Jurnal Rekayasa Elektrika Vol. 18, No. 2, Juni 2022, hal. 102-108
ISSN. 1412-4785; e-ISSN. 2252-620X, Terakreditasi Dikbudristekdikti No. 105/KPT/2022
DOI: 10.17529/jre.v18i2.24047

Seleksi Fitur dan Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa

Junta Zeniarja, Abu Salam, dan Farda Alan Ma'ruf
Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
Jl. Imam Bonjol no.207, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang, 50131
e-mail: junta@dsn.dinus.ac.id

Abstrak—**Mahasiswa merupakan bagian utama di dalam siklus hidup suatu universitas. Jumlah kelulusan suatu universitas sering kali mempunyai perbandingan yang kecil bila dibanding dengan jumlah mahasiswa yang didapat pada tahun akademik yang sama. Tingkatan kelulusan mahasiswa yang kecil ini bisa disebabkan oleh sebagian aspek, seperti banyaknya aktivitas kemahasiswaan yang diiringi oleh aspek ekonomi, serta aspek – aspek lainnya. Perihal ini membuat suatu universitas wajib mempunyai model yang bisa memperhitungkan apakah mahasiswa itu bisa lulus tepat waktu atau tidak. Faktor utama yang menentukan reputasi suatu universitas salah satunya adalah kelulusan mahasiswa tepat waktu. Semakin tinggi tingkat mahasiswa baru pada suatu universitas maka dengan rasio yang sama, juga wajib ada mahasiswa yang lulus tepat waktu. Peningkatan jumlah data mahasiswa dan data akademis terjadi jika banyak mahasiswa yang tidak lulus tepat waktu dari semua mahasiswa yang terdaftar. Sehingga akan mempengaruhi citra dan reputasi dari universitas yang nantinya dapat mengancam nilai akreditasi universitas tersebut. Untuk mengatasi hal tersebut, maka diperlukan model yang dapat memprediksi kelulusan mahasiswa sehingga dapat dijadikan pengambilan kebijakan nantinya. Tujuan dari penelitian ini adalah mengusulkan model klasifikasi terbaik dengan membandingkan tingkat akurasi yang tertinggi dari beberapa algoritma klasifikasi antara lain *Naïve Bayes*, *Random Forest*, *Decision Tree*, *K-Nearest Neighbor* (K-NN) dan *Support Vector Machine* (SVM) untuk memprediksi kelulusan mahasiswa. Selain itu proses seleksi fitur juga digunakan sebelum proses klasifikasi untuk mengoptimalkan model. Penggunaan seleksi fitur pada model ini dengan fitur terbaik menggunakan 12 fitur atribut regular dan 1 atribut sebagai label. Didapatkan bahwa model klasifikasi dengan algoritma *Random Forest* yang terpilih, dengan nilai akurasi tertinggi mencapai 77.35% lebih baik dibandingkan dengan algoritma lainnya.**

Kata kunci: ***Seleksi fitur, klasifikasi, kelulusan, mahasiswa, random forest***

Abstract— **Students are a major part of the life cycle of a university. The number of students graduating from a university often has a small ratio when compared to the number of students obtained in the same academic year. This small student graduation rate can be caused by several aspects, such as the many student activities accompanied by economic aspects, as well as other aspects. This makes it mandatory for a university to have a model that can take into account whether the student can graduate on time or not. One of the main factors that determine the reputation of a university is student graduation on time. The higher the level of new students at a university, with the same ratio, there must also be students who graduate on time. An increase in the number of student data and academic data occurs if many students do not graduate on time from all registered students. So that it will affect the image and reputation of the university which can later threaten the accreditation value of the university. To overcome this, we need a model that can predict student graduation so that it can be used as policy making later. The purpose of this study is to propose the best classification model by comparing the highest level of accuracy of several classification algorithms including Naïve Bayes, Random Forest, Decision Tree, K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM) to predict student graduation. In addition, the feature selection process is also used before the classification process to optimize the model. The use of feature selection in this model with the best features using 12 regular attribute features and 1 attribute as a label. It was found that the classification model using the Random Forest algorithm was chosen, with the highest accuracy value reaching 77.35% better than other algorithms.**

Keywords: ***Feature selection, classification, graduation, student, random forest***

I. PENDAHULUAN

Institusi pendidikan tinggi seperti universitas merupakan inti dari sistem pendidikan di mana penelitian dan pengembangan ekstensif dilakukan dalam lingkungan yang kompetitif. Misi utama dari lembaga-lembaga ini adalah untuk menghasilkan, mengumpulkan, dan berbagi pengetahuan. Secara khusus, universitas biasanya membutuhkan pengetahuan yang dikumpulkan dari kumpulan data masa lalu dan saat ini yang, setelah ditambang, dapat digunakan untuk mewakili dan menyampaikan informasi kepada pihak admin universitas untuk memantau kondisi dan mengambil tindakan dalam menyelesaikan masalah [1].

