

Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, dan Hasil Belajar

Eka Sabna¹⁾, Muhardi²⁾

STMIK Hang Tuah Pekanbaru

Jl. Mustafa Sari No 5 Tangkerang Selatan Pekanbaru, Telp.0761-33815/Fax.0761-863646

e-mail: es3jelita@yahoo.com¹⁾, muhardi.yudie@gmail.com²⁾

Abstrak – STMIK Hang Tuah Pekanbaru memiliki data akademik mahasiswa yaitu data Indeks Prestasi Kumulatif (IPK), data identitas mahasiswa dan data Dosen. Data-data ini belum dimanfaatkan secara mendalam dan luas karena selama ini data-data tersebut masih digunakan hanya sebatas informasi saja. Data yang akan diproses dalam penelitian ini adalah bersumber dari data yang tersedia dan data penyebaran kuesioner. Data yang disebarkan melalui kuesioner terdiri dari 5 variabel yaitu adalah data yang terkait dengan peran dosen, motivasi, kedisiplinan, sosial ekonomi, dan hasil belajar masa lalu. Metode data mining yang digunakan adalah metode klasifikasi dengan algoritma C4.5. Algoritma ini dapat membentuk pohon keputusan yang menjadi alat dalam mendukung keputusan untuk memprediksi prestasi akademik mahasiswa. Dari hasil penelitian diperoleh bahwa variabel nilai rapor (hasil belajar masa lalu) menjadi node awal artinya dari 5 variabel yang menentukan prestasi akademik mahasiswa maka nilai rapor menjadi node yang terpilih sebagai penentu pertama terhadap prestasi akademik mahasiswa.

Kata kunci – data mining, algoritma c4.5, metode klasifikasi

PENDAHULUAN

Di masa sekarang ini teknologi basis data telah berkembang maka muncul kebutuhan untuk mendapatkan informasi yang lebih dari data yang tersimpan dalam basis data yaitu Pengetahuan (*Knowledge*). Informasi yang diinginkan tidak dapat diperoleh dengan mudah sebab volume data yang besar. Sehingga di butuhkan suatu metode untuk mendapatkan pengetahuan yang tidak terlihat di dalam data namun potensial untuk digunakan yaitu metode data mining. Berdasarkan data mahasiswa diterapkan suatu metode data mining klasifikasi dengan Algoritma C4.5 untuk memprediksi prestasi akademik mahasiswa. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan [2]. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan yang tersembunyi antara sejumlah variabel input dengan satu variabel target.

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Hang Tuah Pekanbaru terus berupaya untuk mencapai tujuan pendidikan yang berkualitas agar mahasiswa memiliki kemampuan profesional di bidang teknologi komputer. Salah satu indikator tercapainya tujuan tersebut adalah dari hasil prestasi akademik mahasiswa yang dinyatakan dengan Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK). Indeks Prestasi Semester merupakan nilai prestasi akademik mahasiswa dengan semua mata kuliah yang telah ditempuh untuk tiap semester tertentu. dan Indeks Prestasi Kumulatif merupakan prestasi akademik mahasiswa dengan menggabungkan semua mata kuliah yang telah ditempuh sampai suatu semester tertentu.[1]

Untuk mewujudkan pendidikan yang berkualitas berkaitan dengan peran dosen [5], motivasi mahasiswa, kedisiplinan mahasiswa [7], sosial ekonomi mahasiswa dan juga hasil belajar masa lalu. Hal tersebut di buktikan dengan banyaknya penelitian yang telah dilakukan sebelumnya. STMIK Hang Tuah Pekanbaru memiliki data akademik mahasiswa. Data ini berpotensi untuk menghasilkan informasi baru yang berguna. Salah satu yang dapat dilakukan oleh data mining adalah memprediksi prestasi akademik mahasiswa. Jika prestasi akademik mahasiswa dapat di ketahui lebih dini maka pihak program studi dapat melakukan tindakan tindakan yang diperlukan agar mahasiswa dapat mencapai prestasi akademik yang baik. Harapan akhirnya adalah semua mahasiswa dari berbagai latar belakang faktor dapat maksimal dalam meningkatkan prestasi akademiknya. Berdasarkan penjelasan di atas maka fokus penelitian ini adalah memprediksi prestasi akademik mahasiswa dengan menggunakan metode data mining klasifikasi berdasarkan peran dosen, motivasi mahasiswa, kedisiplinan mahasiswa, sosial ekonomi mahasiswa dan juga hasil belajar masa lalu.

METODOLOGI PENELITIAN

Penelitian ini dilaksanakan di STMIK Hang Tuah Pekanbaru. Waktu penelitian dilaksanakan pada mulai bulan Januari 2016. Target penelitian ini adalah mahasiswa STMIK Hang tuah Pekanbaru. Pada penelitian ini menggunakan 2 metode/pendekatan yaitu (1) Metode Pengumpulan Data dan (2) Metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) untuk proses dalam Data Mining nya. [4]

A. Metode Pengumpulan Data

Untuk mendapatkan data, seorang peneliti dapat melakukan berbagai metode. Data dapat diperoleh melalui kuesioner, wawancara dan dokumentasi.

