

Analisis Prediksi Kelulusan Course Pada E-Learning Menggunakan Model Klasifikasi

Gita Cahyani Lestari^{1*}, Sri Supatmi²

^{1,2} Jurusan Magister Sistem Informasi UNIKOM
Jl. Dipati Ukur No. 112-116, Bandung, Jawa Barat, Indonesia

¹gita.75120016@mahasiswa.unikom.ac.id

Abstrak — Akademi Anti Korupsi merupakan platform belajar dengan basis E-Learning yang mudah diakses oleh semua masyarakat sehingga dapat belajar mengenai korupsi dan cara-cara memerangnya. Pada tahun 2021, jumlah pendaftar E-Learning Akademi Anti Korupsi mengalami penurunan secara drastis sebesar 52% dengan jumlah lulusan sebesar 71%. Akademi Anti Korupsi belum memiliki model yang dapat memprediksi kelulusan course peserta. Solusi untuk memprediksi kelulusan seseorang dapat menggunakan proses data mining. Tujuan dari penelitian ini untuk memprediksi kelulusan course dengan membandingkan metode data mining klasifikasi yaitu Decision Tree dan Random Forest, berdasarkan nilai akurasi menggunakan software orange yang memiliki keunggulan dalam visual programming. Data yang akan digunakan sebagai atribut yaitu data form registrasi yang sudah dilengkapi peserta akademi anti korupsi diantaranya jenis kelamin, pekerjaan, pengalaman pendidikan anti korupsi, umur, institusi, pengalaman E-Learning dan alasan mengikuti E-Learning anti korupsi. Pada penelitian ini, dilakukan prediksi kelulusan course menggunakan dua dataset yaitu dataset missing value dan dataset tanpa missing value. Hasil penelitian dengan menggunakan dua metode klasifikasi ini menunjukkan bahwa metode Random Forest dengan dataset missing value yang paling baik dalam melakukan prediksi kelulusan course peserta akademi antikorupsi dengan nilai akurasi sebesar 74,6% yaitu klasifikasi cukup (fair classification). Faktor signifikan yang mempengaruhi kelulusan course adalah umur, domisili, dan jenis kelamin.

Kata kunci— E-Learning, Data Mining, Klasifikasi, Kelulusan

Abstract— The Anti-Corruption Academy is a learning platform based on E-Learning that is easily accessible by all people so they can learn about corruption and ways to combat it. In 2021, the number of E-Learning registrants for the Anti-Corruption Academy had decreased drastically by 52%, with the number of graduates being 71%. The Anti-Corruption Academy does not yet have a model to predict participants' course graduation. The solution to prediction someone's graduation can use a data mining process. This study aims to prediction course graduation by comparing classification data mining methods, namely Decision Tree and Random Forest, based on accuracy values using orange software, which has advantages in visual programming. The data to be used as attributes are registration form data, including gender, occupation, anti-corruption education experience, age, institution, E-Learning experience and reasons for participating in anti-corruption E-Learning. In this study, predictions of course graduation were made by two datasets, namely the missing value dataset and the dataset without missing values. The study's results using these two classification methods indicate that the Random Forest method with a missing value dataset is the best in predicting the graduation course of anti-corruption academy participants with an accuracy value of 74,6%, which is a fair classification. Significant factors that affect course graduation are age, domicile, and gender.

Index Terms— E-Learning, Data Mining, Classification, Graduation

I. PENDAHULUAN

Kemajuan teknologi dalam berbagai bidang sangatlah pesat [1], salah satunya teknologi di bidang Pendidikan yaitu E-Learning [2]. Saat ini E-Learning merupakan sesuatu yang wajib untuk menunjang proses pembelajaran [3]. Akademi Anti Korupsi merupakan platform belajar dengan basis digital yang mudah diakses oleh semua masyarakat sehingga dapat belajar mengenai korupsi dan cara-cara memerangnya. Akademi Anti Korupsi sudah berjalan sejak tahun 2018 [4]. Terdapat dua puluh mata kuliah yang dapat diakses secara virtual. Tidak terdapat persyaratan khusus untuk menjadi peserta di Akademi antikorupsi. Platform Akademi Anti Korupsi menggunakan website sebagai penunjang utama pelaksanaannya. Langkah pertama yang dilakukan peserta untuk mengakses Akademi Anti Korupsi adalah dengan

melengkapi form registrasi peserta pada halaman akademi.antikorupsi.org. Setiap modul pembelajaran atau *course* berisikan konten video pembelajaran yang menjelaskan materi perkuliahan, kuis, tugas, dan forum yang harus diselesaikan peserta setelah menerima materi pembelajaran hingga dapat mengunduh sertifikat pada aktifitas terakhir. Kurun waktu pembelajaran dapat diselesaikan dalam waktu tiga bulan.