Received 09 April 2022; Revised 31 May 2022; Accepted 08 May 2022

Kinerja mahasiswa merupakan bagian penting dalam institusi pendidikan tinggi. Hal ini karena salah satu kriteria sebagai universitas yang berkualitas tinggi didasarkan pada catatan prestasi akademik yang sangat baik. Ada banyak definisi tentang kinerja siswa berdasarkan literatur sebelumnya. Usamah dkk [2] menyatakan bahwa kinerja siswa dapat diperoleh dengan mengukur penilaian pembelajaran dan ko-kurikulum. Namun, sebagian besar penelitian menyebutkan tentang kelulusan menjadi ukuran keberhasilan siswa.

Masalah tingkat kelulusan banyak dialami oleh perguruan tinggi. Tingkat kelulusan belum dapat diprediksi, sehingga tidak ada manajemen untuk menghindari penurunan kelulusan karena semua ini akan berdampak pada akreditasi universitas, oleh karena itu prediksi kelulusan perlu dilakukan. Banyak faktor yang mempengaruhi kualitas sebuah universitas. Keberhasilan mahasiswa yang lulus dalam waktu yang telah ditentukan merupakan salah satunya. Semakin tingginya jumlah mahasiswa yang lulus tepat waktu berpengaruh terhadap akreditasi jurusan di suatu universitas. Sebagian aspek yang mempengaruhi jumlah lulusan perguruan tinggi butuh dianalisa supaya bisa didapat kebijaksanaan hasil prediksi supaya seluruh mahasiswa bisa lulus tepat waktu. Banyak aspek yang pengaruhi tingkatan kelulusan, salah satunya rendahnya keahlian akademis, program perkuliahan, indeks prestasi, ataupun aspek yang lain [3].

Apalagi dengan adanya wabah pandemi Covid-19 pada awal tahun 2020, yang menyebabkan segala proses pendidikan dilakukan secara daring atau *online*. Sehingga banyak mahasiswa yang mengeluhkan tidak terbiasa untuk mendapatkan materi kuliah secara daring, sehingga proses pendidikan berjalan tidak maksimal [4], [5]. Banyak mahasiswa yang dari pedalaman tidak didukung dengan koneksi internet yang baik, terkendala kuota, dan ada juga terkendala pemahaman materi kuliah yang kurang mendalam, atau dengan kata lain tidak mantep kalau tidak tatap muka secara langsung (*offline*). Akibatnya juga akan berdampak dengan kelulusan mahasiswa nantinya.

Universitas Dian Nuswantoro (UDINUS) ialah salah satu akademi besar swasta yang berakreditasi A yang terletak di Semarang, Indonesia. Berdiri pada tahun 1990, yang dipandu oleh bapak rektor Prof. Dr. Ir. Edi Noersasongko, M.Kom. Prof. Edi berkata kalau Universitas kita bertumbuh cepat serta jadi fasilitator pendidikan tinggi yang bermutu. UDINUS berkomitmen buat membagikan pembelajaran buat tiap industri serta jalur kehidupan [6]. Salah satu fakultas yang ada dan menjadi andalan dari UDINUS adalah Fakultas Ilmu Komputer yang memiliki jumlah mahasiswa paling banyak, terutama prodi Teknik Informatika Strata-1 (S1) yang memiliki mahasiswa yang kompleks dan terbanyak daripada prodi yang lain. Sehingga sesuai jika data kelulusan mahasiswa prodi Teknik Informatika S1 dijadikan sebagai objek untuk diteliti.

Data mining bisa dimaksud sebagai sebutan yang dipakai untuk menguraikan temuan wawasan di dalam database. *Data mining* merupakan cara yang memakai metode statistik, matematika, *artificial intelligence*, serta machine learning buat mengekstraksi serta mengenali data yang berguna serta wawasan yang terpaut dari bermacam database yang besar [7], [8].

Dalam teori *Data mining* ada 2 metode pembelajaran yang biasa dipakai, yaitu *supervised learning* dan *unsupervised learning*. Metode *supervised learning* merupakan metode yang sering dipakai dalam konsep Data Science dibanding dengan *unsupervised learning*. Perbandingan dari kedua metode tersebut yaitu terdapat pada bagaimana mereka berlatih buat memakai suatu konsep prediksi ataupun klasifikasi. Dalam *supervised learning*, metode itu seakan dilatih terlebih dulu supaya bisa melaksanakan prediksi ataupun klasifikasi [9], [10].

