



DATA MINING DENGAN 2 (DUA) MODEL KLASIFIKASI UNTUK PREDIKSI KINERJA MAHASISWA

Eka Sabna¹, Yuda Irawan²

^{1,2}STMIK Hang Tuah Pekanbaru, Teknik Informatika

Email :

es3jelita@yahoo.com, yudairawan89@gmail.com

Abstract

The abundance of student data in higher education can be used optimally according to needs and can be processed into useful information so that it can find out the relationship between data attributes in which it can be analyzed and is expected to have output in the form of student performance related to Learning Outcomes (GPA). The method used is the CRISP-DM method which consists of 6 stages. There are 2 (two) classification methods used, namely Nave Bayes Classifier and Decision Tree Algorithm C4.5 to compare which algorithm is better for predicting student performance. The results of the implementation of data mining using Rapidminer software, were carried out on two classification algorithm models, namely C4.5 and NBC, then entered the dataset as test material for both models which contained training data and test data. Based on the best accuracy value on the Naive Bayes Classifier (NBC) algorithm model is 80%, while based on the best accuracy value on the C4.5 algorithm model is 60%.

Keywords: C4.5, NBC, GPA, performance, student.

Abstrak

Berlimpahnya data mahasiswa pada perguruan tinggi dapat digunakan secara maksimal sesuai dengan kebutuhan dan mampu diolah menjadi informasi yang bermanfaat sehingga dapat mengetahui hubungan antara atribut data yang di dalamnya dapat dianalisis dan diharapkan memiliki keluaran berupa kinerja mahasiswa yang berhubungan Hasil Belajar (IPK). Metode yang digunakan menggunakan metode CRISP-DM yang terdiri dari 6 tahapan. Ada 2 (dua) metode klasifikasi yang digunakan yaitu Naive Bayes Classifier dan Decision Tree Algoritma C4.5 untuk dilakukan perbandingan algoritma mana yang lebih baik untuk memprediksi kinerja mahasiswa. Hasil dari implementasi data mining dengan menggunakan software Rapidminer, dilakukan terhadap dua model algoritma klasifikasi yaitu C4.5 dan NBC kemudian memasukan dataset sebagai bahan uji untuk kedua model tersebut yang di dalamnya terdapat data latihan dan data uji. Berdasarkan nilai akurasi terbaik pada model algoritma Naive Bayes Classifier (NBC) adalah 80% sedangkan berdasarkan nilai akurasi terbaik pada model algoritma C4.5 adalah 60%.

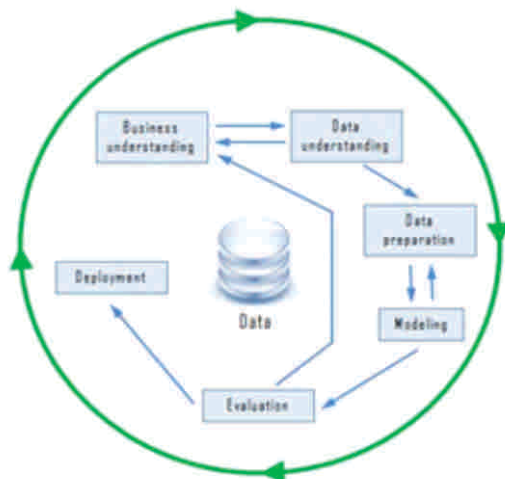
Keywords: C4.5, NBC, IPK, kinerja, mahasiswa.

PENDAHULUAN

Salah satu cara didalam perguruan tinggi untuk mendapat Informasi pola untuk kinerja mahasiswa adalah dengan menggali pengetahuan di dalam database sehingga dapat ditemukan mengenai atribut utama yang dapat berpengaruh terhadap kualitas kinerja mahasiswa . Salah satu atribut yang dapat digunakan sebagai parameter keberhasilan sebuah perguruan tinggi yaitu mahasiswa dapat melaksanakan masa studi secara tepat pada waktunya sesuai jenjang yang diambil. Educational Data Mining (EDM) dapat diterapkan dengan banyak metode seperti menggunakan teknik Pohon Keputusan (Decision Tree), Jaringan Syaraf Tiruan, Naive Bayes Classifier (NBC) dan lain-lain. 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. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan yang tersembunyi antara sejumlah variabel input dengan satu variabel target. Penelitian ini menggunakan kedua algoritma tersebut yaitu Naïve Bayes Classifier dan Decision Tree untuk dilakukan perbandingan algoritma mana yang lebih baik.

