



KILAT

JURNAL KAJIAN ILMU DAN TEKNOLOGI

Endah Lestari;
Irma Wirantina K;
Ranti Hidayawanti

Faisal;
Muhammad Ridwan;
Mardawati

Gita Puspa Artiani;
Indah Handayasari

Kresna Ramanda;
Irmawati Carolina

Ratna Mutu Manikam;
Farid Setiawan

Rayung Wulan

Roni Kartika Pramuyanti

Rr. Mekar Ageng Kinasti;
Djoko Nugroho Notodisuryo

Satria;
Ayu Setiawati Agustini

Syam Gunawan;
Pritasari Palupiningsih

Ali Ridho Gumelar;
Anton;
Ummu Radiyah

M. Yoga Distra Sudirman;
Yessy Fitriani

ANALISA TAMAN ATAP DALAM UPAYA MENGURANGI LIMPASAN AIR HUJAN PADA BANGUNAN PERKOTAAN

OPTIMASI DIAGRAM LAYANAN PEMBELIAN DALAM Mendukung MANAJEMEN HUBUNGAN PELANGGAN

OPTIMALISASI PENGOLAHAN SAMPAH ORGANIK DENGAN TEKNOLOGI BIODIGESTER SEBAGAI UPAYA KONSERVASI LINGKUNGAN

SELEKSI FITUR ALGORITMA *NEURAL NETWORK* MENGGUNAKAN *PARTICLE SWARM OPTIMIZATION* UNTUK MEMPREDIKSI KELAHIRAN PREMATUR

RANCANG BANGUN MEDIA BELAJAR FISIKA DASAR UNTUK MAHASISWA BERBASIS ANDROID

ANALISIS *GLOBAL POSITIONING* INFRASTRUKTUR DAN JARINGAN SOSIAL MEDIA DALAM BISNIS APLIKASI TRANSPORTASI DARAT *ONLINE* DI JAKARTA

PENGARUH INOVASI ANTENA PADA SIARAN TELEVISI MOBIL

PEMANFAATAN LIMBAH PEMBAKARAN BATUBARA (*BOTTOM ASH*) PADA PLTU SURALAYA SEBAGAI MEDIA TANAM DALAM UPAYA MENGURANGI PENCEMARAN LINGKUNGAN

PENERAPAN METODE *GRAPHIC RATING SCALE* (GRS) DALAM PENILAIAN KINERJA KARYAWAN

PEMBENTUKAN MODEL KLASIFIKASI DATA LAMA STUDI MAHASISWA STMIK INDONESIA MENGGUNAKAN *DECISION TREE* DENGAN ALGORITMA NBTree

IMPLEMENTASI *LOAD BALANCING* DENGAN ALGORITMA *EQUAL COST MULTI PATH* (ECMP)

RANCANGAN SISTEM PENILAIAN HASIL KINERJA *MULTI COMPANY* DAN *CROSS BUSINESS SECTOR*

ISSN 2089-1245



9 772089 124519

SEKOLAH TINGGI TEKNIK - PLN (STT-PLN)

KILAT	VOL.6	NO.2	HAL. 81 - 162	OKTOBER 2017	ISSN 2089 - 1245
-------	-------	------	---------------	--------------	------------------

PEMBENTUKAN MODEL KLASIFIKASI DATA LAMA STUDI MAHASISWA STMIK INDONESIA MENGGUNAKAN *DECISION TREE* DENGAN ALGORITMA NBTree

Syam Gunawan¹, Pritasari Palupiningsih²

^{1,2} Program Studi Sistem Informasi, STMIK Indonesia

e-mail: ¹syam@stmik-indonesia.ac.id, ²prita@stmik-indonesia.ac.id

ABSTRAK

Salah satu kriteria penilaian pada akreditasi program studi adalah penilaian terhadap lama studi mahasiswa yang lulus tepat waktu. Tidak sedikit mahasiswa yang menempuh masa studi melebihi standar kelulusan yang telah ditetapkan. Sehingga penting bagi program studi untuk mengetahui mahasiswa mana saja yang memiliki kemungkinan lulus tidak tepat waktu. Untuk itu diperlukan adanya prediksi lama studi mahasiswa. Salah satu cara untuk dapat memprediksi lama studi mahasiswa adalah dengan membangun model klasifikasi. Penelitian ini bertujuan untuk membangun model prediksi lama studi mahasiswa menggunakan *Decision Tree* dengan algoritma NBTree. Data yang digunakan adalah data nilai akademik serta data cuti akademik mahasiswa. Hasil yang diperoleh adalah model klasifikasi berupa *Naive Bayes Decision Tree* dengan akurasi 73,45%.

