



Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa

Rovidatul Hikmah Tanjung¹, Yuhandri Yunus², Gunadi Widi Nurcahyo³

Email: ¹rovidatulh@gmail.com, ²yuhandri.yunus@gmail.com, ³gunadiwidi@yahoo.com

¹Teknik Informatika, Ilmu Komputer, Universitas Putra Indonesia YPTK Padang

Diterima: 05 April 2020 | Direvisi: 05 Mei 2020 | Disetujui: 27 Mei 2020

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Manajemen perguruan tinggi membutuhkan prediksi kelulusan untuk menentukan langkah-langkah pencegahan dini kasus *drop out*. Lama masa studi mahasiswa dapat disebabkan karena berbagai faktor sehingga perlu diketahui siapa saja mahasiswa yang berpotensi lulus tidak tepat waktu. Teknik data mining dapat digunakan untuk menggali suatu pengetahuan baru sehingga dapat menghasilkan prediksi kelulusan mahasiswa. Beberapa Algoritma yang dapat digunakan adalah algoritma C4.5 dan Naive Bayes. Tujuan penelitian ini adalah untuk memprediksi kelulusan mahasiswa Fakultas Ilmu Sosial dan Ilmu Politik Universitas Andalas menggunakan algoritma C4.5 dan Naive Bayes. Atribut yang digunakan yaitu umur masuk kuliah, jenis kelamin, indeks prestasi semester 1-4. Adapun data yang digunakan adalah data mahasiswa FISIP yang lulus di tahun 2022 pada tingkat strata 1. Setelah data diperoleh dilakukan pembersihan dan seleksi sehingga dataset yang diolah sebanyak 378 mahasiswa. Kemudian pengujian dilakukan dengan membagi data *training* dan data *testing* 70%:30%, 80%:20%, dan 90%:20%. Hasil penelitian menunjukkan tingkat akurasi algoritma Naive Bayes lebih baik dari C4.5 dengan akurasi tertinggi 81,58%. Hasil prediksi pada mahasiswa yang lulus terlambat diharapkan dapat menjadi evaluasi sehingga menghasilkan lulusan yang tepat waktu.

Kata kunci: *Prediksi Kelulusan Mahasiswa, Algoritma C4.5, Algoritma Naive Bayes*

Comparison of c4.5 and naive bayes algorithms in predicting student graduation

Abstract

College management requires graduation predictions to determine early prevention measures for drop out cases. The length of a student's study period can be caused by various factors, so it is necessary to know which students have the potential to graduate not on time. Data mining techniques can be used to explore new knowledge so that it can produce predictions of student graduation. Some of the algorithms that can be used are the C4.5 and Naive Bayes algorithms. The purpose of this study was to predict the graduation of students from the Faculty of Social and Political Sciences at Andalas University using the C4.5 and Naive Bayes algorithms. The attributes used are age at college, gender, grade point average 1-4. The data used is data from FISIP students who graduate in 2022 at the strata 1 level. After the data is obtained, cleaning and selection are carried out so that the dataset that is processed is 378 students. Then testing is done by dividing training data and testing data 70%:30%, 80%:20%, and 90%:20%. The results showed that the accuracy of the Naive Bayes algorithm was better than C4.5 with the highest accuracy of 81.58%. The results of predictions on students who graduate late are expected to be an evaluation so as to produce graduates on time.

Keywords: *Prediction of Student Graduation, C4.5 Algorithm, Naive Bayes Algorithm*

1. PENDAHULUAN

Dewasa ini perkembangan ilmu pengetahuan dan teknologi ditandai dengan kemajuan di bidang media informasi dan teknologi. Salah satu pemanfaatan teknologi yang banyak digunakan adalah database yang terkomputerisasi. Pemanfaatan teknologi informasi dianggap sebagai strategi yang sangat jitu untuk unggul dalam bersaing, termasuk dalam dunia pendidikan. Universitas

Andalas termasuk salah satu perguruan tinggi yang menyimpan datanya dalam database yang terkomputerisasi. Data yang disimpan berupa data mahasiswa, data dosen, serta data-data yang ada hubungannya dengan Universitas Andalas.

