

Prediksi Masa Studi Mahasiswa Unama Jambi Menggunakan Metode Algoritma C4.5

Meyer Mega¹, Jasmir²

*Pascasarjana, Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi
Jl. Jend. Sudirman Thehok-Jambi Telp: 0741-35096 Fax : 35093
Email: meyermega170598@gmail.com¹, ijay_jasmir@yahoo.com²*

Abstract

Every year the number of students at The University of Dynamics of the Nation jambi is always increasing but the students who graduate are different from the number of students who enter. Therefore, the author conducts a data mining analysis on student data so that it can be used by academic supervisors to find out the graduation status of students and as a warning so that students can graduate on time so as to reduce the number of delays in graduation. The author uses student data in 2016 and 2017 as training data and 2020 as testing data as many as 120 training data and 109 testing data and has carried out the data cleaning and attribute selection process using the forward selection method. In conducting the analysis, the author used the Tools Weka tool. The author uses the C4.5 algorithm method with 12 attributes, but there are 4 attributes that are most influential after selecting data using the forward selection method on WEKA. In this case, the author uses 4 test options, namely 5 Fold Cross Validation, 10 Fold Cross Validation, 70% Percentage Split, and 80% Percentage Split. The C4.5 Algorithm method produces the largest accuracy value in the training data, namely 5 Fold Cross Validation by 92.5% and in the testing data by 100%.

Keywords : Data Mining, Classification, Prediction, C4.5 Algorithm

Abstrak

Setiap tahunnya jumlah mahasiswa pada Universitas Dinamika Bangsa Jambi selalu bertambah tetapi mahasiswa yang lulus berbeda dengan jumlah mahasiswa yang masuk. Oleh karena itu, penulis melakukan analisis data mining pada data mahasiswa tersebut agar dapat dimanfaatkan oleh pembimbing akademik untuk mengetahui status kelulusan mahasiswa maupun sebagai peringatan agar mahasiswa bisa lulus dengan tepat waktu sehingga dapat menekan angka keterlambatan kelulusan. Penulis menggunakan data mahasiswa tahun 2016 dan 2017 sebagai data training dan 2020 sebagai data testing sebanyak 120 data training dan 109 data testing serta telah dilakukan proses cleaning data dan seleksi atribut menggunakan metode forward selection. Dalam melakukan analisis penulis menggunakan alat bantu Tools Weka. Penulis menggunakan metode algoritma C4.5 dengan 12 atribut, tetapi ada 4 atribut yang paling berpengaruh setelah melakukan seleksi data menggunakan metode forward selection pada WEKA. Dalam hal ini penulis menggunakan 4 options test, yaitu 5 Fold Cross Validation, 10 Fold Cross Validation, 70% Percentage Split, dan 80% Percentage Split. Metode Algoritma C4.5 menghasilkan nilai akurasi terbesar pada data training yaitu 5 Fold Cross Validation sebesar 92,5% dan pada data testing sebesar 100%.

Kata kunci: Data Mining, Klasifikasi, Prediksi, Algoritma C4.5.

© 2023 Jurnal MAGISTER SISTEM INFORMASI.

1. Pendahuluan

Universitas Dinamika Bangsa (UNAMA) Jambi merupakan Perguruan Tinggi yang bernaung di bawah Yayasan Dinamika Bangsa. Pembinaan UNAMA secara fungsional dilakukan oleh Direktorat Jendral Pendidikan Tinggi melalui LLDIKTI Wilayah X. Kampus Universitas Dinamika Bangsa memiliki 2 (dua)

Fakultas serta 1 (Satu) Program Pasca Sarjana dengan total 8 (delapan) program studi yaitu ada fakultas komputer yang memiliki 3 program studi diantaranya: Teknik Informatika (S1), Sistem Informasi (S1), Sistem Komputer (S1), Manajemen Informatika (D3), dan Manajemen akuntansi (D3). Lalu ada Fakultas Ilmu Manajemen dan Bisnis yang memiliki 2 program studi diantaranya: Manajemen (S1) dan Kewirausahaan (S1). Dan yang terakhir ada Fakultas Pasca Sarjana dengan program studi Sistem Informasi (S2) [1]. Dengan perkembangan Sekolah Tinggi Ilmu Komputer Dinamika Bangsa (STIKOM) menjadi Universitas Dinamika Bangsa (UNAMA) Jambi serta bertambahnya fakultas dan program studi yang beraneka ragam, ini tentulah akan menarik bagi calon mahasiswa yang ingin mendaftar di UNAMA.

Pemanfaatan beberapa teknik data mining dalam memanfaatkan tumpukan data yang terdapat pada perguruan tinggi bisa digunakan untuk mendapatkan informasi demi untuk menunjang kegiatan manajemen dan kegiatan operasional harian serta bermanfaat dalam mengambil kebijakan atau keputusan yang strategis.[2]

Berdasarkan dari permasalahan diatas serta beberapa penelitian sejenis yang dapat menguatkan penelitian ini, maka peneliti tertarik untuk melakukan proses data mining dengan metode algoritma C4.5 dikarenakan metode algoritma C4.5 memiliki hasil akurasi yang lebih besar dari pada menggunakan metode lain terutama dalam memprediksi masa studi mahasiswa dengan menggunakan data mahasiswa dan dasar data nilai IPK dari semester sebelumnya.

2. Tinjauan Pustaka

Dari penelitian yang penulis lakukan di dalam pembuatan tesis, penulis melakukan perbandingan dengan tesis ataupun jurnal yang pernah di buat penulis lainnya yang mengangkat masalah data mining untuk mempresiksi kinerja siswa meliputi:

1. Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu) oleh Siska Haryati, Aji Sudarsono, dan Eko Suryana pada Universitas Dehasen Bengkulu yaitu memaparkan bahwa ketatnya persaingan mahasiswa dalam mendapatkan lapangan pekerjaan menuntut ilmu di perguruan tinggi menghasilkan sarjana yang berkualitas dan memiliki daya saing. Untuk itu, setiap perguruan tinggi selalu melakukan evaluasi performansi mahasiswa. Peneliti menerapkan metode algoritma C4.5 dengan menggunakan sepuluh parameter yaitu: NPM, Nama Mahasiswa, Semester, Prodi, Jenjang Pendidikan, Jenis Kelamin, IPK, dan Jumlah SKS. Kemudian untuk pengolahan data penulis menggunakan aplikasi pada Rapid Miner 5. Dari hasil penelitian terbukti bahwa algoritma C4.5 lebih akurat dibandingkan analisa yang dilakukan oleh analisis mahasiswa. Penulis juga menggunakan metode seleksi dalam menentukan atribut dalam menganalisis serta melakukan perbandingan antara metode algoritma C4.5 dengan metode lainnya.
2. Penerapan Algoritma C.45 Dalam Memprediksi Kelulusan Tepat Waktu Pada Perguruan Tinggi (Studi Kasus : STMIK Royal Kisaran) oleh Nanda Dimas Prayoga pada STMIK ROYAL Kisaran yaitu memaparkan bahwa banyak mahasiswa tidak lulus tepat waktu yang tidak sesuai standar masa studi dikarenakan mahasiswa menganggap bahwa ketika kuliah hanya ingin mengambil ijazah sarjana saja, dan juga mahasiswa tidak lulus tepat waktu diakibatkan faktor lainnya seperti biaya, sibuk bekerja, sibuk menjadi aktivis bahkan punya masalah dalam kampus. Metode yang peneliti gunakan adalah metode Klasifikasi dengan Algoritma C.45 dengan tools yang digunakan di dalam pengolahan data adalah Rapidminer 5, atribut yang di gunakan yaitu IPK, prestasi, etika, sks, dan kelayakan. Hasil dari penelitian ini berupa aturan yang akan menjadi landasan mahasiswa kelulusan tepat waktu yang dikatakan lulus tepat waktu atau tidak. Berdasarkan hasil penelitian tersebut, menjadi kesimpulan yang dapat membantu mahasiswa maupun akademik untuk mengevaluasi / prediksi sejak awal setiap menjalani kegiatan akademik. Dari segi persamaan penelitian diatas dengan penulis yaitu pada objeknya yang membahas tentang memprediksi masa studi mahasiswa, Metode yang di gunakan menggunakan metode algoritma C4.5, serta dari sisi tujuan juga memiliki kesamaan guna melihat hasil perkembangan mahasiswa untuk membantu meningkatkan kualitas kampus. Namun memiliki beberapa perbedaan antara lain, terdapat perbedaan pada penggunaan aplikasi rapid miner sedangkan penulis menggunakan weka serta dalam atribut dari hasil kelulusan yaitu : lulus dan tidak lulus. Penulis juga menggunakan metode seleksi dalam menentukan atribut dalam menganalisis serta melakukan perbandingan antara metode algoritma C4.5 dengan metode lainnya.
3. Implementasi Data Mining Dengan Algoritma Decision Tree C4.5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandanaran oleh Abdul Rohman, dan Anief Rufiyanto di Universitas Pandanaran yaitu memaparkan bahwa untuk mempertahankan ataupun mempertahankan kinerja

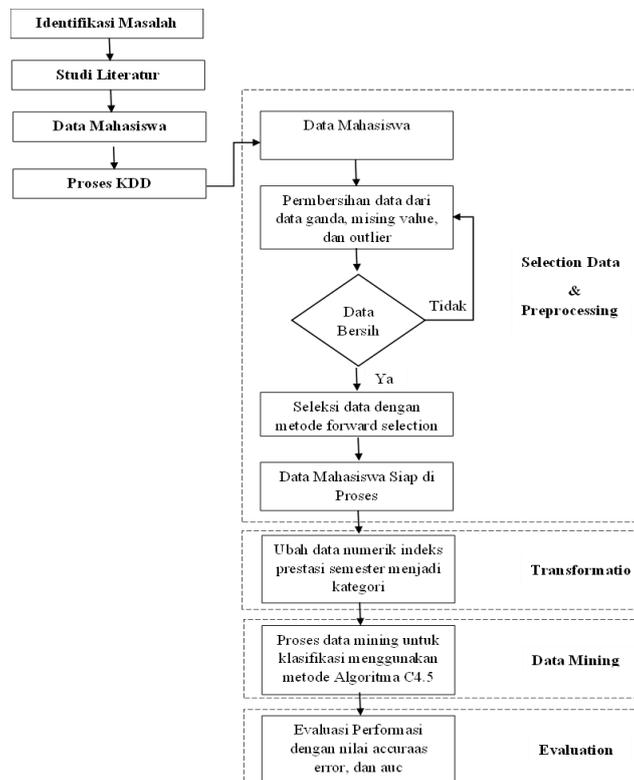
mahasiswa di perlukannya suatu program strategi dengan memprediksi prestasi akademik mahasiswa yang dilakukan oleh penyelenggara pendidikan selama masa studi di perguruan tinggi. Dalam penelitian ini menggunakan algoritma klasifikasi data mining Decision Tree C4.5 serta data yang digunakan pada penelitian ini adalah data kelulusan mahasiswa di Universitas Pandanaran dengan jenjang pendidikan D3 di Fakultas Teknik dari program studi Teknik Sipil, Teknik Mesin, Teknik Elektronika, Teknik Lingkungan, dan Teknik Kimia. Jumlah data yang diolah yaitu 235 mahasiswa, yaitu terdiri dari: Teknik Sipil 53 mahasiswa, Teknik Mesin 95 mahasiswa, Teknik Elektronika 45 mahasiswa, Teknik Lingkungan 25 mahasiswa, Teknik Kimia 17 mahasiswa. Pada penelitian ini parameter yang digunakan adalah Nama Jurusan, Usia, Jenis Kelamin, Status Pekerjaan, Indeks Prestasi Semester Satu Sampai Dengan Indeks Prestasi Semester Empat. Dari data tersebut terdapat 151 mahasiswa yang lulus tepat waktu dan 84 mahasiswa yang terlambat. Hasil dari penelitian ini menghasilkan 10 rule dengan nilai akurasi 65,98% dengan nilai AUC 0,874, dan dapat dikategorikan sebagai klasifikasi data yang baik. Maka hasil tersebut sangat penting untuk dijadikan pengambilan keputusan dalam lembaga. Penulis juga menggunakan metode seleksi dalam menentukan atribut dalam menganalisis serta melakukan perbandingan antara metode algoritma C4.5 dengan metode lainnya.

Dari segi persamaan penelitian diatas dengan penulis yaitu pada objeknya yang membahas tentang memprediksi masa studi mahasiswa, Metode yang digunakan menggunakan metode algoritma C4.5 serta dari sisi tujuan juga memiliki kesamaan guna melihat hasil perkembangan mahasiswa untuk membantu meningkatkan kualitas kampus. Namun memiliki perbedaan antara lain, peneliti merancang sistem yang dapat melakukan prediksi kelulusan tepat waktu dengan menggunakan salah satu model data mining yaitu algoritma C4.5 sedangkan penulis menggunakan aplikasi weka untuk menghasilkan suatu hasil analisis. Peneliti menggunakan satu metode saja tanpa melakukan perbandingan dengan metode lainnya, sedangkan penulis melakukan perbandingan dengan metode lainnya.