Kata Kunci: Lama Studi Mahasiswa, *Decision Tree*, Algoritma NBTree

ABSTRACT

One of the assessment criteria for the accreditation of the study program is the assessment of the duration of the study of students who graduated on time. Not a few students who pursue the study period exceeds the established standard of graduation. So it is important for the study program to know which students have the possibility of passing is not timely. For that it is necessary to predict the length of student study. One way to predict the length of a student's study is to build a classification model. This study aims to build a long prediction model of student study using *Decision Tree* with NBTree algorithm. The data used are academic value data and student academic leave data. The result obtained is a classification model of *Naive Bayes Decision Tree* with 73.45% accuracy.

Keywords: Duration of Student Studies, *Decision Tree*, NBTree Algorithm

1. PENDAHULUAN

Program Studi merupakan garda terdepan dalam penyelenggaraan pendidikan dari sebuah Perguruan Tinggi, sehingga program studi senantiasa melakukan evaluasi guna meningkatkan mutu dan efisiensi perguruan tinggi termasuk peningkatan kualitas lulusan. Selain itu, salah satu kriteria penilaian pada akreditasi program studi adalah penilaian terhadap lama studi mahasiswa yang lulus tepat waktu. Masa studi mahasiswa telah diatur dalam ketetapan Kementerian Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Tinggi tentang Sistem Pendidikan Tinggi yang menyebutkan bahwa untuk memenuhi standar kompetensi lulusan bagi mahasiswa program sarjana (S1) beban wajib yang harus ditempuh adalah paling sedikit 144 - 160 satuan kredit semester (sks) dengan masa studi selama 8 - 10 semester atau 4 - 5 tahun.

Pada institusi pendidikan perguruan tinggi, data mahasiswa dan data jumlah kelulusan mahasiswa dapat menghasilkan informasi yang berlimpah berupa jumlah kelulusan setiap tahunnya, profil, dan hasil akademik mahasiswa selama menempuh proses kegiatan belajar mengajar di perguruan tinggi. Adanya informasi mengenai lama studi mahasiswa tentu akan menjadi pendukung

suatu pengambilan keputusan yang tepat bagi manajemen Perguruan Tinggi dalam mengambil langkah berikutnya.

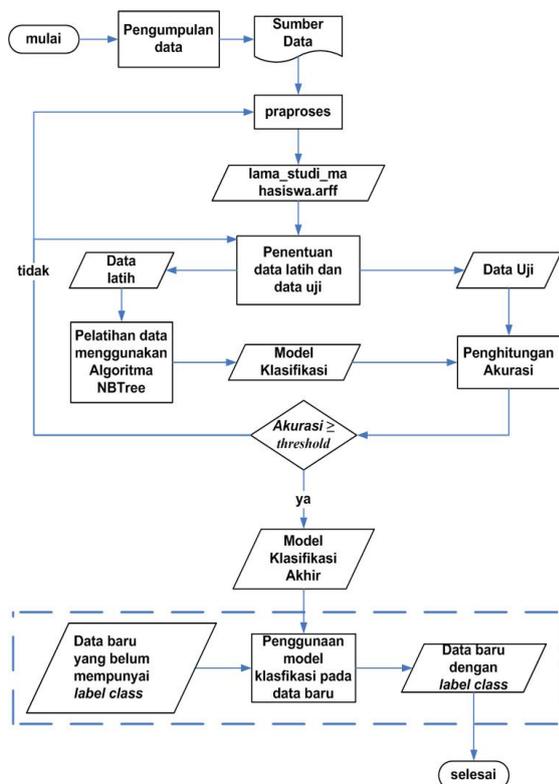
Permasalahan yang sering terjadi adalah masih banyaknya jumlah mahasiswa yang lulus dengan masa studi melampaui waktu yang telah ditetapkan atau tidak tepat waktu. Hal ini dapat mempengaruhi mutu lulusan Perguruan Tinggi. Sehingga penting bagi program studi untuk mengetahui mahasiswa mana saja yang memiliki kemungkinan lulus tidak tepat waktu. Untuk itu diperlukan adanya prediksi lama studi mahasiswa. Salah satu cara untuk dapat memprediksi lama studi mahasiswa adalah dengan membangun model klasifikasi. Selama ini STMIK Indonesia belum memiliki model klasifikasi lama studi mahasiswa yang dapat digunakan sebagai prediksi jumlah lulus tepat waktu. Padahal data mahasiswa sangat berlimpah, hanya saja data-data tersebut belum dimanfaatkan untuk dianalisis lebih jauh.