Data yang disimpan tidak terlalu banyak memiliki kegunaan dan hanya akan disimpan secara terus menerus. Data tersebut hanya diperlukan saat universitas membutuhkan suatu informasi tertentu atau saat proses akreditasi dilakukan. Terlebih jika mahasiswa telah lulus maka data mahasiswa akan semakin jarang digunakan dan akan menjadi arsip yang bertambah setiap tahunnya. Padahal data lulusan sangat berpengaruh terhadap akreditasi jurusan di masa depan.

Berdasarkan standar pengukuran akreditasi program studi di Indonesia yang dilakukan oleh Badan Akreditasi Nasional Perguruan Tinggi atau BAN PT salah satu poin akreditasi adalah mahasiswa yang lulus tepat waktu di mana Program Sarjana (S1) dirancang untuk menyelesaikan studi dalam waktu delapan semester. Oleh karena itu data lulusan perlu dipelajari lebih lanjut sehingga program studi mampu menghasilkan mahasiswa yang lulus tepat waktu dan dapat mempertahankan akreditasi program studi di masa depan.

Berkaitan dengan kondisi tersebut data lulusan bisa dimanfaatkan untuk menggali suatu informasi yang berguna. Salah satu cara untuk memanfaatkan data mahasiswa yang lulus ini adalah dengan mengolahnya dengan Data Mining. Data Mining merupakan sebuah dasar dan langkah penting dalam menemukan pengetahuan baru pada sebuah database sehingga menghasilkan informasi yang sangat berharga [1].

Penelitian tentang Data Mining sudah banyak dilakukan, seperti penelitian yang dilakukan untuk mendeteksi penipuan kartu kredit dengan membandingkan algoritma klasifikasi pohon keputusan Naive Bayes, C4.5 dan Bagging Ensemble Machine Learning, sehingga algoritma dengan performa terbaik adalah C4.5 [2]. Algoritma C4.5 digunakan untuk memprediksi pekerjaan lulusan sehingga diketahui faktor-faktor yang berpengaruh dalam keberhasilan lulusan dalam lapangan kerja [3]. Algoritma C4.5 juga digunakan untuk prediksi hasil belajar siswa pada Masa Pandemi COVID-19 [4]. Kemudian Decision Tree merupakan salah satu metode klasifikasi [5] digunakan untuk memprediksi waktu kelulusan mahasiswa [6] dan dikombinasi dengan Algoritma Naive Bayes dan K-Nearest Neighbor [7][8] [9]

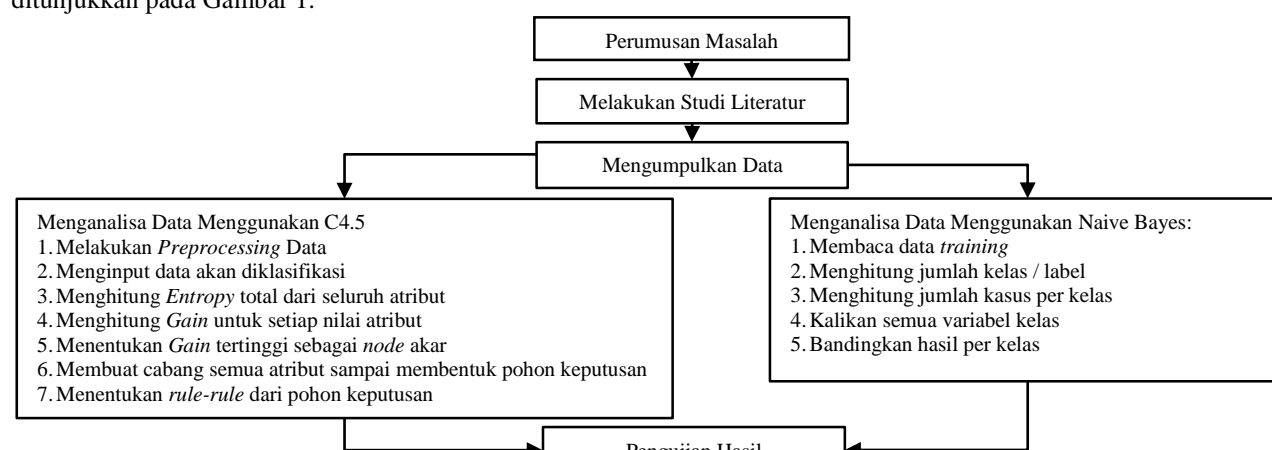
Penelitian lainnya juga memprediksi kelulusan mahasiswa menggunakan algoritma Naive Bayes [10][11][12] lalu dibandingkan dengan Neural Network dan dihasilkan akurasi Neural Network lebih tinggi dari Naive Bayes [13]. Selain itu prediksi kelulusan mahasiswa juga dilakukan dengan menggunakan algoritma Support Vector Machine [14] dan hybrid decision support system [15].

