



Analisis Prediksi Mahasiswa Terhadap Kelulusan Tepat Waktu Menggunakan Metode Data Mining *Decision Tree* (Studi Kasus: FTI UKSW)

Imelda Ruwae Lutunani ^{1*}, Adi Nugroho ²

^{1,2} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

article info

Article history:

Received 4 October 2022

Received in revised form

15 December 2022

Accepted 16 February 2023

Available online April 2023

DOI:

<https://doi.org/10.35870/jti.k.v7i2.781>

Keywords:

Data Mining; Decision Tree;

Student Graduation.

Kata Kunci:

Data Mining; Decision Tree;

Mahasiswa.

abstract

In general, students have the responsibility to complete their studies at a university. For students of the Satya Wacana Christian University Faculty of Information Technology, which every year there are more and more students, the world of work is currently required to become someone who masters the field of technology. In addition, as a student, there are many things that must be done to complete studies by participating in activities on campus, organizations, and being active in the teaching and learning process so that they can complete their studies on time. In this study, a predictive analysis of SWCU FTI students will be conducted on timely graduation using the decision tree data mining method. which will see students who graduate on time and graduate late using the decision tree algorithm which is a decision tree algorithm that has a high level of accuracy in large amounts of data. In this study, the decision tree algorithm was used to run 983 sample data, resulting in a match accuracy of 91.25%. This means that it is very good and effective in predicting student graduation.

abstrak

Pada umumnya mahasiswa memiliki tanggungjawab untuk menyelesaikan studi pada sebuah universitas. Bagi mahasiswa Fakultas Teknologi Informasi Universitas Kristen Satya Wacana yang setiap tahunnya semakin banyak mahasiswanya, saat ini dituntut dunia pekerjaan untuk menjadi seorang yang menguasai bidang teknologi. Selain itu, sebagai mahasiswa banyak hal yang harus dilakukan untuk menyelesaikan studi dengan mengikuti kegiatan di kampus, organisasi, dan aktif dalam proses belajar mengajar sehingga dapat menyelesaikan studi tepat waktu. Pada penelitian ini akan dilakukan suatu analisis prediktif mahasiswa FTI UKSW terhadap kelulusan tepat waktu menggunakan metode data mining decision tree. dimana akan dilihat mahasiswa yang lulus tepat waktu dan terlambat lulus dengan menggunakan algoritma decision tree yang merupakan algoritma pohon keputusan yang mempunyai tingkat akurasi yang tinggi dalam jumlah data yang besar. Dalam penelitian ini, algoritma decision tree digunakan untuk menjalankan 983 data sampel, menghasilkan akurasi kecocokan 91,25%. Artinya sangat baik dan efektif dalam memprediksi kelulusan mahasiswa.

Corresponding Author. Email: imeldalutunani@gmail.com ^{1}.

1. Latar Belakang

Mahasiswa yang belajar di universitas memainkan peran yang sangat penting di universitas [1]. Proses pembelajaran tidak dapat dilaksanakan tanpa adanya mahasiswa. Dalam proses belajar mengajar, siswa harus belajar banyak hal, baik itu pengetahuan umum maupun pengetahuan khusus, harus dikembangkan agar dapat dilakukan berdasarkan keterampilan yang dimiliki siswa di dunia kerja [2]. Mahasiswa Fakultas Teknologi Informasi Universitas Kristen Satya Wacana memiliki enam jurusan: Jurusan Teknik Informatika, Jurusan Sistem Informasi, Jurusan Desain Komunikasi Visual, Jurusan Pendidikan Teknik Informatika dan Komputer, Jurusan Hubungan Masyarakat, dan Jurusan Ilmu Perpustakaan. Selain Program Sarjana, terdapat Program Diploma yaitu Program Studi Diploma Teknik Informatika dan Program Studi Diploma Akuntansi Komputer. Untuk meningkatkan kemampuan komputasi, mahasiswa FTI UKSW juga telah membuka Program Pascasarjana yaitu Program Studi Magister Sistem Informasi dan Program Doktor, dan saat ini sedang dibentuk Program Magister tambahan.

Secara umum mahasiswa, khususnya di fakultas teknologi informasi UKSW, setiap tahunnya semakin banyak yang diterima namun tidak sedikit yang tidak lulus untuk tepat waktu. Berbagai tahapan proses pembelajaran yang dilakukan oleh mahasiswa seperti menyelesaikan beberapa mata kuliah, kerja praktek, bahkan seminar tugas akhir merupakan proses dan tahapan yang menentukan tingkat kelulusan di fakultas teknologi informasi [3]. Untuk melakukan pencarian prediksi dapat digunakan dengan penggunaan teknik data mining yang adalah suatu proses dalam menggumpulkan sebuah informasi penting pada suatu data yang berskala besar yang membantu pihak-pihak dalam membutuhkan informasi. Dalam penelitian ini digunakan metode ID3 untuk memprediksi kelulusan mahasiswa FTI UKSW. *Decision tree* merupakan salah satu algoritma proses pembuatan keputusan yang mempunyai tingkat akurasi yang tinggi dalam mengaplikasikan jumlah data yang besar dibandingkan dengan algoritma pohon keputusan lainnya yang menghasilkan pohon keputusan yang mudah diinterpretasikan [4].