3. Metodologi

3.1 Alur Penelitian

Untuk menyusun penelitian ini tentunya diperlukan alur penelitian berupa kerangka kerja, yang mana kerangka kerja ini akan dijadikan pedoman urutan proses demi proses jalannya penelitian. Berikut ini kerangka kerja penelitian yang akan dilakukan:



Gambar 2. Alur Penelitian

Berdasarkan kerangka kerja penelitian yang telah digambarkan 2, maka dapat diuraikan pembahasan masing-masing tahap dalam penelitian adalah sebagai berikut:

1. Identifikasi Masalah

Pada tahap ini penulis melakukan identifikasi masalah terhadap hal apa saja yang menjadi permasalahan data mining mengenai masa studi pada data mahasiswa di UNAMA Jambi. Identifikasi ini bertujuan untuk menentukan rencana kerja serta menentukan data apa saja yang akan dibutuhkan dalam penelitian ini.

2. Studi Literatur

Pada tahapan ini penulis menambah wawasan guna mendapatkan sebuah topik yang layak diangkat sebagai sebuah penelitian dengan mempelajari dan memahami teori dan konsep dimana penulis banyak melakukan penelitian pada buku, jurnal, paper, dan berbagai sumber diantaranya Perpustakaan Universitas Dinamika Bangsa Jambi.

3. Data Mahasiswa

Pada tahap ini penulis melakukan pengumpulan data mahasiswa dan informasi yang dibutuhkan dalam penelitian. Data mahasiswa yang akan digunakan merupakan dari angkatan 16, 17 dan 20. Pengumpulan data ini dilakukan dengan beberapa metode yaitu :

- a. Pengamatan Langsung (Observation)
Penelitian dengan metode observation ini dilakukan dengan melakukan pengamatan langsung terhadap objek yang akan diteliti yang bertujuan untuk memperkuat data, mengetahui serta mendapatkan informasi secara langsung mengenai tentang masa studi mahasiswa UNAMA Jambi.
- b. Wawancara (Interview)
Metode pengumpulan data yang dilakukan oleh penulis secara tatap muka antara penulis dengan narasumber.

4. Proses KDD

Pada tahap ini peneliti melakukan proses KDD (Knowledge Discovery in Database Process) yang dimana secara garis besar dapat dijelaskan sebagai berikut [5]:

- a. Data Selection dan Preprocessing
Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk Knowledge Discovery in Database (KDD), seperti data atau informasi eksternal lainnya yang diperlukan. Setelah memperoleh data bersih maka melakukan seleksi data lagi dengan menggunakan metode forward selection dengan aplikasi WEKA.
- b. Transformation
Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam Knowledge Discovery in Database (KDD) merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data
- c. Data Mining
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Metode yang akan di gunakan pada penelitian ini adalah algoritma C4.5 dengan menggunakan aplikasi weka.
- d. Evaluation
Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses Knowledge Discovery in Database (KDD) yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya

4. Hasil dan Pembahasan

4.1 Profil Data Penelitian

Data yang digunakan pada penelitian ini menggunakan data mahasiswa dari Universitas Dinamika Bangsa. Data mahasiswa yang akan di teliti, menggunakan data mahasiswa dari tahun 2016, 2017 dan 2020. Mahasiswa tahun 2016 dan 2017 akan menjadi data training dan mahasiswa tahun 2020 akan menjadi data

testing. Data yang di peroleh melalui wawancara kepada pihak BAAK dari kampus universitas dinamika bangsa jambi. Data mahasiswa yang akan digunakan hanya mahasiswa yang memiliki prodi Teknik informatika saja. Data mahasiswa yang di gunakan sebanyak 229 data yang sudah melalui tahap preprocessing dan tahap seleksi.

4.2 Selection Data dan Preprocessing

Data Training di sini berupa kumpulan data mahasiswa/I dari angkatan 2016 dan 2017, yang memiliki jumlah atribut ada 12 yaitu NIM, Nama, Prodi, Angkatan, Jenis Kelamin (JK), Status Kuliah, Asal Sekolah, Shift Kuliah, Status Kerja, IPK Terakhir dan Keterangan Lulus.

#	NIM	Nama	Prodi	Angkatan	JK	S.Kuliah	Tahun Lulus	Asal Sekolah	Shift	Status Kerja	IPK	Ket. Lulus
1	8020160003	Diana Novita S	TI	2016	P	Lulus	2020	SMK	Malam	Sudah Bekerja	3.48	Tepat Waktu
2	8020160008	Indrawan Jam	TI	2016	L	Lulus	2020	SMK	Malam	Belum Bekerja	3.81	Tepat Waktu
3	8020160011	Desy Ratna Sar	TI	2016	P	Lulus	2020	SMA	Pagi	Belum Bekerja	3.62	Tepat Waktu
4	8020160012	Yoga Tanoko	TI	2016	L	Lulus	2020	SMA	Malam	Belum Bekerja	3.94	Tepat Waktu
5	8020160014	Kevin Rahmat	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.73	Tepat Waktu
6	8020160015	Akmal Rustand	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.22	Tepat Waktu
7	8020160016	Angga Trisnan	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	2.97	Tepat Waktu
8	8020160017	Muhtadi Abdill	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.67	Tepat Waktu
9	8020160018	Shofan Rizki Sa	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.45	Tepat Waktu
10	8020160020	Feliks Lourens	TI	2016	L	Lulus	2020	SMA	Malam	Sudah Bekerja	3.92	Tepat Waktu
11	8020160021	Kevin Wenard	TI	2016	L	Lulus	2020	SMA	Malam	Belum Bekerja	3.82	Tepat Waktu
12	8020160022	Andrean Oscar	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.62	Tepat Waktu
13	8020160023	Wahyudin	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.82	Tepat Waktu
14	8020160024	Ambar Wahyu	TI	2016	L	Lulus	2020	SMA	Pagi	Belum Bekerja	3.35	Tepat Waktu
15	8020160026	Juni Ahmad	TI	2016	P	Lulus	2020	SMA	Pagi	Belum Bekerja	3.18	Tepat Waktu
16	8020160027	Martha P	TI	2016	P	Lulus	2020	SMK	Malam	Belum Bekerja	3.33	Tepat Waktu
17	8020160028	Febry Valentin	TI	2016	L	Lulus	2020	SMA	Malam	Belum Bekerja	3.88	Tepat Waktu
18	8020160030	Ricky Fernand	TI	2016	L	Lulus	2020	SMK	Malam	Belum Bekerja	3.66	Tepat Waktu
19	8020160031	Ahmad Suhard	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.22	Tepat Waktu
20	8020160033	Surya Krisman	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.75	Tepat Waktu
21	8020160034	Sri Cahyoning	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.94	Tepat Waktu
22	8020160035	Muklas Bintar	TI	2016	L	Lulus	2020	SMK	Pagi	Belum Bekerja	3.56	Tepat Waktu

Gambar 3. Data Training

Sedangkan Data Testing di sini berupa kumpulan data mahasiswa/I dari angkatan 2020, yang memiliki jumlah atribut ada 12 yaitu NIM, Nama, Prodi, Angkatan, Jenis Kelamin (JK), Status Kuliah, Asal Sekolah, Shift Kuliah, Status Kerja, IPK Terakhir dan Keterangan Lulus. Pada data testing keterangan lulus masih kosong, yang dimana kita akan mecaari keterangan kelulusan mahasiswa angkatan 2020 dengan melakukan perbandingan dengan data Training.

