

ANALISIS PENENTUAN METODE DATA MINING UNTUK PREDIKSI KELULUSAN MAHASISWA SEBAGAI PENUNJANG ANGKA EFISIENSI EDUKASI

Valentinus Roby Hananto¹⁾

1) Program Studi Sistem Informasi, Institut Bisnis dan Informatika Stikom Surabaya, email: valentinus@stikom.edu

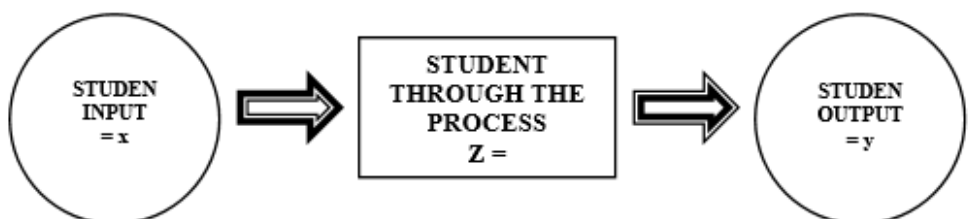
Abstrak

Salah satu tantangan yang dihadapi perguruan tinggi adalah untuk memperbaiki kualitas program pendidikannya. Karena itu, penentuan strategi dan perencanaan untuk perbaikan program pendidikan merupakan suatu hal yang krusial. Angka Efisiensi Edukasi (AEE) merupakan salah satu parameter kualitas program pendidikan. Lamanya masa studi mahasiswa menjadi faktor penentu dari AEE. Apabila mahasiswa memiliki masa studi yang melebihi masa studi normal, salah satunya karena durasi pengerjaan Tugas Akhir yang terlalu panjang, maka nilai AEE dapat menurun. Nilai AEE yang rendah dapat berpengaruh dalam penilaian akreditasi perguruan tinggi ataupun program studi. Oleh karena itu, diperlukan suatu tindakan untuk mengantisipasi masalah ini. Tujuan dari penelitian ini adalah bagaimana melakukan prediksi kelulusan mahasiswa sebagai penunjang pengambilan keputusan dalam upaya meningkatkan AEE menggunakan metode *data mining*. Data yang akan digunakan adalah data historis akademik dari mahasiswa tingkat akhir. Proses pengembangan data mining pada penelitian ini menggunakan metodologi CRISP-DM. Beberapa metode algoritma data mining yang akan dibandingkan dalam penelitian ini adalah *Naïve Bayes*, *Multi Layer Perceptron*, *SMO*, *J48*, dan *REPTree*. Dari penelitian ini, dapat disimpulkan metode data mining dengan akurasi yang terbaik untuk memprediksi kelulusan mahasiswa.

Kata Kunci: *Data Mining, CRISP-DM, Perguruan Tinggi, Angka Efisiensi Edukasi*

PENDAHULUAN

Pada zaman sekarang, perguruan tinggi dituntut untuk selalu memperbaiki kualitas program pendidikannya. Salah satu parameter kuantitatif keberhasilan suatu program pendidikan adalah Angka Efisiensi Edukasi (AEE). Angka ini dibuat dalam bentuk persentasi yang merupakan perbandingan dari jumlah lulusan dibandingkan dengan total jumlah mahasiswa seluruhnya dalam kurun waktu satu tahun (Lubis, 2004). Apabila masa studi normal adalah 8 semester, jumlah penerimaan (input) = X orang, dan output adalah Y orang, sedangkan Z adalah jumlah mahasiswa atau *student body*, maka AEE adalah $(Y / Z \times 100\%)$. Skema ini dapat dilihat pada gambar 1.



Gambar 1. Skema AEE

AEE yang baik untuk program studi dengan masa studi 8 semester adalah 25%, tetapi ini jarang terpenuhi dan umumnya berkisar antara 18% sampai 22%. AEE dapat menunjukkan tingkat produktifitas dan efisiensi dari suatu perguruan tinggi. Semakin bagus nilai AEE, maka semakin baik kinerja perguruan tinggi tersebut. Oleh sebab itu, AEE juga digunakan sebagai salah satu kriteria penilaian akreditasi bagi suatu perguruan tinggi atau program studi (BAN-PT, 2007). Peningkatan nilai AEE ini dapat didukung dengan:

1. Input atau bahan baku yang baik
2. Kualifikasi staf pengajar baik
3. Proses belajar-mengajar telah tertata dengan baik
4. Pengukuran keberhasilan proses pendidikan telah diterapkan
5. Administrasi pendidikan yang berjalan dengan baik
6. Sarana penunjang pendidikan tersedia dan mudah dimanfaatkan

