

Teknik *Data Mining* Menggunakan Algoritma *Decision Tree C4.5* untuk Memprediksi Tingkat Kelulusan Tepat Waktu

Chandra Wirawan

Abstrak—Lulus tepat waktu merupakan salah satu indikator keberhasilan mahasiswa dalam memperoleh gelar sarjana. Lulus tepat waktu merupakan salah satu indikator penilaian terhadap kualitas perguruan tinggi, karena salah satu penilaian akreditasi Badan Akreditasi Nasional Perguruan Tinggi (BAN PT). Sehingga apabila tingkat kelulusan dan mahasiswa tidak seimbang, maka akan mempengaruhi penilaian akreditasi pada Program Studi dan Universitas tersebut. Pada penelitian ini peneliti membahas prediksi tingkat kelulusan tepat waktu menggunakan teknik data mining algoritma C.4.5 dengan studi kasus UIN Syarif Hidayatullah Jakarta. Hasil dari penelitian ini diperoleh Nilai akurasi setelah pengujian menggunakan confusion matrix sebesar 89.82% dengan Precision 52.63% dan Recall 41.67%, mengindikasikan bahwa performa keakurasian pada percobaan tersebut bernilai cukup baik, dan Model yang dihasilkan dari decision tree pada penelitian ini adalah mahasiswa yang IPSmt 2 > 3.145 maka diprediksi mahasiswa tersebut dapat lulus tepat waktu dan untuk mahasiswa yang IPSmt 2 ≤ 3.145 dinyatakan dapat lulus tepat waktu dan tidak dapat lulus tepat waktu apabila memenuhi kondisi sesuai dengan model decision tree yang telah dihasilkan.

Keywords—*Decision tree*, algoritma C.4.5, Prediksi tingkat kelulusan tepat waktu, Teknik data mining

I. PENDAHULUAN

Lulus tepat waktu merupakan salah satu indikator keberhasilan mahasiswa dalam memperoleh gelar sarjana. Dalam praktiknya mahasiswa tidak selalu dapat menyelesaikan pendidikan sarjana dalam kurun waktu empat tahun [1].

Akreditasi perguruan tinggi sebagai satuan penyelenggara pendidikan tinggi menyediakan metode evaluasi terhadap institusi perguruan tinggi dalam mengukur, menetapkan mutu

dan kelayakan programnya. Lulus tepat waktu merupakan salah satu indikator penilaian terhadap kualitas perguruan tinggi, karena salah satu penilaian akreditasi Badan Akreditasi Nasional Perguruan Tinggi (BAN PT) yang terdapat pada standar 3 mengenai mahasiswa dan lulusan adalah profil mahasiswa lulus tepat waktu. Sehingga apabila tingkat kelulusan dan mahasiswa tidak seimbang, maka akan mempengaruhi penilaian akreditasi pada Program Studi dan Universitas tersebut.

UIN Syarif Hidayatullah Jakarta didirikan dengan Surat Keputusan Presiden RI Nomor 031 tahun 2002. Menurut data akreditasi Universitas pada data.uinjkt.ac.id, terdapat data yang tidak signifikan antara mahasiswa yang masuk dan tingkat kelulusan mahasiswa tepat waktu. Jumlah mahasiswa yang lulus tepat waktu, jauh lebih sedikit dibandingkan dengan jumlah lulusan yang masuk UIN Syarif Hidayatullah Jakarta.

Jumlah lulusan mahasiswa S-1 7 tahun terakhir yang lulus tepat waktu tidak sampai 50%. Dalam waktu jangka yang panjang, hal ini dapat menyebabkan penumpukan mahasiswa di UIN Syarif Hidayatullah Jakarta dan tentu saja akan berakibat pada penurunan nilai akreditasi bagi Program Studi dan Universitas. [2]

Salah satu teknik untuk melakukan prediksi dapat dilakukan dengan menggunakan data mining, Metode data mining dapat digunakan untuk memberikan pengetahuan yang sebelumnya tersembunyi didalam *data warehouse* sehingga menjadi informasi yang berharga, data mining digunakan untuk prediksi waktu tempuh studi mahasiswa. [3].

