EVALUASI KINERJA KLASIFIKASI DATA UNTUK LAYANAN AKADEMIK DAN PREDIKSI KELULUSAN MAHASISWA

Meliana O. Meo¹, Donzilio Antonio Meko²

Program Studi Teknik Informatika STIKOMUyelindo Kupang¹, STIMIK Kupang² meliana.oktavia.g@gmail.com¹, donzi.antonio.g@gmail.com²

ABSTRACT

STIKOM Uyelindo Kupang was established in the year 2000 as an information technology-based tertiary institution which has three study programs, namely under graduate of informatics engineering, diploma three informatics engineering and under graduate of information systems. The three study programs always strive to improve the status of accreditation by continuously improving internal quality and making accreditation a strategy to compete with other universities. To maintain quality, STIKOM Uyelindo Kupang, especially the undergraduate informatics engineering study program routinely monitors and evaluates the performance of lecturers. The problem that is often faced in routine monitoring and evaluation of lecturer performance is the performance evaluation process that is still objective so that to overcome these problems, a decision support system is needed that can assist in evaluating the performance of lecturers at STIKOM Uyelindo Kupang. The purpose of this study is to make a decision support system for the assessment of performance of lecturers of the first-degree informatics engineering study program at STIKOM Uyelindo Kupang using TOPSIS method. The results of this study are in the form of a desktop-based application that can facilitate the monitoring and performance evaluation teams of lecturers in evaluating the performance of lecturers of study programs.

Keywords: Decision support system, evaluation, lecturer, TOPSIS, tridharma

1. PENDAHULUAN

Klasifikasi data merupakan salah satu metode dalam data mining yang digunakan untuk menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya kedalam kelas-kelas yang berbeda berdasarkan model klasifikasi yang ditetapkan [1] Pemanfaatan metode data mining ini telah banyak diterapkan dalam berbagai bidang seperti pertanian, bisnis hingga pendidikan.

Bidang pendidikan merupakan salah satu bidang yang kaya akan informasi. Informasi yang tersimpan dalam database pendidikan meningkat pesat setiap tahunnya dan informasi-informasi tersebut berkaitan dengan data mahasiswa [2]. Dari informasi yang ada, salah satu informasi yang dapat digali adalahin formasimengenai data mahasiswa dan data kelulusan mahasiswa.

Informasi mengenai data kelulusan mahasiswa merupakanin formasi yang sangat penting bagi suatu perguruan tinggi. Hal ini dikarenakan salah satu faktor yang menentukan kualitas perguruan tinggi adalah persentasi kemampuan mahasiswa untuk menyelesaikan stu ditepat pada waktunya. Penurunan tingkat kelulusan mahasiswa akan

berpengaruh terhadap akreditasi perguruan tinggi tersebut, untuk itu perlu adanya monitoring serta evaluasi terhadap kecenderungan kelulusan mahasiswa tepat waktu ataut idak.

Dengan memanfaatkan data kelulusan mahasiswa yang diperolehdari data akademik perguruan tinggi maka akan diketahui informasi mengenai tingkat kelulusan mahasiswa melalui teknikklasifikasi pada data mining. Teknik klasifikasi data pada data mining terdiri dari berbagai metode. Di sini digunakan metode Naïve Bayes dan Algoritma C4.5 dalam proses data tersebut dengan maksud agar dapat membandingkan kinerja antara metode Naïve Bayes denganAlgoritma C4.5 sehingga dapat diperoleh metode dengan akurasip rediksi ketepatan kelulusan mahasis water baik berdasar model data yang ada. Data yang telah diuji dengan kedua metode diatas kemudianakan dibandingkan dan dievaluasidengan confusion matrix untuk diketahui tingkat akurasinya.

Ada puntujuan dari penulisan dan penelitian ini adalah

 Dapat membandingkan kinerja dari metode Naïve Bayes dan Algoritma C4.5 sehingga dapat diperoleh metode dengan akurasi prediksi ketepatan kelulusan mahasiswa yang terbaik berdasar model data yang ada.