Akademi Anti Korupsi memiliki jumlah peserta pada kurun waktu 2018-2021 sebanyak 12.385 [5]. Pada tahun 2021 mengalami penurunan jumlah pendaftar secara drastis dari tahun sebelumnya sebesar 52% [6], [5]. Tetapi tidak relevan juga dengan jumlah presentase lulusan yang dihasilkan oleh akademi Anti Korupsi sebesar 71%. Akademi Anti Korupsi memiliki tujuan strategi bisnis untuk meningkatkan jumlah pendaftaran peserta yang dapat lulus [5]. Konsep e-learning

dapat memberikan banyak informasi melalui data yang dihasilkan dalam proses yang ada di dalam sistem untuk mengekstrak pengetahuan tentang faktor-faktor yang mempengaruhi kelulusan course [7]. Beberapa informasi yang bisa didapatkan adalah sasaran atau target adopter dari program Akademi Anti Korupsi. Untuk mengolah data tersebut dilakukan dengan memantau prediksi kelulusan jumlah mata kuliah (*course*) yang dapat diambil melalui form registrasi peserta e-learning. Sehingga merujuk pada data tersebut, dapat diketahui sasaran atau target dalam Akademi Anti Korupsi untuk mengatasi permasalahan menurunnya jumlah pendaftar pada tahun 2021 dan meningkatkan kelulusan peserta.

Masalah ini dapat diselesaikan dengan menggunakan data mining. Data mining dapat dibangun menggunakan teknik perangkaan yang pada dasarnya menggunakan teknik klasifikasi [8], [9]. *Decision Tree* adalah salah satu yang paling terkenal metode klasifikasi karena proses formulasi nya yang relatif cepat dan mudah dan mudah dipahami karena kesederhanaan [10]. *Decision Tree* sebagai metode klasifikasi yang menggunakan representasi struktur pohon dimana node mewakili atribut tertentu dan daun mewakili kelas dengan node paling atas menjadi root [10]. Metode ini dapat digunakan pada masalah seperti kinerja siswa [11], [12] dan mengklasifikasikan mahasiswa putus sekolah [13], dan memprediksi tingkat kelulusan mahasiswa [14]. Kemudian terdapat metode yang memiliki konsep mirip decision tree yaitu random forest. Metode Random Forest ini merupakan kombinasi masing-masing tree dari decision tree yang kemudian digabungkan menjadi satu model [15]. Studi kasus dengan menggunakan Akademi Antikorupsi dipilih sebagai objek untuk menerapkan pendekatan *Decision Tree* dan *Random Forest* dengan fokus pada penggunaan klasifikasi untuk mengetahui faktor kelulusan course dan memprediksi kelulusan course.

Terdapat beberapa penelitian sebelumnya yang menerapkan model klasifikasi dibidang pendidikan. Penelitian yang dilakukan oleh [16] melakukan prediksi potensi siswa putus sekolah dan mengidentifikasi faktor-faktor potensial berkaitan dengan siswa putus sekolah menggunakan *Random Forest*. Utari dkk menghilangkan data *missing value* pada data preparation dengan pendistribusian dataset adalah cross-validation. Hasil penelitian diperoleh metode *Random Forest* mendapat akurasi sebesar 93,43%. Kemudian, peneliti [17] melakukan prediksi kelulusan mahasiswa menggunakan metode *Decision Tree* dengan algoritma C4.5. Budiman dkk menghilangkan data *missing value* pada data preparation dengan pendistribusian dataset adalah *split validation*. Hasil penelitian menunjukkan bahwa Algoritma C4.5 memiliki akurasi (AC) sebesar 78,57% dengan true positive rate (TP) sebesar 76,72% dengan menggunakan data *testing* 90% memiliki nilai akurasi kinerja terbaik. Selanjutnya, peneliti [18] melakukan prediksi prestasi siswa menggunakan metode *Decision Tree* dengan algoritma C5.0 merupakan penyempurnaan algoritma C4.5. Benediktus dan Oetama tidak

menghilangkan data *missing value* pada data preparation dengan pendistribusian dataset secara random. Hasil diperoleh bahwa algoritma C5.0 dibagi menjadi data latih sebesar 75% dan data uji sebesar 25% diperoleh akurasi sebesar 71,667%.

