

Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM

Zaenul Amri ^{1,*}, Kusrini ¹, Kusnawi ¹

¹ Program Studi Teknik Informatika, Universitas Amikom Yogyakarta, Indonesia

* Correspondence: harazukuamry@gmail.com

Copyright: © 2023 by the authors

Received: 17 Juni 2023 | Revised: 3 Juli 2023 | Accepted: 13 Juli 2023 | Published: 20 Desember 2023

Abstrak

Tingkat lulusan mahasiswa di semua perguruan tinggi dapat diukur dengan melihat lama studinya, baik yang tepat waktu maupun yang terlambat. Sehingga, dengan melihat lama studi tersebut dapat mempengaruhi kualitas program studi di perguruan tinggi. Tujuan penelitian ini adalah menerapkan dan mengkomparasi model *algoritma naïve bayes*, *decision tree*, *artificial neural network*, *k-nearest neighbor* (K-NN), *support vector machine* (SVM) dalam melakukan prediksi tingkat kelulusan mahasiswa. Dataset yang digunakan dalam penelitian ini berjumlah 807 data mahasiswa fakultas teknik Universitas Hamzanwadi. Teknik analisis data yang digunakan adalah *statistic deskriptif* dengan menerapkan metode *knowledge discovery in database* (KDD). Pengujian kelima algoritma tersebut dilakukan optimasi data dengan teknik *smotenn* dan *split* data menggunakan data *training* 80% dan data *test* 20% dengan *random state* 42. Hasil temuan kami menunjukkan bahwa *algoritma naïve bayes* memiliki akurasi sebesar 92,37%, *decision tree* 91,60%, KNN 96,95%, SVM 93,13% dan ANN 90,84%. pengujian kelima *algoritma* tersebut, algoritma KNN memiliki tingkat akurasi terbaik sebesar 96,95%. Adapun hasil prediksi tersebut cenderung terlambat yang dipengaruhi oleh IPK. Oleh karena itu Lembaga perlu memberi perhatian lebih terhadap mahasiswa yang diprediksi terlambat untuk meningkatkan IPK di setiap semester, sehingga dapat meningkatkan kelulusan tepat waktu sesuai rentang waktu yang diharapkan.

Kata kunci: ann; decision tree; knn; naïve bayes; svm

Abstract

The student graduation rate in all universities can be measured by looking at their study duration, both on time and delayed. Thus, by observing the study duration, it can affect the quality of study programs in universities. The purpose of this research is to apply and compare the Naïve Bayes, Decision Tree, Artificial Neural Network, K-Nearest Neighbor (K-NN), and Support Vector Machine (SVM) algorithms in predicting the graduation rate of students. The dataset in this research consisted of 807 student data from the Faculty of Engineering, Universitas Hamzanwadi. The data analysis technique used was descriptive statistics by applying the knowledge discovery in a database (KDD) method. The testing of the five algorithms was done by optimizing the data using the SMOTEENN technique, with a data split of 80% for training and 20% for testing, using a random state of 42. Our findings show that the Naïve Bayes algorithm had an accuracy of 92.37%, Decision Tree 91.60%, K-NN 96.95%, SVM 93.13%, and ANN 90.84%. Among the five algorithms tested, the K-NN algorithm had the highest accuracy rate of 96.95%. The predicted results tended to indicate delayed graduation influenced by GPA. Therefore, the institution needs to pay more attention to students predicted to be delayed to improve their GPA each semester, thus promoting timely graduation within the expected time frame.

Keywords: ann; decision tree; knn; naïve bayes; svm



PENDAHULUAN

Perguruan tinggi merupakan Lembaga Pendidikan tinggi yang menyediakan Pendidikan pada tingkat yang tinggi dengan tujuan untuk memberikan Pendidikan lebih mendalam kepada peserta didiknya. Dengan mempersiapkan peserta didik untuk menjadi anggota masyarakat yang memiliki kemampuan baik secara akademis maupun keterampilan. Pendidikan tinggi juga memiliki peran penting dalam perubahan ilmu pengetahuan (Sumardi & Efendi, 2021). Selain itu Lembaga pendidikan tinggi dituntut untuk menjadi penyedia pendidikan yang berkualitas. Salah satu instrumen yang digunakan oleh pemerintah untuk mengukur kualitas penyedia pendidikan adalah jumlah tingkat kelulusan. Oleh karena itu Lembaga Pendidikan diharapkan mampu meningkatkan kelulusan mahasiswanya tepat waktu sesuai rentang waktu yang diharapkan. Kelulusan mahasiswa tepat waktu menjadi suatu kebanggaan bagi Lembaga Pendidikan dikarenakan lulus tepat waktu dapat memberi peluang lebih untuk dapat pekerjaan bagi lulusannya. Untuk itu Lembaga Pendidikan perlu memberi perhatian terhadap lama studi mahasiswa. Lama studi mahasiswa adalah rentang waktu bagi mahasiswa untuk menyelesaikan studinya (Etriyanti et al., 2020). Salah satu indikator keberhasilan Program Studi dapat dilihat dari lama studi mahasiswa. Selain itu lama studi mencerminkan tingkat pencapaian mahasiswa dalam studinya. Dalam perspektif yang lebih luas rata-rata lama studi mahasiswa mempengaruhi kualitas program studi dan oleh karena itu lama studi mahasiswa dijadikan salah satu kriteria untuk menentukan penilaian akreditasi oleh Badan Akreditasi Nasional Perguruan Tinggi (Zainuddin, 2018). Lama studi mahasiswa hampir terjadi di setiap Lembaga Pendidikan, salah satunya pada Lembaga Pendidikan tinggi Universitas Hamzanwadi.

