Perbandingan Algoritma C4.5 Dan Id3 Untuk Prediksi Ketepatan Waktu Lulus Mahasiswa

Titik Faizah*1, Arief Jananto²

^{1,2}Jurusan Sistem Informasi Fakultas Teknologi Informasi Universitas Stikubank (UNISBANK) e-mail: *¹faizah015@gmail.com, ²ajananto09@edu.unisbank.ac.id

Abstrak

Data kelulusan mahasiswa pada sebuah universitas khususnya sebuah program studi sangat berperan dalam mengevaluasi keberhasilan program pendidikan yang diberikan kepada mahasiswa. Program pendidikan dikatakan baik apabila ketepatan kelulusan mahasiswa juga baik. Sehubungan dengan upaya mengevaluasi program pendidikan pada sebuah program studi diperlukan studi untuk menggali informasi dari pola kelulusan mahasiswa dengan teknik data mining Knowledge Discovering Database (KDD) yang meliputi tahap seleksi data, preprocessing, transformasi data, proses mining, dan evaluasi. Proses mining menggunakan algoritma ID3 dan C45 yang digunakan untuk membangun model decission tree atau pohon keputusan. Berdasarkan hasil implementasi algoritma ID3 dan C45 terhadap data testing kelulusan mahasiswa menggunakan RStudio, nilai akurasi tertinggi diperoleh algoritma C4.5 yaitu sebesar 81,88% pada data testing 30%. Secara keseluruhan hasil implementasi terhadap komposisi data 90%, 80%, 70%, 30%, 20%, dan 10% diperoleh informasi bahwa nilai akurasi yang dihasilkan menggunakan algoritma C4.5 rata-rata lebih tinggi dibandingkan dengan nilai akurasi algoritma ID3. Dengan demikian dapat disimpulkan bahwa algoritma terbaik untuk memprediksi ketepatan kelulusan mahasiswa adalah algoritma C45.

Kata kunci — Algoritma ID3, Algoritma C45, Akurasi, Prediksi

Abstract

Student graduation data at a university, especially a study program, plays a very important role in evaluating the success of the educational program given to students. The education program is said to be good if the accuracy of student graduation is also good. In connection with the effort to evaluate educational programs in a study program, a study is needed to explore information from student graduation patterns with the Knowledge Discovering Database (KDD) data mining technique which includes the stages of data selection, preprocessing, data transformation, mining processes, and evaluation. The mining process uses the ID3 and C45 algorithms which are used to build a decision tree model. Based on the results of the implementation of the ID3 and C45 algorithms on student graduation testing data using RStudio, the highest accuracy value obtained by the C4.5 algorithm is 81.88% on the 30% testing data. Overall the results of the implementation of the data composition of 90%, 80%, 70%, 30%, 20%, and 10% obtained information that the accuracy value generated using the C4.5 algorithm is on average higher than the accuracy value of the ID3 algorithm. Thus it can be concluded that the best algorithm to predict the accuracy of student graduation is the C45 algorithm.

Keywords—ID3 Algorithm, C45 Algorithm, Accuracy, Prediction

1. PENDAHULUAN

Teknologi informasi khususnya penyimpanan data yang semakin canggih memberikan banyak dampak terhadap perkembangan pengelolaan data. Penyimpanan data dalam jumlah besar atau yang dikenal dengan big data tentunya tidak hanya dalam bentuk harddisk melainkan dapat berbentuk cloud yang sering digunakan pada sistem informasi. Dengan semakin mudahnya menyimpan data dalam jumlah besar, tentunya tetap saja akan memberikan dampak menumpuknya data, sebagai contoh pada sebuah institusi pendidikan tinggi. Sebuah kumpulan data sangatlah penting bagi institusi pendidikan tinggi dalam hal evaluasi pendidikan. Dengan adanya data besar yang dimiliki oleh sebuah institusi pendidikan tinggi, maka dibutuhkan data mining dalam memnuhi kebutuhan informasi. Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data [1]. Informasi yang dihasilkan diperoleh dengan mengekstraksi dan mengenali pola menarik dari sebuah data. Dalam mencari pola menarik tersebut dibutuhkan sebuah algoritma yang dapat membantu menemukan sebuah informasi dari suatu basis data.