Institut Bisnis dan Informatika Stikom Surabaya merupakan lembaga pendidikan perguruan tinggi swasta yang berada dibawah naungan Yayasan Putra Bhakti, didirikan pertama kali oleh Laksda. TNI (Purn) Mardiono, Ir. Andrian A. T, Ir. Handoko Anindyo, Dra. Suzana Surojo, dan Dra. Rosy Merianti, Ak. Institut Bisnis dan Informatika Stikom Surabaya (selanjutnya disebut Stikom Surabaya) mempunyai 2 fakultas yang telah terdaftar, yaitu Fakultas Ekonomi dan Bisnis (terdiri dari Program Studi S1 Akuntansi, Program Studi S1 Manajemen, Program Studi DIII Komputersasi Perkantoran dan Kesekretariatan) dan Fakultas Teknologi dan Informatika (terdiri dari Program Studi S1 Sistem Informasi, Program Studi S1 Sistem Komputer, Program Studi S1 Desain Komunikasi Visual, Program Studi S1 Desain Grafis, Program Studi DIV Multimedia Komputer, Program Studi DIII Manajemen Informatika).

Program studi S1 Sistem Informasi di Stikom Surabaya merupakan salah satu program studi unggulan dengan jumlah mahasiswa sebanyak 1924 mahasiswa (semester 16.1). Program studi ini terdaftar dengan Akreditasi B di BAN-PT dengan SK No 0102/SK/BAN-PT/Akred/S/III/2016. Sebagai upaya untuk meningkatkan mutu layanan pendidikan, program studi S1 Sistem Informasi Stikom Surabaya juga berusaha untuk meningkatkan nilai akreditasinya. Salah satunya dengan meningkatkan nilai AEE yang saat ini masih kurang. Lamanya masa studi mahasiswa S1 Sistem Informasi merupakan salah satu faktor penyebabnya. Hal ini terjadi karena rata-rata masa pengerjaan Tugas Akhir adalah lebih dari 1 semester, yang idealnya adalah 1 semester agar mahasiswa bisa lulus tepat waktu. Berdasarkan data mahasiswa angkatan 2010, jumlah mahasiswa yang lulus tepat waktu adalah 2,84%, sehingga masih belum mencapai standar ideal AEE antara 18% - 22%. Oleh karena itu, upaya untuk mempercepat masa pengerjaan tugas akhir menjadi fokus utama yang dilakukan program studi S1 Sistem Informasi.

Program percepatan Tugas Akhir ini dilakukan prodi Sistem Informasi Stikom Surabaya sejak tahun 2015. Pada tahun 2016 sudah dilaksanakan beberapa aktivitas untuk mendukung kelulusan tepat waktu mahasiswa Sistem Informasi angkatan 2012: memberikan pengarahan umum untuk mahasiswa angkatan 2012 yang sudah layak menempuh Tugas Akhir, menstrukturkan pemilihan dosen pembimbing Tugas Akhir, serta menjadwalkan siding proposal dan ujian Tugas Akhir secara simultan untuk mencapai kelulusan tepat waktu.

Salah satu cara untuk meningkatkan kualitas pengambilan keputusan dalam dunia pendidikan adalah dengan mengekstrak pengetahuan berdasarkan data historis yang sudah dimiliki. Pengetahuan ini dapat dihasilkan berdasarkan data yang diperoleh dari aktivitas-aktivitas akademik yang sudah dilakukan (perkuliahan, ujian, perwalian, dan sebagainya) menggunakan teknologi *data mining*. *Educational Data Mining* merupakan teknologi yang digunakan untuk memeperolah pengetahuan tentang perilaku sivitas akademika dan dapat digunakan untuk menunjang pengambilan keputusan dalam aktivitas pendidikan (N. Delavari, Alaa M. El-Halees, Dr. M. Reza Beikzadeh, 2005). Dalam penelitian ini akan dilakukan analisis penggunaan metode *data mining* untuk memprediksi kelulusan tepat waktu mahasiswa berdasarkan data akademik. Hasil dari prediksi ini diharapkan dapat membantu manajemen dalam mengambil kebijakan untuk meningkatkan kelulusan tepat waktu mahasiswa. Hal ini sejalan dengan tujuan dari program studi untuk meningkatkan rasio Angka Efisiensi Edukasi (AEE).

Terdapat beberapa algoritma *data mining* yang dapat digunakan untuk prediksi performa akademik mahasiswa, diantaranya yaitu *Naïve Bayes*, *Multi Layer Perceptron*, *SMO*, *J48*, dan *REPTree* (V.Ramesh, 2013). Perbandingan fitur dan batasan dari masing-masing metode dapat dilihat pada tabel 1.