Data lulusan yang diperoleh melalui system informasi akademik (IOSYS) pada fakultas teknik didapatkan 807 data mahasiswa Fakultas Teknik tahun angkatan 2015 sampai 2018 terdapat 501 mahasiswa yang dapat menyelesaikan studinya tepat waktu sedangkan 306 mahasiswa tidak tepat waktu atau terlambat dari jumlah tersebut jika dipersentasikan mencapai 37,92% yang terlambat, angka tersebut jauh dari target kelulusan dan harapan Lembaga khususnya fakultas teknik. Akibatnya hal tersebut dapat menghambat mahasiswa baru untuk bergabung dengan lembaga karena kapasitas mahasiswa yang terbatas, sehingga daya tampung mahasiswa baru dan lama studi mahasiswa perlu diperhatikan (Bisri & Wahono, 2015). Mahasiswa merupakan aspek penting dalam perguruan tinggi yang harus diperhatikan dengan serius dalam melakukan evaluasi akademik pada program studi.

Mengatasi permasalahan keterlambatan tingkat kelulusan maka prediksi perlu dilakukan untuk mengetahui tingkat kelulusan mahasiswa sesuai rentang waktu yang diharapkan. Prediksi mahasiswa yang lulus tepat waktu dan terlambat merupakan proses keilmuan untuk mendapatkan *knowledge* secara berurutan berdasarkan bukti-bukti (Suardika, 2019). Maka dari itu, perlu adanya proses yang menggunakan teknik statistik, matematik, kecerdasan buatan (*Artificial Intelegent*) dan *machine learning* untuk mengestrak pengetahuan atau menemukan pola dari suatu data yang besar (Hasanah et al., 2021). salah satunya adalah teknik penambangan data (data mining).

Data mining adalah proses menemukan korelasi baru yang bermakna, pola dan tren dengan memilah-milah sejumlah data besar yang tersimpan dalam repositori, menggunakan teknologi penalaran pola serta teknik-teknik statistik dan matematika (Hidayat et al., 2021; Uska et al., 2020). Data mining menurut Juliansa (2019) adalah proses untuk mendapatkan ilmu pengetahuan dari sebuah informasi yang berasal dari gudang basis data. Teknik data mining merupakan cara yang mudah dan relatif cepat untuk memperoleh pengetahuan secara otomatis (Pramadhana, 2021; Rahayu et al., 2022; Suyanto, 2017; Takdirillah, 2020) dan pengetahuan abstrak dari sebuah database yang besar (Mulya, 2019; Rizky et al., 2021) yang meliputi bentuk dan/atau hubungan antar data.

Bagian penting dalam data mining adalah teknik klasifikasi untuk mengklasifikasikan data yaitu *knowledge discovery in database* (KDD) salah satu cara dalam menciptakan,

mengganti, serta menyempurnakan informasi serta pola yang bermakna dari database mentah yang digunakan untuk domain atau aplikasi yang berbeda (Rahayu et al., 2022; Sari et al., 2020). KDD sendiri merupakan pendekatan terprogram serta analitis memodelkan informasi dari database untuk mengekstrak pengetahuan yang bermanfaat serta berlaku. Informasi *mining* merupakan pusat dari suatu proses KDD yang memakai algoritma spesial untuk mengecek informasi, membangun model, dan menciptakan pola yang belum ditemui (Zanuardi & Suprayitno, 2018). Pada proses data mining yang biasa disebut *knowledge discovery database* (KDD) terdapat beberapa proses yaitu; seleksi data (*selection*), pemilihan data (*preprocessing/cleaning*), transformasi (*transformation*), *data mining*, dan interpretasi/evaluasi (*interpretation/evaluation*) (Nofriansyah & Nurcahyo, 2015). Adapun model *data mining* yang digunakan untuk klasifikasi data yaitu, *algoritma naïve bayes*, *decision tree*, *artificial neural network*, *K-nearest neighbor* (K-NN), *support vector machine* (SVM) dan lain-lain. Beberapa penelitian sebelumnya yang terkait dengan metode dan algoritma yang digunakan pada penelitian ini.

Penelitian oleh Kusrini & Prasetyo (2020) *prediction of student graduation with naïve bayes algorithm*. Data yang digunakan adalah data siswa tahun 2016. Dengan 2 model variabel pengukuran. Yaitu, model 1 variabel IP mahasiswa dan Model 2 untuk prediksi menggunakan 8 variabel yaitu NIM, Jenis Kelamin, Tempat Tinggal, IPS 1, IPS2, IPS3, IPS4. Penelitian mereka menggunakan teknik data *mining* dengan *algoritma naïve bayes*. Berdasarkan hasil temuan mereka, yang melakukan perbandingan menggunakan 2 model, dimana prediksi pada model 2 memberikan hasil yang lebih baik dibandingkan dengan model 1 dengan tingkat akurasi sebesar 89% sehingga dapat ditunjukkan bahwa model prediksi dengan menggunakan variabel IP, nilai UN, JK, status kependudukan dapat memberikan prediksi yang lebih baik. Oleh karena itu diperoleh nilai akurasi pada *algoritma Naïve Bayes* yang telah diterapkan dengan kategori cukup tinggi. Selain faktor atau variabel utama dalam menentukan kelulusan ini adalah variabel Indeks Prestasi (IP).