Penelitian ini menggunakan metode ID3 untuk memprediksi kelulusan mahasiswa tepat waktu dan menjelaskan aspek-aspek yang mempengaruhi pengukurab tingkat akurasi menggunakan aplikasi rapidminer. Penelitian selanjutnya dilakukan di AI Yarmouk University College salah satu perguruan tinggi yang ada di Iraq dengan judul *Classification of Human Skin Diseases Using Data Mining* [5]. Yang menjadi masalah pada penelitian ini tentang bagaimana mengklasifikasikan penyakit kulit manusia dimana sistem yang dibuat ini agar dapat mendiagnosis penyakit kulit dan dalam administrator sistem ini dapat mengelola informasi dari suatu penyakit kulit, gejala, perawatan medis dan akan diberikan saran atau pernyataan dari sistem untuk menampilkan deskripsi penyakit kulit dimana ini penting dan berguna bagi dokter untuk mendiagnosis penyakit kulit manusia agar dalam menentukan obat yang sesuai atau cocok dengan penyakit. Untuk mengatasi masalah tersebut dilakukan dengan membuat sistem otomatis untuk mengenali penyakit kulit manusia dengan metode diagnosis berbasis gambar yang berfokus pada mendiagnosis penyakit pada kulit yang disebabkan oleh virus dan bakteri. Untuk mendeteksi penyakit kulit ini dapat dilakukan dengan Langkah-langkah yaitu pra-pemrosesan, ekstraksi fitur dan pengklasifikasi pada decision tree. Untuk mengklasifikasikan gambar penyakit kulit maka sistem ini dirancang dan diimplementasikan pada MATLAB dan diuji dengan berbagai gambar pada database.

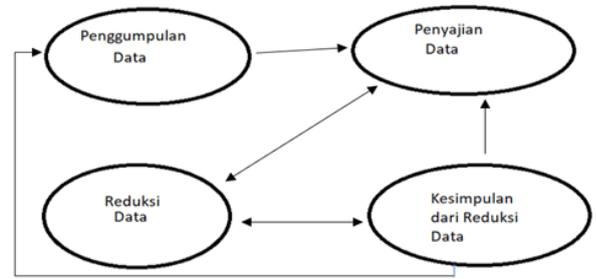
Perbedaan dari penelitian ini adalah penulis menerapkan algoritma decision tree untuk memprediksi kelulusan mahasiswa yang lulus sesuai waktu studi atau tepat waktu dengan RapidMiner sedangkan pada penelitian ini menggunakan algoritma decision tree data mining sebagai memproses data gambar menggunakan pengolahan citra untuk mendeteksi penyakit kulit. Penelitian selanjutnya dilakukan di Universitas of Education salah satu perguruan tinggi yang ada di Vietnam dengan judul *Improving Performance Benchmark of Decision Tree Classifications For E-mail Spam Filtering* [6]. Penelitian ini bertujuan untuk menyajikan tiga pendekatan untuk meningkatkan kinerja deteksi spam e-mail menggunakan klasifikasi dalam hal kualitas parameter fine-tune, pemilihan fitur dan kombinasi pengklasifikasi dan data yang didapat yaitu kumpulan data populer yang disebut dengan spambase. Dalam

dataset ini jumlah instance adalah 4601 dari mana 1813 spam yang sama dengan 39,4% dan jumlah atribut 58 dari yang 57 adalah kontinu dan 1 memiliki label kelas nominal untuk klasifikasi spam. Email spam diklasifikasikan menggunakan validasi silang sepuluh kali lipat dengan menggunakan machine learning aplikasi perangkat lunak Weka yang melibatkan dua belas algoritma yang berbeda. Dari di analisis dan dibandingkan dalam hal akurasi klasifikasi maka ditemukan bahwa kinerja akurasi tertinggi dicapai oleh klasifikasi J48 untuk dataset email spambase. Hasil eksperimen menunjukkan bahwa algoritma Random Forest merupakan pengklasifikasi terbaik dengan tingkat akurasi 94,68% pada dataset spambase ini menunjukkan bahwa pendekatan ini adalah solusi kompetitif untuk masalah tersebut. Perbedaan dengan penelitian ini yaitu penulis menggunakan metode *Decision Tree* dalam data mining untuk mendapat hasil yang akurat dalam memprediksi kelulusan mahasiswa yang tepat waktu dan terlambat, sedangkan. Penelitian ini menggunakan data mining dengan tools Weka dalam mengklasifikasi spam E-mail.

2. Metode Penelitian

Pada metode penelitian ini menggunakan metode klasifikasi ID3, untuk memprediksi kelulusan tepat waktu mahasiswa FTI UKSW. Dalam prediksi, ada beberapa tahapan atau langkah yang dapat dilakukan dalam penelitian ini. Penelitian ini menggunakan Teknik data mining adalah ID3 data tersebut adalah:

- 1) Data Training adalah proses mining dan testing, berupa data sampel mahasiswa angkatan 2014 dan 2015 sebagai data latih. Data ini memiliki atribut nama, progdi, jenjang, angkatan, lulus, ipk, jenis kelamin, daerah asal, tahun masuk dan tahun lulus mahasiswa.
- 2) Data Testing adalah sampel data mahasiswa Angkatan 2016 sebagai data testing. Data ini memiliki atribut nama, progdi, jenjang, angkatan, lulus, ipk, jenis kelamin, daerah asal, tahun masuk dan tahun lulus mahasiswa. Data ini akan dijadikan pengujian berdasarkan data latih.