Pada penelitian Janwata dan Tsai, penelitian ini membangun model lulusan ketenagakerjaan menggunakan klasifikasi, dengan membandingkan beberapa pendekatan data mining, seperti metode Bayesian dan metode Pohon. Metode Bayesian mencakup 5 algoritma, termasuk AODE, BayesNet, HNB, NaviveBayes, WAODE. Metode Tree mencakup 5

C. Wirawan Program Pasca Sarjana Magister Ilmu Komputer, Universitas Budi Luhur Jakarta, Indonesia (e-mail: chandrawirawan50@gmail.com).

Received: 28 Oktober 2019; Revised: 9 Juli 2020 Accepted: 9 Juli 2020

algoritma, termasuk BFTree, NBTree, REPTree, ID3, C4.5. dan ternyata dari perbandingan tersebut algoritma C.4.5 lebih unggul akurasi, yaitu sebesar 98.71%, sedangkan untuk algoritma yang lainnya berada dibawah nilai akurasi algoritma C.4.5 [4].

Menurut Kaur et al dalam Classification and prediction based data mining algorithms to predict slow learners in education sector. mengidentifikasi siswa yang lambat di antara siswa lainnya, menampilkannya dengan prediksi data mining model menggunakan algoritma berbasis klasifikasi, data di bandingkan dengan menggunakan Naïve Bayes, SMO, J48 and REPTree, dan setelah di uji nilai akurasi C.4.5 juga lebih tinggi dibandingkan yang lain, yaitu sebesar 69.73%. [5]

Pada penelitian ini peneliti menggunakan teknik data mining untuk memprediksi tingkat kelulusan tepat waktu menggunakan Algoritma C.4.5 pada UIN Syarif Hidayatullah Jakarta.

II. LITERATURE REVIEW

Dalam Kegiatan evaluasi, perencanaan, dan pengambilan keputusan akan dapat dilakukan dengan lebih baik jika sebuah organisasi memiliki informasi yang lengkap, cepat, tepat, dan akurat. Informasi yang dibutuhkan dapat diekstrak dari data operasional yang tersimpan dalam database yang terintegrasi. [6] Menurut Sabna & Muhandi. (2016) dalam “Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, dan Hasil Belajar” diperoleh bahwa variabel nilai rapor (hasil belajar masa lalu) menjadi node awal artinya dari 5 variabel yang menentukan prestasi akademik mahasiswa maka nilai rapor menjadi node yang terpilih sebagai penentu pertama terhadap prestasi akademik mahasiswa. [7] Menurut Novianti & Aziz (2015) dalam “Aplikasi Data Mining Menggunakan Metode Decision Tree Untuk Menampilkan Laporan Hasil Nilai 18 Akhir Mahasiswa (Studi Kasus Di Fakultas Teknik Umsurabaya)” dihasilkan, aplikasi laporan hasil nilai akhir dengan metode decision tree algoritma C4.5 sebagai acuan dalam membuat kebijakan dan tindakan untuk mengurangi jumlah mahasiswa yang tidak lulus dengan akurasi 100% menggunakan evaluasi Cross Validation. [8] Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Di Universitas Klabat Dengan Metode Klasifikasi” diteliti oleh Mandias (2015). Hasil yang didapat yaitu untuk mencapai semester akhir dalam 4 tahun seorang mahasiswa diawal semester berdasarkan pohon keputusan yang terbentuk yaitu mahasiswa yang mempunyai indeks prestasi 3.5 – 3.87 (magna) serta penempatan bahasa inggris yang berada pada level adv dan inter. [9]

III. METODOLOGI

A. Metode Pemilihan Sample

Menurut hukum statistik dalam menentukan jumlah sample, semakin besar jumlah sample yang diambil maka semakin menggambarkan keadaan populasi dan pemakaian

jumlah sample yang besar sangat di anjurkan, dengan pertimbangan adanya berbagai keterbatasan pada peneliti, sehingga peneliti berusaha mengambil sample minimal dengan syarat dan aturan statistic tetapi terpenuhi sebagaimana dianjurkan oleh Issac dan Michael.