Tabel 1. Perbandingan metode *Data Mining* (Nikam, 2015)

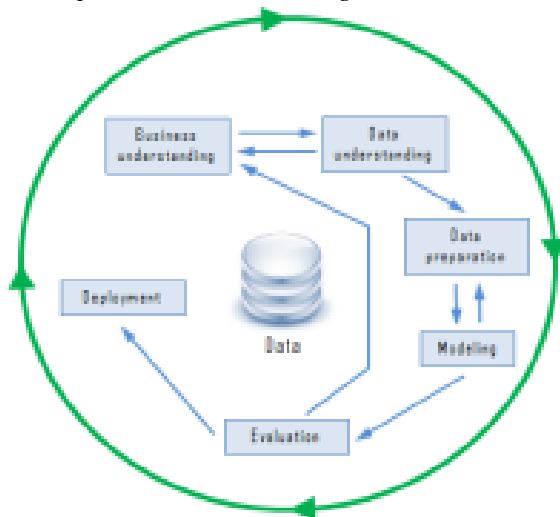
| Nomor | Metode | Fitur | Batasan |
|-------|---|--|---|
| 1 | <i>Decision Tree</i> (J48 dan <i>REPTree</i>) | <ol style="list-style-type: none"> 1. Model yang dibangun mudah untuk diinterpretasikan 2. Mudah untuk diimplementasikan 3. Dapat digunakan untuk data diskrit maupun kontinyu 4. Bisa mengatasi noise | <ol style="list-style-type: none"> 1. Sedikit variasi pada data dapat merubah <i>decision tree</i> 2. Tidak bekerja dengan baik untuk <i>training dataset</i> yang kecil 3. <i>Overfitting</i> |
| 2 | <i>Naïve Bayes</i> | <ol style="list-style-type: none"> 1. Mudah diimplementasikan 2. Efisiensi dalam komputasional dan <i>classification rate</i> 3. Hasil prediksi yang akurat untuk kebanyakan problem klasifikasi dan prediksi | <ol style="list-style-type: none"> 1. Presisi dari algoritma berkurang apabila jumlah data sedikit 2. Untuk mendapatkan hasil yang bagus, dibutuhkan data dalam jumlah yang sangat besar |
| 3 | <i>Neural Network</i> (<i>Multi Layer Perceptron</i>) | <ol style="list-style-type: none"> 1. Mudah digunakan, dengan sedikit parameter yang digunakan 2. Mudah diimplementasikan 3. Aplikatif untuk banyak kasus dalam kehidupan | <ol style="list-style-type: none"> 1. Membutuhkan waktu proses yang lama apabila <i>neural network</i> besar 2. Proses <i>learning</i> mungkin lama 3. Sulit untuk |

| | | | |
|---|-------------------------------------|---|---|
| | | | menentukan berapa <i>neuron</i> dan <i>layer</i> yang dibutuhkan |
| 4 | <i>Support Vector Machine (SMO)</i> | <ol style="list-style-type: none"> 1. Akurasi tinggi 2. Bekerja dengan baik bahkan jika data tidak terpisah secara linear pada <i>feature space</i> | <ol style="list-style-type: none"> 1. Kebutuhan kecepatan dan ukuran yang besar baik untuk <i>training</i> dan <i>testing</i> 2. Kompleksitas yang besar dan kebutuhan memori yang tinggi |

Berdasarkan permasalahan yang telah diuraikan di atas, maka tujuan dari penelitian ini adalah untuk membandingkan metode-metode *data mining* yang dapat diterapkan untuk prediksi kelulusan mahasiswa. Metode data mining dengan akurasi terbaik dapat digunakan sebagai pendukung keputusan bagi manajemen dalam mengambil kebijakan. Dukungan teknologi *data mining* dalam pengambilan keputusan ini diharapkan dapat menunjang upaya manajemen untuk meningkatkan nilai Angka Efisiensi Edukasi.

METODE

Penelitian ini menggunakan tahapan-tahapan CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang banyak digunakan dalam proses data mining. Sebelum memasuki bagian tahapan dari CRISP-DM, terlebih dahulu dilakukan tahap awal yaitu pengumpulan data dan studi literatur. Data diperoleh dari tim akademik yang mempersiapkan program kelulusan tepat waktu dan juga dari bagian administrasi akademik. Di samping pengumpulan data, ada beberapa cara yang telah dilakukan dalam tahap awal penelitian ini, yaitu studi literatur terkait metode-metode pemodelan *data mining*.