Selanjutnya, pada penelitian yang telah dilakukan oleh Samuel, Jonathan, & Naibaho (2019) dengan menerapkan metode *Decision Tree* J48 untuk memprediksi kelulusan mahasiswa tepat waktu. Dimana data yang digunakan berjumlah 206 data dengan 19 atribut. Hasil temuan mereka menunjukkan bahwa dari 19 atribut yang digunakan sebagai dataset, Cuti adalah atribut yang paling dominan atau adalah *root* teratas setelah dilakukan Analisa menggunakan algoritma *Decision Tree* j48 dengan hasil akurasi adalah 90,24%.

Sementara itu, penelitian oleh Putra & Walmi (2020) memprediksi produksi padi menggunakan *artificial neural network algoritma backpropagation*. Temuan yang diperoleh berupa rancangan optimal untuk melakukan prediksi yaitu dengan menggunakan multilayer. Hasil pengujian sistem prediksi produksi padi yang terdiri dari 75 kali pengujian pada di 19 daerah di Sumatera Barat, diperoleh tingkat akurasi mencapai 88,14% atau dengan tingkat error yang relatif rendah yaitu 11,86%. Selanjutnya penelitian yang dilakukan oleh Hakim et al., (2019) memprediksi kelulusan mahasiswa dengan model *K-Nearest Neighbor* (KNN). Data yang digunakan berjumlah 247 dengan 9 variabel diantaranya: NIM, Nama, Jenis kelamin, IPS 1 sampai 5 dan keterangan lulus. Dari hasil pengujian didapatkan 143 orang diprediksi lulus tepat waktu dan 104 orang diprediksi tidak lulus tepat waktu dengan tingkat akurasi sebesar 80%. Selain itu, faktor yang paling penting dalam menentukan kelulusan pada penelitian ini adalah indeks prestasi.

Penelitian selanjutnya oleh Abdullah, Kusrini, & Arief (2022) dengan menggunakan metode *Support Vector Machine* (SVM) untuk prediksi waktu kelulusan mahasiswa. Data yang digunakan pada penelitian ini berjumlah 807 dengan 9 atribut. dari 807 data terklasifikasi dengan benar oleh model sebanyak 697 data termasuk dalam kategori baik karena tingkat kesalahan sedikit, sehingga algoritma SVM dapat diimplementasikan dalam melakukan prediksi dengan akurasi yang di dapatkan sebesar 86,36%.

Berdasarkan hasil temuan sebelumnya, mereka telah menerapkan beberapa model algoritma seperti: *naïve bayes*, *decision tree*, *artificial neural network* (ANN), *k- nearest neighbor* (K-NN), dan SVM dalam melakukan prediksi tingkat kelulusan di berbagai perguruan tinggi. Namun pada penelitian yang telah mereka lakukan, hanya menerapkan beberapa model *algoritma* dan belum dilakukan optimasi pada setiap algoritma yang digunakan serta melakukan perbandingan dengan beberapa *algoritma* untuk mengetahui *algoritma* mana yang memiliki akurasi terbaik dengan menggunakan data yang sama dengan menentukan data training dan data uji dari data yang digunakan. Sehingga penelitian ini bertujuan untuk menerapkan kelima *algoritma* dengan melakukan optimasi pada data serta membandingkan kelima *algoritma* tersebut yang terdiri dari *naïve bayes*, *decision tree*, ANN, KNN, dan SVM untuk mengetahui *algoritma* mana yang terbaik dalam melakukan prediksi.

METODE

Jenis penelitian yang digunakan adalah eksperimen dimana peneliti melakukan sebuah eksperimen penelitian dengan menentukan variabel target dan variabel prediktor pada dataset yang akan digunakan. Dalam hal ini dataset tersebut dilakukan komparasi dengan algoritma Naïve Bayes, Decision Tree, Artificial Neural Network, K-Nearest Neighbor (K-NN), dan Support Vector Machine (SVM) dalam memprediksi tingkat kelulusan mahasiswa. Data diperoleh langsung melalui sistem informasi akademik (IOSYS) pada bagian akademik dan data tersebut termasuk data private yang belum pernah digunakan dalam penelitian-penelitian sebelumnya. Data yang digunakan pada penelitian ini berupa dataset dari data mahasiswa fakultas teknik Angkatan 2015– 2018 berjumlah 807 data dengan 11 variabel diantaranya; NIM, Nama, Jenis kelamin, Usia, Indek Prestasi Semester (IPS) 1 – 4, Indek Prestasi Kumulatif (IPK) semester 4 dan Keterangan Kelulusan. Metode analisis data mengikuti tahapan yang ada pada proses *knowledge discovery in database* (KDD). Proses KDD dimulai dengan menetapkan tujuan dan diakhiri dengan evaluasi (Nofriansyah & Nurcahyo, 2015). Adapun metode analisis data mengikuti tahapan yang ada pada proses *knowledge discovery in database* (KDD) yang digunakan pada penelitian ini dengan menggunakan *tools software excel* dan *jupyter notebook*.