- a. Kuesioner dilakukan dengan cara membagikan pertanyaan kepada para responden. Responden penelitian ini adalah mahasiswa STMIK Hang Tuah Pekanbaru sebanyak 250 orang. Kuesioner yang digunakan terlebih dahulu dilakukan uji Validitas dan Reliabilitas.
- b. Wawancara dilakukan dengan cara melakukan percakapan langsung dengan responden dan narasumber.
- c. Dokumentasi dapat ditempuh dengan cara meminta data pada lembaga atau instansi yang terkait, dalam hal ini adalah data prestasi akademik STMIK Hang Tuah Pekanbaru.

Selain dengan cara diatas juga digunakan metode lain untuk mendukung penelitian yang dilakukan yaitu dengan menggunakan data sekunder. Data Sekunder adalah data yang diperoleh secara tidak langsung bersumber dari dokumentasi, literature, buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti. Data Sekunder pada penelitian ini adalah Buku-buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti. Data sekunder pada penelitian ini adalah buku-buku, jurnal tentang Algoritma Decision Tree C4.5 dan data mining serta data mahasiswa STMIK Hang Tuah Pekanbaru .

B. Variabel (Atribut) Penelitian

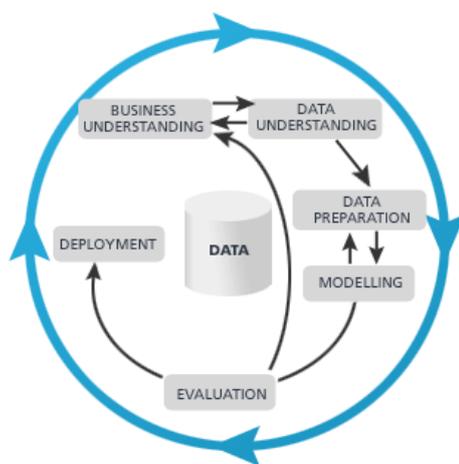
Variabel dalam penelitian terdiri dari 6 variabel yang terdiri dari 5 atribut biasa yaitu status ekonomi mahasiswa (X1) , disiplin (X2), motivasi (X3), peran dosen (X4) dan hasil belajar masa lalu dan 1 atribut sebagai target atribut (label) yaitu Prestasi Akademik.

C. Metode CRISP-DM (*Cross Industry Standard Process for Data Mining*)

Penelitian ini menggunakan metode CRISP-DM dalam proses Data Mining. Metode ini terdiri dari 6 tahapan :

- a. **Pemahaman terhadap bisnis (*Business/Research Understanding Phase*)**
Data yang diperoleh dari database Mahasiswa STMIK Hang Tuah Pekanbaru selama ini belum pernah dilakukan penggalian kekayaan terhadap data tersebut dan belum dimanfaatkan dalam menentukan prediksi prestasi akademik mahasiswa menggunakan metode Data Mining. Oleh karena itu penelitian ini akan menggali data tersebut dengan menggunakan metode Klasifikasi dengan Algoritma C4.5 [3].
- b. **Fase Pemahaman Data(*Data Understanding Phase*)**
Data di peroleh dari database mahasiswa STMIK Hang Tuah Pekanbaru sebanyak 250 mahasiswa. Atribut yang digunakan adalah yaitu NIM, Nama dan IPK. Dan data yang di dapatkan dari penyebaran kuesioner yang terdiri dari data peran dosen, status ekonomi mahasiswa, motivasi, disiplin dan hasil belajar masa lalu. Data data tersebut nanti nya akan diproses untuk memprediksi prestasi akademik mahasiswa. Sebelumnya dilakukan terlebih dahulu uji Validitas dan Reliabilitas terhadap kuesioner yang akan disebarakan kepada Responden (mahasiswa).
- c. **Persiapan data (*Data Preparation*)**
Dari 250 data mahasiswa maka akan dilakukan teknik Data Preparation agar kualitas data diperoleh lebih baik dengan cara:
 1. *Data Validation*, untuk mengidentifikasi dan menghapus data yang ganjil (outlier/noise), data yang tidak konsisten dan data yang tidak lengkap (missing value).
 2. *Data Integration and Transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penelitian ini bernilai kategorikal untuk model klasifikasi, data ditransformasi ke dalam angka menggunakan software Rapidminer.
 3. *Data Size Reduction and Dcretization*, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informative. Di dalam *data training* yang digunakan dalam penelitian ini, dilakukan seleksi atribut dan penghapusan data duplikasi.

- d. **Fase Pemodelan (*Modeling Phase*)**
Pada tahapan ini merupakan tahapan pemrosesan data training yang diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini menggunakan *algoritma C 4.5*.
- e. **Fase Evaluasi (*Evaluation Phase*)**
Pada fase ini dilakukan pengujian terhadap model klasifikasi dengan algoritma C4.5.
- f. **Fase Penyebaran (*Deployment Phase*)**
Setelah pembentukan model selanjutnya dapat digunakan untuk memprediksi prestasi akademik mahasiswa oleh Program Studi di lingkungan STMIK Hang Tuah Pekanbaru.