METODE PENELITIAN

Dalam penelitian ini akan mengimplementasikan model standar dalam proses penggalian data yaitu CRISP-DM dalam memecahkan masalah yang terjadi contohnya pada bidang bisnis ataupun dalam bidang penelitian, adapun proses yang terjadi dalam penggalian data memiliki siklus hidup (life cycle) yang dibagi menjadi enam tingkatan masa atau tahapan proses (Larose, 2006).



Gambar 1. Tahapan CRIPS-DM

Tahapan Pemahaman Bisnis

Ada beberapa hal yang dilakukan antara lain, menentukan tujuan bisnis, menilai situasi, dan menentukan tujuan data mining. Penelitian ini untuk mengklasifikasi data kinerja

mahasiswa dengan 2 (dua) Algoritma yang digunakan yaitu C4.5 dan NBC.

Tahapan Pemahaman data

Pada tahapan pemahaman data, ada beberapa hal yang dilakukan antara lain, mengumpulkan data awal, mendeskripsikan data, mengeksplorasi data, dan memverifikasi kualitas data. Data yang digunakan adalah data kinerja mahasiswa.

Tahapan Persiapan Data

Pada tahapan persiapan data, ada beberapa hal yang dilakukan antara lain, deskripsi data set, memilih data, membangun data, mengintegrasikan data dan membersihkan data. Data mahasiswa yang digunakan terdiri dari atribut yaitu nama, jenis kelamin, Asal sekolah, pekerjaan orang tua dan IPK. Data mahasiswa yang digunakan sebanyak 40 orang.

Tabel 1. Atribut Data Penelitian.

Atribut	Deskripsi	Nilai
Nama	Nama Mahasiswa	Nama
Jenis Kelamin	Jenis Kelamin	Laki-laki Perempuan
Asal Sekolah	Asal-sekolah	SMA SMK
Pekerjaan OrangTua	Pekerjaan Ortu	PNS Petani Pedagang Swasta Nelayan
IPK	IPK	Baik Baik Sekali Sangat Baik

Tahap Pemodelan dan Evaluasi

Pada tahapan pemodelan, ada beberapa hal yang dilakukan antara lain, memilih teknik pemodelan, membangun model, dan menilai model. Dalam penelitian menggunakan 2 (dua) algoritma klasifikasi yaitu C4.5 dan NBC.

Evaluasi dilakukan dengan tujuan agar hasil pada tahap modelling sesuai dengan sasaran yang ingin dicapai dalam tahap business understanding'. Hasil dari 2 algoritma yang digunakan yaitu C4.5 dan NBC memberikan hasil nilai akurasi dari kedua algoritma tersebut.

Tahapan Penyebaran

Diharapkan pada penelitian ini dapat menghasilkan sebuah analisa dengan akurasi yang baik dari salah satu dari 2 algoritma klasifikasi C4.5 dan NBC. Sehingga dapat membantu institusi perguruan tinggi dalam menentukan Algoritma yang digunakan untuk klasifikasi terhadap data kinerja mahasiswa.

HASIL & PEMBAHASAN

Pada proses pengujian validasi terhadap dataset dilakukan menggunakan Rapid Miner Studio seperti pada Gambar 2,3, 4 dan 5.

besar dalam pendataan karyawan. sehingga menyulitkan Divisi HRD dalam melakukan penginputan data karyawan.

2.Melihat informasi data karyawan Divisi HRD masih kesulitan dalam rekap laporan, karena data harus dicari dahulu dan membutuhkan waktu yang lama, sehingga pekerjaan Divisi HRD menjadi tidak efisien.

c.Economy (Ekonomi)

Ekonomi adalah penilaian sistem atas pengurangan dan keuntungan yang akan didapatkan dari sistem yang dikembangkan. Dari hasil analisis ini, akan ditemukan hasil.