Banyak sekali peneliti yang sudah terjun dalam ranah *Data mining* bidang Pendidikan atau bisa dikenal dengan istilah *Educational Data Mining* (EDM) [11]–[19]. Dan banyak survey yang telah dilakukan dalam bidang EDM yang membahas tentang prediksi kelulusan mahasiswa [17], [19] yang sukses membangun model menggunakan metode machine learning. Selain itu juga ada beberapa penelitian yang menggunakan konsep big data kedalam EDM [20]. Sehingga sudah banyak sekali topik tentang EDM dengan prediksi kelulusan mahasiswa [21], [22] yang membuat topik ini menjadi salah satu tren para peneliti dan akademisi.

Klasifikasi ialah suatu metode yang bisa dipakai buat menggambarkan input jadi output diskrit yang dikenal dengan label serta kategori. Sebagai ilustrasi, penampilan dari sesuatu berkas mahasiswa bisa diklasifikasikan sebagai "lulus" ataupun "tidak lulus". Sebagian algoritma klasifikasi antara lain *Support Vector Machine* (SVM), *Decision Tree, Discriminant Analysis*, *K- Nearest Neighbor*, *Neural Network* (K-NN), serta *Naïve Bayes*. Salah satu algoritma yang bisa dipakai buat klasifikasi dengan proses yang maksimal adalah *Random Forest*. Algoritma *Random Forest* adalah salah satu teknik klasifikasi machine learning yang digunakan pada proses *data mining* dengan membentuk suatu pohon keputusan yang diilustrasikan ke dalam bentuk *rule*. Algoritma *Random Forest* juga bagian dari salah satu jenis algoritma dengan menggunakan konsep pohon keputusan yang merupakan metode klasifikasi dan prediksi yang sangat kokoh dan terkini [23]–[25]. Beberapa penelitian yang telah menggunakan algoritma *Random Forest* [26], [27] menunjukkan hasil optimal, yang menandakan bahwa algoritma ini layak digunakan untuk memprediksi data kelulusan mahasiswa UDINUS.

Sehingga dalam penelitian ini mengusulkan untuk membangun model klasifikasi yang terbaik dengan membandingkan tingkat akurasi yang tertinggi dari beberapa algoritma klasifikasi antara lain *Naïve Bayes*, *Random Forest*, *Decision Tree*, K-NN dan SVM untuk memprediksi kelulusan mahasiswa. Selain itu proses seleksi fitur juga digunakan sebelum proses klasifikasi untuk mengoptimalkan model.

II. Studi Pustaka

A. Penelitian yang Relevan

Sistem simulasi prediksi profil kelulusan mahasiswa dengan *Decision Tree* yang diusulkan oleh Bekti Amalia A dan Ridha Sefina S [28], mereka melaksanakan prediksi profil kelulusan mahasiswa yang bersumber pada informasi dikala tercatat selaku mahasiswa. Dimana hasil perkiraan itu sukses diaplikasikan serta bisa dipakai oleh program studi buat melaksanakan aksi prediksi lewat aktivitas pembimbingan akademik, pembimbingan skripsi serta aktivitas lainnya.

Aplikasi prediksi kelulusan mahasiswa berbasis K-NN yang diusulkan oleh Lalu Abd Rahman Hakim dkk [29], mereka menggunakan algoritma K-NN buat bisa mengenali kelulusan mahasiswa pada permasalahan terkini dengan metode mengadopsi pemecahan dari permasalahan yang mempunyai kedekatan dengan permasalahan terkini. Didapatkan hasil pengujian berupa nilai akurasi tertinggi sebesar 98% pada K=1 untuk klasifikasi "Tepat Waktu" dan 98% pada K=2 untuk klasifikasi "Tidak Tepat Waktu".

EDM untuk prediksi kelulusan mahasiswa menggunakan Algoritma *Naïve Bayes Classifier* yang diusulkan oleh Edi Sutoyo dan Ahmad Almaarif [22] dengan melaksanakan penemuan mahasiswa yang beresiko pada langkah dini pembelajaran serta ditopang dengan muat kebijaksanaan yang bisa memusatkan mahasiswa supaya bisa menuntaskan pendidikannya. Data diperoleh dari Universitas Telkom sebanyak 4000 *data instance*, didapatkan hasil berupa nilai akurasi sebesar 73.725%, precision 0.742, recall 0.736 dan *F-measure* sebesar 0.735.

Penerapan *data mining* untuk memprediksi kelulusan mahasiswa menggunakan algoritma *Naïve Bayes* (studi kasus STMIK Primakara) yang diusulkan oleh Putu Sainanda Cahyani Moonallika dkk [30], mereka menggunakan *Naïve Bayes Classifier* untuk mengklasifikasikan data kelulusan mahasiswa, didapatkan hasil pengujian yang sudah baik berupa nilai *recall*, *accuracy* dan *precision* sebesar 80%.

Penerapan Algoritma SVM untuk model prediksi kelulusan mahasiswa tepat waktu yang diusulkan oleh Emy Haryatmi dan Sheila Pramita Hervianti [31], mereka menggunakan algoritma SVM dari data pelatihan dan data pengujian yang didapatkan dari hasil pengujian kelompok pertama dengan jumlah data pelatihan sebanyak 90% dan data pengujian sebanyak 10% membuktikan bahwa algoritma SVM mendapatkan nilai akurasi yang sangat baik yaitu sebesar 94.4%.