Data lama studi mahasiswa berukuran besar dapat dianalisis menggunakan teknik klasifikasi. Salah satu metode dalam klasifikasi adalah *decision tree* yang akan menghasilkan model klasifikasi. Salah satu algoritma yang dapat diterapkan dalam metode *decision tree* adalah NBTree. Model klasifikasi yang terbentuk akan dapat digunakan dalam prediksi. Penelitian ini

bertujuan untuk menerapkan metode *decision tree* dengan algoritma NBTee untuk membentuk model klasifikasi. Hasil dari penelitian ini adalah terbentuknya model klasifikasi data lama studi mahasiswa STMIK Indonesia yang nantinya dapat digunakan untuk prediksi jumlah mahasiswa lulus tepat waktu.

2. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data mahasiswa 2013-2015 di STMIK Indonesia. Penelitian ini dilakukan secara bertahap sesuai tahapan yang telah disusun pada Gambar 2. Nilai *threshold* yang digunakan pada penelitian ini adalah 70%.



Gambar 2. Metode Penelitian

2.1 Praproses Data

Data dari sumber dikumpulkan dan dilakukan tahapan praproses data sebagai berikut :

1. Seleksi data
Pada tahap ini, dilakukan pengelompokan mahasiswa berdasarkan lama studinya dan memilih atribut-atribut yang sesuai dengan kategori permasalahan.
2. Pembersihan data
Pada data dilakukan pembersihan data untuk memperbaiki data yang hilang atau kosong, data yang mengandung *noise*, dan data yang tidak konsisten.
3. Integrasi data
Pada tahap ini dilakukan penggabungan data dari berbagai sumber ke suatu basis data. Kemudian dilakukan proses reduksi data,

dimana data yang tidak relevan dan data yang redundansi dibuang.

4. Transformasi data

Proses perubahan bentuk ke dalam bentuk data yang tepat agar dapat digunakan untuk proses selanjutnya. Proses ini meliputi penyeragaman nama atribut.

2.2 Pembagian Data Latih dan Data Uji

Proses pembagian data menjadi data latih dan data uji dilakukan dengan menggunakan 10 *fold cross validation*. Data latih akan digunakan untuk membentuk model klasifikasi. Sedangkan data uji akan digunakan untuk menghitung akurasi yang diperoleh dari model klasifikasi.

2.2.1 Klasifikasi

Pada proses klasifikasi dilakukan pembentukan model klasifikasi menggunakan metode *decision tree*. Kemudian dilakukan penghitungan akurasi dari model klasifikasi yang terbentuk. Dari proses klasifikasi ini akan diperoleh model klasifikasi yang dapat digunakan untuk mengisi label kelas dari data baru yang belum diketahui label kelasnya.

2.2.2 Model Klasifikasi

Proses pembentukan model klasifikasi dari data mahasiswa menggunakan data tabel gabungan dan kelas targetnya. Pembentukan model klasifikasi ini menggunakan salah satu algoritme pembentukan *decision tree* yaitu algoritme NBTee.

2.2.3 Penghitungan Akurasi

Tahap ini adalah tahap untuk menghitung akurasi dari model klasifikasi yang diperoleh dari proses klasifikasi. Metode yang digunakan dalam proses penghitungan akurasi ini adalah dengan menggunakan *confusion matrix*.

Jika hasil akurasi yang diperoleh sudah memenuhi nilai *threshold*, maka model klasifikasi itu akan digunakan untuk menentukan label kelas dari data baru. Akan tetapi jika akurasi yang diperoleh belum memenuhi nilai *threshold*, maka proses klasifikasi akan diulang dengan menggunakan proporsi data latih dan data uji yang berbeda atau mengulang tahap praproses dengan objek yang berbeda.