Gambar 2. Tahapan metodologi CRISP-DM

Tahapan selanjutnya dalam penelitian ini mengikuti metodologi CRISP-DM. Metodologi ini terdiri dari 6 tahapan yaitu: *business understanding*, *data understanding*, *data preparation*, *modeling*, dan *evaluation* seperti dapat dilihat pada gambar 2. Metodologi CRISP-DM adalah standarisasi *data mining* yang disusun oleh tiga organisasi penggagas *data mining* yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), dan NCR. Lebih dari 300 organisasi ikut berkontribusi dalam pengembangan standar ini dan akhirnya CRISP-DM versi 1.0 dipublikasikan pada 1999 (Prajitno, 2011). Penerapan metodologi CRISP-DM ini juga sudah dilakukan dalam jurnal penelitian tentang prediksi kelulusan studi mahasiswa menempuh mata kuliah (Fadillah, 2015).

Langkah pertama pada metodologi CRISP-DM adalah *Business Understanding*. Tahapan ini berfokus pada pemahaman tujuan dan kebutuhan dari perspektif bisnis atau organisasi. Langkah selanjutnya yaitu *Data Understanding* yaitu pengumpulan data awal yang dapat digunakan dalam proses data mining. Data tersebut kemudian dideskripsikan dan diidentifikasi untuk menentukan apa saja yang bisa dipilih untuk menghasilkan informasi/pengetahuan. Tahapan ketiga yaitu *Data Preparation* yang terdiri dari proses *data cleaning*, *data transformation* dan *data selection* hingga dataset siap untuk digunakan. Selanjutnya adalah tahapan *Modeling* yaitu penerapan metode-metode data mining yang sudah ditentukan. Setelah data mining dimodelkan dan dibangun, maka tahap berikutnya adalah *Evaluation*. Di sini dilakukan analisis apakah model yang sudah dibangun dapat memenuhi kebutuhan bisnis. Tahapan terakhir adalah *Deployment*, yaitu ketika sudah dilakukan aksi untuk menerapkan model yang sudah dikembangkan. Penelitian ini hanya akan membahas sampai tahap *Evaluation*, yaitu menganalisis metode data mining terbaik yang dapat digunakan untuk prediksi kelulusan tepat waktu mahasiswa.

Beberapa metode data mining yang digunakan untuk prediksi klasifikasi pada penelitian ini adalah:

1. *Multilayer Perceptron*
2. *Naïve Bayes*
3. *SMO*
4. *J48*
5. *REPTree*

HASIL DAN PEMBAHASAN

Hasil dari tahap awal penelitian yaitu terkumpulnya data kondisi mahasiswa angkatan 2012 pada awal semester 15.2 (periode semester ketika program tepat waktu untuk mahasiswa 2012 dilaksanakan). Data mahasiswa ini terdiri dari NIM mahasiswa, nama dosen wali, Indeks Prestasi Kumulatif (IPK), SKS Kumulatif (SKSK), status TA mahasiswa, status tempuh KP, dosen pembimbing terpilih, dan dosen pembimbing pilihan.

1) Pemahaman Bisnis (*Business Understanding*)

a) Menentukan Tujuan Bisnis (*Determine Business Objectives*)

Tujuan bisnis dari penelitian ini adalah untuk merancang model *data mining* yang dapat digunakan sebagai prediktor kelulusan tepat waktu mahasiswa berdasarkan data akademik. Hasil dari prediksi ini diharapkan

dapat membantu manajemen dalam mengambil keputusan atau kebijakan untuk meningkatkan kelulusan tepat waktu mahasiswa. Hal ini sejalan dengan tujuan dari program studi untuk meningkatkan rasio Angka Efisiensi Edukasi (AEE).

b) Menilai situasi (*Asses the Situation*)

Program Kerja peningkatan Angka Efisiensi Edukasi sudah dilakukan prodi Sistem Informasi Stikom Surabaya sejak tahun 2015. Pada tahun 2016 sudah dilaksanakan beberapa aktivitas untuk mendukung kelulusan tepat waktu mahasiswa Sistem Informasi angkatan 2012: memberikan pengarahan umum untuk mahasiswa angkatan 2012 yang sudah layak menempuh Tugas Akhir, menstrukturkan pemilihan dosen pembimbing Tugas Akhir, serta menjadwalkan sidang proposal dan ujian Tugas Akhir secara simultan untuk mencapai kelulusan tepat waktu. Sebelum rangkaian aktivitas ini dimulai, tim akademik melakukan pengolahan data mahasiswa untuk menentukan mahasiswa yang sudah layak TA. Data ini di antaranya mencakup NIM, Nama, IPK, SKS, dan status tempuh Kerja Praktik (KP). Data inilah yang digunakan dalam penelitian ini untuk membentuk sebuah model *data mining* prediktor kelulusan mahasiswa.

c) Menentukan tujuan Data Mining (*Detemine the Data Mining Goals*)

Model data mining yang dibuat bertujuan untuk memprediksi apakah mahasiswa dapat lulus tepat waktu atau tidak berdasarkan data akademik mahasiswa. Pengetahuan ini dapat digunakan oleh manajemen untuk mengambil keputusan atau membuat kebijakan untuk mendukung kelulusan tepat waktu mahasiswa.