B. Kelulusan Mahasiswa

Kelulusan mahasiswa di dalam suatu universitas di atur pula di dalam hukum ataupun peraturan menteri. Pada kedua peraturan itu membuktikan kalau kelulusan mahasiswa wajib penuhi standar kompetensi lulusan. Standar kompetensi lulusan ialah patokan minimum mengenai kualifikasi keahlian alumnus yang melingkupi tindakan, wawasan serta keahlian yang diklaim dalam rumusan capaian pembelajaran lulusan [32].

Di dalam capaian pembelajaran, tingkatan kelulusan serta kedalaman materi lulusan program sarjana (S1) sangat sedikit memahami rancangan teoritis aspek wawasan serta ketrampilan khusus dengan cara biasa serta rancangan teoritis bagian spesial dalam aspek wawasan dan ketrampilan itu dengan cara yang mendalam. Sebaliknya bila bersumber pada masa serta beban belajar penyelenggaraan program pendidikan maksimal 7 tahun akademik buat program sarjana (S1) dengan beban belajar mahasiswa paling sedikit 144 sks [28].

C. Data Mining

Data mining bisa dimaksudkan sebagai sebutan yang dipakai untuk menguraikan temuan wawasan di dalam database. *Data mining* merupakan cara yang dipakai pada metode statistik, matematika, *artificial intelligence*, serta machine learning buat mengekstraksi dan mengenali data yang berguna serta wawasan yang terpaut dari bermacam database yang besar [33].

Dalam konsep *data mining* ada 2 metode pembelajaran yang kerap dipakai, yaitu s*upervised learning* dan *unsupervised learning*. Metode *supervised learning* merupakan metode yang sangat kerap dipakai dalam bidang *Data science* dibanding dengan *unsupervised learning*. Perbandingan dari kedua metode itu terdapat pada bagaimana mereka berlatih untuk menghasilkan suatu prediksi ataupun klasifikasi. Dalam *supervised learning*, metode itu seakan dilatih dulu supaya bisa melaksanakan prediksi ataupun klasifikasi [22].

III. Metode

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan metode penelitian yang dipakai pada penelitian ini. Karena sudah banyak para peneliti yang menggunakannya dan terbukti cocok untuk menyelesaikan masalah proyek atau penelitian bidang *data mining* [25], [34]. Dimana alur dari metode ini dimulai dari tahap pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan data (*modeling*), evaluasi model (*evaluation*) dan pengembangan (*deployment*) seperti yang terlihat pada Gambar 1.

A. Pemahaman Bisnis (Data Understanding)

Proses pemahaman bisnis berpusat pada uraian tujuan keinginan yang bersumber pada penilaian bisnis. Selanjutnya hal tersebut dirubah ke dalam sebuah kerangka awal *data mining* yang telah dibuat untuk menggapai tujuan. Pemahaman bisnis merujuk pada proses prediksi kelulusan mahasiswa yang tepat waktu dan tidak, yang telah ditetapkan yaitu tidak lebih dari 4 tahun atau 8 semester. Pada proses ini dibutuhkan pemaparan tentang latar belakang serta tujuan pada rangkaian bisnis yang

berkaitan dengan prediksi kelulusan mahasiswa:

1. Menentukan tujuan bisnis (*determine business objectives*)

Tujuan bisnis pada penelitian ini adalah mengidentifikasi pengetahuan dan karakteristik data mahasiswa untuk dapat menentukan hasil prediksi kelulusan mahasiswa apakah <=8 semester atau >8 semester.

2. Menilai situasi (*assess situation*)

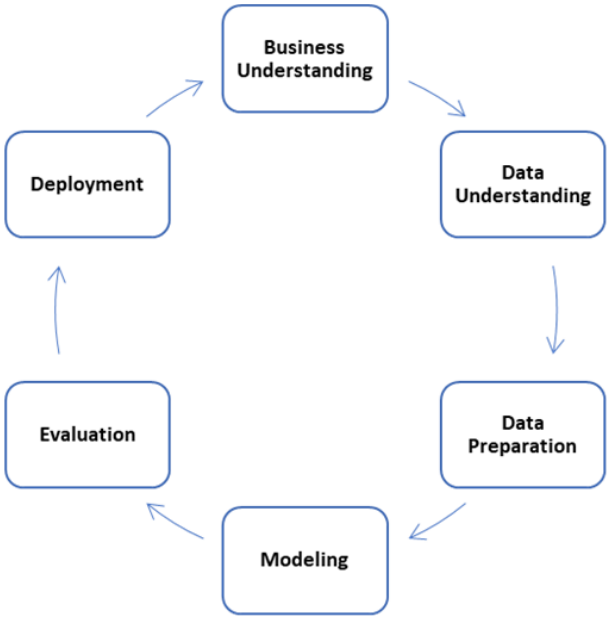
Sistem informasi akademik ini berkaitan dengan suasana akademik mahasiswa pada UDINUS terutama untuk prodi Sarjana Teknik Informatika. Sistem yang berjalan sekarang adalah Sistem Informasi Akademik Dian Nuswantoro (SIADIN) yaitu *software* berbasis web yang dibuat untuk menjalankan manajemen dan pengolahan data terhadap rangkaian kegiatan administrasi dan operasional akademik pada program studi dalam suatu universitas. Data profil mahasiswa, data nilai, dan data pendukungnya digunakan untuk diolah pada sistem informasi akademik tersebut.