#	NIM	Nama	Prodi	Angkatan	JK	S.Kuliah	Asal Sekolah	Shift	Status Kerja	IPK	Keterangan Lulus
1	8020200001	Khairunnisa Putri	TI	2020	P	Aktif	SMK	Pagi	Belum Bekerja	3.84	?
2	8020200003	Rama Rifki Fadli	TI	2020	L	Aktif	SMA	Malam	Belum Bekerja	3.17	?
3	8020200004	Laurent Bellavinsca	TI	2020	P	Aktif	SMA	Malam	Belum Bekerja	3.86	?
4	8020200005	Willie Prakarsa Hartanto	TI	2020	L	Aktif	SMA	Malam	Belum Bekerja	3.80	?
5	8020200006	Fathoni Athallah	TI	2020	L	Aktif	SMK	Pagi	Belum Bekerja	3.81	?
6	8020200007	Cinta Anindya Putri Abdoellah	TI	2020	P	Aktif	SMA	Pagi	Belum Bekerja	3.96	?
7	8020200008	Jenia Valiant Tori	TI	2020	P	Aktif	SMA	Malam	Sudah Bekerja	3.54	?
8	8020200009	Elbert Devino Henata	TI	2020	L	Aktif	SMK	Malam	Belum Bekerja	3.85	?
9	8020200010	Tiko Nur Annisa	TI	2020	P	Aktif	SMA	Pagi	Belum Bekerja	3.96	?
10	8020200011	M. Raihan Mufadhhal	TI	2020	L	Aktif	MA	Pagi	Belum Bekerja	3.92	?
11	8020200012	Adryan	TI	2020	L	Aktif	SMK	Pagi	Belum Bekerja	3.93	?
12	8020200014	Elvi Liza Oktavia	TI	2020	P	Aktif	SMK	Pagi	Belum Bekerja	3.87	?
13	8020200015	Eric Wardata	TI	2020	L	Aktif	SMK	Malam	Belum Bekerja	3.80	?
14	8020200016	Vincent Harianto	TI	2020	L	Aktif	SMK	Pagi	Belum Bekerja	3.56	?
15	8020200017	Defry Nazrian	TI	2020	L	Aktif	SMA	Pagi	Belum Bekerja	3.84	?
16	8020200018	Adrian Pirtama	TI	2020	L	Aktif	SMK	Pagi	Belum Bekerja	3.56	?
17	8020200019	Drenda Scheber	TI	2020	L	Aktif	SMA	Pagi	Belum Bekerja	3.69	?
18	8020200020	Fiki Saputra	TI	2020	L	Aktif	SMK	Pagi	Belum Bekerja	3.57	?
19	8020200021	David Liem	TI	2020	L	Aktif	SMA	Pagi	Belum Bekerja	3.74	?
20	8020200022	Mutoharoh	TI	2020	P	Aktif	SMA	Pagi	Belum Bekerja	3.76	?

Gambar 4. Data Testing

4.2.2.1 Pembersihan Data

Pembersihan data bekerja untuk "membersihkan" data dengan mengisi nilai yang hilang, menghaluskan data yang terdapat noise, mengidentifikasi atau menghapus outlier, dan menyelesaikan inkonsistensi. Selain itu, data kotor dapat menyebabkan kebingungan pada prosedur mining, sehingga menghasilkan keluaran yang tidak dapat dipercaya. Pembersihan data menggunakan Microsoft excel.

Pada data mahasiswa, yang digunakan hanya data mahasiswa yang telah lulus bagi data training, dan mahasiswa yang aktif untuk data testing. Sehingga data mahasiswa yang keluar, cuti, dan non-aktif tidak digunakan dalam perhitungan data mining. Serta jika ada terdapat “koma” pada data mahasiswa itu di hilangkan atau diganti dengan “space” agar dapat diterima oleh aplikasi Weka tersebut.

#	NIM	Nama	Prodi	Angkatan	JK	S.Kuliah	Tgl. Lulus	Asal Sekol	Shift	Status Ker	IPK
1	802016000	Diana Nov	TI	2016	P	Lulus	2020-03-0	SMKS Ung	Malam	Sudah Bel	3.48
2	802016000	Indrawan	TI	2016	L	Lulus	2020-02-2	SMKS Ung	Malam	Belum Bel	3.81
3	802016001	Randitiya	TI	2016	L	Keluar		SMKN 2 K	Pagi	Belum Bel	0.6
4	802016001	Desy Ratn	TI	2016	P	Lulus	2020-03-0	SMA Nege	Pagi	Belum Bel	3.62
5	802016001	Yoga Tano	TI	2016	L	Lulus	2020-03-0	SMAS BIN	Malam	Belum Bel	3.94
6	802016001	Kevin Rah	TI	2016	L	Lulus	2020-08-2	SMK Ungg	Pagi	Belum Bel	3.73
7	802016001	Akmal Rus	TI	2016	L	Lulus	2020-09-0	SMK Nege	Pagi	Belum Bel	3.22
8	802016001	Angga Tris	TI	2016	L	Lulus	2020-08-3	SMK Nege	Pagi	Belum Bel	2.97
9	802016001	Muhtadi A	TI	2016	L	Lulus	2020-02-2	SMK Nege	Pagi	Belum Bel	3.67
10	802016001	Shofan Ri	TI	2016	L	Lulus	2020-08-2	SMKN 2 K	Pagi	Belum Bel	3.45
11	802016001	Rey Hari C	TI	2016	L	Keluar		SMAN 1 K	Pagi	Belum Bel	3.26
12	802016002	Feliks Lou	TI	2016	L	Lulus	2020-02-2	SMAS Xav	Malam	Sudah Bel	3.92
13	802016002	Kevin Wei	TI	2016	L	Lulus	2020-09-0	SMA Swas	Malam	Belum Bel	3.82
14	802016002	Andrean C	TI	2016	L	Lulus	2020-02-2	SMK NEGE	Pagi	Belum Bel	3.62
15	802016002	Wahyudin	TI	2016	L	Lulus	2020-02-2	SMKN 2 K	Pagi	Belum Bel	3.82
16	802016002	Ambar We	TI	2016	L	Lulus	2020-02-2	SMA N 6 B	Pagi	Belum Bel	3.35
17	802016002	Juni Ahm	TI	2016	P	Lulus	2020-08-2	SMA Nege	Pagi	Belum Bel	3.18
18	802016002	Martha P	TI	2016	P	Lulus	2020-02-2	SMKN 2 K	Malam	Belum Bel	3.33
19	802016002	Febry Val	TI	2016	L	Lulus	2020-02-2	SMAS Xav	Malam	Belum Bel	3.88
20	802016002	Indra Gun	TI	2016	L	Non-Aktif		SMK Ungg	Pagi	Belum Bel	3.22
21	802016003	Ricky Ferr	TI	2016	L	Lulus	2020-02-2	SMKS Ung	Malam	Belum Bel	3.66
22	802016003	Ahmad Su	TI	2016	L	Lulus	2020-03-0	SMK PGRI	Pagi	Belum Bel	3.22
23	802016003	Nilla N	TI	2016	L	Non-Aktif		SMK Ungg	Pagi	Belum Bel	3.25