Selain untuk memprediksi kelulusan, data mining juga digunakan untuk klasifikasi jurusan siswa [16]. Untuk meningkatkan keakuratan klasifikasi penentuan jurusan, beberapa peneliti mengkombinasikan algoritma C4.5 dengan Gradient Boosting Trees, Random Forests, dan Deep Learning [17]

Pada penelitian ini akan dilakukan perbandingan algoritma akurasi Naive Bayes dan C4.5 untuk memprediksi kelulusan mahasiswa. Algoritma pohon keputusan menunjukan prediksi yang lebih tinggi dari pada yang lain [18] [19]. Selain itu beberapa penelitian terdahulu juga sudah banyak menggunakan Naive Bayes dan C4.5 dalam mengekstrak pengetahuan baru.

2. METODE PENELITIAN

Penelitian ini menggunakan klasifikasi algoritma C4.5 dan Naive Bayes. Adapun tahapan dalam melakukan penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Metodologi penelitian

Langkah-langkah yang dilakukan di metodologi penelitian pada Gambar 1 dapat dijelaskan sebagai berikut:

2.1 Mengidentifikasi Masalah

Tahap identifikasi masalah merupakan tahapan awal penelitian. Hal ini dilakukan dengan mengidentifikasi masalah dimulai dengan cara mengamati lebih dalam dan menggali permasalahan yang ada dari masalah yang ditemukan pada objek penelitian.

2.2 Menganalisa Masalah

Tahap ini peneliti melakukan beberapa cara dan metode dalam menganalisa masalah. Pada tahap ini data yang akan dikumpulkan, disusun, dikelompokkan, dianalisa sehingga diperoleh beberapa gambaran yang jelas pada masalah penelitian dan diharapkan masalah dapat dipahami dengan baik dan benar.

2.3 Melakukan Studi Literatur

Pada tahap ini dilakukan pencarian studi literatur mengenai materi yang berhubungan dengan penelitian seperti data mining, klasifikasi, algoritma Naive Bayes, algoritma C4.5 serta cara-cara pengujian tingkat akurasi. Pencarian literatur didasarkan pada penelitian-penelitian terdahulu, buku, artikel, jurnal ilmiah serta bacaan lain yang berkaitan dengan penelitian ini.

2.4 Mengumpulkan Data

Data dikumpulkan dengan melakukan observasi langsung di Fakultas Ilmu Sosial dan Ilmu Politik Universitas Andalas. Pengamatan secara langsung di lokasi penelitian untuk melihat data yang akan di proses. Data penelitian ini bersumber dari data mahasiswa yang lulus di tahun 2022 pada tingkat Strata 1 kemudian digabungkan dengan data Indeks Prestasi Semester yang bersumber dari Sistem Informasi Akademik.