Berdasarkan penelitian-penelitian yang dilakukan sebelumnya dalam melakukan prediksi kelulusan menggunakan model klasifikasi, terdapat perbedaan dalam melakukan persiapan data. Beberapa penelitian ada yang menghilangkan data *missing value* [16], [17]. Tetapi ada juga yang tidak menghilangkan data *missing value* [18]. Kemudian terdapat perbedaan dalam melakukan pendistribusian dataset. Beberapa peneliti ada yang menggunakan cross validation [16], secara random [18], dan menggunakan *split validation* [19]. Hal tersebut menjadi salah satu fokus pada tujuan penelitian ini, yaitu untuk menentukan bagaimana pengaruh *missing value* pada data dalam melakukan prediksi kelulusan dan mengetahui konsistensi hasil perhitungan data *training* masing-masing metode klasifikasi menggunakan *split validation*. Selain itu, belum terdapat penelitian terkait perbandingan metode *Decision Tree* dan *Random Forest* untuk melakukan prediksi kelulusan E-Learning. Sehingga peneliti juga akan membandingkan keakurasian dari model klasifikasi menggunakan metode *Decision Tree* menggunakan algoritma C5.0 dan metode *Random Forest*. Banyak *Tools* atau *software* yang ditawarkan untuk melakukan proses klasifikasi salah satunya adalah aplikasi Orange [20]. Orange digunakan untuk menganalisa data atau *Visual Programming (Free Coding)* pada data mining Orange dikembangkan menggunakan Bahasa pemrograman Python [21].

Tujuan penelitian ini untuk mengetahui faktor – faktor yang mempengaruhi kelulusan course peserta Akademi Anti Korupsi lulus. Kemudian, untuk mengetahui perbandingan hasil akurasi antara metode *Decision Tree* dan *Random Forest* dalam memprediksi kelulusan course. Selanjutnya untuk mengetahui pengaruh *missing value* dalam melakukan prediksi kelulusan course dengan model klasifikasi. Serta untuk mengetahui konsistensi perhitungan masing-masing metode *Decision Tree* dan *Random Forest* pada tingkatan jumlah data training.

II. METODOLOGI PENELITIAN

Metode penelitian ini menggunakan salah satu metodologi standar dalam penelitian Data Mining yaitu metode CRISP-DM. [22] menyatakan bahwa CRISP-DM sebagai defacto menjadi standar untuk pengembangan proyek data mining dan knowledge discovery. Dalam prosesnya CRISP-DM mempunyai enam tahapan yaitu:

1. Business Understanding

Pada tahap ini, memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menterjemahkan pengetahuan ini ke dalam pendefinisian masalah dalam data mining. Pada penelitian ini, kebutuhan dari sudut pandang bisnis Akademi Anti Korupsi adalah prediksi kelulusan peserta berdasarkan jumlah course.

2. Data Understanding

Pada tahap ini, dimulai dengan pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data dan mengidentifikasi masalah kualitas data. Dalam penelitian ini, data yang digunakan berasal dari data set pada LMS Moodle Akademi Anti Korupsi data set ini terdiri 12.385 data. Berikut merupakan penjelasan dari masing-masing atribut yang terdapat pada tabel data kelulusan yang dapat dilihat pada tabel 1 di bawah ini:

Tabel 1. Atribut Kelulusan Course

Atribut	Indikator	Unit Analisis
Course	0	Tidak Lulus
	1	Lulus dan dapat mengambil 1 Course
	2	Lulus dan mengambil lebih dari 2 course
Jenis Kelamin	Jenis kelamin peserta	Perempuan atau laki-laki
Usia	Usia peserta	16-25 tahun
		26 - 45 tahun
		>=46 tahun
Pengalaman E-Learning	Riwayat pengalaman E-Learning peserta	Belum pernah Pernah sekali Sudah sering
Pengalaman Pendidikan Antikorupsi	Riwayat pendidikan pendidikan antikorupsi	Belum pernah atau pernah
Pekerjaan	Pekerjaan peserta	ASN/Pegawai Negeri
		Aktivis
		Lain – lain
		Mahasiswa
		Masyarakat
		Pelajar
Alasan Mengikuti	Alasan peserta mengikuti elearning akademi antikorupsi	Ingin lebih mendalami tentang kajian korupsi
		Ingin memperoleh alat untuk berpartisipasi dalam gerakan pemberantasan korupsi
		Menambah wawasan secara umum
		Ingin tahu
		Lain - lain
Institusi	Jenis institusi pekerjaan peserta	Akademik, Pemerintah, Umum
Domisili	Alamat peserta berdasarkan kepulauan	Jawa
		Kalimantan
		Maluku & Papua
		Nusa Tenggara & Bali
		Sulawesi
		Sumatera

Pada tabel 1 merupakan atribut yang akan digunakan untuk melakukan klasifikasi kelulusan menggunakan 8 atribut yang terdiri dari jenis kelamin, pekerjaan, pengalaman pendidikan anti korupsi, umur, institusi, domisili, pengalaman e-learning dan alasan mengikuti e-learning anti korupsi.

3. Data Preparation

Pada tahap ini, data preparation dilakukan menggunakan software untuk melakukan proses pembersihan data dengan menghilangkan data missing value dan transformasi data dengan mengubah format data kedalam bentuk numeric. Untuk kemudian dijadikan masukan dalam tahap pemodelan.

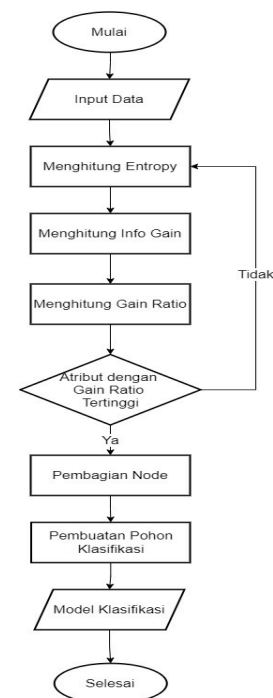
4. Modelling

Pada tahap ini dilakukan analisis deskriptif terlebih dahulu. Kemudian dilakukan pemodelan dengan dua metode klasifikasi sebagai berikut:

a. Analisis Deskriptif

Analisis deskriptif dilakukan untuk mendeskripsikan karakteristik peserta akademi anti korupsi berdasarkan variabel dependen dan variabel independen.

b. Memodelkan dengan metode Decision Tree

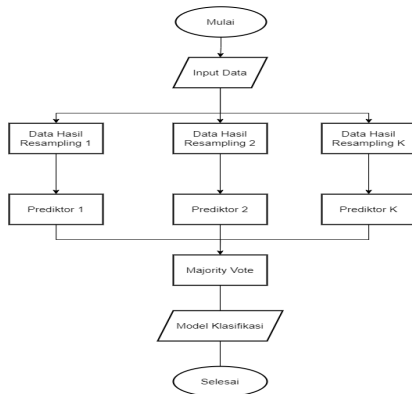


Gambar 1. Pemodelan Dengan Metode *Decision Tree*

Langkah-langkah untuk memodelkan dengan metode Decision Tree menggunakan algoritma C5.0 dapat dilihat pada gambar 1. Langkah-langkah klasifikasi menggunakan algoritma C5.0 yang harus dilakukan pertama kali adalah input data dengan memilih atribut yang akan digunakan pada saat proses klasifikasi. Pada penelitian ini menggunakan 8 atribut terdiri dari jenis kelamin, pekerjaan, pengalaman pendidikan anti korupsi, umur, institusi, domisili, pengalaman e-learning dan alasan mengikuti e-learning anti korupsi. Langkah selanjutnya yaitu memilih node akar dengan menghitung entropy, gain,

dan gain ratio. Atribut yang memiliki nilai gain ratio tertinggi dibanding atribut lainnya maka atribut tersebut yang dijadikan node akar. Setelah mendapat node akar, hitung kembali entropy, gain, dan gain ratio pada atribut sisanya untuk dijadikan node cabang. Lakukan perhitungan entropy, gain, dan gain ratio hingga tidak ada atribut yang tersisa. Setelah semuanya selesai, maka akan terbentuk pohon klasifikasi menggunakan algoritma C5.0.