2) Pemahaman Data (*Data Understanding*)

1) Mengumpulkan Data Awal (*Collect the Initial Data*)

Pengumpulan data awal dilakukan oleh tim akademik AEE dengan menggali data dari bagian *data center*. Data yang dikumpulkan adalah data akademik mahasiswa Sistem Informasi angkatan 2012 pada awal semester 15.2. Jumlah data adalah 89 mahasiswa yang sudah memenuhi persyaratan layak TA, dengan kriteria layak TA adalah SKS Kumulatif (SKSK) ≥ 114 dan sudah/ sedang menempuh Kerja Praktik. Kemudian, tiap *record* akan ditambahkan atribut status lulus tepat waktu nya berdasarkan data mahasiswa wisuda pada periode Oktober 2016.

2) Mendeskripsikan Data (*Describe the Data*)

Deskripsi dari atribut data yang akan digunakan dalam membangun model *data mining* dapat dilihat pada tabel 2.

Tabel 2. Deskripsi Data

| Atribut | Deskripsi |
|------------|---------------------------|
| NIM | Nomor Induk Mahasiswa |
| Nama | Nama lengkap mahasiswa |
| Dosen Wali | Dosen wali mahasiswa |
| IPK | Indeks Prestasi Kumulatif |

| | |
|-------------------|---|
| | mahasiswa (pada awal semester 15.2) |
| SKSK | SKS Kumulatif (Jumlah SKS yang sudah ditempuh mahasiswa) |
| STS | Status mahasiswa (Tugas Akhir) |
| STS_Tempuh_KP | Status menempuh Kerja Praktik (Sudah tempuh / sedang tempuh / belum tempuh) |
| Dosen 1 | Nama dosen pembimbing 1 |
| Dosen 2 | NIK dosen pembimbing 2 |
| Lulus Tepat Waktu | Status lulus tepat waktu (Ya/tidak) |

3) Persiapan Data (*Data Preparation*)

Pada proses ini dilakukan pemilihan data (*Select Data*) dan membersihkan data (*Clean Data*).