Pada tahap pertama dilakukan seleksi data mahasiswa fakultas Teknik Angkatan 2015 sampai 2018 dengan data yang di dapatkan melalui sistem informasi akademik (IOSYS) pada bagian akademik yaitu, data awal mahasiswa perangkatan, data nilai mahasiswa perangkatan, dan data keterangan kelulusan mahasiswa sesuai dengan rentang waktu yang diharapkan yaitu 8 (delapan) semester atau 4 (empat) tahun. Dari data tersebut dilakukan tahap seleksi data dengan mengambil variabel yang akan digunakan dalam proses prediksi diantaranya: NIM, nama, jenis kelamin, usia, indeks prestasi semester 1, 2, 3, 4, indek prestasi kumulatif (IPK) 4 dan Keterangan Lulus. Selanjutnya pada tahap kedua dilakukan pemilihan data atau *preprocessing/cleaning* data. Dari data tersebut dilakukan pembersihan data apabila terdapat data yang hilang, data ganda atau bersifat *outlier* dan melakukan penghapusan data variabel yang tidak digunakan seperti variable NIM dan Nama karena dirasa tidak *kontributif* untuk digunakan dalam melakukan prediksi. Setelah itu dilakukan transformasi data dengan dengan mengubah data variabel Jenis Kelamin dan keterangan lulus yang awalnya berbentuk teks (*string*) menjadi angka (*numerik*). Kemudian dilakukan teknik data mining dengan menggunakan algoritma *naïve bayes*, *decision tree*, *artificial neural network*, KNN, dan SVM. Terakhir dilakukan Interpretasi/evaluasi (*interpretation/evaluation*) dengan menggunakan metode *confusion matrix* dengan nilai *performansi* yang digunakan yaitu *accuracy* dan *error*.

HASIL DAN PEMBAHASAN

Hasil

Data diperoleh langsung melalui system informasi akademik (IOSYS) pada bagian akademik dan data tersebut termasuk data private yang belum pernah digunakan dalam penelitian-penelitian sebelumnya. Data yang digunakan pada penelitian ini berupa dataset dari data mahasiswa fakultas teknik Angkatan 2015 – 2018 berjumlah 807 data dengan 11 variabel diantaranya; NIM, Nama, Jenis kelamin, Usia, Indek Prestasi Semester (IPS) 1 – 4, Indek Prestasi Kumulatif (IPK) semester 4 dan Keterangan Kelulusan. Metode analisis data untuk penelitian ini adalah kuantitatif, Adapun metode analisis data mengikuti tahapan yang ada pada *proses knowledge discovery in database (KDD)* yang digunakan pada penelitian ini dengan menggunakan *tools software excel* dan *jupyter notebook*.

Pada tahap pertama dilakukan seleksi data mahasiswa fakultas Teknik Angkatan 2015 sampai 2018 dengan data yang di dapatkan melalui system informasi akademik (IOSYS) pada bagian akademik yaitu, data awal mahasiswa perangkatan, data nilai mahasiswa perangkatan, dan data keterangan kelulusan mahasiswa sesuai dengan rentang waktu yang diharapkan yaitu 8 (delapan) semester atau 4 (empat) tahun. Dari data tersebut dilakukan tahap seleksi data dengan mengambil variable yang akan digunakan dalam proses prediksi diantaranya: NIM, Nama, Jenis Kelamin, Usia, Indeks Prestasi Semester 1, 2, 3, 4, Indek Prestasi Kumulatif 4 dan Keterangan Lulus bisa dilihat pada gambar 2 dataset penelitian.

DATA PENELITIAN										
FAKULTAS TEKNIK UNIVERSITAS HAMZANWADI 2015-2018										
NO	NIM	Nama	Jenis Kelamin	Usia	IPS1	IPS2	IPS3	IPS4	IPK4	Status Kelulusan
1	0103150001	NITA ARNITA NAFIANI	Perempuan	19	2.77	2.55	2.80	3.00	2.78	Tepat Waktu
2	0103150002	BQ. DEVI SRIKANDI	Perempuan	19	3.32	3.90	3.55	3.43	3.55	Tepat Waktu
3	0103150003	ASTUTI HERLINA	Perempuan	19	0.00	0.00	0.00	0.00	0.00	Terlambat
4	0103150004	ZAINUL MAJDI	Laki-laki	18	2.59	2.80	2.95	2.86	2.80	Tepat Waktu
5	0103150005	RIZKIKA ZULHIMAYANI	Perempuan	18	3.18	3.20	2.95	3.14	3.12	Tepat Waktu
6	0103150006	ZULDYA KUMALASARI	Perempuan	20	3.32	3.05	2.95	3.14	3.12	Tepat Waktu
7	0103150007	NURUL HIDAYATI	Perempuan	19	2.55	2.85	2.45	2.71	2.64	Tepat Waktu
8	0103150008	AHMAD MUKRI	Laki-laki	19	2.27	2.65	2.65	2.71	2.57	Tepat Waktu
9	0103150009	NURHIDAYATI	Perempuan	17	2.68	3.10	2.65	2.71	2.79	Tepat Waktu
10	0103150010	HIRWAN HADI ASYAIR	Laki-laki	25	2.95	2.75	3.10	3.14	2.99	Tepat Waktu
11	0103150011	RONI MASYUDI EFFENDI	Laki-laki	23	2.86	2.50	2.70	2.71	2.69	Tepat Waktu
12	0103150012	MUHAMMAD HANAFI	Laki-Laki	21	2.05	1.75	0.00	0.00	0.95	Terlambat
13	0103150013	TAJUDIN	Laki-Laki	20	2.77	2.65	2.30	0.00	1.93	Tepat Waktu
14	0103150014	M. MASHUN	Laki-Laki	19	3.00	3.55	3.40	3.14	3.27	Terlambat

