

PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN TEOREMA *TEOREMA BAYES* (STUDI KASUS: UNIVERSITAS PENDIDIKAN GANESHA)

I Gede Bendesa Subawa

Program Studi Pendidikan Teknik Informatika, Fakultas Teknik dan Kejuruan,
Universitas Pendidikan Ganesha
Singaraja, Indonesia

e-mail: bendesa.subawa@undiksha.ac.id

Abstrak

Penelitian ini bertujuan untuk melakukan prediksi terhadap data mahasiswa program sarjana melalui proses *data mining* menggunakan algoritme *teorema bayes*. Data mahasiswa Universitas Pendidikan Ganesha menghasilkan data yang sangat berlimpah berupa data profil mahasiswa dan data akademik. Informasi yang tersembunyi dapat diketahui dengan cara melakukan penggalian dan analisis terhadap data mahasiswa sehingga lebih berguna bagi pihak universitas. Fitur yang digunakan adalah jurusan, asal sekolah, jalur masuk perguruan tinggi, indek prestasi kumulatif, sks kumulatif, dan tahun lulus mahasiswa. Akurasi algoritme *teorema bayes* untuk memprediksi kelulusan mahasiswa sebesar 80.96% dengan data latih sebanyak 3892 data. Hasil prediksi menunjukkan bahwa terdapat 1205 mahasiswa lulus tepat waktu dan 634 mahasiswa lulus tidak tepat waktu per Agustus 2019. Hasil penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan kebijakan oleh pihak jurusan.

Kata kunci: *data mining*, kelulusan, *teorema bayes*, prediksi

Abstract

This study aimed to prediction student document of undergraduate program by utilizing data mining process using teorema bayes algorithm. The data of students on Ganesha University of Education generated abundant documents in the form of student profile and academic data. The hidden information can be discovered by using data mining and analyzing student's data so it can be more useful for the university. The elements used were department, the origin of high school, the selection of college entrance, achievement index, cumulative of university credit unit, and graduate year. The accuration of teorema bayes algorithm to predict length of study is 80.96% with 3892 training data. The result of prediction show that 1205 students will graduate on time and 634 students will not graduate on time per August 2019. The results of this study is used as one of the basic decision-making to determine the policy by the department.

Keywords : *data mining, graduation, teorema bayes, prediction*

PENDAHULUAN

Kelulusan mahasiswa merupakan salah satu aspek penting dalam evaluasi keberhasilan penyelenggaraan program studi pada perguruan tinggi^[1]. Menganalisis lama studi, kinerja mahasiswa (*students performance*), mengidentifikasi keunikan-keunikan yang ada pada mahasiswa dan membangun suatu strategi pengembangan lebih lanjut, serta tindakan-tindakan yang dapat dilakukan untuk masa mendatang, merupakan tantangan utama bagi universitas modern saat ini. Kelulusan merupakan fitur penting bagi pengelola akademik, melakukan klasifikasi kelulusan mahasiswa dapat membantu pihak universitas dapat meminimalisir kegagalan kelulusan mahasiswa dengan membuat perencanaan, pengawalan studi dan bimbingan lebih intensif.

Berdasarkan data mahasiswa dan data kelulusan mahasiswa, informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data mahasiswa sehingga berguna bagi pihak universitas^[2]. Selama ini UNDIKSHA belum memiliki pola klasifikasi kelulusan tepat waktu sebagai acuan untuk memklasifikasikan jumlah lulus tepat waktu. Beberapa penelitian telah banyak dilakukan dengan menggunakan teknik *data mining* untuk menggali berbagai informasi dari sebuah *database* mahasiswa seperti (1). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Di Universitas Klabat Dengan Metode Klasifikasi^[3], (2) Sistem Evaluasi Dan Klasifikasi Kinerja Akademik Mahasiswa Universitas Madura Menggunakan Teorema bayes Dengan Dirichlet Smoothing^[4]. (3) Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritme Teorema bayes^[5], (4) *Prediction Of Students Academic Performance Using Classification and Clustering Algorithms*^[6], dan (5) *Analyzing the Student's Academic Performance by using Clustering Methods in Data Mining*^[7].