2.5 Menganalisa Data Menggunakan C4.5

Langkah yang dilakukan pada tahap ini adalah menganalisa data yang sudah dikumpulkan sehingga terbentuk *rule-rule* yang akan menjadi keputusan. Sebelum dilakukan pengolahan, data tersebut dilakukan *preprocessing* terlebih dahulu dengan melakukan *cleaning* dan *transformation*. *Cleaning data* dilakukan untuk menentukan data-data mana saja yang akan digunakan karena pada data awal terdapat banyak data dan tidak semua atribut yang ada akan diproses. Untuk itu penulis melakukan *cleaning data*, seperti pada data tersebut terdapat NIM, tempat lahir, tanggal lahir, kab/kota asal, angkatan, program studi, jenjang, lama studi, dan IPK dilakukan penghapusan karena atribut tersebut tidak diperlukan dalam penelitian ini. Selain itu *cleaning data* juga dilakukan pada data mahasiswa yang bukan lulusan S1 sehingga selain lulusan S1 akan dihapus. Pada tahap *transformation* jenis data diubah agar data dapat diolah pada *tools* yang digunakan.

Secara umum tahapan dalam algoritma C4.5 untuk membangun sebuah pohon keputusan dengan memilih atribut sebagai akar, kemudian cabang untuk tiap-tiap nilai, membagi kasus dalam cabang, dan mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama dan membentuk pohon keputusan.

Pemilihan atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan berikut:

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

Di mana S adalah himpunan kasus, A adalah atribut, n adalah jumlah partisi atribut A, $|S_i|$ adalah jumlah kasus pada partisi ke-I, dan $|S|$ adalah jumlah kasus S.

Sementara itu, sebelum mencari nilai gain terlebih dahulu mencari nilai entropi. Penghitungan nilai entropi pada algoritma C4.5 dapat dilihat pada persamaan berikut:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Di mana p_i adalah porsi atau rasio antara jumlah S_i dengan jumlah semua kasus S

2.6 Menganalisa Data Menggunakan Naive Bayes

Data yang digunakan pada tahap ini sama dengan data yang digunakan pada analisa menggunakan algoritma C4.5. Naive Bayes adalah suatu model independen yang membahas mengenai klasifikasi sederhana berdasarkan teorema Bayes. Naive Bayes merupakan suatu algoritma yang dapat mengklasifikasikan suatu variable tertentu dengan menggunakan metode probabilitas dan statistik [20]. Secara garis besar algoritma dijelaskan seperti persamaan berikut:

$$P(H | X) = \frac{P(X|H)}{P(X)} P(H) \quad (3)$$

Di mana X adalah data dengan kelas yang belum diketahui, H adalah hipotesis data X merupakan suatu kelas spesifik, $P(H|X)$ adalah probabilitas hipotesis H berdasarkan kondisi X, $P(H)$ adalah probabilitas hipotesis H, $P(X|H)$ adalah probabilitas X berdasarkan kondisi H, dan $P(X)$ adalah probabilitas dari X.

2.7 Pengujian Hasil

Tahap selanjutnya pada penelitian ini adalah melakukan pengujian hasil penelitian yang telah dilakukan dengan software Rapid Miner. Pengujian dilakukan dengan menghitung akurasi algoritma C4.5 dan Naive Bayes menggunakan data *training* dan *testing* 70%:30%, 80%:20%, dan 90%:10%. Nilai akurasi didapatkan dari Confusion Matrix dengan menghitung jumlah data benar terhadap jumlah seluruh data. Setelah akurasi dari masing-masing algoritma didapatkan maka dibandingkan hasil akurasi algoritma C4.5 dan Naive Bayes sehingga didapatkan algoritma dengan akurasi terbaik dalam memprediksi kelulusan mahasiswa.

3. HASIL DAN PEMBAHASAN

Data dalam penelitian ini bersumber dari data mahasiswa Fakultas Ilmu Sosial dan Ilmu Politik pada tingkat Strata 1 pada tahun 2022 sebanyak 436 mahasiswa. Data tersebut adalah data mentah yang belum dilakukan *preprocessing*. Pada *preprocessing* terdapat dua tahap yang dilalui yaitu *cleaning* dan *transformation*. Pada tahap *cleaning* dilakukan penghapusan pada data mahasiswa selain jenjang S1 dan penghapusan atribut yang tidak diperlukan dalam penelitian sehingga menyisakan atribut Umur Masuk Kuliah, Jenis Kelamin, Indeks Prestasi Semester (IPS)1, IPS2, IPS3, dan IPS4. Pada tahap *transformation* dilakukan pada beberapa jenis data yang sifatnya numerik yaitu umur masuk kuliah, IPS1, IPS2, IPS3, dan IPS4. Berikut adalah dataset pada Tabel 1 yang akan diolah sebanyak 378 data mahasiswa.