1.Aplikasi data karyawan Rumah Sakit Ibu dan Anak Budhi Mulia yang dibangun dapat mempercepat dalam pencarian dan pengelolaan data-data laporan, sehingga memudahkan pekerjaan Divisi HRD dan waktu lebih efektif.

2.Aplikasi yang dibangun dapat mengurangi dari segi penggunaan kertas, sehingga pengeluaran biaya akan berkurang dalam pembelian buku besar.

d.Control (Pengendalian)

Analisis kontrol digunakan untuk meningkatkan kinerja sistem dalam mencegah dan mendeteksi secara dini terhadap kesalahan sistem atau penyalahgunaan oleh pihak-pihak yang tidak berkepentingan dan menjamin keamanan dan keakuratan data serta informasi yang akan disampaikan.

1.Dalam perancangan aplikasi yang dibuat adanya hak akses siapa saja yang berhak untuk menggunakan sistem. Untuk mencegah terjadinya kesalahan terhadap sistem agar menjamin keamanan dan keakuratan data dan informasi yang ada.

e.Efficiency(Efisiensi)

Efisiensi ini adalah memahami bagaimana menghasilkan output sebanyak banyaknya dengan input yang sekecil mungkin. Ini sangat berhubungan dengan sumber daya yang ada, guna meminimalkan pemborosan. Efisiensi dari Sistem yang dikembangkan adalah pemakaian secara maksimal atas sumber daya yang tersedia yang meliputi manusia, informasi, waktu, uang, peralatan, ruang dan keterlambatan pengolahan data.

1.Divisi HRD membutuhkan waktu yang lama dalam pencarian data karyawan, sehingga mengurangi efisiensi dan keefektifan pekerjaan.

2.Dalam membuat laporan akan membutuhkan waktu yang lama. Karena data terlebih dahulu dicari didalam buku pendataan karyawan.

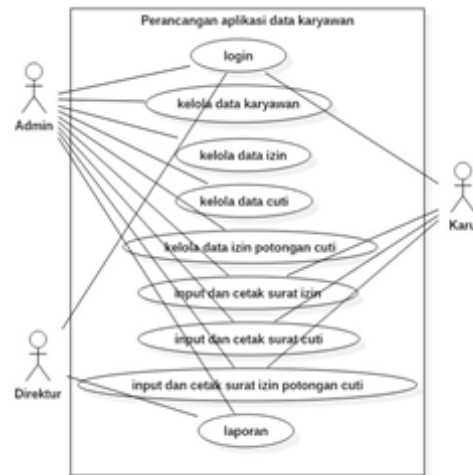
f.Service (Service)

Suatu instansi dapat dikatakan lebih berhasil jika selalu memberikan peningkatan pelayanan yang sangat baik. Dari peningkatan pelayanan terhadap sistem yang dikembangkan akan memberikan suatu akurasi dalam pengolahan data, kehandalan terhadap konsistensi dalam pengolahan input dan output serta kehandalan dalam menangani pengecualian, sistem yang mudah dan enak dipakai, dan mampu mengkoordinasi aktifitas untuk mencapai tujuan dan sasaran.

1.Membutuhkan waktu yang lama dengan sistem pendataan yang manual yang ada pada saat ini.

Use Case Diagram

Perancangan Use Case Diagram dapat menggambarkan kebutuhan fungsional dari aplikasi yang dibuat. Dengan dirancangnya use case diagram ini, maka dapat dideskripsikan interaksi antara aktor dengan sistem, dapat dilihat pada gambar di bawah ini:



Gambar 1. Use Case Diagram

Tahap Pemodelan dan Evaluasi

Pada tahapan pemodelan, ada beberapa hal yang dilakukan antara lain, memilih teknik pemodelan, membangun model, dan menilai model. Dalam penelitian menggunakan 2 (dua) algoritma klasifikasi yaitu C4.5 dan NBC.