c. Memodelkan dengan metode Random Forest



Gambar 2. Pemodelan Dengan Metode *Random Forest*

Langkah-langkah klasifikasi menggunakan metode *Random Forest* adalah input data, pada penelitian ini menggunakan 8 atribut. Menggunakan teknik resampling bootstrap untuk mengambil data sampel. Dilakukan resample data dengan berulang-ulang dalam membuat banyak simulasi sampel. Membuat pohon klasifikasi dari hasil resampling bootstrap. Variabel predictor yang diambil secara acak menjadi acuan dalam penentuan pemilahan terbaik. Melakukan prediksi klasifikasi data sampel berdasarkan pohon klasifikasi yang terbentuk. Mengulangi langkah tersebut untuk memperoleh sejumlah pohon klasifikasi. Melakukan prediksi klasifikasi menggunakan aturan majority vote. Setelah semuanya selesai, maka akan terbentuk model klasifikasi menggunakan metode *Random Forest*.

5. Evaluation/ Interpretation

Pada tahap ini, akan dilakukan evaluasi untuk mengetahui apakah model decision tree dan random forest dapat mencapai tujuan yang ditetapkan pada fase awal [23]. Terdapat beberapa cara untuk mengukur performa klasifikasi, beberapa cara yang sering digunakan adalah dengan menghitung accuracy, precision, dan recall.

- Accuracy merupakan adalah ukuran untuk mengukur ketepatan prediksi pengklasifikasi pada kelas tertentu. Rumus untuk menghitung akurasi klasifikasi sesuai persamaan 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Precision merupakan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model klasifikasi. Rumus untuk menghitung akurasi klasifikasi sesuai persamaan 2.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- Recall adalah ukuran untuk mengukur berapa banyak data dari kelas tertentu yang dapat diprediksikan secara benar. Rumus untuk menghitung akurasi klasifikasi sesuai persamaan 3

$$Recall = \frac{TN}{TN + FP} \quad (3)$$

- Skor F1 adalah perbandingan rata-rata presisi dan Recall yang dibobotkan. Rumus untuk menghitung F1 dapat dilihat pada persamaan 4.

$$Skor F1 = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (4)$$

TP adalah True Positive, FP adalah False Positive, TN adalah True Negative, dan FN adalah False Negative.

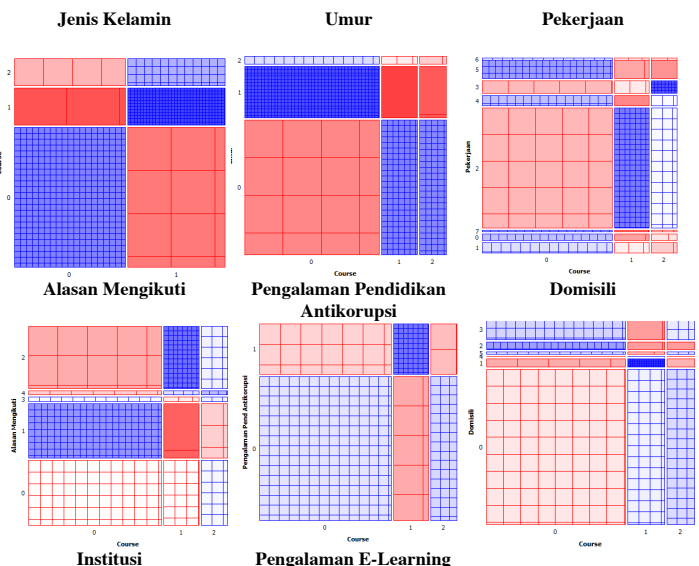
Kemudian, menggunakan Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. Apabila garis linear ROC mendekati nilai 1 maka lebih baik dalam memodelkan klasifikasi.

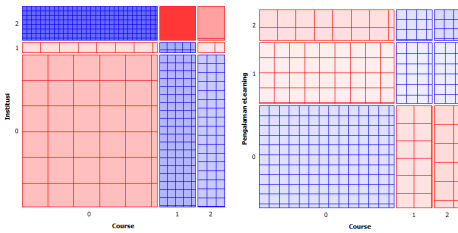
6. Perbandingan Metode Decision Tree dan Random Forest Hasil pengujian masing-masing metode akan dibandingkan untuk mendapatkan model prediksi terbaik.