Gambar 2. Dataset penelitian

Pada tahap selanjutnya dilakukan pemilihan data atau *preprocessing/cleaning* data. Dari data tersebut dilakukan pembersihan data apabila terdapat data yang hilang, data ganda atau bersifat *outlier* dan melakukan penghapusan data variabel yang tidak digunakan seperti variabel NIM dan Nama karena dirasa tidak kontributif untuk digunakan dalam melakukan prediksi dapat dilihat pada gambar 3 setelah dilakukan *preprocessing* data.

	JenisKelamin	Usia	IPS1	IPS2	IPS3	IPS4	IPK4	StatusKelulusan
0	Perempuan	19	2.77	2.55	2.80	3.00	2.7800	Tepat Waktu
1	Perempuan	19	3.32	3.90	3.55	3.43	3.5500	Tepat Waktu
2	Perempuan	19	0.00	0.00	0.00	0.00	0.0000	Terlambat
3	Laki-laki	18	2.59	2.80	2.95	2.86	2.8000	Tepat Waktu
4	Perempuan	18	3.18	3.20	2.95	3.14	3.1175	Tepat Waktu

Gambar 3. Dataset setelah dilakukan preprocessing

Setelah itu dilakukan transformasi data dengan dengan mengubah data variabel Jenis Kelamin dan keterangan lulus yang awalnya berbentuk teks (*string*) menjadi angka (*numerik*). Dilihat pada gambar 4 dataset setelah dilakukan tranformasi.

	JenisKelamin	Usia	IPS1	IPS2	IPS3	IPS4	IPK4	StatusKelulusan
0	0	19	2.77	2.55	2.80	3.00	2.7800	1
1	0	19	3.32	3.90	3.55	3.43	3.5500	1
2	0	19	0.00	0.00	0.00	0.00	0.0000	0
3	2	18	2.59	2.80	2.95	2.86	2.8000	1
4	0	18	3.18	3.20	2.95	3.14	3.1175	1

Gambar 4. Dataset setelah dilakukan transformasi data

Tabel 1. Hasil klasifikasi algoritma

Algoritma Klasifikasi	Status Kelulusan	Precision	Recall	F1-score	Support	Accuracy
Naïve Bayes	Terlambat	0,97%	0,90%	0,94%	80	92,37%
	Tepat waktu	0,86%	0,96%	0,91%	51	
Decision Tree	Terlambat	0,96%	0,90%	0,93%	80	91,60%
	Tepat waktu	0,86%	0,94%	0,90%	51	
KNN	Terlambat	0,99%	0,96%	0,97%	80	96,95%
	Tepat waktu	0,94%	0,98%	0,96%	51	
SVM	Terlambat	1,00%	0,89%	0,94%	80	93,13%
	Tepat waktu	0,94%	0,92%	0,93%	51	
ANN	Terlambat	0,97%	0,89%	0,93%	80	91,60%
	Tepat waktu	0,84%	0,96%	0,90%	51	

Kemudian pada tahap selanjutnya dilakukan teknik *data mining* dengan menggunakan algoritma *Naïve Bayes*, *Decision Tree*, *Artificial Neural Network*, KNN, dan SVM untuk melakukan prediksi dengan bantuan *tools jupyter notebook* Bahasa pemrograman *python*. Selanjutnya dilakukan pemisahan data antara variabel independen (prediktor) dan atribut dependen (target). Yaitu Jenis kelamin, Usia, IPS1 – IPS4, dan IPK4 sebagai variabel predictor dan variabel target yaitu Status Kelulusan dengan menggunakan fungsi `iloc` yaitu: $X = \text{dataset.iloc}[:, :7]$ dan $y = \text{dataset.iloc}[:, 7]$. X adalah variabel independent dan y adalah variabel dependen. Berikutnya dilakukan optimasi data dengan teknik *smoteenn* pada data status kelulusan antara data tepat waktu dan terlambat agar data tersebut menjadi balance/seimbang diperoleh jumlah data terlambat sebanyak 345 data dan tepat waktu sebanyak 306 data. Setelah itu dilakukan split data untuk data training 80% sebanyak 520 data dan data uji 20% sebanyak 131 data dengan *random state* 42. Selanjutnya dilakukan pengujian model algoritma menggunakan bantuan *tools jupyter notebook*.