Mahasiswa merupakan salah satu alat ukur evaluasi program pendidikan sebuah perguruan tinggi. Masa studi mahasiswa merupakan salah satu faktor yang berhubungan dengan evaluasi program pendidikan, di mana keberhasilan program pendidikan dilihat dari masa studi mahasiswanya. Apabila mahasiswa dapat lulus tepat waktu maka bisa dikatakan bahwa program pendidikan yang dilaksanakan sudah baik. Sebagai contoh, Universitas Y memiliki beban studi yang harus ditempuh paling cepat 7 semester dan paling lama 14 semester. Untuk itu, diperlukan analisa terhadap pola kelulusan (masa studi) mahasiswa sebagai informasi dalam hal evaluasi program pendidikan. Berdasarkan hal tersebut di atas, peneliti tertarik untuk menggali informasi data kelulusan mahasiswa Program Studi X Universitas Y yang bertujuan untuk melihat pola kelulusan mahasiswa dengan teknik data mining decission tree.[10]

Pembentukan model *decission tree* menggunakan algoritma ID3 dan algoritma C45 yang diimplementasikan pada data kelulusan mahasiswa. Dengan membandingkan nilai akurasi yang diperoleh dari hasil implementasi algoritma ID3 dan C45, akan dipilih algoritma yang terbaik untuk memprediksi ketepatan kelulusan mahasiswa.

1.1 Penelitian Terdahulu

Referensi [2] menunjukkan bahwa penelitian dilakukan untuk menganalisis data Jurusan Teknik Komputer menggunakan teknik data mining. Penerapan teknik tersebut akan menghasilkan pohon keputusan untuk melihat kemungkinan mahasiswa lulus lebih dari 8 semester. Pohon keputusan tersebut merupakan keluaran dari aplikasi yang dibangun menggunakan algoritma C4.5 untuk memprediksi masa studi mahasiswa. Data yang digunakan dalam penelitian ini adalah data kelulusan Jurusan Teknik Komputer angkatan 2007, dengan menggunakan atribut nilai beberapa mata kuliah.

Penelitian ini menggunakan metode yang meliputi lima aspek utama yaitu studi pustaka, analisis algoritma C4.5 dalam memprediksi masa studi mahasiswa, perancangan aplikasi berdasarkan hasil yang diperoleh pada tahap analisis, implementasi perangkat lunak, pengujian dan evaluasi hasil penelitian. Dari hasil penelitian tersebut, diperoleh kesimpulan bahwa aplikasi untuk memprediksi masa studi mahasiswa berhasil dirancang dan diimplementasikan menggunakan algoritma C4.5. Hasil penelitian juga menunjukkan bahwa jumlah data training berpengaruh terhadap presentase kecocokan [7].

Penelitian [3] bertujuan untuk meningkatkan keakuratan analisa masa studi mahasiswa menggunakan algoritma C4.5 dengan bantuan software RapidMiner 5. Parameter yang digunakan dalam penelitian ini antara lain NPM, nama mahasiswa, semester, prodi, jenjang pendidikan, jenis kelamin, IPK, dan jumlah SKS. Berdasarkan penelitian tersebut dapat diketahui bahwa penerapan algoritma C4.5 lebih akurat dalam memprediksi masa studi mahasiswa.

Referensi [4] menunjukkan bahwa penelitian yang dilakukan bertujuan untuk mencari hubungan data awal masuk mahasiswa dengan prestasi akademik. Adapun data awal yang digunakan antara lain jalur pendaftaran, jurusan, jenis kelamin, pekerjaan orang tua, asal daerah dan IPK. Metode yang digunakan dalam penelitian tersebut adalah data mining dengan algoritma C4.5 dan ID3. Hasil dari penelitian tersebut diperoleh rule yang memperlihatkan bahwa algoritma C4.5 lebih sederhana dibandingkan dengan algoritma ID3.

Penelitian lain seperti dalam [5] yang bertujuan untuk mengevaluasi kinerja akademik menggunakan teknik data mining algoritma Naïve Bayes Classifier (NBC). Data training yang digunakan pada penelitian tersebut yaitu 1115 sampel data untuk akademik mahasiwa yang sudah dinyatakan lulus, dengan pengklasifikasian 781 mhasiswa lulus tepat waktu dan 334 mahasiswa tidak lulus tepat waktu. Atribut data yang digunakan adalah NIM, nama, jenis kelamin, IPS 1, IPS2, IPS3, IPS 4, IPK Semester 4, dan keterangan Lulus. Sedangkan data testing yang digunakan yaitu data akademik mahasiswa angkatan 2013-2014. Hasil yang diperoleh dari penelitian tersebut yaitu implementasi data mining menggunakan algoritma Naïve Bayes Classifier dapat mengklasifikasikan ketepatan waktu lulus mahasiswa.