Berdasarkan uraian di atas, pada penelitian ini akan dilakukan prediksi kelulusan mahasiswa di UNDIKSHA dengan cara mengevaluasi kinerja pada tahun kedua. Pada penelitian ini, digunakan

teknik data mining untuk menemukan pola kelulusan mahasiswa yang sudah lulus berdasarkan informasi mahasiswa (sekolah asal, jalur masuk perguruan tinggi, jurusan/fakultas), indeks prestasi kumulatif (ipk) tahun kedua, sks yang sudah ditempuh, kemudian dijadikan dasar untuk mengklasifikasikan kelulusan mahasiswa pada tahun ke-2. *Data mining* adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar pengambilan keputusan. Metode yang digunakan adalah *teorema bayes* dikarenakan metode ini mampu mengolah data kualitatif maupun kuantitatif dengan akurasi yang tinggi.

Data Mining

Kecerdasan buatan, dan mesin pembelajaran untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar^[7]. Istilah *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. *Data mining*, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar^[9]. Teknik-teknik data mining telah digunakan untuk menemukan pola yang tersembunyi dan memprediksi tren masa depan. Keuntungan kompetitif dari data mining termasuk dengan meningkatnya pendapatan, berkurangnya pengeluaran, dan kemampuan pemasaran meningkat.

Secara garis besar metode pelatihan yang digunakan dalam teknik-teknik *data mining* dibedakan ke dalam dua pendekatan, yaitu:

1. *Unsupervised learning*, metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*). Guru di sini adalah label dari data. Algoritme *data mining* mencari pola dari semua variabel (atribut). Variabel (atribut) yang menjadi target/label/class

tidak ditentukan (tidak ada). Algoritme klasterisasi adalah algoritme *unsupervised learning*.

2. *Supervised learning*, yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai *output* atau label selama proses *training*. Sebagian besar algoritme *data mining* (*estimation, prediction/forecasting, klasifikasi*) adalah *supervised learning*. Variabel yang menjadi target/label/class sudah ditentukan dan algoritme melakukan proses belajar berdasarkan nilai dari variabel target yang terasosiasi dengan nilai dari *variable predictor*.

Klasifikasi

Klasifikasi merupakan sesuatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan yaitu: (1) pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan / klasifikasi/prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang mudah disimpan.

Contoh aplikasi yang sering ditemui adalah pengklasifikasian jenis hewan, yang mempunyai sejumlah atribut tertentu. Dengan atribut tersebut, jika ada hewan baru, ielas hewannya bisa langsung diketahui. Contoh lainnya adalah bagaimana melakukan klasifikasi kecenderungan penyakit mata di Jawa Barat^[10]. Peneliti melakukan pembangunan model tertentu berdasarkan data latih yang ada, kemudian menggunakan metode *teorema bayes* untuk mengidentifikasi penyakit pasien baru sehingga diketahui pasien kecenderungan menderita penyakit mata di kelas yang mana.

Teorema Bayes

NBC merupakan salah satu algoritme dalam teknik data mining yang menerapkan teori *Bayes* dalam klasifikasi^[9].

Matriks klasifikasi memberikan informasi perbandingan klasifikasi aktual dengan klasifikasi yang dihasilkan melalui sistem pengklasifikasian.

Evaluasi yang dilakukan menggunakan parameter *F-Measure* yang terdiri dari perhitungan *precision* dan *recall*. *Recall, precision* merupakan metode pengukuran efektifitas yang biasa dilakukan pada proses klasifikasi. Dalam penelitian ini, nilai *precision* dan *recall* didasarkan pada hasil *confusion matrix*, seperti pada Tabel 1.

Recall dan *Precision* adalah dua kriteria yang digunakan untuk mengevaluasi tingkat efektifitas kinerja sistem temu kembali informasi.

Tabel 1 Matriks Klasifikasi

Matrik	Prediksi		
	Positif	Negatif	
Aktual	Positif	TP	FN
	Negatif	FP	TN

Keterangan:

TP : Benar Positif

TN : Benar Negatif

FP : Salah Positif

FN : Salah Negatif

a. Precision

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Rumus dari *precision* seperti pada persamaan 3:

$$precision(p) = \frac{TP}{TP+FP} \quad (3)$$

b. Recall

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Rumus dari *Recall* seperti pada persamaan 4

$$Recall(r) = \frac{TP}{TP+FN} \quad (4)$$

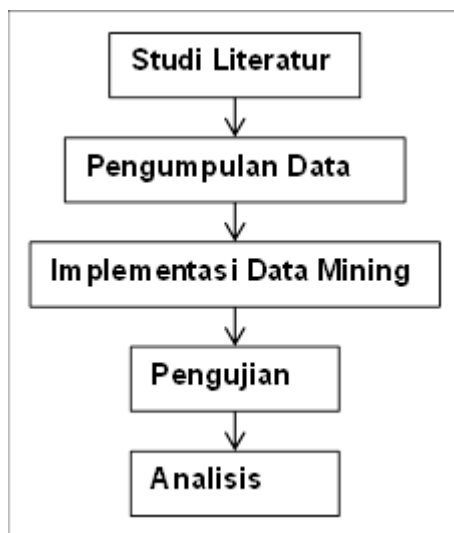
METODE

Pada penelitian ini, yang menjadi obyek penelitian adalah Universitas

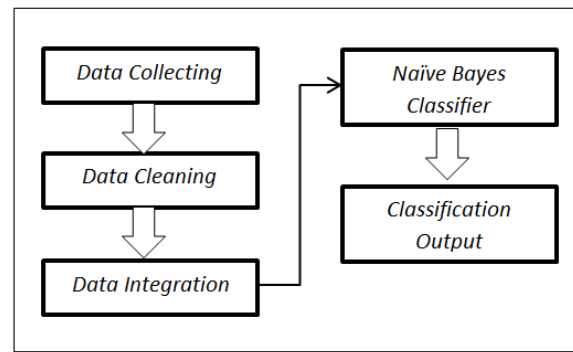
Pendidikan Ganesha yang berlokasi di Singaraja. Data yang akan digunakan berasal dari Unit Pelaksanaan Teknis TI Undiksha. Langkah atau tahapan penelitian ini dilakukan sebagaimana Gambar 1.

Studi Literatur

Melakukan studi literatur terhadap berbagai referensi yang berkaitan dengan penelitian yang dilakukan. Dalam tahap ini dipelajari literatur tentang konsep klasifikasi prediksi kelulusan mahasiswa dengan menggunakan metode *teorema bayes* dalam berbagai sumber berupa buku, artikel dan jurnal. Pada studi literatur diketahui bahwa klasifikasi masa studi mahasiswa lebih cocok menggunakan algoritme *teorema bayes*, hal ini dikarenakan salah satu karakteristik dari *teorema bayes* yaitu memiliki fitur/atribut yang *independent* (tidak terikat dengan atribut atau fitur lainnya). Dalam penelitian ini fitur/atribut seperti sekolah asal, ipk, jalur masuk, sks, dan jurusan mahasiswa (X) bersifat *independent* yang hanya mempengaruhi variabel *dependent* yaitu masa studi (Y). Jadi tidak ada fitur X yang mempengaruhi fitur X yang lainnya.



Gambar 1. Tahapan-Tahapan Penelitian



Gambar 2 Tahapan Data Mining

Pengumpulan Data / Data Collecting

Berdasarkan studi literatur data yang diperlukan dalam penelitian ini adalah data yang meliputi data kelulusan mahasiswa. Adapun data kelulusan yang digunakan adalah data dari mahasiswa tahun ajar 2012/2013 dan 2013/2014 dikarenakan penggunaan algoritme *teorema bayes* bersifat *supervised* yang berarti memerlukan data latih terlebih dahulu. Sehingga dengan data mahasiswa tahun ajar 2012/2013 dan 2013/2014 diharapkan sudah memiliki data kelulusan (lulus tepat waktu ≤ 8 semester atau tidak tepat waktu). Data mahasiswa tersebut nantinya akan diperoleh dari bagian Unit Pelaksana Teknis (UPT) TI Undiksha.

Adapun fitur yang digunakan dalam penelitian ini adalah jurusan, asal sekolah, jalur masuk perguruan tinggi, indeks prestasi kumulatif (IPK), SKS Kumulatif, dan tahun kelulusan.

Implementasi

Setelah data tersebut terkumpul barulah masuk ke tahap data mining:

- a. Data Cleaning / Pembersihan Data
Data yang didapat dari UPT TIK Undiksha merupakan data kasar. Pada data tersebut masih banyak data yang hilang ataupun kosong. Pada tahap ini dilakukan pembersihan data sehingga data yang digunakan adalah data yang utuh.
- b. Data Integration / Integrasi data
Setelah diperoleh data yang utuh selanjutnya adalah menghubungkan seluruh data menjadi satu kesatuan data pada sebuah basis data. Selanjutnya adalah mengkonversi data tersebut menjadi

data dengan format yang bisa diterima oleh aplikasi data mining yang digunakan.

c. Implementasi Teorema Bayes

Data yang sudah dikonversi diimplementasikan ke dalam alat bantu *data mining* WEKA 3.9. Algoritme yang akan diimplementasikan adalah *Teorema bayes*.