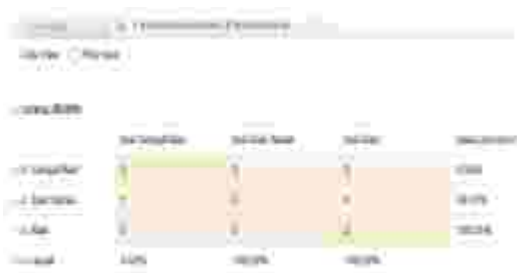
Evaluasi dilakukan dengan tujuan agar hasil pada tahap modelling sesuai dengan sasaran yang ingin dicapai dalam tahap business understanding'. Hasil dari 2 algoritma yang digunakan yaitu C4.5 dan NBC memberikan hasil nilai akurasi dari kedua algoritma tersebut.

Tahapan Penyebaran

Diharapkan pada penelitian ini dapat menghasilkan sebuah analisa dengan akurasi yang baik dari salah satu dari 2 algoritma klasifikasi C4.5 dan NBC. Sehingga dapat membantu institusi perguruan tinggi dalam menentukan Algoritma yang digunakan untuk klasifikasi terhadap data kinerja mahasiswa.



Gambar 2. Desain Model NBC



Gambar 3. Nilai Akurasi dari Model NBC



Gambar 4. Desain Model C4.5



Gambar 5. Nilai Akurasi dari Model C4.5

SIMPULAN DAN SARAN

Berdasarkan pengujian pada kedua model algoritma tersebut dengan menggunakan data training adapun nilai akurasi terbaik pada model algoritma Naive Bayes Classifier (NBC) adalah 80 % sedangkan pada model algoritma C4.5 adalah 60 %.

Dapat disimpulkan nilai akurasi terbaik dari hasil komparasi kedua model algoritma diperoleh oleh model algoritma Naive Bayes dengan nilai akurasi 80%.

TERIMAKASIH

Terima kasih saya haturkan kepada Yayasan Hang Tuah Pekanbaru dan Ketua STMIK Hang Tuah Pekanbaru atas segala dukungan terhadap penyelesaian penelitian yang saya lakukan ini. Semoga penelitian ini dapat bermanfaat sebagai penambahan referensi bagi peneliti selanjutnya

DAFTAR PUSTAKA

Irawan, Y., Linarta, A., & Febriani, A. (2021, March). Smart Home Light Based Service Oriented Architecture and IoT. In *Journal of Physics: Conference Series* (Vol. 1845, No. 1, p. 012070). IOP Publishing.

Larose, D. T. (2006). DATA MINING METHODS AND MODELS. In *Contemporary Psychology: A Journal of Reviews* (Vol. 21). <https://doi.org/10.1037/014836>

Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data. In *Discovering Knowledge in Data*. <https://doi.org/10.1002/9781118874059>

Linarta, A., & Nurhadi, N. (2019). Aplikasi Bel Sekolah Otomatis Berbasis Arduino Dilengkapi Dengan Output Suara. *Informatika*, 10(2), 1-7.

Rapidminer Inc. (2019). Split Validation (RapidMiner Studio Core). Retrieved May 8, 2019, from https://docs.rapidminer.com/latest/studio/operators/validation/split_validation.html

Rahmalisa, U., Febriani, A., & Irawan, Y. (2021). Detector Leakage Gas Lpg Based On Telegram Notification Using Wemos D1 and Mq-6 Sensor. *Journal of Robotics and Control (JRC)*, 2(4), 287-291.

Sabna, E. (2019). Pemanfaatan Data Mining Untuk Penempatan Buku Di Perpustakaan Menggunakan Metode Association Rule: Pemanfaatan Data Mining Untuk Penempatan Buku Di Perpustakaan Menggunakan Metode Association Rule. *Jurnal Ilmu Komputer*, 8(2), 59-63.

Sabna, E., & Muhardi, M. (2016). Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, Dan Hasil Belajar. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi*

Tair, M. M. A., & El-halees, A. M. (2012). Mining Educational Data to Improve Students 'Performance : A Case Study. *Journal of Theoretical and Applied Information Technology*, 2(2), 140–146.

Tempola, F., Muhammad, M., & Khairan, A.

(2018). Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation Comparison of Classification Between Knn and Naive Bayes At the Determination of the Volcanic Status With K-Fold Cross. 5(5),577–584.

<https://doi.org/10.25126/jtiik20185983>