3. HASIL DAN PEMBAHASAN

3.1 Praproses Data

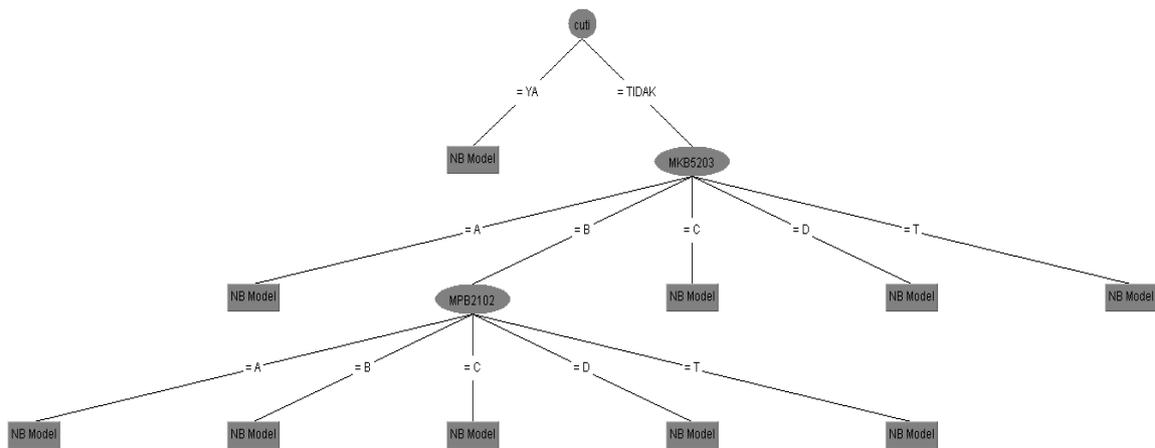
Data yang digunakan pada penelitian ini adalah data nilai akademik mahasiswa, data cuti akademik, dan data ketepatan waktu lulus mahasiswa program studi sistem informasi tahun 2013 – 2015. Tidak semua data yang terdapat pada data nilai akademik mahasiswa digunakan pada penelitian ini. Setelah melalui tahap pemilihan atribut, terdapat 36 atribut yang digunakan pada penelitian ini. Atribut itu terdiri dari 34 mata kuliah semester 1 sampai semester 4, cuti kuliah dan ketepatan lulus studi. Atribut ketepatan lulus studi menjadi kelas dari data yang digunakan pada penelitian ini.

Proses selanjutnya adalah proses pembersihan data. Salah satu tujuan proses pembersihan data adalah untuk mengganti data yang kosong. Jika terdapat nilai atribut yang kosong untuk suatu record, akan diganti dengan nilai T. Dimana nilai T ini berarti mahasiswa tersebut tidak mengambil mata kuliah tersebut. Hal ini terjadi karena terdapat perbedaan kurikulum antara mahasiswa yang lulus tahun 2013 dengan mahasiswa yang lulus tahun 2014 – 2015. Selain itu nilai masing-masing atribut mata kuliah terdiri dari A, B, C, D, dan T. Tidak terdapat nilai E, dikarenakan salah satu persyaratan sidang skripsi adalah tidak terdapat nilai E. Sehingga mahasiswa yang sudah lulus, otomatis tidak memiliki nilai E.

Kemudian masing-masing data tersebut digabungkan menjadi satu.

3.2 Klasifikasi

Proses klasifikasi dilakukan dalam dua tahap, yaitu pembentukan model klasifikasi berupa *decision tree* dan penghitungan akurasi dari *decision tree* yang terbentuk. Pembentukan *decision tree* dilakukan dengan menggunakan algoritme NBTree. Proses pembentukannya dilakukan dengan perangkat lunak Weka. Gambar 3 merupakan *decision tree* yang dibentuk. Pada *decision tree* yang dihasilkan, diketahui bahwa tidak semua atribut yang digunakan muncul sebagai *node* pada *decision tree*.