Gambar 5. Data Sebelum Dibersihkan

Data yang di hilangkan seperti mahasiswa yang keluar, yang non-aktif / cuti, serta yang aktif tapi belum lulus untuk angkatan 16 dan 17.

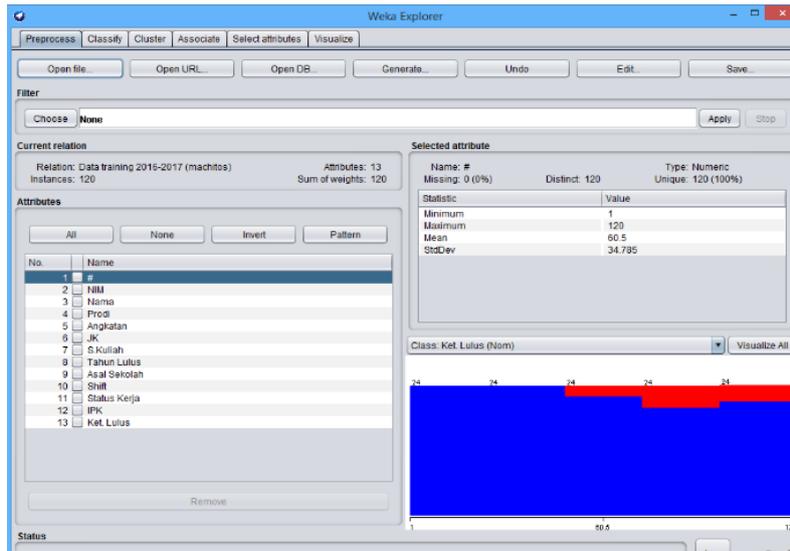
#	NIM	Nama	Prodi	Angkatan	JK	S.Kuliah	Tgl. Lulus	Asal Sekol	Shift	Status Ker	IPK
1	802016000	Diana Novita Sari	TI	2016	P	Lulus	2020-03-0	SMKS Ung	Malam	Sudah Bel	3.48
2	802016000	Indrawan Jamino	TI	2016	L	Lulus	2020-02-2	SMKS Ung	Malam	Belum Bel	3.81
3	802016001	Desy Ratna Sari	TI	2016	P	Lulus	2020-03-0	SMA Nege	Pagi	Belum Bel	3.62
4	802016001	Yoga Tanoko	TI	2016	L	Lulus	2020-03-0	SMAS BIN	Malam	Belum Bel	3.94
5	802016001	Kevin Rahmat Chr	TI	2016	L	Lulus	2020-08-2	SMK Ungg	Pagi	Belum Bel	3.73
6	802016001	Akmal Rustandi	TI	2016	L	Lulus	2020-09-0	SMK Nege	Pagi	Belum Bel	3.22
7	802016001	Angga Trisnanda	TI	2016	L	Lulus	2020-08-3	SMK Nege	Pagi	Belum Bel	2.97
8	802016001	Muhtadi Abdillah	TI	2016	L	Lulus	2020-02-2	SMK Nege	Pagi	Belum Bel	3.67
9	802016001	Shofan Rizki Setia	TI	2016	L	Lulus	2020-08-2	SMKN 2 K	Pagi	Belum Bel	3.45
10	802016002	Feliks Lourensus	TI	2016	L	Lulus	2020-02-2	SMAS Xav	Malam	Sudah Bel	3.92
11	802016002	Kevin Wenardy	TI	2016	L	Lulus	2020-09-0	SMA Swas	Malam	Belum Bel	3.82
12	802016002	Andrean Oscar	TI	2016	L	Lulus	2020-02-2	SMK NEGE	Pagi	Belum Bel	3.62
13	802016002	Wahyudin	TI	2016	L	Lulus	2020-02-2	SMKN 2 K	Pagi	Belum Bel	3.82
14	802016002	Ambar Wahyu Ba	TI	2016	L	Lulus	2020-02-2	SMA N 6 B	Pagi	Belum Bel	3.35
15	802016002	Juni Ahmad	TI	2016	P	Lulus	2020-08-2	SMA Nege	Pagi	Belum Bel	3.18
16	802016002	Martha P	TI	2016	P	Lulus	2020-02-2	SMKN 2 K	Malam	Belum Bel	3.33
17	802016002	Febry Valentino	TI	2016	L	Lulus	2020-02-2	SMAS Xav	Malam	Belum Bel	3.88
18	802016003	Ricky Fernando	TI	2016	L	Lulus	2020-02-2	SMKS Ung	Malam	Belum Bel	3.66
19	802016003	Ahmad Suhardi	TI	2016	L	Lulus	2020-03-0	SMK PGRI	Pagi	Belum Bel	3.22
20	802016003	Surya Krismanto	TI	2016	L	Lulus	2020-02-2	SMKS YAD	Pagi	Belum Bel	3.75
21	802016003	Sri Cahyoning Sap	TI	2016	L	Lulus	2020-03-0	SMKS Yadi	Pagi	Belum Bel	3.94
22	802016003	Muklas Bintar Aw	TI	2016	L	Lulus	2020-02-2	SMK Nege	Pagi	Belum Bel	3.56
23	802016003	Aufi Gusman	TI	2016	L	Lulus	2020-02-2	SMAN 1 K	Pagi	Belum Bel	3.44