Pengujian

Proses pengujian yang akan dilakukan adalah pengujian akurasi. Yaitu seberapa akurasi pengimplementasian algoritme *teorema bayes* dalam melakukan klasifikasi masa studi mahasiswa di Undiksha. *Testing data* adalah data mahasiswa angkatan tahun ajar 2012/2013 dan 2013/2014 juga yang diasumsikan belum diketahui kelasnya. Selanjutnya berdasarkan data latih ini akan dilakukan prediksi kelulusan mahasiswa angkatan tahun ajar 2015/2016.

Analisis

Setelah semua hasil pengujian didapatkan, tahap selanjutnya adalah melakukan analisis dari hasil-hasil pengujian tersebut. Analisis yang dimaksud antara lain, menganalisis keakurasian penerapan algoritme *teorema bayes* dalam pengklasifikasian. Kemudian menganalisis pengaruh masing-masing atribut independent terhadap atribut target.

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan mengumpulkan data dari *database* Sistem Informasi Akademik Universitas Pendidikan Ganesha. Data yang diperoleh adalah 4488 data akademis mahasiswa dari angkatan tahun ajar 2012/2013 dan 2013/2014. Adapun hasil dari penelitian yang telah dilakukan peneliti adalah sebagai berikut.

Pembelajaran (*Learning*) Data Latih / *Training Data*

Algoritme *teorema bayes* merupakan metode klasifikasi dan prediksi yang bersifat *supervised learning* yang artinya dalam pengklasifikasian kelas *teorema bayes* memerlukan pelatihan atau pembelajaran terlebih dahulu. Sehingga

diperlukan data latih / *training data* yang nantinya digunakan sebagai sumber pengetahuan atau dasar dalam dalam klasifikasi atau prediksi. Dari data yang diperoleh dilakukan *cleaning* data dan pemilahan data mahasiswa program sarjana (S1) saja sehingga diperoleh 3892 data. Data tersebut akan digunakan untuk menentukan fitur kelulusan mahasiswa.

Analisis Data Penuh / *Full Testing*

Analisis data penuh merupakan penggunaan 100% data latih/*training* sebagai data uji/*testing*. Menggunakan algoritme *teorema bayes* dalam menentukan klasifikasi masing-masing data dari keseluruhan data. Hasil yang didapatkan dengan menggunakan data seperti Gambar 2.

<i>Correctly Classified Instances</i>	80.96%
<i>Incorrectly Classified Instances</i>	19.04%
<i>Total Number of Instances</i>	3892
<i>Avg. Precision</i>	80.90 %
<i>Avg. Recall</i>	81.00 %
<i>F-Measure</i>	80.80 %
<i>ROC Area</i>	86.00 %

Gambar 3. Hasil Analisis Data Penuh

Dari hasil perhitungan data menggunakan Tool Weka 3.4.12 diatas dapat dibaca informasi sebagai berikut :

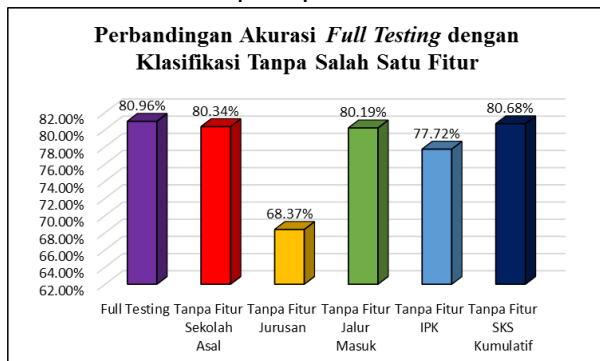
- Correctly Classified Instances*, dimana dari masing-masing baris data mendapatkan hasil yang cocok atau sesuai dengan hasil klasifikasi yaitu sejumlah 3151 dari 3892 data atau tingkat kebenaran klasifikasi adalah 80.96 %
- Incorrectly Classified Instances*, kebalikan dari penjelasan point (a) terdapat kesalahan klasifikasi dari masing-masing baris data yaitu sejumlah 741 dari 3892 data atau 21.18 %
- Total Number of Instances*, merupakan jumlah total data yaitu 3892 data
- Avg. Precision*, rata-rata tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem sebesar 80.90%