Gambar 1. Tahapan Penelitian

Pada Gambar 1 dapat dijelaskan sebagai berikut:

- 1) Tahap pengumpulan data adalah tahap pengumpulan data dengan meminta data dari lokasi penelitian yaitu Bagian Administrasi Registrasi Akademik (BARA) dimana data yang diberikan berupa data mahasiswa FTI UKSW pada tahun 2014 sampai tahun 2020.
- 2) Tahap reduksi data adalah proses pemilihan data pada saat pengambilan sampel untuk digunakan sample dalam penelitian
- 3) Tahap penyajian data adalah tahap penyajian setelah proses pemilihan dan perolehan sampel dan data uji selesai
- 4) Tahap penarikan kesimpulan adalah tahap akhir dari analisis data telah diolah dan dapat diselesaikan sesuai dengan penelitian.

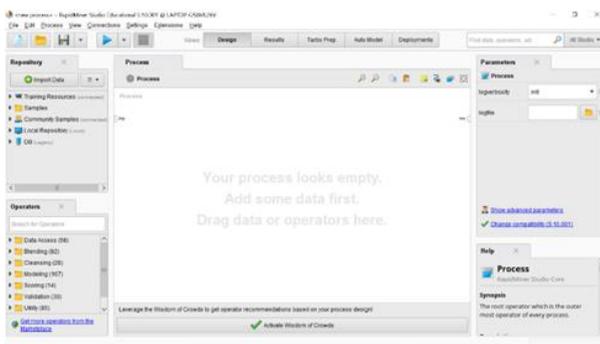
Alat dan Bahan yang diperlukan

dalam penelitian ini adapun alat dan bahan yang dibutuhkan diantaranya:

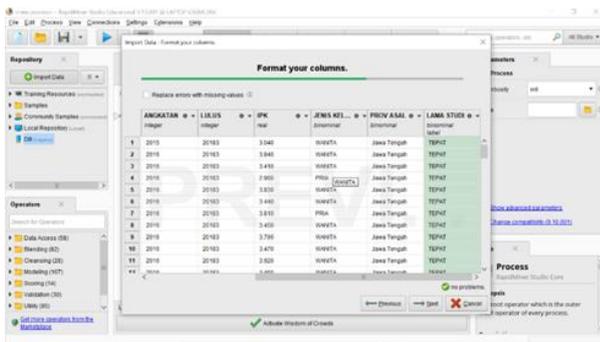
- 1) kebutuhan Hardware yaitu sebuah computer PC/leptop dimana menggunakan leptop acer swift Sf314 AMD Ryzen 5 4500U dengan memory leptop 5 GB.
- 2) Kebutuhan Software yaitu sebuah tools rapidminer yang digunakan dalam penelitian ini untuk mengolah data.
- 3) Kebutuhan Data yaitu data yang dibutuhkan dari data mahasiswa FTI UKSW meliputi beberapa atribut yaitu Nama, Progdi, Jenjang, Angkatan, Lulus, IPK, Jenis Kelamin, Daerah Asal, tahun Masuk dan Tahun lulus mahasiswa dengan jumlah data 2.424 data mahasiswa.



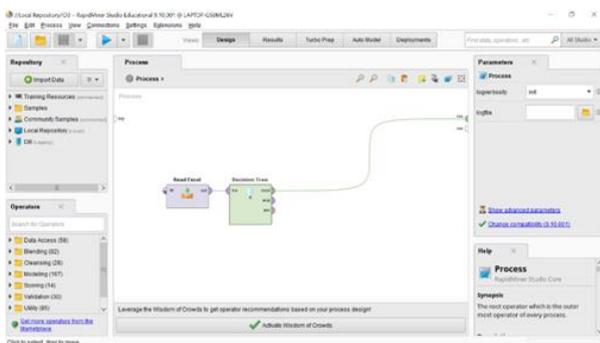
Gambar 2. Tools Rapidminer



Gambar 3. Data dimasukkan pada Software Rapidminer



Gambar 4. Data Setelah Penggantian Type



Gambar 5. Proses Pohon Keputusan (Decision Tree)

Pengumpulan data ini memiliki sumber data: data primer dan data sekunder. Data primer adalah data yang dikumpulkan langsung oleh peneliti [7]. Data sekunder adalah data yang diperoleh melalui tangan kedua dan data yang digunakan peneliti saat ini merupakan data sekunder dimana diperoleh dari database mahasiswa Universitas Kristen Satya Wacana terlebih khusus yang diambil adalah data mahasiswa Fakultas Teknologi Informasi Tahun Akademik 2014 – 2020 Tahun. Dalam pengumpulan data ini termasuk yang bersifat akademik dengan kelulusan mahasiswa. Dimana data tersebut mencari nilai probabilitistik yang dijadikan pola dalam membentuk model ID3 yang menunjukkan atribut mana yang paling baik dan efektif dampaknya terhadap kelulusan. Dalam pengujian data pada penelitian ini adalah nim, nama, jenis kelamin, asal daerah dan ipk setelah itu dilakukan integrasi data yang akan dijadikan satu tabel sehingga akan terjadi transformasi data yang akan digabungkan untuk proses dalam data mining.