Gambar 3. Model Klasifikasi Data Lama Studi Mahasiswa Program Studi Sistem Informasi STMIK Indonesia

Dari gambar 3 di atas, dapat dilihat bahwa dari 35 atribut yang digunakan, hanya 3 atribut yang muncul pada *decision tree* tersebut, yaitu cuti kuliah, MKB5203 (Mata kuliah Sistem Operasi) dan MPB2102 (Mata kuliah Komputer dan Masyarakat). *Leaf node* yang dihasilkan dari *decision tree* yang dibentuk dari algoritme NBTree merupakan sebuah model naïve bayes, dimana model ini berisi peluang untuk masing-masing kelas, dan peluang setiap atribut terhadap masing-masing kelas.

Berdasarkan *decision tree* yang terbentuk, dapat dibuat 10 aturan. Sebagai contoh, aturan yang terbentuk dari *decision tree* pada Gambar 3, adalah sebagai berikut :

- Aturan 1:
Jika nilai atribut cuti adalah YA, maka penentuan kelas dapat dihitung menggunakan model naïve bayes 1.
- Aturan 6:
Jika nilai atribut cuti adalah TIDAK, nilai atribut MKB5203 adalah B, dan nilai atribut MPB2102 adalah A, maka penentuan kelas dapat dihitung menggunakan model naïve bayes 6.
Nilai yang dihitung dengan menggunakan model naïve bayes adalah *conditionally independent* antara satu atribut dan atribut lainnya dengan menggunakan persamaan (2). Kemudian dilakukan penghitungan peluang suatu *record* termasuk ke dalam setiap kelas yang dihitung

dengan Persamaan (3). Penghitungan ini dilakukan untuk semua kelas yang ada. Untuk penentuan kelas, diambil berdasarkan nilai peluang suatu record termasuk ke dalam suatu kelas yang paling tinggi.

3.3 Penghitungan Akurasi

Untuk menghitung akurasi dari model klasifikasi yang terbentuk digunakan *confusion matrix*. *Confusion matrix* yang diperoleh dari model klasifikasi tersebut terdapat pada Tabel 2.

Tabel 2. *Confusion matrix* dari model klasifikasi

		Kelas yang diprediksi	
		Kelas = Tepat	Kelas = Tidak Tepat
Kelas Aktual	Kelas = Tepat	270	35
	Kelas = Tidak Tepat	115	145

Penghitungan akurasi dengan menggunakan *confusion matrix* adalah sebagai berikut:

$$\text{akurasi} = \frac{\text{banyaknya prediksi yang benar}}{\text{total banyaknya prediksi}}$$

Dengan menggunakan data pada tabel *confusion matrix*, dapat dihitung akurasi dari model klasifikasi.

$$\begin{aligned} \text{akurasi} &= \frac{270 + 145}{565} \\ &= \frac{415}{565} = 0,7345 \end{aligned}$$

Hasil akurasi yang diperoleh adalah 73,45%. Sedangkan nilai *threshold* yang digunakan adalah 70%. Sehingga nilai akurasi yang diperoleh telah memenuhi *threshold* yang diberikan.

4. KESIMPULAN

4.1 Kesimpulan

Berdasarkan penelitian yang dilakukan dalam membentuk model klasifikasi untuk data lama studi mahasiswa, dapat diambil kesimpulan sebagai berikut :

1. Telah terbentuk model klasifikasi untuk yang memiliki 10 aturan klasifikasi dengan akurasi 73,45%.
2. Lama studi mahasiswa dapat dideskripsikan oleh nilai akademik dan data cuti akademik mahasiswa.

4.2 Saran

1. Data yang digunakan bisa ditambah lebih dari 3 tahun, agar penelitiannya agar model data klasifikasinya lebih variatif.
2. Gunakan metode penelitian yang lain, agar bisa membandingkan hasil penelitian NBTre

DAFTAR PUSTAKA

- [1] Han J, Kamber M. 2006. *Data Mining : Concepts and Techniques*. Morgan Kaufman Publisher, San Francisco .
- [2] Hastuti K. 2012. *Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif*. Semarang : Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012.
- [3] Kohavi R. 1996. *Scaling Up the Accuracy of Naive-Bayes Classifiers : a Decision-Tree Hybrid*. AAAI.
- [4] Tan P, Michael S, dan Vipin K. 2005. *Introduction to Data mining*. Pearson Education, Inc, Boston.
- [5] Williams N, Zander S, & Armitage G. 2006. *Evaluating Machine Learning Algorithms for Automated Network Application Identification*. CAIA.