III. HASIL DAN PEMBAHASAN

A. Analisis Deskriptif

Analisis deskriptif digunakan untuk menjelaskan karakteristik peserta akademi antikorupsi berdasarkan tiga status kelulusan course dan tujuh faktor kelulusan course dapat dilihat pada gambar 1.



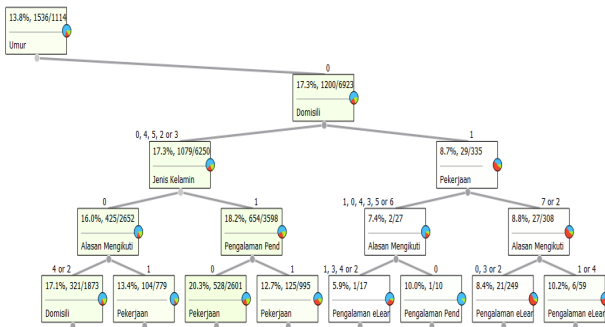


Gambar 3. Analisis Deskriptif Untuk Faktor Kelulusan

Berdasarkan gambar 3, untuk warna menunjukkan apakah penyimpangan dari independensi positif (biru) atau negatif (merah). Besar kecilnya kotak menunjukkan proporsi data yang menunjukkan banyaknya frekuensi terjadi dalam dataset. Untuk jenis kelamin memiliki proporsi lulus lebih tinggi dengan jenis kelamin perempuan. Pada umur diatas usia 16-25 tahun memiliki proporsi lulus lebih tinggi. Pada pengalaman e-learning memiliki proporsi lulus lebih tinggi apabila belum mempunyai pengalaman e-learning. Pengalaman pendidikan antikorupsi memiliki proporsi lulus lebih tinggi apabila mempunyai pengalaman. Pekerjaan memiliki proporsi lulus lebih tinggi apabila memiliki pekerjaan sebagai mahasiswa. Alasan mengikuti memiliki proporsi lulus lebih tinggi apabila alasannya menambah wawasan secara umum. Institusi memiliki proporsi lulus lebih tinggi dengan peserta berasal dari akademik. Domisili memiliki proporsi lulus lebih tinggi dari Kepulauan Jawa.

B. Faktor Kelulusan Course

Pada penelitian ini, untuk mengetahui faktor yang mempengaruhi kelulusan course dengan menggunakan metode decision tree. Hasil analisis algoritma decision tree untuk menentukan faktor kelulusan course dapat dilihat pada gambar 2.



Gambar 4. Metode Decision Tree

Dari Gambar 4, terdapat banyak simpul untuk mendapatkan faktor yang dapat mempengaruhi kelulusan course dengan target class lulus. Simpul akar dari pohon ini adalah umur yang artinya umur merupakan atribut yang paling informatif untuk membangun sebuah pohon. Sehingga diperoleh kesimpulan bahwa faktor pertama yang mempengaruhi kelulusan seorang yaitu peserta yang berumur 16-25 tahun. Kedua, peserta yang berdomisili di kepulauan jawa. Ketiga, peserta yang berjenis kelamin perempuan.

Keempat, peserta yang belum pernah memiliki pengalaman pendidikan antikorupsi. Kelima, peserta yang memiliki pekerjaan pelajar, pegawai negeri dan masyarakat. Keenam, peserta yang memiliki alasan mengikuti yaitu ingin lebih mendalami tentang kajian korupsi, ingin memperoleh alat untuk berpartisipasi dalam gerakan pemberantasan korupsi, dan ingin tahu. Ketujuh, peserta yang memiliki pengalaman E-learning sudah sering. Kedelapan, peserta yang berasal dari institusi akademik.

C. Proporsi Jumlah Data Training

Berdasarkan penelitian sebelumnya mengenai model klasifikasi pada prediksi kelulusan seseorang, terdapat perbedaan dalam menangani pendistribusian. Pengaruh persentase split validation pada dataset kelulusan adalah adanya perbedaan nilai keakurasian yang dihasilkan, baik pada metode Decision Tree, dan Random Forest. Berikut nilai keakurasian model klasifikasi dapat dilihat pada gambar 5.