3. Menentukan tujuan *data mining* (*determine the data mining goals*)

Tujuan *data mining* atau tujuan dari penelitian ini adalah membangun model yang terbaik dari beberapa metode klasifikasi dan mengekstraksi pengetahuan (*knowledge*) tentang pola studi mahasiswa mana yang bisa lulus ≤ 8 semester dan pola studi mahasiswa mana yang lulus > 8 semester. Sehingga model yang terbaik dapat digunakan untuk memperlakukan dan memprediksi kelulusan mahasiswa baru nantinya.

B. *Pemahaman Data (Data Understanding)*

Pada proses pemahaman data, langkah – langkahnya dimulai dengan mengumpulkan data awal, mendeskripsikan data, mengeskplorasi data dan memverifikasi kualitas data. Jumlah Dataset yang digunakan untuk penelitian berjumlah 2293 records dimana teridiri dari data mahasiswa lulusan program studi Sarjana Teknik Informatika dengan kode A11 pada tahun masuk antara 2012 – 2017. Untuk jumlah pada label 1 sekitar 1356 records (terdiri dari mahasiswa pada tahun masuk 2012 – 2017, dengan masa studi antara 38 – 50 bulan). Sedangkan pada label 2 terdapat 937 records (terdiri dari mahasiswa pada tahun masuk antara 2012 – 2016, dengan masa studi antara 52 – 88 bulan).



Gambar 1. Metode penelitian dengan CRISP-DM

C. *Persiapan Data (Data Preparation)*

Pada proses persiapan data, terdiri dari rangkaian agenda yang digunakan dalam membuat dataset akhir (data yang akan dimodelkan nantinya) dari data mentah awal yang terdiri dari mendeskripsikan data set, menentukan data, menggunakan data, menggabungkan data, menghilangkan data yang tidak sesuai dan memformat data.

Pada Tabel 1 terlihat untuk sampel dataset mahasiswa yang digunakan. Dimana terdiri dari banyak atribut mahasiswa yang digunakan untuk mendukung proses training model nantinya.

Sebelum masuk ke dalam tahap modeling, akan dilakukan seleksi fitur terlebih dahulu, supaya mendapatkan gambaran atribut apa saja nantinya yang akan digunakan.

Untuk atribut yang digunakan, dapat dilihat dan disimak pada Gambar 2.

D. *Pemodelan (Modeling)*

Pada sesi modeling ini, dilakukan dengan cara memilih model yang tepat dan sesuai untuk data mahasiswa dengan tahapan diantaranya: memilih metode pemodelan, membangun model dan menilai model. Teknik *Data mining* yang digunakan adalah algoritma klasifikasi menggunakan *Random Forest*, setelah memilih yang terbaik dari beberapa algoritma klasifikasi seperti *Naïve Bayes, Decision Tree*, K-NN dan SVM.

Random Forest ialah metode learning yang dipakai untuk mengoptimalkan nilai akurasi pada kasus klasifikasi data. Pengoptimalan akurasi itu, digunakan untuk melewati proses penggabungan banyak pemilah dari

```
Data columns (total 14 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   sex                         2293 non-null   int64
 1   kota_asal                   2293 non-null   int64
 2   jml_ajuan_cuti              2293 non-null   int64
 3   jml_tunggakan               2293 non-null   int64
 4   usia                        2293 non-null   int64
 5   beasiswa                    2293 non-null   int64
 6   marital                     2293 non-null   int64
 7   jml_aktivitas_kemahasiswaan 2293 non-null   int64
 8   jml_prestasi                2293 non-null   int64
 9   ips1                        2293 non-null   int64
 10  ips2                        2293 non-null   int64
 11  ips3                        2293 non-null   int64
 12  ips4                        2293 non-null   int64
 13  label                       2293 non-null   int64
dtypes: int64(14)
memory usage: 250.9 KB
```