| No | 1: NIM | 2: DOSEN_WALI | 3: IPK | 4: SKSK | 5: STS | 6: STS_TEMPUH_KP | 7: DOSEN 1 | 8: DOSEN 2 | 9: WISUDA |
|----|---------|------------------|---------|---------|---------|------------------|--------------|---------------|-----------|
| | Nominal | Nominal | Numerik | Numerik | Nominal | Nominal | Nominal | Nominal | Nominal |
| 1 | 1.24... | Endra Rahma... | 3,67 | 122,0 | x | SUDAH TEMPUH | M.J. Dewi... | Erwin Sut... | Y |
| 2 | 1.24... | Endra Rahma... | 3,44 | 123,0 | x | SUDAH TEMPUH | M.J. Dewi... | Haryanto ... | Y |
| 3 | 1.24... | Endra Rahma... | 3,8 | 120,0 | x | SEDANG TEMPUH | M.J. Dewi... | Valentinu... | T |
| 4 | 1.24... | Endra Rahma... | 3,15 | 122,0 | x | SUDAH TEMPUH | M.J. Dewi... | Jusak | Y |
| 5 | 1.24... | Endra Rahma... | 3,2 | 122,0 | x | SEDANG TEMPUH | M.J. Dewi... | Pantjawa... | T |
| 6 | 1.24... | Endra Rahma... | 3,16 | 120,0 | x | SEDANG TEMPUH | Sri Hana... | A. E. Tjan... | Y |
| 7 | 1.24... | Endra Rahma... | 3,3 | 122,0 | x | SEDANG TEMPUH | Sri Hana... | A. E. Tjan... | T |
| 8 | 1.24... | Endra Rahma... | 3,68 | 123,0 | x | SEDANG TEMPUH | Sri Hana... | M. J. Dewi... | T |
| 9 | 1.24... | Endra Rahma... | 3,32 | 120,0 | x | SEDANG TEMPUH | Erwin Sut... | Haryanto ... | Y |
| 10 | 1.24... | Endra Rahma... | 3,43 | 123,0 | x | SEDANG TEMPUH | Sri Hana... | Erwin Sut... | Y |
| 11 | 1.24... | Endra Rahma... | 3,35 | 123,0 | x | SEDANG TEMPUH | Tan Amel... | Erwin Sut... | T |
| 12 | 1.24... | Endra Rahma... | 3,64 | 124,0 | x | SUDAH TEMPUH | Tan Amel... | Erwin Sut... | Y |
| 13 | 1.24... | Henry Bamba... | 3,32 | 123,0 | x | SEDANG TEMPUH | Tan Amel... | Tegar He... | Y |
| 14 | 1.24... | Henry Bamba... | 3,5 | 122,0 | x | SEDANG TEMPUH | Tan Amel... | Tegar He... | Y |
| 15 | 1.24... | Henry Bamba... | 3,19 | 114,0 | x | SEDANG TEMPUH | Tan Amel... | | T |
| 16 | 1.24... | Henry Bamba... | 3,29 | 125,0 | x | SEDANG TEMPUH | Endra Ra... | Tan Amel... | Y |
| 17 | 1.24... | Henry Bamba... | 3,2 | 118,0 | x | SEDANG TEMPUH | Endra Ra... | Sri Hana... | T |
| 18 | 1.24... | Henry Bamba... | 3,58 | 126,0 | x | SEDANG TEMPUH | Endra Ra... | Sri Hana... | Y |
| 19 | 1.24... | Jusak | 3,36 | 125,0 | x | SEDANG TEMPUH | Erwin Sut... | Haryanto ... | Y |
| 20 | 1.24... | Jusak | 3,57 | 130,0 | x | SUDAH TEMPUH | Erwin Sut... | Haryanto ... | Y |
| 21 | 1.24... | Jusak | 3,35 | 125,0 | x | SEDANG TEMPUH | Valentinu... | Jusak | Y |
| 22 | 1.24... | Jusak | 3,29 | 125,0 | x | SEDANG TEMPUH | Valentinu... | Yopy Mir... | T |
| 23 | 1.24... | Jusak | 2,87 | 119,0 | x | SEDANG TEMPUH | Valentinu... | | T |
| 24 | 1.24... | M.J. Dewyani ... | 3,43 | 123,0 | x | SEDANG TEMPUH | Valentinu... | M. J. Dewi... | Y |
| 25 | 1.24... | M.J. Dewyani ... | 3,33 | 123,0 | x | SEDANG TEMPUH | Valentinu... | Sri Hana... | missing |
| 26 | 1.24... | M.J. Dewyani ... | 3,55 | 123,0 | x | SEDANG TEMPUH | Agus Dwi... | | Y |
| 27 | 1.24... | M.J. Dewyani ... | 3,67 | 123,0 | T | SEDANG TEMPUH | Agus Dwi... | Haryanto ... | Y |
| 28 | 1.24... | M.J. Dewyani ... | 3,41 | 123,0 | T | SEDANG TEMPUH | Agus Dwi... | | Y |
| 29 | 1.24... | M.J. Dewyani ... | 3,5 | 123,0 | T | SEDANG TEMPUH | Haryanto ... | Erwin Sut... | Y |
| 30 | 1.24... | M.J. Dewyani ... | 3,59 | 123,0 | T | SEDANG TEMPUH | Haryanto ... | Endra Ra... | T |
| 31 | 1.24... | M.J. Dewyani ... | 3,48 | 125,0 | T | SEDANG TEMPUH | Haryanto ... | Endra Ra... | Y |
| 32 | 1.24... | Antok Supriya... | 3,54 | 119,0 | x | SEDANG TEMPUH | Achmad ... | Ajuningb... | Y |
| 33 | 1.24... | Antok Supriya... | 3,84 | 125,0 | x | SEDANG TEMPUH | Bamban... | Endra Ra... | Y |
| 34 | 1.24... | Henry Bamba... | 3,83 | 131,0 | T | SUDAH TEMPUH | Pantjawa... | Ignatius ... | Y |
| 35 | 1.24... | Henry Bamba... | 3,58 | 131,0 | x | SUDAH TEMPUH | Haryanto ... | Sri Hana... | Y |

Gambar 3. Proses persiapan data pada Weka

Pada tahap ini, dataset dimasukkan dalam aplikasi Weka untuk kemudian dilakukan proses *data cleaning* dengan mengganti setiap *missing value* (0 dan *Null*)

4) Pemodelan (*Modeling*)