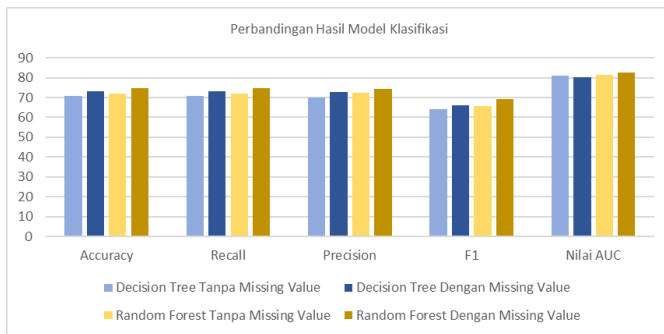
Gambar 5. Hasil Perhitungan Klasifikasi Berdasarkan Jumlah Data Training

Hasil analisa dari gambar 5, dapat diketahui bahwa hasil pengklasifikasian dataset dengan model klasifikasi menggunakan orange, split validation berpengaruh pada hasil pengujian. Hasil pengujian untuk metode Decision Tree dan Random Forest memiliki akurasi tertinggi diperoleh dari pengujian menggunakan dataset dengan missing value pada data training sebesar 90% jumlah record 11.147 dan data testing sebesar 10% dengan jumlah record 1.238. metode Random memperoleh Accuracy sebesar 75,7 % sedangkan dengan menggunakan Decision Tree diperoleh Accuracy sebesar 74%. Berdasarkan nilai variabilitas, hasil pengujian untuk metode Random Forest dan Decision Tree memiliki variability kurang dari 1. Oleh karena itu dapat menggunakan semua percentage split. Pada penelitian ini akan menggunakan pengujian dengan data training sebesar 70%.

D. Performa Kelulusan Course

Berdasarkan penelitian sebelumnya mengenai model klasifikasi pada prediksi kelulusan seseorang, terdapat perbedaan dalam menangani data missing value yaitu menggunakan data missing value atau menghilangkannya. Atribut yang berkorelasi terbesar dengan kelulusan yaitu umur, jenis kelamin, dan institusi. Pengaruh missing value pada

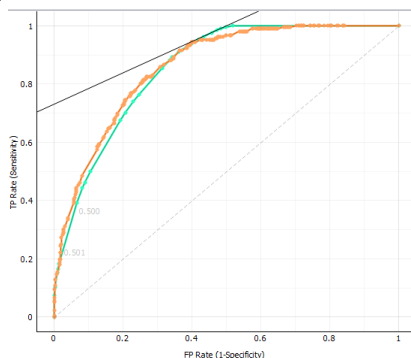
dataset kelulusan adalah adanya perbedaan nilai keakurasian yang dihasilkan, baik pada metode Decision Tree, dan Random Forest. Berikut nilai keakurasian yang dihasilkan dapat dilihat pada gambar 6.



Gambar 6. Hasil Uji Metode Klasifikasi

Dari gambar 6, dapat terlihat bahwa dalam penerapan model klasifikasi berdasarkan tingkat keakurasian yaitu metode Random Forest dan Decision Tree menggunakan dataset dengan missing value memiliki keakurasian lebih baik dibandingkan menggunakan dataset yang dihilangkan dahulu data missing value. Hasil analisa dari gambar 6, dapat diketahui bahwa hasil pengklasifikasian dataset dengan model klasifikasi menggunakan dataset missing value berpengaruh pada hasil pengujian. Sehingga model klasifikasi dapat bekerja secara baik dalam mengatasi permasalahan missing value tanpa harus dilakukan imputasi data.

Kemudian melakukan pengujian menggunakan Kurva ROC untuk menunjukkan akurasi dan membandingkan klasifikasi secara visual. Kurva ROC masing-masing metode ditandai dengan warna grafik yang berbeda. Untuk metode Decision Tree ditandai dengan warna hijau, metode SVM warna biru dan metode Random Forest dengan warna orange. Kurva ROC untuk skema terbaik klasifikasi disajikan pada Gambar 7.