Algoritma ID3 adalah metode yang dapat digunakan untuk Entropy (S) = $\sum_{i=1}^n -p_i * \log_2 p_i$ untuk pohon ke-mprediksi pada suatu penelitian. Entropy adalah ukuran seberapa acak isi sebuah kelas. Entropy bernilai nol bila kelas isinya 100% homogen (sejenis semua) dan bernilai 1 bila isi kelas sepenuhnya acak (tidak ada yang sejenis) Entropy sebagai ukuran seberapa acak suatu kelompok data entropy yang semakin rendah menunjukkan bahwa kelompok data semakin homogen (berisi data yang sejenis berasal dari kelas yang sama), dan nilai entropy nol berarti kelompok data sudah homogen sepenuhnya [8]. Sementara itu, nilai entropy 1 berarti kelompok data berisi data acak 50-50. Contoh bila kelompok data jumlahnya sama banyak maka entropynya adalah 1. Rumus matematika Entropy sebagai berikut:

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

Keterangan:

- S : Himpunan kasus
- A : Atribut
- N : Jumlah partisi S
- Pi : Proposi dari Si terhadap S

Metode yang dapat digunakan selain algoritma ID3 ada algoritma K-NN, Naïve Bayes, Logistic Regression, Neural Network dll. ini juga menjadi metode yang sering dipakai dalam mencari akurasi dalam model klasifikasi dengan berbagai prediksi. Sebagai contoh peneliti akan melakukan penilaian akurasi untuk model klasifikasi yang digunakan dengan metode K-NN, Naïve Bayes dan Logistic Regression menggunakan data mahasiswa FTI UKSW tahun kelulusan 2014-2020. Berikut ini adalah Tabel perbandingan performance algoritma

Tabel 1. Perbandingan *Performance Algoritma*

Algoritma	Accurarcy
K-NN	91,05 %
Naïve Bayes	58,99 %
Logistic Regression	91,46%

Berdasarkan tabel 1 dapat dilihat nilai akurasi untuk metode klasifikasi yang menjadi nilai akurasi terbaik adalah Algoritma Logistic Regression yang memiliki nilai akurasi tertinggi dengan nilai 91,46% lalu K-NN memiliki nilai akurasi 91,05% sedangkan Naïve Bayes yang menjadi nilai akurasi terendah yaitu 58,99%.

Tabel 2. Parameter Akurasi Pada Rapidminer

Nilai Akurasi (%)	Hasil Prediksi
90-100	Sangat Baik
80-90	Baik
70-80	Sedang
60-70	Lemah
50-60	Sangat Lemah

Sumber : Galih. 2019. Data Mining di Bidang Pendidikan untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Klasifikasi pada STMIK Jabar

3. Hasil dan Pembahasan

Dalam menggunakan data ini. Pertama melakukan pengumpulan data dimana data ini didapat dengan namanya data sekunder. Data sekunder yaitu data yang diperoleh melalui tangan kedua ini dikarenakan data yang diperoleh dari database mahasiswa Universitas Kristen Satya Wacana yang datanya yaitu data mahasiswa Universitas Kristen Satya Wacana Fakultas Teknologi Informasi Tahun Akademik 2014, 2015 dan 2016 yang telah terkumpul data sebanyak 1000 data mahasiswa.

Pada dasarnya, algoritma ID 3 melakukan *splitting*(pemecahan) data ke dalam dua kelompok berdasarkan atribut-atribut yang ada di dalam dua kelompok berdasarkan atribut-atribut yang ada didalam data, dengan mengukur suatu angka yang disebut entropy [9]. Proses *splitting* dilakukan bertingkat dari atas ke bawah, menghasilkan kelompok data yang semakin kecil. Setiap kelompok data dibuat sedemikian rupa agar entropinya serendah mungkin. Masing-masing kelompok data masuk ke kelompok masing-masing contoh mobil sedan semua masuk kedalam kelompok sedan, minibus semua masuk ke dalam kelompok minibus [10].

Disetiap tingkatan yaitu atribut data yang ada, diuji untuk mendapatkan entropy yang serendah mungkin. Semakin banyak tingkatan dalam decision tree, artinya pohon semakin “tinggi” dan proses *splitting* semakin kompleks, model akan menjadi lebih fit. Adapun begitu, untuk mencegah decision tree menjadi terlalu Panjang yang dapat mengakibatkan overfitting (model nampak bagus pada saat training namun tidak bisa bekerja baik saat dipakai dengan data baru), suatu proses yang disebut pruning (pemangkasan) perlu dilakukan. Cabang-cabang yang terlalu Panjang dipangkas dan node-node ditingkat atas dijadikan leaf node [11].