Gambar 1 Tahapan dari CRISP-DM

HASIL DAN PEMBAHASAN

Hasil penelitian berdasarkan tahapan tahapan proses dalam CRISP-DM. Tahapannya adalah (1) Pemahaman terhadap bisnis (*Business/Research Understanding Phase*) yaitu pemahaman terhadap permasalahan yang diteliti. Dalam penelitian permasalahannya adalah prediksi prestasi akademik mahasiswa menggunakan metode Data Mining. (2) Fase Pemahaman Data (*Data Understanding Phase*) yaitu pemahaman terhadap data yang akan diteliti. Data diperoleh dari database mahasiswa dan data yang di dapatkan dari penyebaran kuesioner yang terdiri dari data peran dosen, status ekonomi mahasiswa, motivasi, disiplin dan hasil belajar masa lalu. Sebelum kuesioner di sebar nya dilakukan terlebih dahulu uji Validitas dan Reliabilitas terhadap kuesioner yang akan disebar kepada Responden (mahasiswa) [6]. Untuk uji validitas dan reliabilitas di gunakan data sampel sebanyak 30 data yang di peroleh menggunakan teknik Purposive Sampling. (3) *Data Preparation* yaitu

tahap persiapan data. Data penelitian terdiri dari 6 variabel yang terdiri dari 5 atribut biasa yaitu status ekonomi mahasiswa (X1), disiplin (X2), motivasi (X3), peran dosen (X4) dan hasil belajar masa lalu dan 1 atribut sebagai target atribut (label) yaitu Prestasi Akademik. Data yang diperoleh kemudian dilakukan pembersihan data (cleaning data), transformasi data sehingga data yang digunakan selanjutnya tidak terdapat *missing*, lihat tabel berikut ini :

Tabel 1 Hasil Tahap Preparation

Nama Variabel	Type	Missing
IPK	Polynomial	0
Sosial	Integer	0
Disiplin	Polynomial	0
Motivasi	Polynomial	0
Dosen	Polynomial	0
Rapor	Polynomial	0

Kemudian berikutnya tahap (4) Fase Pemodelan (*Modeling Phase*) Pada tahapan ini merupakan tahapan pemrosesan data training yang diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini menggunakan *algoritma C 4.5*. Dari hasil proses data mining tersebut diperoleh *rule/aturan* sbb :

1. Jika hasil belajar masa lalu = sangat baik maka Prestasi Akademi = sangat memuaskan
2. Jika hasil belajar masa lalu = baik dan peran dosen = sangat baik maka Prestasi Akademik = sangat memuaskan.
3. Jika hasil belajar masa lalu = baik dan peran dosen = baik dan disiplin = baik maka Prestasi Akademik = sangat memuaskan

Berikutnya (5) *Fase Evaluasi (Evaluation Phase)*. Pada fase ini dilakukan pengujian terhadap model klasifikasi dengan algoritma C4.5. Pengujian ini dilakukan dengan analisis tingkat Performansinya dengan metode Area Under Curve (AUC) memperoleh nilai 65%. (6) Fase Penyebaran (*Deployment Phase*) Setelah pembentukan model selanjutnya dapat digunakan untuk memprediksi prestasi akademik mahasiswa oleh Program Studi di lingkungan STMIK Hang Tuah Pekanbaru.

KESIMPULAN

Berdasarkan analisis data menggunakan algoritma *Decision Tree* untuk memprediksi prestasi akademik berdasarkan sosial ekonomi, motivasi, peran dosen, disiplin dan hasil belajar masa lalu diperoleh hasil : (1) variabel hasil belajar masa lalu adalah variabel yang menentukan

potensi seseorang berhasil atau tidak dalam prestasi akademik. Hal ini dibuktikan bahwa Hasil Belajar menjadi node yang terpilih/awal. (2) Variabel Peran Dosen menjadi variabel kedua menentukan Prestasi akademik (3). Variabel Disiplin menjadi variabel ketiga menentukan Prestasi Akademik. (4) Hasil Akurasi klasifikasi menggunakan metode Area Under Curve (AUC) memperoleh nilai 65%.

REFERENSI

- [1] Hamalik.1991. *Manajemen Belajar di Perguruan Tinggi Bandung* : Sinar Baru .1991. hlm 16
- [2] Han,J and Kamber, M. 2012. *Data Mining Concept and Techniques Third Edition*. Morgan Kauffman. San Francisco.
- [3] Kusriani dan Luthfi, E.T.2009.*Algoritma Data Mining*. Andi Offset . Yogyakarta
- [4] Larose, Daniel T. 2005. *Discovering Knowledge in Data : An Introduction to Data Mining*. John Wiley & Sons.Inc Publication.
- [5] Margaretha Purwanti. *Peran Pengajaran Dosen, Konsep Pembelajaran, Konsep Diri Akademik, Dan Pendekatan Belajar Dalam Menentukan Hasil Belajar*.2006
- [6] Priyatno. Duwi. 2009. *5 jam belajar olah data dengan SPSS 17*. Andi Yogyakarta
- [7] Tulus. 2004. *Peran Disiplin pada perilaku dan prestasi siswa*. Jakarta : Grasindo