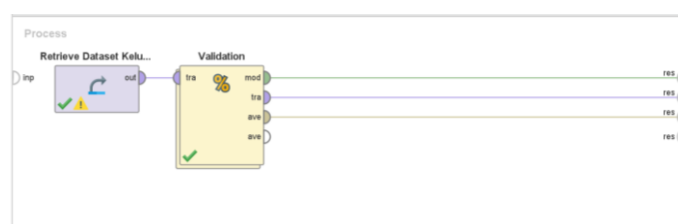
Tabel 1. Dataset prediksi kelulusan mahasiswa

No	Umur Masuk Kuliah	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	Status Lulus
1	Lebih	Laki-laki	Kecil	Kecil	Cukup	Kecil	Terlambat
2	Lebih	Laki-laki	Kecil	Cukup	Cukup	Cukup	Terlambat
3	Lebih	Laki-laki	Cukup	Kecil	Cukup	Kecil	Terlambat
4	Lebih	Perempuan	Cukup	Cukup	Cukup	Tinggi	Terlambat
5	Lebih	Perempuan	Tinggi	Cukup	Cukup	Tinggi	Terlambat
..
377	Sesuai	Perempuan	Cukup	Cukup	Tinggi	Tinggi	Tepat
378	Sesuai	Perempuan	Tinggi	Cukup	Tinggi	Tinggi	Tepat

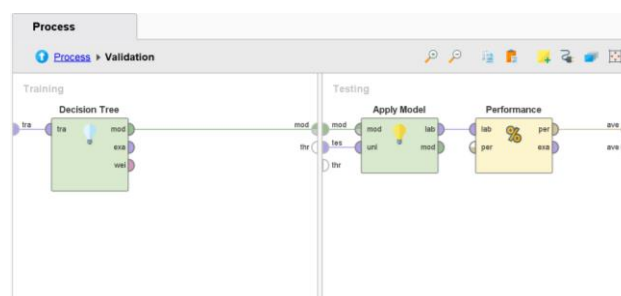
Berdasarkan Tabel 1 dapat dilihat atribut Umur masuk kuliah dihitung dari tahun lahir sampai tahun mahasiswa mendaftarkan diri di Universitas, Jenis kelamin yang digunakan adalah Laki-laki dan Perempuan, atribut IPS1-4 dibagi menjadi tiga kategori yaitu tinggi bagi Indeks prestasi semester diatas 3,50, cukup antara 2,75 – 3,50, dan kecil dibawah 2,75. Sedangkan atribut status lulus mejadi label target dengan kategori Tepat bagi masa studi kurang atau sama dengan 4 tahun dan Terlambat bagi masa studi lebih dari 4 tahun.

3.1 Menganalisa Data Menggunakan C4.5

Setelah melakukan *preprocessing* data, dilakukan analisa menggunakan algoritma C4.5 dengan *tools* RapidMiner. Model algoritma C4.5 diuji dengan *split validation*, yaitu teknik yang digunakan untuk membagi data menjadi data *training* dan data *testing*.



Gambar 1. Proses pengambilan data untuk algoritma C4.5 dan Naive Bayes



Gambar 2. Pemrosesan data dengan algoritma C4.5

Gambar 1 *retrieve dataset* diambil dari *repository* yang digunakan untuk proses pengujian, kemudian dihubungkan dengan operator *split validation* yang memiliki dua subproses: subproses pelatihan dan subproses pengujian. Subproses pelatihan

digunakan untuk melatih model. Model yang terlatih kemudian diterapkan dalam subproses pengujian. Kinerja model diukur selama fase pengujian.

Pada Gambar 2 dibagi menjadi dua bagian yaitu *training* dan *testing*. Pada proses *training* model klasifikasi yang digunakan *Decision Tree*. Selanjutnya pada proses *testing* operator *apply model* digunakan untuk menerapkan model pada proses *training* dan *performance* digunakan untuk mengukur kinerja akurasi dari model.