Kemudian dari data awal yang berisi data total 1000 data mahasiswa namun data tersebut semua tidak semua dipakai atau digunakan hanya beberapa yang akan dipilih dan atribut [12] juga tidak semua digunakan hal ini disebabkan karena beberapa tahapan dalam pengolahan data awal (*preparation data*) dimana untuk memperoleh data yang berkualitas dengan cara melakukan beberapa tahapan diantaranya:

- 1) Data validation, melakukan identifikasi dan menghapus data yang ganjil (outlier/noise) dengan data yang tidak lengkap (missing value).
- 2) Data intergration and transformation. Melakukan dalam meningkatkan akurasi yang efisiensi pada algoritma dan kemudian data tersebut ditransformasikan ke dalam software RapidMiner.

Tabel 3. Rekap Data Mahasiswa Lulus Tepat dan Terlambat Berdasarkan Angkatan dan Jenis Kelamin

Angkatan	Jenis kelamin				Jumlah
	Laki-laki		Wanita		
	Tepat	Terlambat	Tepat	Terlambat	
2014	226	51	112	11	400
2015	173	30	88	9	300
2016	174	35	79	12	300
	573	116	279	32	1000

Setelah melakukan tahap data preparation maka total data menjadi berkurang sehingga total data menjadi 983 record.

Tabel 4. Rekap Data Mahasiswa Lulus Tepat dan Terlambat Berdasarkan Angkatan dan Jenis Kelamin Setelah Melakukan Data Preparation.

Angkatan	Jenis kelamin				Jumlah
	Laki-laki		Wanita		
	Tepat	Terlambat	Tepat	Terlambat	
2014	223	49	110	7	389
2015	172	30	86	9	297
2016	171	35	79	12	297
	566	114	275	28	983

Pada pengolahan data ini menggunakan algoritma ID3 sehingga membutuhkan data training dan data testing untuk selanjutnya dilakukan perhitungan dari masing-masing atribut. Data training yang digunakan dalam penelitian sebanyak 686 dataset, berasal dari mahasiswa Angkatan tahun 2014 sampai tahun 2015. Data testing yang digunakan sebanyak 297 dataset berasal dari mahasiswa Angkatan tahun 2016 dimana data training dan data testing untuk pengolahan data dapat dilihat pada gambar 6 berikut ini.

1	NAMA	PROGDI	JENIANG	ANGKATAN	LULUS	IPK	JENIS KELAMIN	PROV ASAL	LAMA STUDI
2	ARTHA CAHYA ZULFKARANI	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,04	WANITA	Jawa Tengah	TEPAT
3	DIANINI AZZA FAKRATI	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,84	WANITA	Jawa Tengah	TEPAT
4	ANIS SETIA WATI	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,41	WANITA	Jawa Tengah	TEPAT
5	DHOKHA CHRISTIAN WIGAS PUTRA	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	2,98	PRIA	Jawa Tengah	TEPAT
6	NOFIRA DIANA WATI	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,83	WANITA	Jawa Tengah	TEPAT
7	DANI SETIANI	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,44	WANITA	Jawa Tengah	TEPAT
8	DARMA SETIAB	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,81	PRIA	Jawa Tengah	TEPAT
9	WITWA AGUSTINA	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,45	WANITA	Jawa Tengah	TEPAT
10	MERY KENY SRIYANTI	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,79	WANITA	Jawa Tengah	TEPAT
11	EVA DIANA KOSTALANI	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,47	WANITA	Jawa Tengah	TEPAT
12	YUNIA NINGSH	SISTEM INFORMASI AKUNTANSI	D3	2015	20183	3,92	WANITA	Jawa Tengah	TEPAT
13	DAWI WILANSARI	TEKNIK INFORMATIKA	D3	2015	20183	3,46	WANITA	Jawa Tengah	TEPAT
14	RICHARDO KURNIAWAN	TEKNIK INFORMATIKA	D3	2015	20183	3,84	PRIA	Jawa Tengah	TEPAT
15	SHERLY TOONG	TEKNIK INFORMATIKA	D3	2015	20183	3,55	WANITA	Jawa Tengah	TEPAT
16	SEPTIAN ADE PRABOWO	TEKNIK INFORMATIKA	D3	2015	20183	3,25	PRIA	Jawa Tengah	TEPAT
17	LAHENDY ALEKANDRO DESTIAN	TEKNIK INFORMATIKA	D3	2015	20183	3,6	PRIA	Jawa Tengah	TEPAT
18	GLAR RAMADGA SENDENA	TEKNIK INFORMATIKA	D3	2015	20183	3,43	PRIA	Jawa Tengah	TEPAT
19	FA BAGUS KUNCORO	TEKNIK INFORMATIKA	D3	2015	20183	3,66	PRIA	Jawa Tengah	TEPAT
20	KACHABADI MELINDHARI	TEKNIK INFORMATIKA	D3	2015	20183	3,46	PRIA	Jawa Tengah	TEPAT
21	RICKO PERMANA PUTRA	TEKNIK INFORMATIKA	D3	2015	20183	3,59	PRIA	Jawa Tengah	TEPAT
22	LUCKY LABASATI	TEKNIK INFORMATIKA	D3	2015	20183	3,19	WANITA	Jawa Tengah	TEPAT