- e. *Avg. Recall*, rata-rata tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi sebesar 81.00%.
- f. *F-Measure*, menunjukkan tingkat keseimbangan/balance antara precision dan recall yaitu 80.80%. Semakin tinggi nilainya maka semakin bagus performance klasifikasi tersebut.
- g. *ROC Area*, merupakan salah satu cara mengukur kinerja / performance suatu klasifikasi. Nilai *ROC Area* adalah 0.860. Itu berarti klasifikasi yang dilakukan memiliki kinerja / performance yang bagus (Mohanty et al., 2011).

Analisis Pengaruh Fitur Terhadap Akurasi Klasifikasi

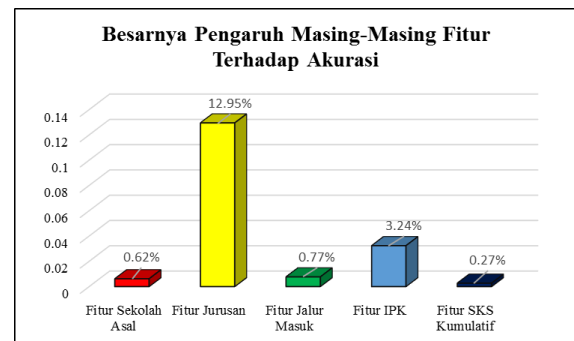
Pada tahap ini dilakukan pengujian pengaruh masing-masing fitur dalam menentukan prosentase kebenaran / akurasi klasifikasi pada data uji. Pengujian dilakukan pada testing data penuh (*full testing*) dengan keakuratan sebesar 80.96%. Pada masing-masing tahap dilakukan penghapusan salah satu fitur kemudian kemudian dilakukan proses *mining* dengan algoritme *teorema bayes*. Tingkat akurasi dari proses *mining* tersebut dibandingkan dengan akurasi pada data penuh (*full testing*).

Adapun perbandingan hasil dari pengujian akurasi *full testing* dengan akurasi yang dihasilkan dari penghapusan salah satu fitur seperti pada Gambar 3.



Gambar 4. Perbandingan Akurasi *Full Testing* dengan Klasifikasi Tanpa Salah Satu Fitur

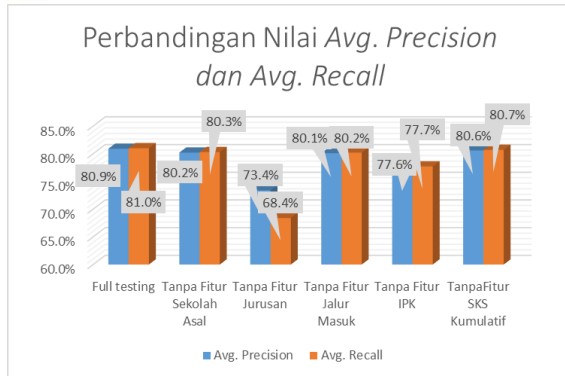
Berdasarkan grafik pada Gambar 3 diperoleh informasi bahwa seluruh fitur yang digunakan dalam klasifikasi berpengaruh terhadap besarnya akurasi. Tanpa salah satu fitur maka akan menyebabkan penurunan akurasi dalam pengklasifiasian. Adapun besarnya perubahan akurasi atau besarnya pengaruh masing-masing fitur terhadap akurasi dalam prediksi kelulusan adalah seperti yang disajikan pada Gambar 4.



Gambar 5 Besarnya Pengaruh Masing-Masing Fitur Terhadap Akurasi

Berdasarkan Gambar 4 diketahui bahwa fitur yang paling berpengaruh terhadap prediksi kelulusan mahasiswa adalah fitur jurusan sebesar 12.95% dan IPK sebesar 3.24%. Jadi tanpa adanya fitur jurusan maka akurasi akan menurun hingga 12.95%.

Selain nilai akurasi, nilai *precision* dan *recall* juga perlu diperhatikan untuk melihat *performance* / kinerja dalam klasifikasi sudah tergolong baik atau belum. Semakin tinggi dan *balance* / seimbang nilai *precision*/presisi dengan *recall* / temu kembali maka kinerja suatu sistem dalam melakukan klasifikasi juga semakin bagus.