Gambar 2. Atribut data yang digunakan

Tabel 1. Sampel data mahasiswa yang digunakan

Sex	Kota Asal	Jml Ajuan Cuti	Jml Tunggakan	Usia	Beasiswa	Marital	Jml Aktivitas Kemahasiswaan	Jml Prestasi	IPS1	IPS2	IPS3	IPS4	LABEL
1	2	2	2	3	2	2	2	2	0	2	2	3	2
2	2	2	2	3	2	2	2	2	0	1	3	3	2
1	2	2	2	3	2	2	2	2	2	3	2	2	3
1	2	2	2	3	2	2	2	2	2	2	2	2	3
1	2	2	2	3	2	2	2	2	2	2	3	2	3
1	2	2	2	3	2	2	2	2	3	3	3	2	1
1	2	2	2	3	2	2	2	2	2	2	2	2	3
1	2	2	2	3	2	2	2	1	3	3	3	3	3
1	2	2	2	3	2	2	2	2	2	2	2	3	3
1	2	2	2	3	2	2	2	2	0	3	2	2	3
1	2	2	2	3	2	2	2	2	3	3	3	3	3
1	2	2	2	3	2	2	2	2	0	3	2	2	3
1	1	2	2	3	2	2	2	2	3	3	2	3	3
1	2	2	2	3	2	2	2	2	3	2	0	2	3
1	2	2	2	3	2	2	2	2	2	2	2	2	3
2	1	2	2	3	2	2	2	1	3	3	3	3	2

metode yang sejenis dan menghasilkan prediksi klasifikasi akhir menggunakan proses *voting*.

Untuk kasus *Random Forest*, banyak *tree* yang dipakai sehingga tercipta suatu *forest* atau hutan yang selanjutnya masing – masing *tree* dianalisa secara bertahap. Untuk tahapan dari algoritma *Random Forest* dapat dilihat pada Gambar 3.

Untuk *modeling* dengan algoritma *Random Forest*, menggunakan bahasa pemrograman Python yang diset parameter sebagai berikut: nilai *random_state* = 1 dan nilai *test_size* = 0.25 atau menggunakan pembagian data training sebesar 75% dan *data testing* sebesar 25%.

E. *Evaluasi (Evaluation)*

Untuk pengolahan seluruh data juga menggunakan Bahasa pemrograman Python, dengan *diagram confussion matrix* untuk mendapatkan nilai akurasinya. Sedangkan validitas dan keakuratan hasil dari model yang telah didapatkan akan dievaluasi berdasarkan nilai – nilai tersebut.

Gambar 3. Alur tahapan algoritma *RandomForest*

F. *Pengembangan (Deployment)*

Setelah didapatkan hasil pemodelan yang terbaik dengan didapatkan nilai akurasi dan pendukungnya yang optimal, maka pengambilan keputusan atau kebijakan dalam membangun pengetahuan tentang pola mahasiswa di dalam prediksi kelulusan mahasiswa berupa pohon keputusan (*tree*) yang dihasilkan dari *Random Forest* akan digunakan dan diterapkan pada sistem akademik nantinya.

IV. Hasil dan Pembahasan

A. *Seleksi Fitur*

Berdasarkan hasil eksperimen untuk seleksi fitur dari penelitian yang dilakukan proses seleksi fitur tidak menambah nilai akurasi, atribut lengkap untuk 12 atribut regular dan 1 atribut label tetap memiliki nilai akurasi yang terbaik. Untuk nilai akurasi (dengan jumlah *data testing* 25%), didapatkan nilai akurasi training tertinggi sebesar 74.1% dan nilai akurasi testing tertinggi sebesar 77.4% setelah dilakukan percobaan menggunakan kombinasi 5 atribut terbaik sampai dengan kombinasi 12 atribut terbaik.

B. *Analisa Perbandingan Algoritma Klasifikasi*

Berdasarkan hasil eksperimen menggunakan beberapa algoritma klasifikasi seperti *Naïve Bayes*, *Random Forest*, *Decision Tree*, K-NN dan SVM. Maka didapatkan hasil akurasi terbaik menggunakan algoritma *Random Forest* dengan nilai akurasi tertinggi mencapai 77.35%. Dimana menurut Goronescue bahwa model tersebut sudah dikategorikan kedalam model klasifikasi yang adil (tidak baik dan tidak buruk) atau *fair classification* [35].

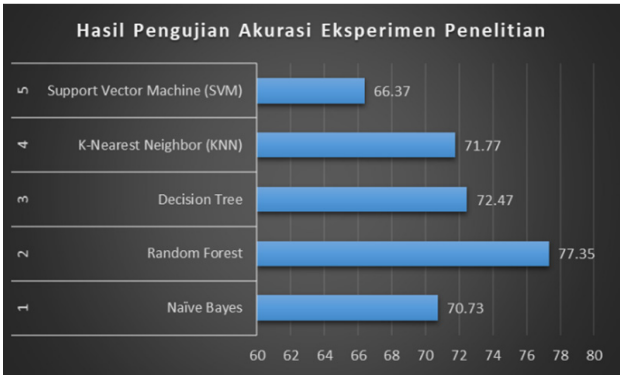
Tabel 2. Sampel data mahasiswa yang digunakan

No	Algoritma	Akurasi (%)
1	Naïve Bayes	70.73
2	Random Forest	77.35
3	Decision Tree	72.47
4	K-NN	71.77
5	SVM)	66.37

Berdasarkan hasil uji akurasi seperti yang terlihat pada Tabel 2 dan Gambar 4, menggunakan algoritma *Random Forest* didapatkan nilai akurasi yang terbaik pada 77.35% dengan seting parameter *random_state* = 42 dan *max_depth* = 4.