Gambar 6. Setelah Pembersihan Data

4.2.1. Seleksi menggunakan Metode Forward Selection:

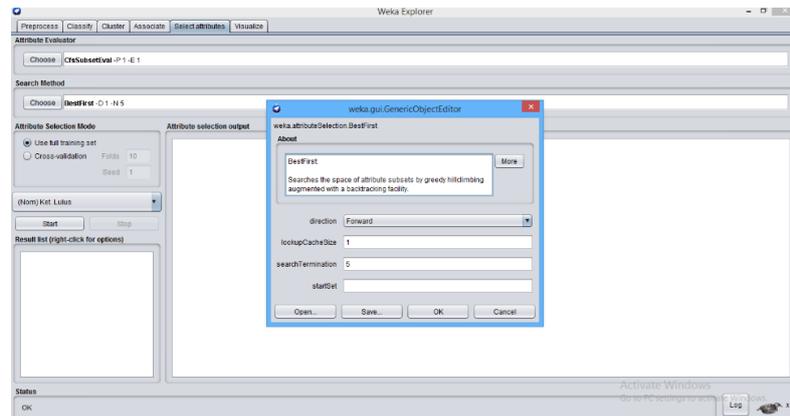
Setelah melakukan pembersihan pada data yang tidak sesuai, selanjutnya melakukan seleksi atribut menggunakan metode forward Selection pada aplikasi weka seperti berikut:

1. Memasukkan data training yang telah bersih dari missing value yang dapat mempengaruhi hasil dari data.

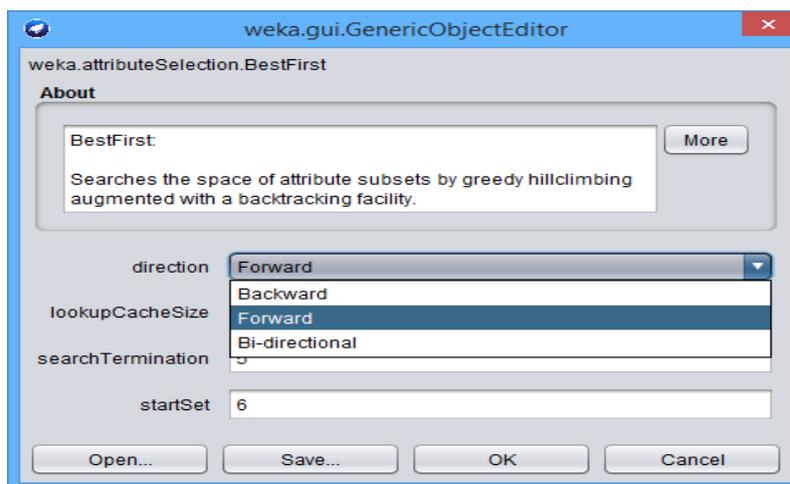


Gambar 7. Data Training pada WEKA

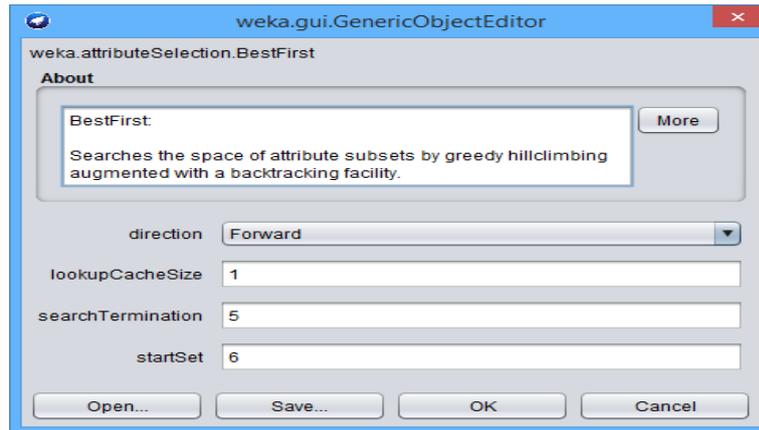
2. Memilih Seleksi atribut dan tentukan metode seleksinya



Gambar 8. Melakukan Seleksi Atribut

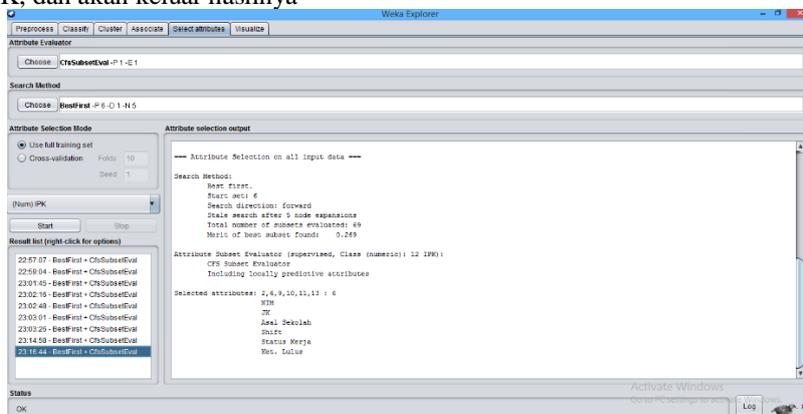


Gambar 9. Menentukan Metode Seleksi Atribut

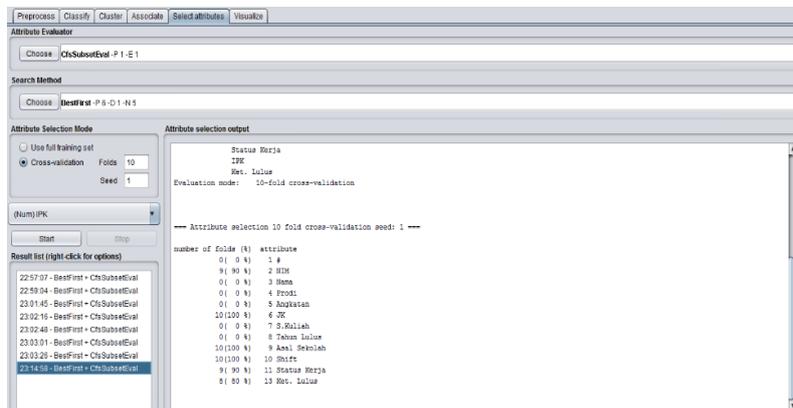


Gambar 10. Menentukan Metode Seleksi

3. Lalu pilih OK, dan akan keluar hasilnya



Gambar 11. Hasil Seleksi Metode Forward Selection



Gambar 12. Hasil Seleksi Metode Forward Selection dengan Cross-validation

Hasil dari menentukan atribut dengan menggunakan Forward Selection yaitu: NIM, JK (jenis Kelamin), asal sekolah, shift kuliah, status kerja dan keterangan lulus, lalu melakukan seleksi forward selection menggunakan cross-validation dengan folds 10 dan seed 1 dengan hasil NIM 9 (90%), JK 10 (100%), asal sekolah 10 (100%), shift kuliah 10 (100%), status kerja 10 (100%) dan keterangan lulus 10 (100%). Data hasil seleksi ini yang akan digunakan untuk proses data mining atau perhitungan manual dengan metode algoritma C4.5, dikarenakan NIM tidak dapat dihitung pada perhitungan manual maka atribut yang akan digunakan adalah JK (jenis Kelamin), asal sekolah, shift kuliah, dan status kerja lalu disimpan dalam suatu berkas, terpisah dari basis data operasional.