Hasil analisis pada tabel 1 menunjukkan bahwa tingkat kelulusan diperoleh adalah terlambat setelah dilakukan analisis menggunakan kelima algoritma tersebut. Didapatkan

akurasi pada algoritma *naïve bayes* sebesar 92,37%, *Decision tree* sebesar 91,60%, KNN sebesar 96,95%, SVM sebesar 93,13% dan ANN sebesar 90,84%. Hasil pengujian algoritma pada tabel 1, dengan menggunakan kelima algoritma diatas diketahui hasil akurasi dengan menggunakan algoritma *naïve bayes* diketahui sebesar 92,37%, algoritma *decision tree* sebesar 91,60%, algoritma KNN sebesar 96,95, *algoritma SVM* sebesar 93,13%, dan algoritma ANN sebesar 91,60%.

Selanjutnya, tahap terakhir dilakukan evaluasi dengan menggunakan metode *confusion matrix*. Hasil evaluasi kelima *algoritma* tersebut dapat dilihat pada tabel 2.

Tabel 2. Hasil Evaluasi Confusion Matrix

<i>Algoritma Klasifikasi</i>	<i>True Positive</i>	<i>True Negative</i>	<i>False Positive</i>	<i>False Negative</i>	<i>Total</i>	<i>Accuracy</i>	<i>Error</i>
<i>Naïve Bayes</i>	72	49	8	2	131	0,923	0,076
<i>Decision Tree</i>	72	48	8	3	131	0,916	0,083
KNN	77	50	3	1	131	0,969	0,030
SVM	71	51	9	0	131	0,931	0,068
ANN	70	49	10	2	131	0,908	0,091

Berdasarkan hasil pengujian kelima algoritma tersebut setelah dilakukan evaluasi *confusion matrix* pada tabel 2. Diketahui algoritma *naïve bayes* memiliki akurasi sebesar 0,923 dan *error* sebesar 0,076 dimana data yang diprediksi benar sebanyak 121 data dan data yang diprediksi salah sebanyak 10 data, algoritma *decision tree* memiliki akurasi sebesar 0,916 dan *error* sebesar 0,083 dimana data prediksi benar sebanyak 120 data, prediksi salah sebanyak 11 data, algoritma KNN memiliki akurasi sebesar 0,969 dan *error* sebesar 0,030 dimana data prediksi benar sebanyak 127 data dan prediksi salah sebanyak 4 data, algoritma SVM memiliki akurasi sebesar 0,931 dan *error* sebesar 0,068 dimana data prediksi benar sebanyak 122 data dan prediksi salah sebanyak 9 data, algoritma ANN memiliki akurasi sebesar 0,908 dan *error* sebesar 0,091 dimana data prediksi benar sebanyak 119 dan prediksi salah sebanyak 12 data.

Pembahasan

Data yang digunakan pada penelitian ini berupa dataset dari data mahasiswa fakultas teknik Angkatan 2015 – 2018 berjumlah 807 data dengan 11 variabel diantaranya; NIM, Nama, Jenis kelamin, Usia, Indek Prestasi Semester (IPS) 1 – 4, Indek Prestasi Kumulatif (IPK) semester 4 dan Keterangan Kelulusan. Berdasarkan hasil pengujian kelima algoritma yang digunakan yaitu algoritma *naïve bayes*, *decision tree*, KNN, SVM, dan ANN dalam melakukan prediksi dengan total data uji berjumlah 131 dan melakukan optimasi pada data uji.

Pada algoritma *naïve bayes* didapatkan nilai akurasi sebesar 92,37% dimana tingkat kelulusan mahasiswa didominasi terlambat dengan nilai *support* data sebanyak 80 dibandingkan dengan tingkat kelulusan mahasiswa tepat waktu yang hanya memiliki nilai *support* data sebanyak 50, serta diketahui dari hasil evaluasi *confusion matrik* bahwa data yang diprediksi benar berjumlah 121 data dan prediksi salah sebanyak 10 data dengan akurasi *error* sebesar 0,076%. terbukti dengan melakukan optimasi pada data dapat meningkatkan akurasi dibandingkan dengan hasil penelitian sebelumnya oleh Kusrini & Prasetio (2020) tanpa melakukan optimasi pada data didapatkan akurasi sebesar 89,00%.

Pada algoritma *decision tree* didapatkan nilai akurasi sebesar 91,60% dimana tingkat kelulusan mahasiswa didominasi terlambat dengan nilai *support* data sebanyak 80 dibandingkan dengan tingkat kelulusan mahasiswa tepat waktu yang hanya memiliki nilai *support* data sebanyak 50, serta diketahui dari hasil evaluasi *confusion matrik* bahwa data yang diprediksi benar berjumlah 120 data dan prediksi salah berjumlah 11 data dengan akurasi *error* sebesar

0,083%. selain itu dengan melakukan optimasi pada data terbukti dapat meningkatkan akurasi dibandingkan dengan hasil penelitian sebelumnya oleh Samuel, Jonathan, & Naibaho (2019) tanpa melakukan optimasi pada data didapatkan akurasi sebesar 90,24%.