1.2 Landasan Teori

Klasifikasi [6] adalah sebuah model dalam data mining di mana *classifier* dikontruksi untuk memprediksi label berkategori, seperti "aman" atau "beresiko" untuk data aplikasi peminjaman uang; "ya" atau "tidak" untuk data marketing; atau "treatment A" atau "treatmen B" untuk data medis. Kategori tersebut dapat dipresentasikan dengan nilai yang sesuai dengan kebutuhan. Tahapan proses klasifikasi yang pertama adalah *Learning*, yaitu menganalisa data training menggunakan sebuah algoritma klasifikasi. Sedangkan tahap yang kedua adalah *Classification*, yaitu mengestimasi ketepatan dari *classification rules* (aturan klasifikasi) menggunakan data test.

Decission tree atau pohon keputusan merupakan sebuah metode klasifikasi yang digunakan untuk membantu menentukan pengambilan keputusan dalam pemecahan masalah [9]. Pohon keputusan berfungsi memperlihatkan faktor-faktor yang dapat mempengaruhi alternatif keputusan tersebut dan disertai dengan estimasi hasil akhir yang akan didapat jika mengambil alternatif keputusan tersebut [1]. Adapun tahapan pembentukan decission tree meliputi: memilih atribut sebagai root node (akar pohon keputusan) dengan cara menentukan nilai Gain dari semua variabel, menghitung nilai Gain dari semua variabel kemudian memilih variabel dengan nilai Gain terbesar yang menjadi root node level 0, dan melakukan langkah kedua tersebut untuk menentukan variabel pada node level 1, level 2, dan seterusnya sampai pohon terbentuk.

Algoritma ID3 atau *Iterative Dichotomiser 3* merupakan sebuah metode untuk membuat pohon keputusan yang dikembangkan oleh J.Ross Quinian. Algoritma ID3 dapat menghaslkan *decision tree* sederhana yang terbaik dan konsisten dari sekumpulan data [1]. Mekanisme kontruksi dilakukan secara top-down dengan memilih atribut mana yang harus diperiksa yang akan menjadi root dari decision tree. Pembentukan model keputusan menggunakan algoritma ID3, yang pertama kali dilakukan adalah mencari *Gain Informaion* dari masing-masing atribut untuk menentukan *root node*. Untuk mendapatkan Gain Information dari masing-masing atribut, perlu menghitung nilai *Entropy* dengan formula sebagai berikut:

$$Entropy(S) = \sum_{i=1}^{n} -p_i (\log_2 p_i) \quad (1)$$

Entropy (S) bernilai nol jika semua data berada dalam kelas yang sama, bernilai satu jika semua data positif dan negatif dalam S adalah sama. Gain(S,A) adalah Gain Information dari sebuah atibut A dari sebuah koleksi data S. Nilai gain dihitung dengan rumus sebagai berikut:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropy(S_i)$$
 (2)

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 [11]. Pada algoritma C4.5 hasilnya lebih akurat karena algoritma ini mampu menghitung data kosong. Proses pembentukan pohon keputusan dengan algoritma C4.5 mirip dengan proses pembentukan pohon

keputusan dengan algoritma ID3. Sebelum menghitung nilai Gain untuk menentukan root node, terlebih dahulu mencari nilai Entropy. Adapun rumus yang digunakan dalam menghitung Gain variabel (atribut) pada algoritma C4.5 sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropy(S_i)$$
 (3)

 $Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropy(S_i)$ (3) S merupakan himpunan kasus yang akan dibentuk pohon keputusan. Sedangkan A adalah atribut atau variabel yang terdapat dalam himpunan kasus dan n merupakan jumlah partisi atribut. Sedangkan nilai Entropy dihitung dengan formula sebagai berikut:

$$Entropy(S) = \sum_{i=1}^{n} -p_i (log_2 p_i)$$
 (4)

di mana n adalah jumlah nilai yang ada pada atribut target dan p_i merupakan jumlah sampel pada kelas i.