B. Metode Pengumpulan Data

Pengumpulan data digunakan dalam proses pengolahan data dan analisa selanjutnya pada penelitian, karenanya dalam pengumpulan data, harus memperhatikan bahwa data yang diperoleh harus akurat dan dapat dipertanggung jawabkan secara ilmiah

C. Teknik Analisa

Teknik analisa yang digunakan dalam pengembangan teknik data mining pada penelitian ini yaitu menggunakan metode CRISP DM (Cross Standart Industries for Data Mining). Terdapat 6 fase pada metode ini, adapun tahapannya sebagai berikut: [10]

1) Bussines Understanding

Menurut data akreditasi universitas pada data.uinjkt.ac.id, terdapat data yang tidak signifikan antara mahasiswa yang masuk dan keluar. Jumlah lulusan mahasiswa S-1 7 tahun terakhir yang lulus tepat waktu tidak sampai 50%. Dalam jangka panjang, hal tersebut dapat menyebabkan penumpukkan mahasiswa di UIN Syarif Hidayatullah Jakarta. Maka dari itu penelitian ini akan melakukan prediksi tingkat kelulusan tepat waktu untuk mahasiswa di UIN Syarif Hidayatullah Jakarta

2) Data Understanding

Untuk memprediksi kelulusan mahasiswa tepat waktu, peneliti menggunakan data akademik mahasiswa tahun 2012-2014 sebanyak 754 record, data tersebut diperoleh dari data Pusat Informasi dan Data (PUSTIPANDA) UIN Syarif Hidayatullah Jakarta, data yang digunakan terdiri dari 9 Atribut prediksi dan 1 atribut hasil, jadi total atribut yang digunakan 10 atribut parameter. Terdiri dari jenis_kelamin, Program_studi, jenis_sekolah, jurusan_SLTA, daerah_asal, IPSmt 1-4, dan kelulusan.

3) Data Preparation

Pengolahan data awal perlu dilakukan untuk menyiapkan data yang benar dan valid sebelum diproses. Data akademik yang telah diperoleh akan dilakukan preprocessing data.

4) Modelling Phase

Pada tahapan ini merupakan tahap learning karena pada tahap ini dilakukan pemrosesan data training yang diklasifikasikan oleh model kemudian menghasilkan sejumlah aturan. Pada penelitian ini peneliti menggunakan algoritma C.45.dan menggunakan tools RapidMiner 7.3 untuk memproses data tersebut.

5) Evaluation Phase

Tahap ini digunakan untuk melakukan pengujian terhadap model yang bertujuan untuk mendapatkan model yang paling akurat kedalam frameworks RapidMiner 7.3. Evaluasi dan validasi pada penelitian ini menggunakan

metode confusion matrix, metode ini digunakan untuk mengukur precision, Recall dan Akurasi.

6) Deployment Phase

Pada tahapan ini peneliti menerapkan model data mining menggunakan decision tree Algoritma C.4.5 untuk memprediksi tingkat kelulusan tepat waktu mahasiswa UIN Syarif Hidayatullah Jakarta.