4.3 Perhitungan Manual

Pada tahap ini akan dilakukan perhitungan manual menggunakan data training. Alur kerja dalam perhitungan algoritma C4.5 adalah sebagai berikut:

1. Menyiapkan data training.
Data training biasanya dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon.
Akar akan diambil dari atribut yang terpilih dengan cara menghitung nilai Gain dari masing-masing atribut, nilai Gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai Gain dari atribut, hitung dahulu nilai entropy yaitu:

$$Entropy(S) = \sum_{j=1}^k - p_j \log_2 p_j \dots\dots (4.1)$$

Keterangan:
 S : Himpunan kasus
 A : Atribut
 n : Jumlah partisi S
 pi : Proporsi dari Si terhadap S

3. Kemudian hitung nilai Gain dengan metode information gain:

$$Gain(A) = Entropi(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i) \dots\dots (4.2)$$

Keterangan:
 S : Himpunan kasus
 A : Atribut
 n : Jumlah partisi atribut A
 |Si| : Jumlah kasus pada partisi ke-i
 |S| : Jumlah kasus dalam S

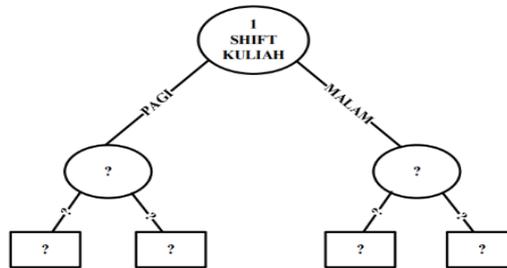
4. Ulangi proses perhitungan entropy dan gain untuk atribut yang lain, yaitu JK, asal sekolah, shift kuliah, dan status kerja. Masukkan seluruh nilai perhitungan ke dalam tabe l (opsional).

Tabel 1. Perhitungan Entropy dan Gain Akar Pertama

Node	Atribut	Data (jumlah kasus)	Cepat	Tepat	Lambat	Entropy	Gain
1.1	TOTAL	199	91	83	25	1,418363	
	JK		Cepat	Tepat	Lambat		0,026363
	L	162	67	72	23	1,446612	
	P	37	24	11	2	1,15289	
	Asal Sekolah						0,020083
	SMA	87	35	39	13	1,457181	
	SMK	97	48	38	11	1,388024	
	MA	13	7	5	1	1,295738	
	PKBM	2	1	1	0	0	
	PELITA RAYA	1	0	1	0	0	
	Status Kerja						0,02523
	Belum Bekerja	185	80	81	24	1,426943	
	Sudah Bekerja	14	11	2	1	0,946373	

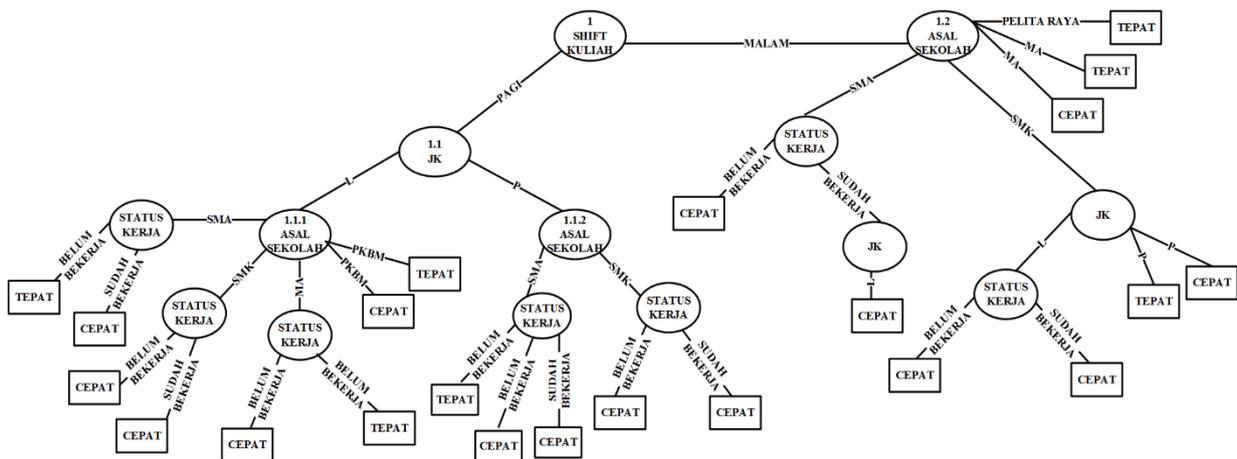
1. Perhitungan Node 1 (root)
 Jumlah Kasus (s) = 251
 Jumlah kasus (Cepat) = 128
 Jumlah kasus (Tepat) = 94
 Jumlah Kasus (Lambat) = 29

Dari perhitungan di atas diperoleh bahwa atribut dengan Gain tertinggi adalah “SHIFT KULIAH” 0,031757. maka SHIFT KULIAH menjadi node akar (Pertama), SHIFT KULIAH memiliki dua jenis yaitu “PAGI” dan “MALAM” dimana nilai ini masih perlu dilakukan perhitungan lagi (karena belum mendapat “TEPAT” dan “LAMBAT”).



Gambar 13. Cabang Pohon Keputusan Pertama

Gambar diatas merupakan Pohon Keputusan pertama dari perhitungan data *training*, dimana *root* 1 atau akar pertama adalah SHIFT KULIAH, serta urutan atribut yang berpengaruh dalam kelulusan mahasiswa adalah nilai JK, sekolah asal dan status bekerja. Lakukan perhitungan sampai memperoleh hasil akhir yang berupa lulus tepat waktu atau tidak tepat waktu. dan terlihat hasilnya seperti gambar di bawah ini



Gambar 14. Pohon Keputusan Algoritma C4.5

a. Hasil Perbandingan Evaluasi Akurasi

Setelah dilakukan analisis klasifikasi algoritma C4.5 pada tool WEKA menggunakan 5 Fold Cross Validation, 10 Fold Cross Validation, 70% Percentage Split dan 80% Percentage Split, maka diperoleh akurasi tertinggi yaitu dengan menggunakan 70% Percentage Split dengan persentasi akurasi yaitu 100% untuk Correctly Classified Instances dan 0% untuk Incorrectly Classified Instances. Sedangkan untuk data testing, dilakukan analisis klasifikasi algoritma C4.5 pada tool WEKA menggunakan 5 Fold Cross Validation, 10 Fold Cross Validation, 70% Percentage Split dan 80% Percentage Split, maka diperoleh akurasi tertinggi yaitu dengan menggunakan Cross Validasi 5 dan Cross Validasi 10 dengan persentasi akurasi yaitu 100% untuk Correctly Classified Instances dan 0% untuk Incorrectly Classified Instances. Perbandingan hasil analisis keduanya dapat dilihat pada tabel 2.