Setelah nilai Gain diketahui, selanjutnya perlu menghitung Gain Ratio. Dalam menghitung gain ratio perlu diketahui adanya suatu term baru yang disebut split information. Adapun formula untuk menghitung Split Information sebagai berikut:

Split information =
$$-\sum_{i=1}^{n} \frac{|S_{\nu}|}{|S|} \log 2 \frac{|S_{\nu}|}{|S|}$$
 (5)

Gain ratio dihitung dengan rumus:

$$Gain\ ratio = \frac{Gain}{Split\ information} \quad (6)$$

2. METODE PENELITIAN

Penelitian ini menggunakan data kelulusan mahasiswa sebuah program studi sebanyak 532 data record yang akan diolah menggunakan Microsoft Office Excel dan RStudio. Metode yang digunakan dalam penelitian ini menggunakan pendekatan klasifikasi, di mana penelitian ini digunakan untuk menentukan model pohon keputusan yang diperoleh dari analisa pengklasifikasian sebuah data. Model pohon keputusan tersebut diharapkan dapat membantu untuk melihat pola kelulusan mahasiswa. Proses pengklasifikasian menggunakan teknik data mining, yaitu serangkaian proses untuk menggali informasi yang selama ini tidak diketahui secara manual dari sebuah basis data yang berukuran besar sehingga sering disebut Knowledge Discovery Database (KDD) [1]. Ada 5 (lima) tahap rangkaian proses yang terdapat dalam Knowledge Discovery Database (KDD) yaitu: Seleksi Data, *Preprocessing*, Transformasi Data, Proses Mining, dan Evaluasi.

2.1 Seleksi Data

Tidak semua data yang terdapat dalam database digunakan dalam proses data mining. Tahap seleksi data merupakan tahap menentukan atau memilih atribut yang akan digunakan dalam proses data mining [8]. Struktur database yang tersedia adalah Nama, NIM, Jurusan, Jenjang, IPK, TOEFL, Jenis kelamin, Kota Asal, Tanggal Lahir, dan Agama, Nama Orang Tua, Status Beasiswa, dan data-data pribadi mahasiswa lainnya. Adapun data/atribut yang dipilih adalah IPK, Jenis Kelamin, Asal Daerah, TOEFL, Beasiswa, Tanggal Lahir, dan menambahkan satu atribut sebagai atribut kelas yaitu Masa Studi. Sedangkan atribut Jenjang, Jurusan, Status Kelulusan, dan Judul KTI tidak terlalu bermakna dalam proses data mining. Atribut Nama, NIM, Agama, Nama Orang Tua, No. HP, dan data pribadi lainnya juga tidak bermakna dalam proses data mining dan bersifat rahasia.

2.2 Preprocessing

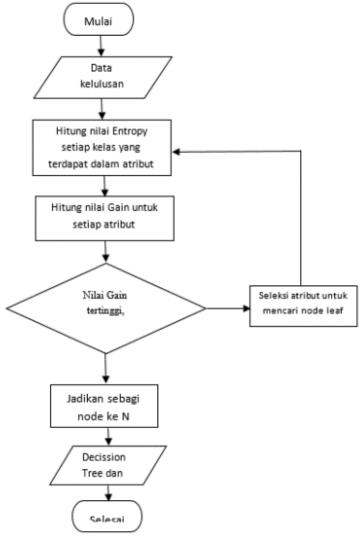
Proses cleaning pada data yang akan dilakukan data mining perlu dilakukan untuk medapatkan data yang relevan. Yang dilakukan pada cleaning data antara lain memeriksa, memperbaiki, dan membuang data yang tidak relevan. Selain itu dilakukan pula proses enrichment, yaitu proses memperkaya data yang sudah ada dengan informasi lain yang relevan agar bisa diproses data mining. Proses *cleaning* pada penelitian ini menggunakan metode data mean interval, yaitu mengisi data yang kosong dengan rata-rata data yang dibagi menjadi tiga interval (rata-rata awal, rata-rata tengah, dan rata-rata akhir).

2.3 Transformasi Data

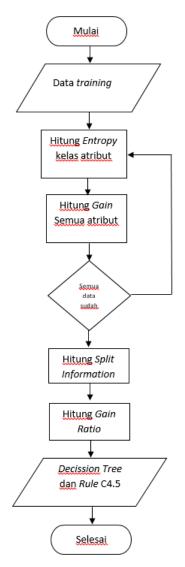
Pada tahap ini dilakukan transformasi data yaitu *coding* untuk setiap atribut agar mudah dilakukan proses mining. Namun pada penelitian ini proses *coding* sudah dilakukan pada tahap data *cleaning* untuk mempermudah proses pada tahap tersebut. Ada beberapa atribut yang perlu dilakukan pengkategorian sebelum dilakukan transformasi seperti IPk, JK, dan Asal Daerah.