Untuk mendapatkan hasil akurasi yang baik peneliti menguji akurasi dengan pembagian data *training* dan *testing* 70%:30%, 80%:20%, dan 90%:10. Pengujian akurasi didasarkan pada nilai *Confussion Matrix* untuk melihat jumlah data yang diprediksi benar dan salah. Berikut adalah hasil akurasi dari data testing 30%.

Tabel 2. *Confussion matrix* data 70%:30%

Label	True Terlambat	True Tepat
Prediksi Terlambat	26	10
Prediksi Tepat	16	62

Berdasarkan Tabel 2 dapat dilihat bahwa pada nilai terlambat sebanyak 26 yang diprediksi dengan benar dan 10 yang tidak. Sedangkan pada nilai tepat terdapat 16 yang diprediksi dengan benar dan 62 data yang tidak. Untuk mencari akurasi digunakan rumus sebagai berikut:

$$Akurasi = \frac{Jumlah\ data\ uji\ benar}{Jumlah\ semua\ data\ uji} * 100\% \quad (3)$$

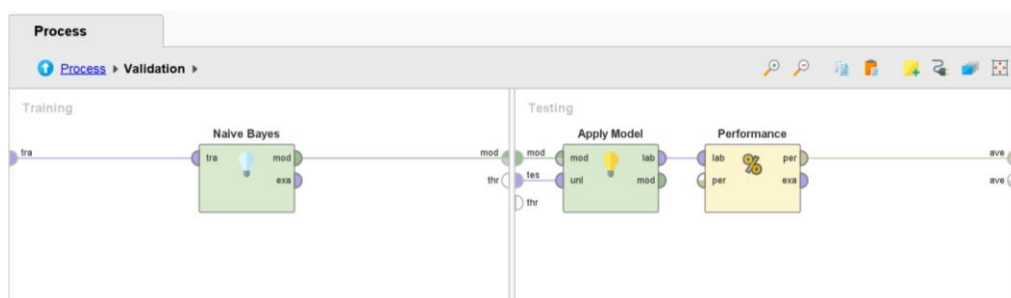
$$Akurasi = \frac{88}{114} * 100\%$$

$$Akurasi = 77,19\%$$

Akurasi yang dihasilkan untuk data *training* dan data *testing* 70%:30% adalah 77,19% dengan *error* 22,81%. Nilai error didapatkan dari selisih akurasi. Hasil pengujian akurasi pada data 80%:20%, dan 90%:10 dapat dilihat pada Tabel 4.

3.2 Menganalisa Data Menggunakan Naive Bayes

Data yang digunakan pada tahap ini sama dengan tahap sebelumnya pada algoritma C4.5 yang dianalisa menggunakan *tools* RapidMiner. Model algoritma Naive Bayes diuji dengan *split validation* dengan pembagian data *training* dan data *testing* yang sama yaitu 70%:30%, 80%:20%, dan 90%:10.



Gambar 3. Pemrosesan data dengan algoritma Naive Bayes

Sebagaimana yang sudah dijelaskan pada tahap sebelumnya bahwa pada proses *validation* terbagi menjadi dua bagian yaitu proses *training* dan *testing*. Pada Gambar 3 proses *training* disesuaikan dengan algoritma yang digunakan, dalam hal ini peneliti menggunakan algoritma Naive Bayes.

Tabel 3. *Confussion matrix* data 70%:30%

Label	True Terlambat	True Tepat
Prediksi Terlambat	28	9
Prediksi Tepat	14	63

Berdasarkan Tabel 3 diperoleh nilai terlambat yang diprediksi dengan benar adalah 28 dan 9 yang tidak. Sedangkan pada nilai tepat yang diprediksi dengan benar adalah 63 dan 14 data yang tidak. Untuk mencari akurasi digunakan perhitungan jumlah data benar terhadap jumlah data yang diuji. Rumus untuk mencari akurasi sama dengan sebelumnya sehingga dihasilkan sebagai berikut:

$$Akurasi = \frac{91}{114} * 100\%$$

$$Akurasi = 79,82\%$$

Hasil akurasi yang diperoleh dari algoritma Naive Bayes sebesar 79,82%, sedangkan *error* yang diperoleh dalam pengujian sebesar 20,18%. Selain pada pembagian data *training* 70% dan data *testing* 30%, pengujian akurasi juga dilakukan pada pembagian 80%:20% dan 90%:10% dengan rumus yang sama. Hasil pengujian akurasi akan dipaparkan pada Tabel 4.