2. Dapat mengevaluasi hasil klasifikasi dari penerapan metode Naïve Bayes dan Algoritma C4.5 sehingga dapat dijadikan bahan pertimbangan dalam layanan akademik untuk mengatasi permasalahan ketepatan waktu kelulusan mahasiswa

2. TINJAUAN PUSTAKA

a. Klasifikasi Data

Klasifikasi data merupakan suatu proses yang menemukan properti-properti yang sama pada sebuah himpunanobyek di dalam sebuah basis data dan mengklasifikasikannya kedalamkelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Tujuan dari klasifikasi adalah untuk menemukan model dari training set yang membedakan atribut kedalam kategor iatau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya [1].

b. Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma Decision Tree. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan digunakan sebagai parameter dalamm elakukan klasifikasi data [3].

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 [3] yaitu:

- 1. Mempersiapkan data training
- 2. Menghitungakar
- 3. Menghitung Entropy dan Gain

$$Entropy(S) = \sum_{i=1}^{n} -pi * log_2Pi$$

Keterangan:

S :Himpunan Kasus n :Jumlah partisi S P_i:Proporsi dari S₁ ke S

$$Gain(S, A) = S - \sum_{i=1}^{n} \frac{|s_i|}{|S|} * S_i$$

- 4. Ulang i langkah ke 2 dan langkah ke 3 hingga semua record terpatisi
- 5. Proses partis i pohon keputusan akan berhenti saat semua tupel dalam node N mendapat kelas yang sama dan atau tidak ada atribut didalam tupel yang dipartisi lagi dan atau tidak ada tupel di dalam cabang yang kosong

c. Metode Naïve Bayes

Naïve Bayes merupakan sebuah pengklasifikasian probabilistic sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi n i lain dari dataset yang diberikan. Algoritma mengunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variable kelas [4]

Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi paremeter yang diperlukan dalam proses pengklasifikasian.

PersamaanMetode Naïve Bayes

Prediksi Bayes didasarkan pada teorema Bayes dengan formula umumsebagaiberikut:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Keterangan:

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

P(H|X): Probabilitas hipotesis H berdasar kondisi X (Posteriori Probability)

P(H) : Probabilitas hipotesis H (Prior Probability)

P(X|H): Probabilitas X berdasarkan kondisi hipotesis H

P(X): Probabilitas X

Formulasi Naïve Bayes untuk klasifikasi adalah

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^{q} P(X_i|Y)}{P(X)}$$

Keterangan:

P(Y|X) : probabilitas data dengan

vector X pada kelas Y.

P(Y) : Probabilitas awal kelas

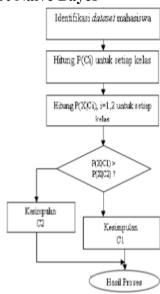
 $\prod_{i=1}^{q} P(X_i|Y)$: Probabilitas independen kelas

Y dari semua fitur dalam

vector X.

P(X): Probabilitas awal kelas X

Flowchart Naive Bayes



Gambar 1.Flowchart Naive Bayes

3. METODE PENELITIAN

Pada penelitian ini data yang digunakan data akademik kelulusan mahasiswa STIKOM Uyelindo Kupang tahun 2015-2017 yang didapat dari pelaporan data Forlap, sedangkan untuk pengujian model menggunakan aplikasi Rapid Miner. Tahapantahapan dalam penelitian ini sebagai berikut :

Pengumpulan Data

Tahapan awal yang dilakukan dalam penelitian ini adalah menyiapkan data, dimana data diperoleh dari bagian akademik. Data ini akan digunakan sebagai proses mining dan pengujian. Data set yang digunakan adalah data kelulusan mahasiswa untuk jenjang S1 yang terdiri dari beberapa atribut seperti prodi, jenis kelamin, IPK semester 1, IPK semester 2, IPK semester 3, IPK semester 4 dan tahun kelulusan. Jumlah data yang digunakan adalah 613 data dengan kelas "Tepat" dan "Tidak Tepat"