IV. HASIL DAN PEMBAHASAN

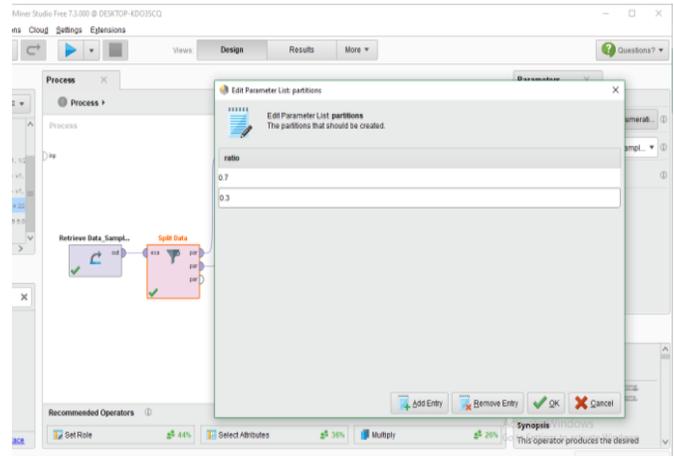
A. Pengelompokan dan Analisa Data

Data yang di gunakan dalam penelitian ini terdiri dari data training dan data testing yang berjumlah 754 data mahasiswa. Sebelum diolah menggunakan aplikasi RapidMiner data terlebih dahulu dilakukan validasi dengan menghapus data yang tidak lengkap atau kosong (null), data sampel yang telah dilakukan validasi selanjutnya dibagi kedalam dua bagian data training dan data testing dengan rasio 70% data training dan 30% data testing menggunakan *Stratified random sampling*, data yang di gunakan pada penelitian ini memiliki 10 atribut 9 atribut menjadi predictor dan 1 atribut hasil

B. Hasil Pengujian Dengan Tools

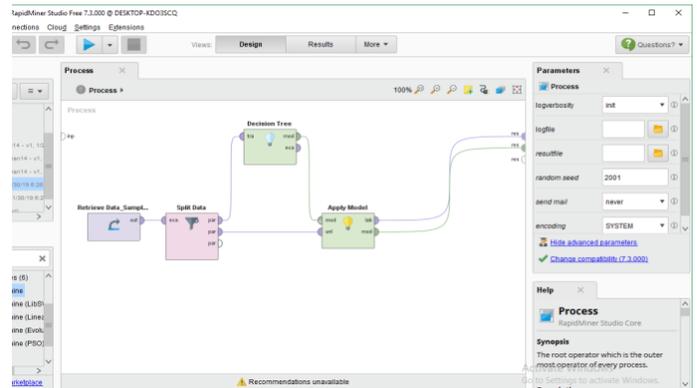
Tahapan-tahapan proses yang akan dilakukan dalam pengujian data set diawali dengan proses memasukkan data set yang telah di verifikasi kedalam program RapidMiner adapaun tahapan prosesnya sebagaimana berikut:

- 1) Mengumpulkan dataset untuk training dibuat dengan format Excel(.xlsx) kemudian data Excel tersebut di impor kedalam aplikasi RapidMiner, berikut dataset yang telah di impor kedalam program RapidMiner
- 2) Menentukan sample dengan cara membagi dataset menjadi dua bagian yaitu data training sebesar 70% dan data testing sebesar 30% menggunakan operator Split Data seperti pada Gambar 1.



Gambar 1. Penerapan Operator Split Data Pada Data Dataset

- 3) Menerapkan operator Algoritma terpilih, peneliti menggunakan Algoritma Decision Tree C.4.5 pada data training, kemudian dilanjutkan dengan tahapan penerapan model pada data testing menggunakan operator Apply Model. Desain dapat dilihat seperti pada Gambar 2.

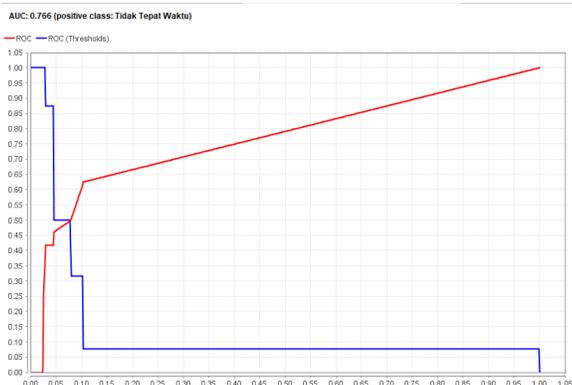


Gambar 2. Penerapan Operator Algoritma Decision Tree C.4.5

recall: 41.67% (positive class: Tidak Tepat Waktu)

	true Tepat Waktu	true Tidak Tepat Waktu	class precision
pred. Tepat Waktu	193	14	93.24%
pred. Tidak Tepat Waktu	9	10	52.63%
class recall	95.54%	41.67%	