Gambar 6: Perbandingan Nilai Avg. Precision dan Avg. Recall Full Testing dengan Klasifikasi Tanpa Salah Satu Fitur

Berdasarkan grafik pada Gambar 5 diperoleh informasi bahwa klasifikasi data yang memiliki *performa* terbaik adalah klasifikasi yang diperoleh dari *full testing* (menggunakan seluruh fitur) karena memiliki nilai *precision*/presisi dan *recall*/temu kembali yang paling tinggi serta keseimbangan/*balances value*. Sedangkan klasifikasi tanpa menggunakan fitur jurusan merupakan klasifikasi yang kurang bagus karena selain memiliki tingkat akurasi yang rendah, juga memiliki nilai *precision* dan *recall* yang rendah dan tidak seimbang.

Prediksi Kelulusan pada Data Uji Mahasiswa Angkatan 2015/2016

Pada bagian ini akan dilakukan prediksi kelulusan mahasiswa angkatan tahun ajaran 2015/2016. klasifikasi yang dilakukan dengan menggunakan algoritme *teorema bayes* berdasarkan data latih data mahasiswa angkatan 2012/2013 dan 2013/2014. Data latih ini digunakan oleh algoritme *teorema bayes* sebagai acuan pelatihan/pembelajaran dalam pembentukan pola klasifikasi yang akan dijadikan dasar untuk klasifikasi data uji. Klasifikasi dilakukan dengan melibatkan seluruh fitur dikarenakan berdasarkan analisis yang telah dilakukan bahwa dengan menggunakan seluruh fitur (*full testing*) akan memberikan akurasi yang paling tinggi, serta kinerja /*performance* terbaik dalam klasifikasi.

Data uji / data mahasiswa yang akan diklasifikasikan lama studinya adalah mahasiswa angkatan tahun ajaran 2015/2016. Jumlah data uji yang digunakan

setelah melalui proses pembersihan data berupa penghapusan data yang tidak bersesuaian, perbaikan kesalahan ejaan dan penghapusan data kosong adalah sebanyak 1839 *record* data mahasiswa.

Hasil Prediksi Kelulusan Mahasiswa Angkatan 2015/2016

Mahasiswa yang di kategorikan lulus tepat waktu adalah mahasiswa yang diklasifikasikan bisa menempuh lama studi paling lambat Agustus 2019 (8 semester). Mahasiswa yang akan diuji untuk prediksi kelulusan adalah mahasiswa angkatan tahun ajar 2015/2016 pada semester 4 (empat). Hasil klasifikasi menunjukkan bahwa dari 1839 mahasiswa, 66 % atau sebanyak 1205 mahasiswa diklasifikasikan bisa menyelesaikan masa studinya paling lambat Agustus 2019. Sedangkan sisanya sebanyak 34% atau 634 mahasiswa diklasifikasikan tidak bisa menyelesaikan studinya hingga Agustus 2019. Adapun sebaran data hasil klasifikasi sebagai berikut.

a. Jurusan

Berdasarkan hasil klasifikasi dari 634 mahasiswa yang diklasifikasikan lulus tidak tepat waktu, 30 % nya merupakan mahasiswa yang berasal dari Jurusan Pendidikan Bahasa Inggris. Untuk mahasiswa yang diklasifikasikan lulus tepat waktu, dari 1205 mahasiswa yang diklasifikasikan lulus tepat waktu sebanyak 35% nya merupakan mahasiswa dari Jurusan Pendidikan Guru Sekolah Dasar.

b. Asal Sekolah

Berdasarkan hasil klasifikasi dari 634 mahasiswa yang diklasifikasikan lulus tidak tepat waktu, sebanyak 71% nya merupakan mahasiswa yang berasal dari SMA Negeri. Sedangkan dari 1205 mahasiswa yang diklasifikasikan lulus tepat waktu sebanyak 73% nya merupakan mahasiswa yang berasal dari SMA Negeri.

c. Jalur Masuk Perguruan Tinggi

Berdasarkan hasil klasifikasi dari 634 mahasiswa yang diklasifikasikan lulus tidak tepat waktu, sebanyak 273 mahasiswa atau 51% berasal dari SNMPTN. Sedangkan dari 1205 mahasiswa yang diklasifikasikan lulus tepat waktu sebanyak 55% nya

merupakan mahasiswa yang lolos jalur masuk SNMPTN.