2.4 Data Mining

Proses data mining pada penelitian ini menggunakan dua algoritma yaitu algoritma ID3 dan Algoritma C4.5. Pada dasarnya kedua algoritma tersebut merupakan algoritma yang sama digunakan untuk pembentukan model *decision tree* (pohon keputusan). Cara kerja algoritma ID3 dan algoritma C4.5 ditunjukkan gambar 1 dan gambar 2 berikut.



Gambar 1. Flowchart Algoritma ID3



Gambar 2. Flowchart Algoritma C4.5

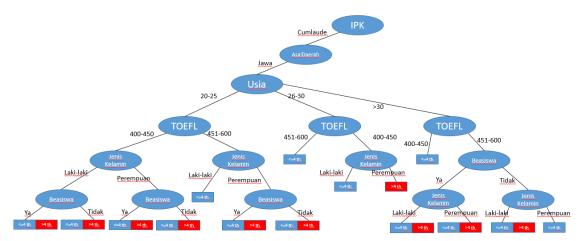
2.5 Evaluasi

Evaluasi dilakukan untuk mengetahui apakah pola atau informasi yang diketahui bertentangan dengan fakta yang ada sebelumnya. Pada tahap evaluasi akan dilakukan pengujian terhadap data testing. Evaluasi dilakukan dengan melihat informasi dari *Confussion Matrix* untuk mengetahui tingkat keakuratan prediksi yang diperoleh dari algoritma yang digunakan.

3. HASIL DAN PEMBAHASAN

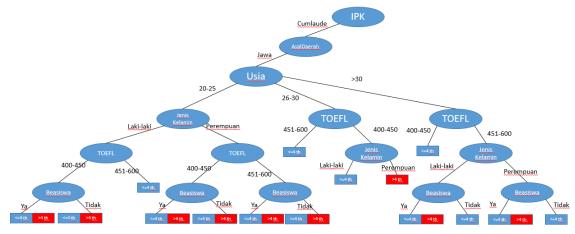
3.1 Implementasi Algoritma dengan Microsoft Office Excel

Implementasi algoritma ID3 dan C4.5 pada penelitian ini menggunakan dataset kelulusan mahasiswa sebanyak 532 data record dengan data training sebanyak 372 data record dan data testing sebanyak 160 data record. Hasil implementasi algoritma ID3dan algoritma C4.5 terhadap data training diperoleh model pohon keputusan seperti pada gambar 3 dan gambar 4.



Gambar 3. Potongan Pohon keputusan ID3

Gambar 3 di atas merupakan salah satu model pohon keputusan yang dihasilkan dari implementasi algoritma ID3 terhadap data training sebanyak 372, diperoleh *root node*-nya adalah atribut IPK. Dari gambar tersebut diperoleh beberapa rule pengklasifikasian ketepatan lulus mahasiswa. Salah satu contoh *rule* sebagai berikut, jika IPK Cumlaude dan asal daerah jawa, maka klasifikasi kelas ketepatan lulus mahasiswa masih dikelompokkan berdasarkan usia, skor TOEFL, dan Jenis Kelamin.

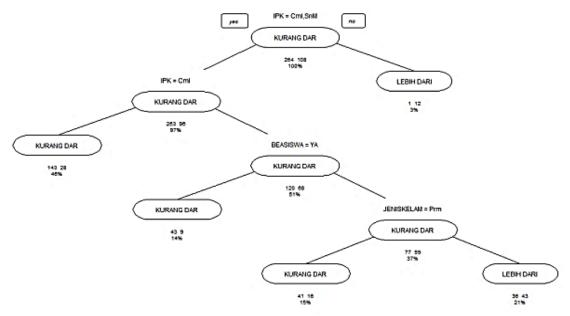


Gambar 4. Potongan Pohon Keputusan C4.5

Sedangkan gambar 4 merupakan bagian dari model pohon keputusan yang dihasilkan dari implementasi algoritma C4.5 terhadap data training dengan root node IPK. Pohon keputusan yang diperoleh berdasarkan perhitungan manual terhadap data training menggunakan algoritma ID3 dan C4.5 terlihat hampir mirip. Perbedaannya hanya terletak pada urutan node cabang yang terbentuk.