Tabel 2. *Perbandingan Evaluasi Akurasi Algoritma C4.5*

Model Evaluasi	Akurasi	Jumlah Kelas		Persentasi	
		Data Training	Data Testing	Data Training	Data Testing
5 Fold Cross Validation	<i>Correctly Classified Instances</i>	159	249	63.3466%	99.2032%
	<i>Incorrectly Classified Instances</i>	92	2	36.6534%	0.7968%
10 Fold Cross Validation	<i>Correctly Classified Instances</i>	160	249	63.745%	99.2032%
	<i>Incorrectly Classified Instances</i>	91	0	36.255%	0%
70% Percentage Split	<i>Correctly Classified Instances</i>	52	75	69.3333%	100%
	<i>Incorrectly Classified Instances</i>	23	0	30.6667%	0%
80% Percentage Split	<i>Correctly Classified Instances</i>	29	50	58%	100%
	<i>Incorrectly Classified Instances</i>	21	0	42%	0%

5. Kesimpulan

5.1 Simpulan

Untuk kesimpulan yang dapat diambil dari hasil yang penelitian telah dilakukan pada bab-bab sebelumnya yaitu:

1. Penelitian ini menggunakan data Mahasiswa angkatan 2016, 2017 dan angkatan 2020 dengan jumlah 502 data. Dengan atribut NIM, Nama, Prodi, Angkatan, Jenis Kelamin (JK), Status Kuliah, Asal Sekolah, Shift Kuliah, Status Kerja, IPK Terakhir dan Keterangan Lulus.
2. Hasil klasifikasi data Training dan data testing menggunakan algoritma C4.5 pada tool WEKA. Hasil Klasifikasi Algoritma C4.5 menggunakan data training, Pada tes 5 Fold Cross Validation dengan jumlah kelas Classified Instances 159 dan 249, Incorrectly Classified Instances 92 dan 2, dan persentasi akurasi Correctly Classified Instances sebesar 63.3466% dan 99.2032%, Incorrectly Classified Instances 36.6534% dan 0.7968%. Pada tes 10 Fold Cross Validation dengan jumlah kelas Correctly Classified Instances 160 dan 249, Incorrectly Classified Instances 91 dan 0, dan persentasi akurasi Correctly Classified Instances sebesar 63.745% dan 99.2032%, Incorrectly Classified Instances 36.255% dan 0%. Pada tes 70% Percentage Split dengan jumlah kelas Correctly Classified Instances 52 dan 75, Incorrectly Classified Instances 23 dan 0, serta persentasi akurasi Correctly Classified Instances sebesar 69.3333% dan 100%, Incorrectly Classified Instances 30.6667% dan 0%. Pada tes 80% Percentage Split dengan jumlah kelas Correctly Classified Instances 29 dan 50, Incorrectly Classified Instances 21 dan 0, serta persentasi akurasi Correctly Classified Instances sebesar 58% dan 100%, Incorrectly Classified Instances 42% dan 0%.
3. Hasil dari seleksi atribut menggunakan forward selection menghasilkan bahwa atribut yang paling berpengaruh terhadap masa studi mahasiswa adalah
 - a. Jenis Kelamin
 - b. Asal Sekolah
 - c. Shift kuliah
 - d. Status Kerja
 - e. Keterangan Lulus

5.2 Saran

Adapun saran yang dapat penulis berikan dari hasil penelitian ini adalah sebagai berikut :

1. Diharapkan kedepannya penelitian ini menggunakan data mahasiswa lebih banyak lagi serta mencakup atribut lainnya agar memiliki presentasi akurasi lebih baik.
2. Diharapkan data ini dapat diujikan menggunakan metode dan algoritma Data Mining lainnya.
3. Diharapkan kedepannya adanya penelitian yang melakukan perbandingan dengan metode algoritma lainnya.

6. Daftar Rujukan

- [1] Jambi, T. dan Jambi, K. B. “Buku Pedoman Akademik Universitas Dinamika Bangsa.”
- [2] Jasmir et al. (2018) “Prediksi Mahasiswa Drop Out dengan menggunakan Algoritma Klasifikasi Data Mining,” *Prosiding Annual Research Seminar*, 4(1), hal. 82–87. Tersedia pada: <http://seminar.ilkom.unsri.ac.id/index.php/ars/article/view/1864>.
- [3] Kantardzic, Mehmed (2020), *Data Mining: Concepts, Models, Methods, And Algorithms*, Edisi ke-3, New Jersey: John Wiley & Sons, Inc
- [4] Widaningsih, S. (2019) “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm,” *Jurnal Tekno Insentif*, 13(1), hal. 16–25. doi: 10.36787/jti.v13i1.78.
- [5] Mardi, Y. (2017) “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *Edik Informatika*, 2(2), hal. 213–219. doi: 10.22202/ei.2016.v2i2.1465.
- [6] Saefudin, S. dan Fernando, D. (2020) “Penerapan Data Mining Rekomendasi Buku Menggunakan Algoritma Apriori,” *JSiI (Jurnal Sistem Informasi)*, 7(1), hal. 50. doi: 10.30656/jsii.v7i1.1899.
- [7] Amelia, M. winny, Lumenta, A. S. . dan Jacobus, A. (2017) “Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes,” *Jurnal Teknik Informatika*, 11(1). doi: 10.35793/jti.11.1.2017.17652.
- [8] Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, F. A. D. (2018) “Metode-metode Klasifikasi,” *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 3(1), hal. 134
- [9] Ramadhan, G. et al. (2020) “Penerapan Data Mining Menggunakan Algoritma C4.5 Dalam Mengukur Tingkat Kepuasan Pasien BPJS,” *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2, hal. 376–385.
- [10] Azis, N. (2021) *Perbandingan dan Prediksi Kelulusan Mahasiswa Dengan Weka*, *Angewandte Chemie International Edition*, 6(11), 951–952.
- [11] Haryati, S., Sudarsono, A. dan Suryana, E. (2015) “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu),” *Jurnal Media Infotama*, 11(2), hal. 130–138.
- [12] Prayoga, N. D. (2018) “Penerapan Algoritma C.45 Dalam Memprediksi Kelulusan Tepat Waktu Pada Perguruan Tinggi (Studi Kasus : Stmik Royal Kisaran).”
- [13] Rohman, A. dan Rufiyanto, A. (2019) “Implementasi Data Mining Dengan Algoritma Decision Tree C4 . 5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandaran,” *Proceeding SINTAK 2019*, hal. 134–139.