d. **Indek Prestasi Kumulatif (IPK)**

Berdasarkan hasil klasifikasi dari 634 mahasiswa yang diklasifikasikan lulus tidak tepat waktu, 59% nya merupakan mahasiswa yang dengan IPK kategori sangat memuaskan dan 19% merupakan mahasiswa dengan IPK kategori memuaskan dan sisanya kategori kurang. Sedangkan dari 1205 mahasiswa yang diklasifikasikan lulus tepat waktu sebanyak 71% nya merupakan mahasiswa dengan IPK kategori sangat memuaskan dan sisanya 29% merupakan mahasiswa yang memperoleh IPK kategori *cumlaude*.

e. **SKS Kumulatif**

Berdasarkan hasil klasifikasi dari 634 mahasiswa yang diklasifikasikan lulus tidak tepat waktu, sekitar 73% nya merupakan mahasiswa yang sudah mengambil beban studi (SKS) lebih dari 75 sks atau kategori D, namun berdasarkan seluruh mahasiswa yang mengambil beban studi kurang dari 75 sks atau selain kategori D diklasifikasikan lulus tidak tepat waktu. Sedangkan dari 1205 mahasiswa yang diklasifikasikan lulus tepat waktu sebanyak hampir 100% nya merupakan mahasiswa yang sudah mengambil beban studi (SKS) lebih dari 75 sks atau kategori D.

Pembahasan

Setelah dilakukan analisa data diketahui bahwa dari 634 mahasiswa, 30% berasal dari Jurusan Pendidikan Bahasa Inggris, kemudian 14% nya berasal Jurusan Pendidikan Teknik Informatika. Peneliti telah melakukan *interview* / wawancara singkat terhadap beberapa ketua jurusan terkait kendala yang kiranya menyebabkan banyaknya mahasiswanya berpotensi lulus tidak tepat waktu disamping jumlah mahasiswa yang banyak. Pertama peneliti melakukan wawancara kepada Ketua Jurusan Pendidikan Bahasa Inggris didapatkan informasi bahwa di Jurusan Pendidikan Bahasa Inggris penyebab utama mahasiswa banyak yang lulus tidak tepat waktu adalah dikarenakan kurikulum yang diterapkan mengharuskan pelaksanaan PPL-Real di semester 8.

Pelaksanaan PPL-Real yang biasanya dilaksanakn di awal semester 7 namun dilaksanakan di semester 8 akan mempengaruhi mahasiswa dalam mengerjakan tugas akhir/skripsi sehingga adanya keterlambatan dalam menyelesaikan studi.

Kedua peneliti melakukan wawancara terhadap Ketua Jurusan Pendidikan Teknik Informatika (PTI) dimana salah satu penyebabnya adalah jenis skripsi yang harus dikerjakan mahasiswa PTI berbeda dengan jurusan yang lain. Di Jurusan PTI sebagian besar mahasiswa mengambil jenis skripsi pengembangan perangkat lunak sehingga dituntut agar menghasilkan suatu produk / perangkat lunak. Hal ini menyebabkan mahasiswa PTI memerlukan waktu lebih lama dalam pengerjaannya. Kedua, di Jurusan PTI terdapat mata kuliah peminatan dimana bertujuan agar mahasiswa bisa menggali potensinya. Tapi kenyataannya banyak mahasiswa PTI yang mengambil topik skripsi diluar peminatan yang dipilih sehingga mahasiswa tersebut harus belajar lagi terkait topik skripsi yang diambil. Ketiga, banyak mahasiswa PTI yang mengajukan sebuah topik skripsi pada semester 7 dan 8, padahal mereka harus mengikuti program PPL Real dan KKN seperti jurusan yang lain, dan mengingat topic yang dikerjakan berbasis produk.

Ketiga peneliti melakukan wawancara kepada Ketua Jurusan Pendidikan Matematika. Menurut beliau tidak ada hal khusus yang menghambat studi mahasiswa selain beban materi perkuliahan itu sendiri. Ketua jurusan sendiri mengakui bahwa materi yang diajarkan di Pendidikan Matematika cukup sulit sehingga banyak dari mahasiswa harus mengulang kembali mengikuti perkuliahan sebelumnya. Hal ini sangat berpengaruh bagi kelulusan mahasiswa karena sks yang harusnya direncanakan untuk mengambil mata kuliah semester tersebut digunakan untuk melakukan perbaikan.

