatif hasil mining oleh pakar di bidangnya. Interpretasi: Decision Tree dan Rule dapat dibaca langsung; Clustering harus digambarkan dan ditabelkan. Pemakaian Visualization Tools: analisis sensitive; histogram untuk distribusi nilai; animasi dan plotting time-series. Sedangkan pada tahapan Interpretation and Evaluation terdapat Evaluasi: dilakukan secara validasi statistic dan pengujian signifikannya; dilakukan review secara kualitatif hasil mining oleh pakar di bidangnya. Interpretasi: Decision Tree dan Rule dapat dibaca langsung; Clustering harus digambarkan dan ditabelkan. Pemakaian Visualization Tools: analisis sensitive; histogram untuk distribusi nilai; animasi dan plotting time-series.

III. HASIL DAN PEMBAHASAN

Data yang diperoleh dari Universitas Graha Nusantara berupa

Tabel 1. Analisa data berdasarkan asal sekolah

Simbol	Keterangan	Simbol	Keterangan
IPK	IPK	K	Jalur Masuk Kemitraan
A	$IPK > 3,5$	J.K	Jenis Kelamin
B	$3 \leq IPK \leq 3,5$	L	Jenis kelamin laki-laki
C	$2,5 \leq IPK < 3$	P	Jenis kelamin perempuan
D	$IPK < 2,5$	Jurusan	Jurusan
LL	Lama Lulus	Q	Jurusan biologi
E	Lama lulus ≤ 4 tahun	R	Jurusan Matematika
F	$4 \text{ tahun} < \text{lama lulus} \leq 5 \text{ tahun}$	S	Jurusan Kimia

G	Lama lulus > 5 tahun	T	Jurusan Fisika
JM	Jalur masuk	U	Jurusan Farmasi
H	Jalur masuk SPMB	Sekolah	Asal Sekolah
I	Jalur masuk Sumikola	X	Asal sekolah Kabupaten
J	Jalur masuk T2	Y	Asal sekolah Kecamatan

Data tersebut diolah dan menghasilkan kombinasi itemset sampai kombinasi 6-itemset. Berikut merupakan beberapa kombinasi itemset seperti pada tabel 2 dan tabel 3.

Tabel 2. Beberapa Kombinasi 2- Itemset

2-Itemset	Support	Confidence
J.K=P -> LL=F	25,2%	36,9%
J.K=L -> LL=G	13,8%	43,5%
SEKOLAH=X -> LL=G	21,8%	38,3%
SEKOLAH=Y -> LL=G	16%	37,2%
J. M=J -> LL=G	33%	51,2%
J. M=I -> LL=F	4,2%	53,2%
J. M=H -> LL=F	3,2%	42,2%
J. M=K -> LL=E	18,3%	90,9%
JURUSAN=S -> LL=G	14,3%	44,8%
JURUSAN=Q -> LL=G	13,3%	43,2%
JURUSAN=R -> LL=G	7,5%	38,5%
JURUSAN=U -> LL=F	5%	52,6%
JURUSAN=T -> LL=E	3,8%	46,9%
IPK=C -> LL=G	18%	48,2%
IPK=B -> LL=F	17,5%	43,4%
IPK=D -> LL=G	6,5%	65%
IPK=A -> LL=E	5%	40,5%

3.1 Pembentukan Aturan Asosiasi

Untuk pembentukan aturan asosiasi, frequent itemset yang telah didapat akan dibentuk menjadi pola “jika A, maka B ($A \Rightarrow B$)” dan akan dicari nilai confidence masing- masing pola tersebut dengan rumus :

Support dan Confidence dari kombinasi 2-itemset diatas diperoleh menggunakan rumus seperti berikut :

$$\text{Support } (X \cap G) = \frac{\text{Jumlah lulusan mengandung } X \text{ dan } G}{\text{Total Kelulusan}}$$

$$\text{Confidence} = P(G|X) = \frac{\text{Jumlah kelulusan mengandung } X \text{ dan } G}{\text{Jumlah Kelulusan Mengandung } X}$$

$$\text{Confidence} = 1 = 0,383 = 38,3 \%$$

Pada rumus yang sama digunakan untuk mencari nilai Support dan confidence dari kombinasi 2-itemset lainnya.