Gambar 6. Menunjukkan Nilai Recall Algoritma C45



Gambar 7. hasil performa Kurva AUC

Kurva pada Gambar 7. Menunjukkan visualisasi dari akurasi model dan perbandingan antar model klasifikasi, dimana kurva ini menguji kinerja pengklasifikasian berdasarkan performance, dan hasil performa dari kurva AUC ini adalah sebesar 76.6%. hasil tersebut menunjukkan hasil performa dari data yang diuji adalah bernilai cukup baik

Penelitian tentang Prediksi Tingkat Kelulusan Tepat Waktu Menggunakan Teknik data Mining Decision Tree Algoritma C.4.5 mahasiswa UIN Syarif Hidayatullah Jakarta, data set yang digunakan untuk pengujian adalah sebanyak 754 mahasiswa, dengan atribut yang digunakan untuk prediksi adalah Terdiri dari jenis_kelamin, Hasil penelitian prediksi tingkat kelulusan tepat waktu.

V. KESIMPULAN

Menggunakan teknik data mining decision tree algoritma C.4.5 dengan menggunakan tools RapidMiner dan di uji menggunakan confusion matrix menghasilkan nilai akurasi 89.82% dengan Precision 52.63% dan Recall 41.67%, dan nilai AUC sebesar 76.6% nilai 50 ini mengindikasikan bahwa performa keakurasian pada percobaan tersebut bernilai cukup baik.

Model yang dihasilkan dari decision tree pada penelitian ini adalah mahasiswa yang IPSmt 2 >3.145 maka diprediksi mahasiswa tersebut dapat lulus tepat waktu dan untuk mahasiswa yang IPSmt 2 <= 3.145 dinyatakan dapat lulus tepat waktu dan tidak dapat lulus tepat waktu apabila memenuhi

kondisi sesuai dengan model decision tree yang telah dihasilkan.

REFERENSI

- [1] F. Nurahmah, Indahwati and Y. Angraini, "Analisis ketepatan waktu lulus berdasarkan karakteristik mahasiswa fem dan faperta menggunakan metode chart," *Xplore*, vol. 2(1), no. Vol. 2(1):e3(1-8), pp. 1-8, 2013.
- [2] C. Wirawan, E. Khudzaeva, T. H. Hasibuan, Karjono and Y. H. K. Lubis, "Application of Data mining to Prediction of Timeliness Graduation of Students (A Case Study)," *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1-4, 2019.
- [3] J. Han and M. Kamber, "Data Mining : Concepts and Techniques," San Francisco., Morgan Kaufmann Publishers , 2006.
- [4] B. Jantawan and C.-F. Tsai, "The Application of Data Mining to Build Classification Model for Predicting Graduate Employment," *IJCSIS International Journal of Computer Science and Information Security*, Vols. Vol. 11, No. 10, October 2013, no. Vol. 11, No. 10, October 2013, 2013.
- [5] P. Kaur, M. Singh and G. Singh, "Classification and prediction based data mining algorithms to predict slow learners in education sector," *Prodia Computer Science*, vol. 57 (2015) , no. 57 (2015) , pp. 500-508, 2015.
- [6] K. Y. N. Andri and S. Murniati, " Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang," *Seminar Nasional Informatika 2013 (SemnasIF 2013)*, vol. 2013(A), no. 2013(A), p. 56-63, 2013.
- [7] E. Sabna and Muhardi, "Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen , Motivasi , Kedisiplinan ,Ekonomi," *CoreIT*, vol. 2(2), p. 41-44, 2016.
- [8] T. Novianti and A. Aziz, "Aplikasi Data Mining Menggunakan Metode decision Tree untuk menampilkan laporan Hasil Nilai Akhir Mahasiswa Fakultas Teknik UMSURABAYA," *Network Engineering Research Operation [NERO]*, vol. 1(3), p. 198-203, 2015.
- [9] G. F. Mandias, " Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Di Universitas Klabat Dengan Metode Klasifikasi," *Konferensi Nasional Sistem & Informatika*, p. 9-10, 2015.
- [10] D. T. Larose, "Discovering Knowledge In Data An Introduction to Data," in *Automotive Industries AI* , Hoboken, New Jersey, New Jersey: John Wiley &, 2005.