Gambar 6. Data Training

1	NAMA	PROGDI	JENIANG	ANGKATAN	LULUS	IPK	JENIS KELAMIN	PROV ASAL	LAMA STUDI
2	HAZEL ALBERTA	HUBUNGAN MASYARAKAT	S1	2017	20203	3,8	WANITA	Jawa Tengah	TEPAT
3	ELE MELINDA CHRISTIAN	HUBUNGAN MASYARAKAT	S1	2017	20203	3,62	WANITA	Jawa Tengah	TEPAT
4	TRIFOSA IOAN ERWANI	HUBUNGAN MASYARAKAT	S1	2017	20203	3,76	WANITA	Jawa Tengah	TEPAT
5	ANGGULU LILIANA PUTRI	HUBUNGAN MASYARAKAT	S1	2017	20203	3,43	WANITA	Jawa Tengah	TEPAT
6	JASON YOEL CHRISTIANO	HUBUNGAN MASYARAKAT	S1	2017	20203	3,64	PRIA	Jawa Tengah	TEPAT
7	NOVY RIZALI	HUBUNGAN MASYARAKAT	S1	2017	20203	3,9	WANITA	Jawa Tengah	TEPAT
8	LILAH POGIA ANGGRAENI	TEKNIK INFORMATIKA	S1	2014	20203	3,16	WANITA	Jawa Tengah	TEPAT
9	MERY PUTRA SALOMON MBAU	TEKNIK INFORMATIKA	S1	2014	20203	3,08	PRIA	Jawa Tengah	TEPAT
10	ANTONNIUS	TEKNIK INFORMATIKA	S1	2014	20203	2,78	PRIA	Jawa Tengah	TEPAT
11	CHRISTIAN BARUNA ARI	TEKNIK INFORMATIKA	S1	2015	20203	3,06	PRIA	Jawa Tengah	TEPAT
12	EXCEL BARTHEL MAITIMU	TEKNIK INFORMATIKA	S1	2016	20203	3,05	PRIA	Jawa Tengah	TEPAT
13	DEVARA PUTRA ARYASA	TEKNIK INFORMATIKA	S1	2016	20203	3,56	PRIA	Jawa Tengah	TEPAT
14	AHEL HERNANDEZ	TEKNIK INFORMATIKA	S1	2016	20203	3,45	PRIA	Jawa Tengah	TEPAT
15	RIVAL EKSON BELORA	TEKNIK INFORMATIKA	S1	2016	20203	3,15	PRIA	Jawa Tengah	TEPAT
16	STEFANUS DANIEL HARTONO	TEKNIK INFORMATIKA	S1	2016	20203	3,59	PRIA	Jawa Tengah	TEPAT
17	DEGLORIANI TUHANY	TEKNIK INFORMATIKA	S1	2016	20203	3,15	PRIA	Jawa Tengah	TEPAT
18	RESTIAWAN YOGA PRATAMA	TEKNIK INFORMATIKA	S1	2016	20203	3,59	PRIA	Jawa Tengah	TEPAT
19	SETYO HADI DARWINTO	TEKNIK INFORMATIKA	S1	2016	20203	3,25	PRIA	Jawa Tengah	TEPAT
20	ARNOLDUS	TEKNIK INFORMATIKA	S1	2016	20203	3,06	PRIA	Jawa Tengah	TEPAT
21	JEREMIA ARI NUGROHO	TEKNIK INFORMATIKA	S1	2016	20203	3,38	PRIA	Jawa Tengah	TEPAT

Gambar 7. Data Testing

Saat pengklasifikasian data ini dengan menggunakan metode ID3 ini sangat berpengaruh dalam proses prediksi adalah melihat hubungan antara data dan variable input dan target, karena metode menggabungkan eksplorasi data dan pemodelan. Dari tabel 5 terlihat bahwa mahasiswa yang lulus tepat waktu sebanyak 834 record dan sebanyak 149 record mahasiswa yang terlambat lulus maka entropinya adalah sebagai berikut:

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

$$Entropy(S) = (-834/983 * \log_2(834/983)) + (-149/983 * \log_2(149/983))$$

$$Entropy(S) = 0,613772$$

Tabel 5. Rekap Hasil Perhitungan ID3 Tiap Atribut Kategorial

	Node	Jumlah Data	Lulus Tepat	Terlambat	Entropy	Gain
Lama Studi	Total	983	834	149	0,613772	
Jenis Kelamin	Wanita	303	274	29	0,455246	0,008381
	Pria	680	560	120	0,672295	
Asal Daerah	Jawa Tengah	614	531	83	0,571473	0,002468
	Luar Jawa Tengah	369	303	66	0,677581	
Progdi	Sistem Informasi Akuntansi D3	36	35	1	0,183122	
	Teknik Informatika D3	55	54	1	0,131107	
Hubungan Masyarakat S1	Teknik Informatika S1	55	48	7	0,549911	
	S1	477	416	61	0,551602	
Sistem Informasi S1	Desain Komunikasi Visual S1	219	196	23	0,484718	
	Pendidikan TI Dan Komputer S1	11	4	7	0,94566	1,137448
Perpustakaan & Sains	Informasi S1	7	4	3	0,985228	
	Magister Sistem Informasi S2	41	38	3	0,377646	
Angkatan	2014	126	85	41	0,910154	0,642659
	2015	276	255	21	0,388239	
	2016	270	265	5	0,13304	

1) Pengujian akurasi prediksi
 Algoritma yang dihasilkan oleh perhitungan menggunakan algoritma ID3 di uji menggunakan aplikasi Rapidminer adalah sebuah software dalam mengolah data salah satu bentuk machine learning data mining, text mining dan *predictive analytics*. Berikut ini adalah Langkah-langkah dalam pengolahan data menggunakan ID3 pada Software Rapidminer [13].