Seleksi dan Transformasi Data

Untuk mempermudah dalam proses mining selanjutnya maka data yang ada dilakukan proses transformasi data kedalam bentuk yang dapat diolah dengan alat bantu data mining yang ada. Adapun proses konversi data data yang telah diperoleh dari tahap persiapan data adalah sebagai berikut:

Tabel 1.Variabel data kelulusan			
Variabel	Deskripsi	Nilai	
Prodi	Program Studi	SI = SistemInformasi TI = Teknik Informatika	
JK	JenisKelam in	P = Perempuan L = laki-laki	
IPK 1	IndeksPrest asiKumulati	Tinggi = IPK >=3.50	
	f semester 1	Sedang = IPK >=3.00 Cukup = IPK >= 2.75	
IPK 2	IndeksPrest asiKumulati	Tinggi = IPK >=3.50	
	f semester 2	Sedang = IPK >=3.00 Cukup = IPK >= 2.75	
IPK 3	IndeksPrest asiKumulati	Tinggi = IPK >=3.50	
	f semester 3	Sedang = IPK >=3.00	
		Cukup = IPK >= 2.75	
IPK 4	IndeksPrest asiKumulati	Tinggi = IPK >=3.50	
	f semester 4	Sedang = IPK >=3.00	
		Cukup = IPK >= 2.75	
Tahun Lulus	TahunKelul usan	Tahunkelulusandia mbildaritahun 2015 - 2017	
Keterang an	Keterangan Kelulusan	Tepat Waktu TidakTepat	

Model yang diusulkan e.

Metode yang diusulkan dalam penelitian ini adalah metode klasifikasi data mining algoritma C4.5 dan Naive Bayes. Pengujian model menggunakan Cross Validation, evaluasi dengan Confusion Matrix

f. Evaluasi dan Validasi Hasil

Evaluasi dilakukan dengan menganalisa hasil klasifikasi. Pengukuran data dilakukan dengan confusion matrix untuk mengevaluasi hasil dari algoritma Decision Tree C4.5 dan Naive Bayes. Confusion matrix merupakan sebuah tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi. Tabel ini diperlukan untuk mengukur kinerja suatu model klasifikasi.

Tabel 2. Confusion Matrix

		Predicted Class	
		Class = 1	Class = 0
Actual Class	Class = 1	A	В
	Class = 0	C	D

Keterangan:

- A :Jumlah prediksi yang tepat bahwa instance bersifat negatif
- B :Jumlah prediksi yang salah bahwa instance bersifat positif
- C :Jumlah prediksi yang salah bahwa instance bersifat negatif
- D :Jumlah prediksi yang tepat bahwa instance bersifat positif.

Perhitunga nakurasi dengan tabel confusion matrix adalah sebagai berikut :

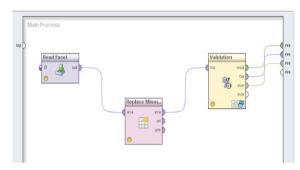
$$Akurasi = \frac{A+D}{A+B+C+D}$$

Precision didefinisikan sebagai rasio item relevan yang dipilih terhadap semua item yang terpilih. Precision merupakan probabilitas bahwa sebuah item yang dipilih adalah Sedangkan Recall didefinisikan relevan. sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. Recall merupakan probabilitas bahwa suatu item yang relevan akan dipilih. Precision dan Recall dapat diberinilai dalam bentuk angka dengan menggunakan perhitungan (1-100%)presentase atau dengan menggunakan bilangan antara 0-1. Evaluasi pengukuran dilakukan dengan Rapid Miner yaitu dengan membandingkan nilai akurasi, nilai precision, dan nilai recall antara algoritma Decision Tree C4.5 dengan Algoritma Naive Bayes

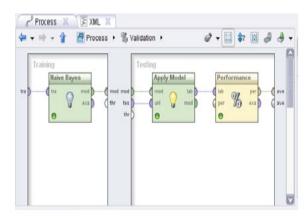
4. HASIL DAN PEMBAHASAN

g. Implementasi Algoritma Naive Bayes dan C4.5

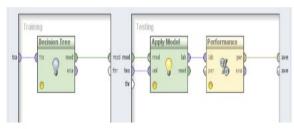
Pengolahan data dengan model Naive Bayes pada Rapidminer sebagai berikut :