V. KESIMPULAN

Berdasarkan hasil eksperimen dari 2293 *record* data kelulusan mahasiswa, fitur yang terbaik tetap menggunakan 12 fitur atribut regular dan 1 atribut sebagai label, dengan *data testing* sebesar 25%, didapatkan nilai akurasi *training* tertinggi sebesar 74.1% dan nilai akurasi *testing* tertinggi sebesar 77.4%. Sedangkan berdasarkan hasil eksperimen menggunakan beberapa algoritma klasifikasi seperti *Naïve Bayes*, *Random Forest*, *Decision Tree*, K-NN dan SVM, didapatkan hasil akurasi terbaik menggunakan algoritma *Random Forest* dengan nilai akurasi tertinggi mencapai 77.35% yang dapat dikatakan sebagai *fair classification*.



Gambar 4. Hasil pengujian akurasi eksperimen penelitian

REFERENSI

[1] A. Tekin, "Early Prediction of students' grade point averages at graduation: a data mining approach," *Eurasian J. Educ. Res.*, vol. 14, no. 54, pp. 207–226, Feb. 2014.

[2] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," in *Procedia Computer Science*, vol. 72, pp. 414–422, Dec .2015.

[3] E. Purnamasari, D. P. Rini, and Sukemi, "The combination of Naive Bayes and Particle Swarm optimization methods of student's graduation prediction," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 5, no. 2, pp. 112–119, Dec. 2019.

[4] M. Babar *et al.*, "Psychological impacts of Covid-19 and satisfaction from online classes : disturbance in daily routine and prevalence of depression , stress , and anxiety among students of Pakistan," *Heliyon*, vol. 7, no. 5, May 2021.

[5] T. A. Birtch, F. F. T. Chiang, Z. Cai, and J. Wang, "Am I choosing the right career? The implications of Covid-19 on the occupational attitudes of hospitality management students," *Int. J. Hosp. Manag.*, vol. 95, no. 102931, May 2021.

[6] M. K. Prof. Dr. Ir. Edi Noersasongko. (view May 2021). "Greetings from our Rector. [Online]. Available: https://dinus.ac.id/greetings.

[7] A. Luthfiarta, J. Zeniarja, E. Faisal, and W. Wicaksono, "Prediction on deposit subscription of customer based on bank telemarketing using Decision Tree with entropy comparison," *J. Appl. Intell. Syst.*, vol. 4, no. 2, pp. 57–66, Dec. 2019.

[8] J. Zeniarja, K. Widia, and R. R. Sani, "Penerapan algoritma Naive Bayes dan forward selection dalam pengklasifikasian status gizi stunting pada Puskesmas Pandanaran Semarang," *J. Inf. Syst.*, vol. 5, no. 1, pp. 1–9, May 2020.

[9] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, "Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization," *Procedia Eng.*, vol. 53, pp. 453–462, March 2013.

[10] A. Khalaf Hamoud *et al.*, "Supervised learning algorithms in educational data mining: a systematic," *Southeast Eur. J. Soft Comput.*, vol. 10, no. 1, pp. 55–70, March 2021.

[11] A. K. Das and E. Rodriguez-Marek, "A predictive analytics system for forecasting student academic review performance: Insights from a pilot project at eastern Washington university," in *Proc. 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR),* June 2019, pp. 255–262.

[12] G. A. Agarkov, A. A. Tarasyev, and A. D. Sushchenko, "Optimization of students' graduation by the university taking into account the needs of the labor market," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 17399–17404, April 2020.

[13] A. Gonzalez-Nucamendi, J. Noguez, L. Neri, V. Robledo-Rella, R. M. G. García-Castelán, and D. Escobar-Castillejos, "The prediction of academic performance using engineering student's profiles," *Comput. Electr. Eng.*, vol. 93, no. 107288, July 2021.

[14] X. Lu, Y. Zhu, Y. Xu, and J. Yu, "Learning from multiple dynamic graphs of student and course interactions for student grade predictions," *Neurocomputing*, vol. 431, pp. 23–33, Jan 2021.

[15] X. Wang, C. Zhou, and X. Xu, "Application of C4.5 decision tree for scholarship evaluations," *Procedia Comput. Sci.*, vol. 151, no. 2018, pp. 179–184, May 2019.