Adapun kendala yang dihadapi dalam penelitian ini adalah keutuhan data yang diperoleh. Data merupakan hal yang sangat penting dalam melakukan penelitian berkaitan dengan *data mining*. Keutuhan

suatu data merupakan prioritas yang wajib dipenuhi. Dalam penelitian ini cukup banyak data yang diperoleh dalam keadaan tidak utuh / kosong. Seperti misalnya data sekolah dan data jalur masuk yang tidak terisi sehingga data tersebut tidak bisa digunakan sebagai data latih (tidak lolos proses *cleaning data*). Kendala yang kedua adalah perkembangan data. Data semakin hari semakin berkembang, seperti contohnya data jurusan. Tahun 2012 Undiksha hanya terdiri dari 26 jurusan program sarjana, namun setiap tahunnya terdapat penambahan jurusan seperti Jurusan IPA, Jurusan Hukum, dan lain-lain.

SIMPULAN

Berdasarkan hasil analisis yang dilakukan dalam penelitian, dapat disimpulkan hal-hal berikut.

1. Akurasi tertinggi yang dihasilkan *teorema bayes* dalam prediksi lama studi mahasiswa adalah 80.96% yang melibatkan seluruh fitur klasifikasi (*full testing*). Nilai rata-rata *precision* atau presisinya adalah 80.2 % dan nilai rata-rata *recall* / temu kembalinya adalah 80.3%.
2. Hasil prediksi kelulusan terhadap 1839 mahasiswa angkatan 2015/2016 adalah sebanyak 1205 mahasiswa yang diklasifikasikan lulus tepat waktu (per Agustus 2019) dan sisanya 634 mahasiswa diprediksi lulus tidak tepat waktu. Dari 634 mahasiswa yang lulus tidak tepat waktu, sebanyak 30% merupakan mahasiswa jurusan Bahasa Inggris.
3. Untuk penelitian selanjutnya prediksi kelulusan bisa dilakukan dengan lebih banyak data latih karena tiap tahun data mahasiswa semakin banyak dan kriteria yang digunakan bisa lebih bervariasi. Metode data mining yang baru juga bisa diterapkan untuk memperoleh hasil prediksi yang lebih akurat dan optimal

REFERENSI

- [1] Nasution, N et.al. 2016. "Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritme Teorema bayes (Studi Kasus: Fasilkom Unilak)".

Jurnal Teknologi Informasi & Komunikasi Digital Zone, Volume 6 (halaman 1-11).

- [2] Johan, Oscar Ong. 2013. "Implementasi Algoritme K-Means Clustering Untuk Menentukan Strategi Marketing President University," Jurnal Ilmiah Teknik Industri, Volume 12, Nomor 1 (Halaman 10-13).
- [3] Mandias, G. F. 2015. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Di Universitas Klabat Dengan Metode Klasifikasi. Konferensi Nasional Sistem & Informatika (pp. 351-354). Denpasar: STMIK STIKOM Bali.
- [4] Prasetyowati, E., & Ramadhani, N. 2018. Sistem Evaluasi Dan Klasifikasi Kinerja Akademik Mahasiswa Universitas Madura Menggunakan Teorema bayes Dengan Dirichlet Smoothing. JUTI: Jurnal Ilmiah Teknologi Informasi, 192-202.
- [5] Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2017). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritme Teorema bayes . Citec Journal, 151-162.
- [6] Venkatesan, E., & Selvaragini, S. 2017. *Prediction Of Students Academic Performance Using Classification And Clustering Algorithms . International Journal of Pure and Applied Mathematics*, 327-332.
- [7] Kadiyala, S., & Potluri, C. S. 2014. Analyzing the Student's Academic Performance by using Clustering Methods in Data Mining. International Journal of Scientific & Engineering Research, 198-202.
- [8] Turban, E. J et.al. 2005. Sistem Pendukung Keputusan dan Sistem Cerdas. Yogyakarta: ANDI.
- [9] Santosa, B. 2007. Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.
- [10] Budiman, Z.T et.al. 2017. "Klasifikasi Kecenderungan Penyakit Mata di Jawa Barat dengan Association Rule Dan Teorema bayes ". Seminar Nasional Teknologi Informasi dan Multimedia.



[11] Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.