2) Implementasi Pada Software Rapidminer.
 Pada rapidminer terdapat langkah-langkah untuk melakukan pengolahan data [14] menggunakan ID 3 pada software rapidminer:

- a) Buka software Rapidminer. Lalu pilih Blank proses.
- b) Klik Add Data dan pilih penyimpanan data yang datanya disimpan. Dimana data tersebut di simpan pada My Computer dengan nama TRAINING. Setelah itu, klik Next.
- c) Setelah itu, Klik Next dan diubah type data termasuk pada kolom provinsi asal, jenis kelamin dan kemudian diberi type pada kolom lama studi binominal khususnya lama studi akan menjadi role sehingga diberikan label.
- d) Pada Local Repository menjadi tempat penyimpanan data yaitu data training dan data testing yang sudah dibuat setelah itu klik Finish sehingga akan terlihat jendela rapidminer. Hal ini juga sama dilakukan untuk melihat hasilnya.
- e) Pada lembar kerja rapidmainer lakukan drop panel untuk data training (data latih) dan data testing. kemudian pilih dan lakukan drop apply model lalu lakukan hal yang sama pada decision Tree begitu juga pada performance kemudian hubungkan atau sambungkan semua panel ke kolom training ada algoritma decision tree yang akan diterapkan kemudian pada kolom data testing memiliki Apply Model untuk menjalankan model algoritma ID3 dan performance untuk mengukur performance dari model ID3 tersebut.
- f) Hasil uji akurasi akan terlihat pada step terakhir yaitu dengan melakukan klik run setelah semua sudah dihubungkan.



Gambar 8. Panel Terhubung

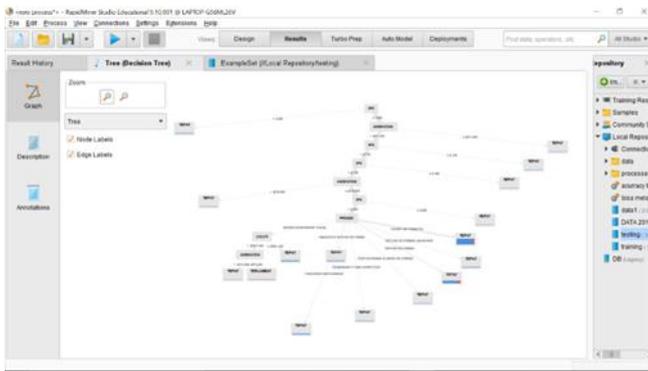
Criterion	True TPAT	True TERLAMBAT	Overall precision
accuracy	271	24	91.81%
prec TPAT	271	0	0.00%
prec TERLAMBAT	2	0	0.00%
data recall	99.27%	0.00%	

Gambar 9. Hasil Uji Akurasi Menggunakan Rapidminer

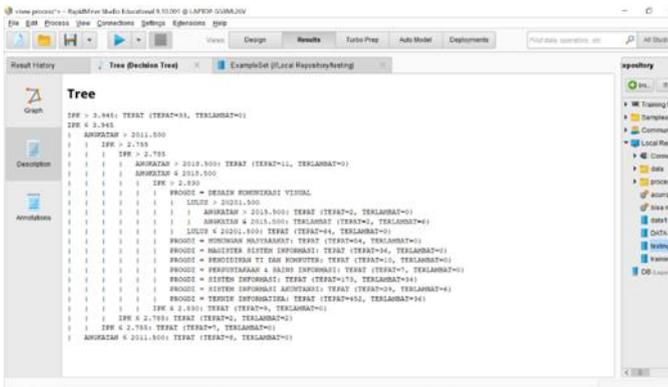
Setelah melakukan tahapan ini hasil akurasi model decision tree menunjukkan bahwa akurasi sebesar 91,25%. Artinya model klasifikasi kelulusan akurasinya mencapai 91,25%, prediksi terlambat sebesar 91,86% dan prediksi tepat sebesar 0,00%. Ini dinilai sangat tinggi dan efektif dalam hal kompleksitas dan jumlah datasetnya.

3) Rule dalam mengambil Keputusan dengan menggunakan metode Pohon Keputusan (Decision Tree)

Pohon keputusan yaitu salah satu juga metode klasifikasi dimana metode ini memrepresentasikan aturan untuk diinterpretasi oleh manusia [15]. Dalam penelitian ini, membangun pohon keputusan dilakukan dengan menggunakan rapidminer dapat di lihat pada gambar 4 berikut ini. *Decision Tree* yang dihasilkan menampilkan cukup banyak leaf sebagai atribut yang digunakan pada data kelulusan mahasiswa [16]. Penjelasan mengenai decision tree kelulusan mahasiswa secara rinci dapat dilihat pada *description result* berikut ini.