3.2 Implementasi Algoritma ID3 dan C4.5 dengan RStudio

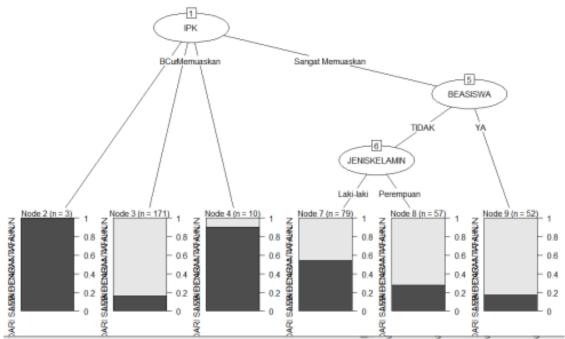
Implementasi algoritma ID3 dengan RStudio pada penelitian ini menggunakan komposisi data 70:30. Dimana diperoleh data training sebanyak 372 record dan data testing sebanyak 160 record. Hasil model pohon keputusan ditunjukkan pada gambar 5 berikut.



Gambar 5. Model Pohon Keputusan ID3

Berdasarkan gambar 5 diperoleh atribut IPK terpilih sebagai *root node* yang berarti bahwa atribut IPK merupakan atribut yang paling berpengaruh dalam menentukan ketepatan kelulusan mahasiswa. Selanjutnya, untuk melihat keakuratan prediksi dari algoritma ID3 terhadap ketepatan kelulusan mahasiswa, dapat dilihat pada nilai akurasi yang diperoleh dari *Confussion Matrix* yaitu sebesar 75,81 %.

Dengan menggunakan komposisi data yang sama, model pohon keputusan yang diperoleh dari implementasi algoritma C4.5 terhadap data training adalah sebagai berikut.



Gambar 6. Model Pohon Keputusan C4.5

Berdasarkan gambar 6, diketahui model pohon keputusan yang dihasilkan dari implementasi algoritma C4.5 terhadap data kelulusan mahasiswa memiliki *root node* dengan atribut IPK dan atribut yang terlibat sebagai node cabang yaitu atribut Beasiswa dan Jenis Kelamin. Hal ini menunjukkan bahwa atribut IPK paling berpengaruh dalam menentukan prediksi ketepatan kelulusan mahasiswa. Sedangkan untuk melihat keakuratan prediksi ketepatan kelulusan mahasiswa menggunakan algoritma C4.5 dapat melihat nilai akurasi pada tabel *Confussion Matrix* yang diperoleh dari hasil implementasi di program R. Nilai akurasi yang diperoleh yaitu sebesar 75,81%.

Nilai akurasi algoritma C4.5 sebesar 75,81% ternyata sama besarnya dengan nilai akurasi algoritma ID3 untuk data training 70%. Kemudian dilakukan pengujian terhadap data testing sebanyak 210 record atau data testing 30% untuk masing-masing algoritma dan diperoleh nilai akurasi menggunakan algoritma ID3 sebesar 78,75%. Sedangkan nilai akurasi pada data testing yang diimplementasikan algoritma C4.5 yaitu sebesar 81,88%. Selanjutnya melakukan uji coba implementasi terhadap dataset dengan komposisi 90:10 dan 80:20 sebagai pembanding. Kemudian diperoleh nilai akurasi seperti yang ditunjukkan pada tabel 1.

Tabel I. Perbandingan Ni	ilai Akurasi Model

Metode	Pengujian Data Training (%)		Pengujian Data Testing (%)			
	90	80	70	30	20	10
ID3	76,78	76,94	75,81	78,75	74,77	72,22
C4.5	77,62	76,94	75,81	81,88	77,57	77,78

4. KESIMPULAN

Berdasarkan hasil penelitian klasifikasi menggunakan algoritma ID3 dan algoritma C4.5 diperoleh beberapa kesimpulan sebagai berikut.