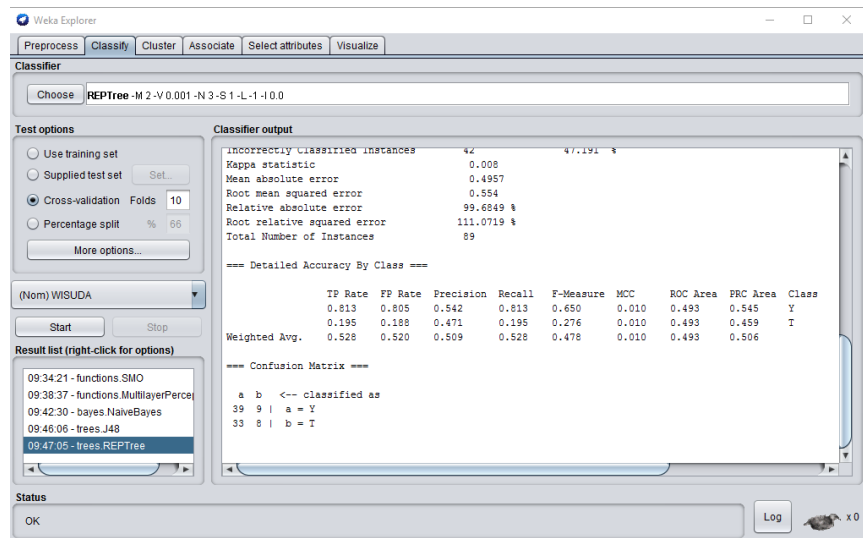
1) Memilih Teknik Pemodelan (*Select Modelling Technique*)

Pada tahapan ini dilakukan pemilihan metode *data mining* yang akan digunakan untuk membangun model. Pada penelitian ini terdapat 5 metode *data mining* yang digunakan: *Multi Layer Perceptron*, *J48*, *SMO*, *REPTree*, dan *Naïve Bayes*.

2) Membangun Model (*Build Model*)

Proses membangun model dilakukan dengan bantuan aplikasi Weka. Weka salah satu aplikasi *data mining* yang bersifat *open source* dan sudah banyak digunakan dalam pengembangan model *data mining*. Weka dibangun berbasis Java oleh tim dari University of Waikato, New Zealand (Witten, 2011). Fitur kunci yang menjadikan Weka populer adalah:

- a. menyediakan banyak algoritma yang berbeda untuk *data mining* dan *machine learning*
- b. merupakan aplikasi *open source* dan tersedia secara bebas
- c. *platform-independent*
- d. mudah digunakan oleh orang yang bukan spesialis *data mining*
- e. menyediakan fasilitas yang fleksibel untuk eksperimen
- f. selalu *up-to-date* dengan penambahan algoritma baru seperti yang muncul dalam literatur penelitian.



Gambar 4. Proses membangun model *Data Mining* pada Weka

3) Menilai Model (*Assess Model*)

Perbandingan akurasi model dari kelima metode *data mining* yang sudah dibangun menggunakan Weka dapat dilihat hasilnya pada tabel 3. Accuracy adalah seberapa akurat model dapat memprediksi *True Positive* (TP) dan *True Negative* (TN) dibandingkan total dataset. Untuk *error rate* yang digunakan pada penelitian ini ada tingkat kesalahan dalam memprediksi jumlah lulusan tepat waktu. Sebagai contoh, algoritma Multilayer Perceptron memprediksi 46 mahasiswa lulus tepat waktu (27 TP dan 19 FP). Dengan data aktual sebanyak 48 mahasiswa lulus tepat

waktu, maka error rate dapat dihitung: $(48-46) / 48 = 0.04 = 4\%$. Perbandingan analisis statistikal dari model yang sudah dibangun juga dapat dilihat hasilnya pada tabel 4.

Tabel 3. Perbandingan akurasi dan error

| <i>Classification Algorithm</i> | <i>Accuracy Rate</i> | <i>Error Rate</i> |
|---------------------------------|----------------------|-------------------|
| <i>Naïve Bayes</i> | 57,3% | 0% |
| <i>Multilayer Perceptron</i> | 55,1% | 4% |
| SMO | 56,2% | 14% |
| J48 | 53,9% | 85% |
| <i>REPTree</i> | 52,8% | 50% |

Tabel 4. Perbandingan analisis statistik

| Classification Algorithm | Class | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area |
|---------------------------------|--------------|----------------|----------------|------------------|---------------|------------------|-----------------|
| <i>Naïve Bayes</i> | Y | 0.604 | 0.463 | 0.604 | 0.604 | 0.604 | 0.555 |
| | T | 0.537 | 0.396 | 0.537 | 0.537 | 0.537 | 0.555 |
| <i>Multilayer Perceptron</i> | Y | 0.563 | 0.463 | 0.587 | 0.563 | 0.574 | 0.572 |
| | T | 0.557 | 0.438 | 0.512 | 0.537 | 0.524 | 0.572 |
| SMO | Y | 0.521 | 0.390 | 0.610 | 0.521 | 0.562 | 0.565 |
| | T | 0.610 | 0.479 | 0.521 | 0.610 | 0.562 | 0.565 |
| J48 | Y | 1 | 1 | 0.539 | 1 | 0.701 | 0.473 |
| | T | 0 | 0 | 0 | 0 | 0 | 0.473 |
| <i>REPTree</i> | Y | 0.813 | 0.805 | 0.542 | 0.813 | 0.650 | 0.493 |
| | T | 0.195 | 0.188 | 0.471 | 0.195 | 0.276 | 0.493 |