3.3 Pengujian Hasil

Setelah penulis melakukan semua tahapan dalam pengujian menggunakan algoritma C4.5 dan Naive Bayes, maka penulis akan membandingkan hasil dari analisis dalam bentuk tabel. Adapun yang akan dibandingkan adalah nilai akurasi dari algoritma C4.5 dan Naive Bayes dengan RapidMiner dalam memprediksi kelulusan mahasiswa. Berikut adalah hasil akurasi pada Tabel 4 di bawah ini:

Tabel 4. Perbandingan hasil akurasi algoritma

No	Algoritma	Data Training (%)	Data Testing (%)	Akurasi
1	C4.5	70	30	77.19%
		80	20	78.95%
		90	10	73.68%
2	Naive Bayes	70	30	79.82%
		80	20	81.58%
		90	10	81.58%

Berdasarkan Tabel 4 dihasilkan nilai akurasi C4.5 dan algoritma Naive Bayes dengan perbandingan data *training* dan data *testing* 70%:30%, 80%:20%, dan 90%:20%. Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual yang dimasukkan ke dalam sistem. Akurasi algoritma C4.5 dengan data *training* dan data *testing* 70%:30% adalah sebesar 77.19%. Pada data 80%:20% dihasilkan akurasi sebesar 78,95%. Untuk data selanjutnya 90%:10% terjadi penurunan akurasi yaitu 73,68%.

Selanjutnya pada algoritma Naive Bayes dengan data *training* dan data *testing* 70%:30% dihasilkan nilai akurasi sebesar 79,82%, pada data 80%:20% dan 90%:10% dihasilkan akurasi yang sama yaitu 81,58%. Setelah mendapatkan hasil perbandingan akurasi maka dapat dinyatakan bahwa algoritma C4.5 dan Naive Bayes layak digunakan sebagai penentuan status kelulusan mahasiswa Fakultas Ilmu Sosial dan Ilmu Politik Universitas Andalas. Data yang diprediksi lulus terlambat diharapkan dapat dilakukan evaluasi agar mahasiswa dapat lulus tepat waktu.

4. KESIMPULAN

Hasil implementasi algoritma C4.5 dan Naive Bayes layak digunakan untuk menentukan kelulusan mahasiswa Fakultas Ilmu Sosial dan Ilmu Politik Universitas Andalas. Dari pengujian akurasi menggunakan *confussion matrix* dengan data 70%:30%, 80%:20% dan 90%:10% dihasilkan akurasi terbaik adalah algoritma Naive Bayes. Hal ini dibuktikan dari nilai akurasi yang lebih tinggi dibandingkan C4.5 pada pengujian data 70%:30%, 80%:20% dengan nilai 79.82%, 81,58% dan pada data 90%:10% akurasi algoritma Naive Bayes tetap sedangkan C4.5 mengalami penurunan.