Hasil implementasi algoritma ID3 dan C4.5 terhadap dataset dengan komposisi data 90:10, 80:20, dan 70:30 menunjukkan bahwa model pohon keputusan ID3 memiliki lebih banyak atribut yang terlibat dibandingakan model pohon keputusan yang dihasilkan menggunakan algoritma C4.5 yang lebih sederhana dan sedikit atribut yang terlibat. Hasil pengujian terhadap dataset dengan 3 komposisi berbeda diperoleh nilai akurasi terbaik yaitu pada data testing 30% sebesar 78,75% pada algoritma ID3 dan sebesar 81,88% pada algoritma C4.5. Secara keseluruhan dari hasil implementasi, Algoritma C4.5 memiliki tingkat akurasi yang lebih tinggi dibandingkan algoritma ID3. Dengan demikian dapat dikatakan bahwa algoritma C4.5 adalah algoritma terbaik untuk menentukan prediksi ketepatan lulus mahasiswa.

5. SARAN

Pengujian terhadap kinerja algoritma ID3 dan C4.5 bisa dilakukan lebih lanjut dengan melihat nilai presisi dan recall. Dataset yang digunakan untuk menguji tingkat akurasi algoritma ID3 dan C4.5 bisa ditambah dengan atribut lain yang sesuai dengan permasalahan dan dengan jumlah data yang lebih banyak.

UCAPAN TERIMA KASIH

Segala puji bagi Allah SWT yang telah memberikan karunia, hidayah, dan kemudahan bagi penulis sehingga jurnal ini dapat disusun dengan baik. Ucapan terima kasih disampaikan kepada segenap keluarga besar dosen Fakultas Teknologi Informasi Universitas Stikubank (UNISBANK) Semarang yang telah memberikan arahan dan bimbingan atas penelitian ini. Terima kasih disampaikan pula kepada petugas akademik di Fakultas X Universitas Y yang telah membantu memberikan data sebagai bahan utama penelitian ini. Tak lupa ucapan terima kasih untuk kedua orang tua yang selalu memberikan dukungan moril dan doa sehingga penelitian ini berjalan dengan lancar.

DAFTAR PUSTAKA

- [1] Vulandari, R.T. 2017, *Data Mining Teori dan Aplikasi Rapidminer*, Gava Media, Yogyakarta.
- [2] Ginting, S.L.Br., Zarman, W., Hamidah, I. 2014, Analisis dan Penerapan Algoritma C4.5 dalam Data Mining untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik, Prosiding Seminar Nasional Aplikasi Sains & Teknolgi (SNAST), Yogyakarta, 15 November, A-263 A-272.
- [3] Haryati, S., Sudarsono, A., Suryana, E. 2015, Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu), Jurnal Media Infotama, 2, 11, pp.130-138.
- [4] Arista, Novyana, Martaleli Bettiza, *Hendra Kurniawan, Penerapan Algoritma ID3 dan C45 Dalam Menemukan Hubungan Data Awal Masuk Mahasiswa Dengan Prestasi Akdemik*, http://jurnal.umrah.ac.id/wp-content/uploads/gravity_forms/1-ec61c9cb232a03a96d0947c6478e525e/2016/08/jurnal-upload-fix.pdf.
- [5] Mustafa, M.S., Ramadhan, M.R., Thenata, A.P. 2017, Implementasi Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier, Cite c Journal, 2, 4, pp.151-162.
- [6] Han, J., & Kamber, M. 2006, *Data Mining: Concept and Technique Second Edition, Data Mining Teori dan Aplikasi Rapidminer*, Gava Media, Yogyakarta.
- [7] Jananto, A., (1), Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa.

 Dinamik, 18(1),https://www.unisbank.ac.id/ojs/index.php/fti1/article/view/1669
- [8] Kusrini dan Emha Taufiq Luthfi. 2009, Algoritma Data Mining, Andi Offset.
- [9] Mashlahah, Susi. 2013, *Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5*, Undergraduate Thesis, Universitas Islam Negeri Maulana Malik Ibrahim.
- [10] Rohman, Abdul, Anief Rufiyanto. 2019, Penerapan Algoritma Decission Tree ID3 Untuk Prediksi Kelulusan Mahasiswa Jenjang Pendidikan D3 Fakultas Teknik Universitas Pandanaran, Jurnal NeoTeknika, Vol. 5, No.2A.

[11] Yahya, N., & Jananto, A. 2019, Komparasi Kinerja Algoritma C.45 dan Naive Bayes Untuk Prediksi Kegiatan Penerimaanmahasiswa Baru (Studi Kasus: Universitas Stikubank Semarang), Proceeding SENDI_U, 221-228, https://www.unisbank.ac.id/ojs/index.php/sendi_u/article/view/7389.