[16] F. M. Almutairi, N. D. Sidiropoulos, and G. Karypis, "Context-aware recommendation-based learning analytics using tensor and coupled matrix factorization," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 5, pp. 729–741, Aug. 2017.

[17] B. Albreiki, N. Zaki, and H. Alashwal, "A Systematic literature review of student' performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, Sept. 2021.

[18] A. M. Olalekan, O. S. Egwuche, and S. O. Olatunji, "Performance evaluation of machine learning techniques for prediction of graduating students in Tertiary Institution," in *Proc. Int. Conf. Math. Comput. Eng. Comput. Sci.* March 2020, pp. 1-7.

[19] S. Alturki, I. Hulpuş, and H. Stuckenschmidt, *Predicting Academic Outcomes: A Survey from 2007 Till 2018*, no. 0123456789. Springer Netherlands, Sept. 2020.

[20] X. Bai *et al.*, "Educational big data: predictions, applications and challenges," *Big Data Res.*, vol. 26, p. 100270, Sept. 2021.

[21] T. F. Prasetyo, D. Susandi, and I. S. Widianingrum, "Prediksi kelulusan mahasiswa pada Perguruan Tinggi Kabupaten Majalengka berbasis knowledge based system," *Semin. Nas.*

Telekomun. dan Inform. Nov. 2016, pp 1-7.

[22] E. Sutoyo and A. Almaarif, "Educational Data mining untuk prediksi kelulusan mahasiswa menggunakan algoritma Naïve Bayes Classifier," *J. Rekayasa Sist. dan Teknol. Informasi*, vol. 4, no. 1, pp. 95–101, Feb. 2020.

[23] Y. P. Chiu, "Social recommendations for facebook brand pages," *J. Theor. Appl. Electron. Commer. Res.*, vol. 16, no. 1, pp. 71–84, Jan 2020.

[24] W. Wiguna and D. Riana, "Diagnosis of coronavirus disease 2019 (Covid-19) surveillance using C4.5 algorithm," *J. Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 71–80, March 2020.

[25] T. Hardiani, "Comparison of Naive Bayes method, K-NN (K-Nearest Neighbor) and Decision Tree for predicting the graduation of 'Aisyiyah University Students of Yogyakarta," *Int. J. Heal. Sci. Technol.*, vol. 2, no. 1, Jan. 2021.

[26] Y. Long, J. Liu, M. Fang, T. Wang, and W. Jiang, "Prediction of employee promotion based on personal basic features and post features," in *Proc. of the Inter. Conf. on Data Proces. and App.*, May 2018, pp. 5–10,

[27] Y. Nieto, V. Gacia-Diaz, C. Montenegro, C. C. Gonzalez, and R. Gonzalez Crespo, "Usage of machine learning for strategic decision making at Higher Educational Institutions," *IEEE Access*, vol. 7, pp. 75007–75017, May 2019.

[28] B. A. Arifiyani and R. S. Samosir, "Sistem simulasi prediksi profil kelulusan mahasiswa dengan Decison Tree," *J. Sains dan Teknol.*, vol. 5, no. 2, pp. 115–123, Aug. 2018.

[29] L. A. R. Hakim, A. A. Rizal, and D. Ratnasari, "Aplikasi Prediksi kelulusan mahasiswa berbasis K-Nearest Neighbor (K-NN)," *J. Teknol. Inf. dan Multimed.*, vol. 1, no. 1, pp. 30–36, May 2019.

[30] P. S. C. Moonallika, K. Q. Fredlina, and I. B. K. Sudiatmika, "Penerapan data mining untuk memprediksi kelulusan mahasiswa menggunakan algoritma Naive Bayes Classifier (studi kasus STMIK Primakara)," *J. Ilm. Komput.*, vol. 6, no. 1, pp. 47–56, Feb. 2020.

[31] E. Haryatmi and S. P. Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 386–392, April 2021.

[32] I. P. Astuti, "Prediksi Ketepatan waktu kelulusan dengan algoritma data mining C4.5," *Fountain Informatics J.*, vol. 2, no. 2, p. 5, Nov. 2017.

[33] L. O. M. Zulfiqar, N. Renaningtias, and M. Y. Fathoni, "Educational data mining in graduation rate and grade predictions utilizing hybrid decision tree and Naïve Bayes Classifier," no. Conrist 2019, pp. 151–157, Dec. 2020.

[34] M. Munawir and T. Iqbal, "Prediksi kelulusan mahasiswa menggunakan algoritma Naive Bayes (studi kasus 5 PTS di Banda Aceh)," *J. Teknol. Inf. dan Komunikasi*, vol. 3, no. 2, p. 59, Dec. 2019.

[35] F. Gorunescu, *Data Mining - Concepts, Models and Techniques*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, June 2011.