4) Evaluasi (*Evaluation*)

Evaluasi pada penelitian ini akan difokuskan pada model dengan akurasi yang tertinggi dan didukung dengan error rate yang kecil. Berdasarkan kriteria tersebut, maka algoritma *Naïve Bayes* merupakan prediktor terbaik dengan tingkat akurasi 57,3%. *Error rate* yang dihasilkan oleh algoritma *Naïve Bayes* adalah 0%, yang artinya prediksi jumlah mahasiswa yang lulus tepat waktu sesuai dengan data yang ada (48 mahasiswa lulus tepat waktu).

SIMPULAN

Setelah melakukan tahapan CRISP-DM, penelitian ini berhasil menentukan metode *data mining* terbaik untuk memprediksi kelulusan mahasiswa. Setelah dilakukan uji coba terhadap beberapa metode *data mining*, maka dapat ditarik kesimpulan bahwa metode dengan akurasi tertinggi dan *error rate* terkecil adalah *Naïve Bayes*. Sebenarnya dengan tingkat akurasi 57,3% model ini masih perlu ditingkatkan lagi apabila akan benar-benar diimplementasikan. Akurasi dari algoritma *Naïve Bayes* berkurang apabila jumlah data sedikit. Untuk mendapatkan hasil yang bagus, maka dibutuhkan data dalam jumlah yang lebih besar.

SARAN

Berdasarkan simpulan di atas, berikut ini adalah beberapa hal yang dapat dikembangkan untuk penelitian berikutnya:

- 1) Untuk meningkatkan akurasi *data mining* diperlukan jumlah data yang besar. Oleh sebab itu diperlukan pengumpulan data yang lebih banyak dari data mahasiswa untuk beberapa angkatan. Dengan demikian, akurasi dari model *data mining* yang dibangun dapat menjadi lebih baik.
- 2) Peningkatan akurasi juga dapat dilakukan dengan merubah parameter-parameter yang digunakan dalam membangun model *data mining*, yang mana belum dilakukan pada penelitian ini. Untuk penelitian selanjutnya dapat dilakukan uji coba pemilihan parameter yang digunakan dalam membangun model *data mining* agar akurasi yang dihasilkan meningkat.
- 3) Model yang dibangun menggunakan metodologi CRISP-DM pada penelitian ini belum sampai pada tahap *deployment* atau implementasi. Penelitian berikutnya dapat melakukan implementasi model yang sudah dibangun menjadi suatu sistem atau aplikasi. Selanjutnya dapat dilakukan evaluasi sistem dalam kaitannya untuk menunjang program peningkatan Angka Efisiensi Edukasi (AEE).

DAFTAR PUSTAKA

- BAN-PT. 2007. Buku V Pedoman Penilaian Portofolio Akreditasi Institusi Perguruan Tinggi. Jakarta: Departemen Pendidikan Nasional.
- Fadillah, Anisa Paramitha. 2015. Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah. Unikom Bandung: Jurnal Teknik Informatika dan Sistem Informasi Vol 1 No 3.
- Lubis, Chairuddin P. 2004. Sistem Kuliah Simultan di Universitas Sumatera Utara. E-USU Repository Unibersitas Sumatera Utara.
- N.Delavari, Alaa M. El-Halees, Dr. M. Reza Beikzadeh. 2005. Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System. Proceedings of 6th International Conference ITHET IEEE.
- Nikam, Sagar S. A Comparative Study of Classification Techniques in Data Mining Algorithms. Oriental Journal of Computer Science & Technology Vol 8 No 1.
- Prajitno, A. A. dan Purwarianti, A. 2011. Prediksi Kinerja Penjualan Karya Musik Menggunakan Framework CRISP-DM (Studi Kasus: X Music Indonesia). Jurnal Sarjana Institut Teknologi Bandung bidang Teknik Elektro dan Informatika.
- V.Ramesh, P.Parkavi, K.Ramar. 2013). Predicting student performance: A statistical and data mining approach. International journal of computer applications, Volume 63 no 8.
- Witten, Ian H. 2011. Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition. San Francisco: Morgan Kaufmann.