Gambar 10. *decision tree* kelulusan Mahasiswa



Gambar 11. *Description Result*

Pada gambar 8 dapat dilihat bahwa penjelasan hasil *decision tree* dengan atribut-atribut seperti IPK, Angkatan dan Progdi dapat dilihat bahwa mahasiswa yang lulus tepat waktu paling banyak terdapat pada progdi Teknik Informatika dengan lulus tetap berjumlah 452 dan yang terlambat berjumlah 36 maka akan semakin kompleks jika memiliki banyak atribut yang digunakan *decision tree* yang dihasilkan juga semakin optimal.

4. Kesimpulan dan Saran

Kesimpulan yang diperbolehkan dari tentang Analisis Prediksi Kelulusan Mahasiswa FTI UKSW Tepat Waktu Menggunakan Metode *Data Mining Decision Tree* yaitu: Data training yang digunakan dalam penelitian sebanyak 686 dataset, berasal dari mahasiswa Angkatan tahun 2014 sampai tahun 2015. Data testing yang digunakan sebanyak 297 dataset berasal dari mahasiswa Angkatan tahun 2016. Atribut yang digunakan adalah nama, progdi, jenjang, angkatan, lulus, ipk, jenis kelamin, daerah asal, tahun masuk dan tahun lulus mahasiswa. Hasil pengujian 983 sampel data menggunakan algoritma decision

tree menunjukkan akurasi kecocokan sebesar 91,25% yang sangat efektif untuk prediksi kelulusan mahasiswa. Model yang dilakukan untuk memprediksi kelulusan mahasiswa berhasil menggunakan algoritma *Decision tree*. Adapun saran-saran yang dapat diberikan adalah: Penelitian selanjutnya diharapkan dapat menggunakan data yang lebih besar dan menggabungkan dengan algoritma lain untuk mencapai akurasi yang lebih baik dan dapat melakukan pengembangan atau membuat aplikasi prediksi dengan menggunakan metode yang berbeda dan platform yang berbeda.

5. Daftar Pustaka

- [1] Mashlahah, S., 2013. *Prediksi kelulusan mahasiswa menggunakan metode decision tree dengan penerapan algoritma C4. 5* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- [2] Thaniket, R., Kusriani, K. and Luthf, E.T., 2020. Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma Support Vector Machine. *Jurnal FATEKSA: Jurnal Teknologi dan Rekayasa*, 5(2), pp.20-29.
- [3] Rohman, A., 2015. Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 1(1).
- [4] Setiyani, L., Wahidin, M., Awaludin, D. and Purwani, S., 2020. Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review. *Faktor Exacta*, 13(1), pp.35-43.
- [5] Kadhim, Q.K., 2017. Classification of human skin diseases using data mining. *International Journal of Advanced Engineering Research and Science*, 4(1), p.237008.
- [6] Van Truong, N., Thai, D.T.M. and Le Bich Lien, N.T.T., 2017. Improving Performance Benchmark of Decision Tree Classifications for E-mail Spam Filtering. *seed*, 1(100), p.20.

- [7] Bisri, A. and Wahono, R.S., 2015. Penerapan Adaboost untuk penyelesaian ketidakseimbangan kelas pada Penentuan kelulusan mahasiswa dengan metode Decision Tree. *Journal of Intelligent Systems*, 1(1), pp.27-32.
- [8] Kristanto, O., 2014. Penerapan algoritma klasifikasi data mining ID3 untuk menentukan penjurusan siswa SMAN 6 Semarang. *Universitas Dian Nuswantoro, Semarang*.
- [9] Tyasti, A.E., Ispriyanti, D. and Hoyyi, A., 2015. Algoritma Iterative Dichotomiser 3 (Id3) Untuk Mengidentifikasi Data Rekam Medis (Studi Kasus Penyakit Diabetes Mellitus Di Balai Kesehatan Kementerian Perindustrian, Jakarta). *Jurnal Gaussian*, 4(2), pp.237-246.
- [10] Wang, Y. and Priestley, J.L., 2017. Binary classification on past due of service accounts using logistic regression and decision tree. *Grey Lit. from PhD Candidates*, vol. 4, 2017.
- [11] Romadhona, A., Suprapedi, S. and Himawan, H., 2017. Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree. *Jurnal Cyberku*, 13(1), pp.8-8.
- [12] Salmu, S. and Solichin, A., 2017, April. Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta. In *Prosiding Seminar Nasional Multidisiplin Ilmu* (Vol. 22).
- [13] Bukhori, A. and Pratiwi, N., 2018. Implementasi Metode Decision Tree dengan Algoritma ID3 dan C4. 5 untuk Mengklasifikasikan Partisipasi Perempuan Nikah dalam Kegiatan Ekonomi Rumah Tangga di DIY. *Jurnal Statistika Industri dan Komputasi*, 3(02), pp.22-32.
- [14] Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S. and Zhou, Z.H., 2008. Top 10 algorithms in data mining. *Knowledge and information systems*, 14, pp.1-37.
- [15] Maimon, O.Z. and Rokach, L., 2014. *Data mining with decision trees: theory and applications* (Vol. 81). World scientific.
- [16] Andie, A., 2016. Penerapan Decision Tree Untuk Menganalisis Kemungkinan Pengunduran Diri Calon Mahasiswa Baru. *Technologia: Jurnal Ilmiah*, 7(1).