Gambar 2. Pengujian 10-Fold Cross



Gambar 3. Proses Training dan Testing Algoritma Naive Bayes



Gambar 4. Proses Training dan Testing Algoritma C4.5

h. Hasil Performance Vector Algoritma Naive Bayes dan C4.5

Berdasarkan hasi limplementasi dari pengolahan data dengan menggunakan algoritma Naive Bayes dan C4.5, maka diperoleh hasil performance vector seperti terlihat pada gambar berikut.

	true Tidak Tepat	true Tepat	class precision
pred. Tidak Tepat	15	19	44.12%
pred. Tepat	27	552	95.34%
class recall	35.71%	96.67%	

Gambar 5. Performance Vector Algoritma Naive Bayes

	accuracy: 94.47% +/- 3.22% (mikro: 94.45%)		
	true Tidak Tepat	true Tepat	dass precision
pred. Tidak Tepat	13	5	72.22%
pred. Tepat	29	566	95.13%
class recall	30.95%	99.12%	

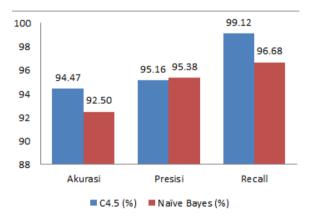
Gambar 6. Performance Vector Algoritma C4.5

i. Perbandingan Hasil Validasi

Dari hasil eksperimen menggunakan data set yang sama dengan menggunakan algoritma Naive Bayes dan C4.5, maka diperoleh hasil perbandingan validasi sebagaimana table berikut:

Tabel 3. Perbandingan Validasi Hasil Prediksi Kelulusan

	C4.5 (%)	Naïve Bayes (%)
Akurasi	94.47	92.50
Presisi	95.16	95.38
Recal1	99.12	96.68



Gambar 7. Grafik Perbandingan Hasil Validasi Prediksi Kelulusan

5. KESIMPULAN

Berdasarkan hasil implementasi yang telah dilakukan maka dapat diambil beberapa kesimpulan, antara lain :

prediksi 1. Dalam melakukan tingkat ketepatan kelulusan mahasiswa, dengan menggunakan pemodelan metode C4.5 didapatkan tingkat akurasi tertinggi sebesar 94.47%, sedangkan pada pemodelan Naïve Bayes metode memperoleh tingkat akurasi rata-rata 92.50%. tertinggi sebesar Dengan demikian disimpulkan bahwa dapat

- Algoritma Decision Tree C4.5 memiliki nilai akurasi yang lebih tinggi dan baik dalam memprediksi tingkat ketepatan kelulusan mahasiswa.
- 2. Dengan adanya penerapan Decision Tree C4.5 diharapkan mampu memberikan dapat membantu pihak akademik dalam memprediksi tingkat ketepatan kelulusan mahasiswa selama studi sehingga peluang untuk sukses dalam studi di perguruaan tinggi semakin besar
- 3. Penentuan data training dapat mempengaruhi hasil pengujian, karena pola data training tersebutakan dijadikan sebagai rule untuk menentukan kelas pada data testing. Sehingga besar atau kecilnya prosenta setingkat precision, recall, dan accuracy dipengaruhi juga oleh penentuan data training.

DAFTAR PUSTAKA

- Ginting, S. L. BR, Zarman, W., Hamidah, I.. Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)
- [2] Bhardwaj B. K. and Pal S. 2011 Data Mining: A prediction for performance improvement using classification (IJCSIS) International Journal of Computer Science and Information Security
- [3] Larose, D. T. 2005. Discovering Knowledge in Databases. New Jersey: John Willey & Sons Inc.
- [4] Patil, T. R., Sherekar, M. S., 2013, Performance Analysis of Naive Bayes and J48Classification Algorithm for Data Classification, International Journal of Computer Scienceand Applications, Vol. 6, No. 2, Hal 256-261
- [5] Bustami., 2013, Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi, TECHSI: Jurnal Penelitian Teknik Informatika, Vol. 3, No.2, Hal. 127-146..