Pada algoritma KNN didapatkan nilai akurasi sebesar 96,95% dimana tingkat kelulusan mahasiswa didominasi terlambat dengan nilai *support* data sebanyak 80 dibandingkan dengan tingkat kelulusan mahasiswa tepat waktu yang hanya memiliki nilai *support* data sebanyak 50, serta diketahui hasil evaluasi *confusion matrix* bahwa data yang diprediksi benar sebanyak 127 data dan prediksi salah sebanyak 4 data dengan akurasi *error* sebesar 0,030%. Selain itu dengan melakukan optimasi pada data terbukti dapat meningkatkan akurasi dibandingkan dengan hasil penelitian sebelumnya oleh Hakim, Rizal, & Ratnasari (2019) tanpa melakukan optimasi pada data didapatkan akurasi sebesar 80,00%.

Pada algoritma SVM didapatkan nilai akurasi sebesar 93,13% dimana tingkat kelulusan mahasiswa didominasi terlambat dengan nilai *support* data sebanyak 80 dibandingkan dengan tingkat kelulusan mahasiswa tepat waktu yang hanya memiliki nilai *support* data sebanyak 50, serta diketahui hasil evaluasi *confusion matrix* bahwa data prediksi benar sebanyak 122 data dan prediksi salah sebanyak 9 data dengan akurasi *error* sebesar 0,068%. selain itu dengan melakukan optimasi pada data terbukti dapat meningkatkan akurasi dibandingkan dengan hasil penelitian sebelumnya oleh Abdullah, Kusri & Arief (2022) tanpa melakukan optimasi pada data didapatkan akurasi sebesar 86,36%.

Pada algoritma ANN didapatkan nilai akurasi sebesar 90,84% dimana tingkat kelulusan mahasiswa didominasi terlambat dengan nilai *support* data sebanyak 80 dibandingkan dengan tingkat kelulusan mahasiswa tepat waktu yang hanya memiliki nilai *support* data sebanyak 50, serta diketahui dari hasil evaluasi *confusion matrix* bahwa data prediksi benar sebanyak 119 dan prediksi salah sebanyak 12 data dengan akurasi *error* sebesar 0,091. selain itu dengan melakukan optimasi pada data terbukti dapat meningkatkan akurasi dibandingkan dengan hasil penelitian sebelumnya oleh Putra & Walmi (2020) tanpa melakukan optimasi pada data didapatkan akurasi sebesar 88,14%.

Hasil pengujian dan evaluasi kelima algoritma tersebut, menegaskan bahwa dengan melakukan optimasi data pada model algoritma *naïve bayes*, *decision tree*, KNN, SVM, ANN dapat meningkatkan keakuratan hasil prediksi daripada hanya menerapkan model algoritma tanpa melakukan optimasi pada data dikarenakan dengan melakukan optimasi pada data dapat membuat data lebih teratur dan seimbang sehingga dapat meningkatkan akurasi yang maksimal.

Adapun hasil terbaik dari pengujian algoritma tersebut dalam melakukan prediksi tingkat kelulusan mahasiswa didapatkan model algoritma KNN adalah model algoritma terbaik yang memiliki akurasi mencapai 96,95% dengan akurasi *error* sebesar 0,030 dan termasuk ke dalam kategori sangat baik. dikarenakan pada penelitian ini dilakukan pembagian data antara data training dan data uji sehingga algoritma KNN memiliki hasil yang lebih akurat dibandingkan dengan algoritma yang lain. Secara teoritis model algoritma KNN juga merupakan metode yang mengklasifikasikan objek didasarkan pada nilai k dengan melihat jarak terdekat suatu objek berdasarkan data latih atau data uji (Setiyorini & Asmono 2019).

SIMPULAN

Hasil temuan kami dengan memprediksi tingkat kelulusan mahasiswa dengan menggunakan lima model algoritma didapatkan hasil terbaik atau hasil yang paling akurat dalam melakukan prediksi yaitu dengan model algoritma KNN dengan melakukan optimasi pada data tingkat akurasi yang didapatkan sebesar 96,95% termasuk kedalam kategori sangat baik. Selain itu hasil prediksi yang didapatkan lebih dominan kepada mahasiswa yang terlambat dibandingkan dengan mahasiswa yang lulus tepat waktu, Adapun faktor keterlambatan mahasiswa dipengaruhi oleh Indeks Prestasi Kumulatif (IPK). Oleh karena itu

dengan mengetahui hasil prediksi tingkat kelulusan mahasiswa dengan menggunakan model algoritma terbaik diharapkan mampu memberikan nilai maksimal dalam melakukan evaluasi akademik, sehingga Lembaga/Program Studi dapat mengambil tindakan lebih awal terhadap mahasiswa yang diprediksi akan terlambat untuk dapat meningkatkan nilai IPK pada setiap semester, dan dapat meningkatkan kelulusan mahasiswa tepat waktu sesuai rentang waktu yang diharapkan.