- Kombinasi 2-itemset berdasarkan asal sekolah
 - Jika X (asal sekolah kabupaten) maka G (lama lulus > 5 tahun) dengan nilai support 21,8% dan confidence 38,3% artinya sebanyak 21,8% lulusan yang memiliki asal sekolah kabupaten menempuh studi selama lebih dari 5 tahun dengan nilai confidence 38,3%.
 - Jika Y (asal sekolah kecamatan) maka G (lama lulus > 5 tahun) dengan nilai support 16% dan confidence 37,2% artinya sebanyak 16% lulusan yang memiliki asal sekolah kecamatanmenempuh studi selama lebih dari 5 tahun dengan nilai confidence 37,2%.
 - Berdasarkan pola diatas dapat dilihat bahwa mahasiswa Universitas Graha Nusantara yang berasal dari kabupaten dan kecamatan cenderung memiliki lama studi yang sama yaitu lebih dari 5 tahun.

2. Kombinasi 2-itemset berdasarkan Jenis Kelamin
 - a. Jika P (jenis kelamin perempuan) maka F (4 tahun $<$ lama lulus \leq 5 tahun) dengan nilai support 25,2% dan confidence 36,9% artinya sebanyak 25,2% lulusan yang berjenis kelamin perempuan menempuh studi selama lebih dari 4 tahun dengan nilai confidence 36,9%.
 - b. Jika L (jenis kelamin laki-laki) maka G (lama lulus $>$ 5 tahun) dengan nilai support 13,8% dan confidence 43,5% artinya sebanyak 13,8% lulusan yang berjenis kelamin laki-laki menempuh studi selama lebih dari 5 tahun dengan nilai confidence 43,5%
 - c. Berdasarkan pola diatas dapat dilihat bahwa alumni Universitas Graha Nusantara yang berjenis kelamin perempuan cenderung lebih cepat lulus di bandingkan dengan alumni Universitas Graha Nusantara yang berjenis kelamin laki-laki.

IV. KESIMPULAN

Berdasarkan pembahasan diatas, Kesimpulan yang dapat diambil dari hasil proses pencarian association rules menggunakan algoritma Foldarm antara lain: Pencarian association rules menggunakan algoritma Foldarm menghasilkan association rules atau aturan asosiasi dengan menggunakan confidence 38,3 %.

REFERENSI

- [1] Putra, R. F., Zebua, R. S. Y., Budiman, B., Rahayu, P. W., Bangsa, M. T. A., Zulfadhilah, M., ... & Andiyan, A. (2023). *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia.
- [2] Mandopa, A. S. (2023). Penerapan Association Rule Untuk Menentukan Lama Studi Mahasiswa di Universitas Graha Nusantara Padangsidimpuan. *SMART: Jurnal Teknologi Informasi dan Komputer*, 2(1), 13-22.
- [3] Rahayu, P. W., Sudipa, I. G. I., Suryani, S., Surachman, A., Ridwan, A., Darmawiguna, I. G. M., ... & Maysanjaya, I. M. D. (2024). *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia.
- [4] Depianti, D., Basri, B., & Syarli, S. (2021). Teknik Data Mining Untuk Menentukan Faktor Kelulusan Mahasiswa Dengan Metode Apriori Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Al Asyariah Mandar. *Journal Peqguruang*, 3(1), 179-182.
- [5] Yudhatama, M. R., Wijaya, A., Abdullah, D., & Sonita, A. (2024). Penerapan Metode Association Rule Menggunakan Apriori untuk Rekomendasi Penyusunan Rak Buku di Perpustakaan Universitas Muhammadiyah Bengkulu. *Jurnal Multidisiplin Dehasen (MUDE)*, 3(3), 163-168.
- [6] Randy, R. (2023). Penerapan Data Mining Untuk Menentukan Pola Penjualan Menggunakan Metode Algoritma Apriori: Studi Kasus: Coffe Shop Kopi Bonjera Jakarta. *LOGIC: Jurnal Ilmu Komputer dan Pendidikan*, 1(6), 1522-1531.
- [7] Handayanto, R. T. (2020). Data mining dan machine learning menggunakan matlab dan python.
- [8] MART, Algoritma Apriori Pada Toko Berkah. "Jurnal Comasie."
- [9] Widaningsih, S. W. (2021). Analisis Perbandingan Metode Klasifikasi Data Mining Dalam Memprediksi Pembayaran Kredit

Halaman ini sengaja dikosongkan