DAFTAR PUSTAKA

- [1] Suntoro, Joko. DATA MINING: Algoritma dan Implementasi dengan Pemrograman php. Elex Media Komputindo, 2019.
- [2] Ademi Husejinovic, "Credit Card Fraud Detection Using Naive Bayesian and C4.5 Decision Tree Classifiers," *Periodicals of Engineering and Natural Sciences*, Jan. 01, 2020. <https://doi.org/10.21533/pen.v%25vi%25i.300>.
- [3] F. Zhou, L. Xue, Z. Yan, and Y. Wen, "Research on college graduates employment prediction model based on C4.5 algorithm," *J Phys Conf Ser*, vol. 1453, p. 12033, 2020, doi: 10.1088/1742-6596/1453/1/012033.
- [4] Y. Fitriani, S. Defit, and G. W. Nurcahyo, "Prediksi Hasil Belajar Siswa Secara Daring pada Masa Pandemi COVID-19 Menggunakan Metode C4.5," *Jurnal Sistim Informasi dan Teknologi*, pp. 120–127, Aug. 2021, doi: 10.37034/jsisfotek.v3i3.54.
- [5] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, "Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 2, no. 2, pp. 91–97, Dec. 2021, doi: 10.37859/coscitech.v2i2.2851.
- [6] C. N. Dengen, K. Kusri, and E. T. Luthfi, "Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *SISFOTENIKA*, vol. 10, no. 1, p. 1, Jan. 2020, doi: 10.30700/jst.v10i1.484.
- [7] A. Budiyantera, E. Prengki, P. Ahmad Pratama, N. Wiliani, T. Informatika, and S. Widuri, "Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu," *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, vol. 5, 2020, doi: <https://doi.org/10.33480/jitk.v5i2.1214>.
- [8] Hozairi, Hozairi, Anwari Anwari, and Syarif Alim. "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes." *Network Engineering Research Operation* 6, no. 2, 2021.: 133-144. doi: <http://dx.doi.org/10.21107/nero.v6i2.237>.
- [9] Sutoyo, Edi, and Ahmad Almaarif. "Educational data mining for predicting student graduation using the naive bayes classifier algorithm." *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)* 4, no. 1, 2020, 95-101. doi: <https://doi.org/10.29207/resti.v4i1.1502>.
- [10] Nurul Khasanah, Agus Salim, Nurul Afni, Rachman Komarudin, and Yana Iqbal Maulana, "Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes," 2022. doi: <http://dx.doi.org/10.31602/tji.v13i3.7312>.
- [11] J. Galopo Perez and E. S. Perez, "Modern Education and Computer Science," *Modern Education and Computer Science*, vol. 3, pp. 57–67, 2021, doi: 10.5815/ijmecs.2021.03.05.

- [12] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Sattriedi Wahyu Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 234–241, Dec. 2022, doi: 10.37859/coscitech.v3i3.4343.
- [13] A. Armansyah and R. K. Ramli, "Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 6, no. 1, pp. 1–10, Jun. 2022, doi: 10.29408/edumatic.v6i1.4789.
- [14] D. Rosita, dan Syamsuddin Mallala, S. Informasi, S. Widya Cipta Dharma, T. Informatika, and P. Korespondensi, "Komparasi Data Mining Naive Bayes Dan Neural Network Memprediksi Masa Studi Mahasiswa S1," vol. 7, no. 3, pp. 443–452, 2020, doi: 10.25126/jtiik.202072093.
- [15] E. Haryatmi and S. Pramita Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 386–392, Apr. 2021, doi: 10.29207/resti.v5i2.3007.
- [16] D. Guswandi, M. Yanto, M. Hafizh, and L. Mayola, "Analisis Hybrid Decision Support System dalam Penentuan Status Kelulusan Mahasiswa," *Jurnal Resti*, 2021. <http://jurnal.iaii.or.id/index.php/RESTI/article/view/3587/517>.
- [17] Putra, M.Y. and Putri, D.I., "Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI" *Jurnal Tekno Kompak*, 16(2), pp.176-187, 2022. doi:10.33365/jtk.v16i2.2002.
- [18] S. Mutrofin, M. Mughniy Machfud, D. H. Satyareni, R. Venantius, H. Ginardi, and C. Fatichah, "Komparasi Kinerja Algoritma C4.5, Gradient Boosting Trees, Random Forests, Dan Deep Learning Pada Kasus Educational Data Mining" *Jurnal Teknologi Informasi dan Ilmu Komputer* 7, no. 4, 2020. doi: 10.25126/jtiik.2020732665.
- [19] L. I. P. Aji and A. Sunyoto, "An Implementation of C4.5 Classification Algorithm to Analyze student's Performance," in *2020 3rd International Conference on Information and Communications Technology, ICOIACT 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 126–130. doi: 10.1109/ICOIACT50329.2020.9332088.