REFERENSI

- Abdullah, M. F., Kusrini, K., & Arief, M. R. (2022). Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Menggunakan Metode Svm (Studi Kasus: Universitas KH A Wahab Hasbullah Jombang. *SAINTEKBU*, 14(01), 35–44. <https://doi.org/10.32764/sainstekbu.v14i01.1096>
- Bisri, A., & Wahono, R. S. (2015). Penerapan Adaboost untuk penyelesaian ketidakseimbangan kelas pada Penentuan kelulusan mahasiswa dengan metode Decision Tree. *Journal of Intelligent Systems*, 1(1), 27–32.
- Etriyanti, E., Syamsuar, D., & Kunang, N. (2020). Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4. 5 untuk Memprediksi Kelulusan Mahasiswa. *Telematika*, 13(1), 56–67. <https://doi.org/10.35671/telematika.v13i1.881>
- Hakim, L. A. R., Rizal, A. A., & Ratnasari, D. (2019). Aplikasi prediksi kelulusan mahasiswa berbasis k-nearest neighbor (k-nn). *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 1(1), 30–36. <https://doi.org/10.35746/jtim.v1i1.11>
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103–108. <https://doi.org/10.30871/jaic.v5i2.3200>
- Hidayat, W., Utami, E., Iskandar, A. F., Hartanto, A. D., & Prasetio, A. B. (2021). Perbandingan Performansi Model pada Algoritma K-NN terhadap Klasifikasi Berita Fakta Hoaks Tentang Covid-19. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 167–176. <https://doi.org/10.29408/edumatic.v5i2.3664>
- Juliansa, H. (2019). Data Mining Rough Set Dalam Menganalisa Kinerja Dosen Stmik Bina Nusantara Jaya Lubuklinggau. *JUSIM (Jurnal Sist. Inf. Musirawas)*, 4(1), 11–17. <https://doi.org/10.32767/jusim.v4i1.440>
- Kusrini, K., & Prasetio, A. B. (2020). Prediction of Student Graduation with Naive Bayes Algorithm. *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1–5.
- Mulya, D. P. (2019). Analisa Dan Implementasi Association Rule Dengan Algoritma Fp-Growth Dalam Seleksi Pembelian Tanah Liat (Studi Kasus Di Pt. Anveve Ismi Berjaya). *Jurnal Teknologi Dan Sistem Informasi Bisnis-JTEKSIS*, 1(1), 47–57. <https://doi.org/10.47233/jteksis.v1i1.6>
- Nofriansyah, D., & Nurcahyo, G. W. (2015). *Algoritma Data Mining dan Pengujian*. Yogyakarta: Deepublish.
- Pramadhana, D. (2021). Klasifikasi Penyakit Diabetes Menggunakan Metode CFS Dan ROS dengan Algoritma J48 berbasis Adaboost. *Edumatic: Jurnal Pendidikan Informatika*, 5(1), 89-98. <https://doi.org/10.29408/edumatic.v5i1.3336>
- Putra, H., & Walmi, N. U. (2020). Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 6(2), 100–107. <https://doi.org/10.25077/TEKNOSI.v6i2.2020.100-107>
- Rahayu, S., Yumarlin, M. Z., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 98–106.

- <https://doi.org/10.29408/edumatic.v6i1.5433>
- Rizky, M., Ridha, A. A., & Prihandani, K. (2021). Penentuan Paket Promosi Pakaian PT. D&C Production dengan Menggunakan Algoritma FP-Growth. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 177–186. <https://doi.org/10.29408/edumatic.v5i2.3714>
- Samuel, Y. T., Jonathan, B., & Naibaho, J. (2019). Memprediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree J48 Di Universitas Advent Indonesia. *TeIka*, 9(1), 43–52. <https://doi.org/10.36342/teika.v9i01.790>
- Sari, V., Firdausi, F., & Azhar, Y. (2020). Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 1–9. <https://doi.org/10.29408/edumatic.v4i2.2202>
- Setiyorini, T., & Asmono, R. T. (2019). Penerapan Metode K-Nearest Neighbor Dan Information Gain Pada Klasifikasi Kinerja Siswa. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, 5(1), 7-14. <https://doi.org/10.33480/jitk.v5i1.613>
- Suardika, I. G. I. (2019). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus Fakultas Ekonomi Dan Bisnis Universitas Pendidikan Nasional. *Jurnal Ilmu Komputer Indonesia*, 4(2), 37-44. <https://doi.org/10.23887/jik.v4i2.2775>
- Sumardi, R., & Efendi, S. (2021). Upaya Meningkatkan Kinerja Dengan Pendekatan Aplikasi Balance Scorecard Pada Perguruan Tinggi Swasta Di Jakarta. *Populis: Jurnal Sosial Dan Humaniora*, 4(2), 220–232. <https://doi.org/10.47313/pjsh.v4i2.681>
- Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Penjualan Bisnis Ritel. *Edumatic: Jurnal Pendidikan Informatika*, 4(1), 37–46. <https://doi.org/10.29408/edumatic.v4i1.2081>
- Uska, M., Wirasasmita, R., Usuluddin, U., & Arianti, B. (2020). Evaluation of Rapidminer-Application in Data Mining Learning using PeRSIVA Model. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 164–171. <https://doi.org/10.29408/edumatic.v4i2.2688>
- Zainuddin, M. (2018). Perbandingan 4 Algoritma Berbasis Particle Swarm Optimization (pso) Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa. *Jurnal Ilmiah Teknologi Informasi Asia*, 13(1), 1-12. <https://doi.org/10.32815/jitika.v13i1.247>
- Zanuardi, A., & Suprayitno, H. (2018). Analisa karakteristik kecelakaan lalu lintas di jalan ahmad yani surabaya melalui pendekatan knowledge discovery in database. *Jurnal Manajemen Aset Infrastruktur & Fasilitas*, 2(1), 45-55. <https://doi.org/10.12962/j26151847.v2i1.3767>