Gambar 7. Kurva ROC untuk klasifikasi

Pada kedua metode terlihat bahwa nilai tidak lebih dari 1 untuk sensitivity dan specificity. Semakin dekat kurva mengikuti batas kiri dan kemudian batas atas ruang ROC, semakin akurat classifier tersebut. Dari ROC Analysis tersebut

dapat diketahui bahwa menggunakan dataset dengan missing value metode Random Forest memiliki keakuratan dari classifier lebih baik dibandingkan Decision Tree. Hal ini dapat dilihat bahwa kurva dari metode Random Forest rata-rata mendekati sumbu axis atau sumbu Y yang menandai bahwa metode Random Forest memiliki keakuratan lebih baik.

IV. KESIMPULAN

Berdasarkan analisis terhadap hasil pengujian yang telah dilakukan terhadap metode Decision Tree dan Random Forest dengan 12.385 data kelulusan, maka faktor yang mempengaruhi tingkatan kelulusan seseorang berdasarkan model klasifikasi secara berurutan adalah umur, domisili, jenis kelamin, pengalaman pendidikan anti korupsi, pekerjaan, alasan mengikuti e-learning, dan institusi. Perbandingan metode Decision Tree dan Random Forest memberikan kinerja yang cukup dalam memprediksi kelulusan seseorang. Metode Random Forest memiliki performance lebih baik dibandingkan Decision Tree. Menggunakan data training sebesar 70% dengan jumlah record 8.670 dan data testing sebesar 10% dengan jumlah record 3.715 diperoleh hasil Accuracy 74.6%, Recall 74.6, Precision 74.3, F1 69.1 dan nilai AUC 82.6. Penanganan missing value dapat mempengaruhi nilai keakurasian model klasifikasi. Pada metode Decision Tree dan Random Forest penanganan data missing value dengan tetap menggunakannya menghasilkan nilai keakurasian lebih baik dari pada menghilangkan missing value. Sehingga metode Decision Tree dan Random Forest mempunyai kelebihan mampu mengatasi missing value tanpa harus melakukan imputasi data. Pendistribusian data training dapat mempengaruhi nilai keakurasian model klasifikasi. Hasil pengujian untuk metode Decision Tree dan Random Forest memperoleh akurasi tertinggi didapat dari pengujian dengan nilai split validation yaitu data training sebesar 90% dengan jumlah record 11.147 dan data testing sebesar 10% dengan jumlah record 1.238. 3. Rekomendasi untuk E-Learning Akademi Anti Korupsi dalam melakukan sasaran atau target bisnisnya yaitu perempuan dengan rentan usia 16-25 tahun yang berdomisili di Jawa.

UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terima kasih kepada Universitas Komputer Indonesia dan Akademi Anti Korupsi yang telah memberikan kesempatan dan bimbingan yang sangat membantu dalam penelitian.

DAFTAR PUSTAKA

- [1] APJII, 'Bulletin APJII 2020', 2020.
- [2] T. McCue, 'E Learning Climbing To \$325 Billion By 2025 UF Canvas Absorb Schoology Moodle', *Forbes.com*, 2018.
<https://www.forbes.com/sites/tjmccue/2018/07/31/e-learning-climbing-to-325-billion-by-2025-uf-canvas-absorb-schoology-moodle/?sh=642fb843b395>.

- [3] W. Hartanto, 'Penggunaan E-Learning sebagai Media Pembelajaran', *J. Pendidik. Ekon.*, vol. 10, no. 1, pp. 1–18, 2016.
- [4] Indonesia Corruption Watch, 'Laporan Akhir Tahun ICW 2018', 2019.
- [5] Indonesian Corruption Watch, 'Laporan Akhir Tahun ICW 2021', 2022.
- [6] Indonesian Corruption Watch, 'Laporan Akhir Tahun ICW 2020', p. 13, 2021.
- [7] S. Wahyuddin, F. I. Estiko, and E. Rijanto, 'Analysis of Factors Affecting Tuition Fee in a Private University: A Data Mining Using VAR Model', *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 662, no. 2, 2019, doi: 10.1088/1757-899X/662/2/022050.
- [8] S. Suyanto, *Data Mining Untuk Klasifikasi dan Klusterisasi Data*, Edisi Revi. Bandung: Informatika Bandung, 2019.
- [9] S. F. Fabrianne, A. Triayudi, I. D. Sholihati, P. Bachhal, S. Ahuja, and S. Gargrish, 'A Comparative Study to Predict Student 's Performance Using Educational Data Mining Techniques', *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 215, no. 012036, 2017, doi: 10.1088/1757-899X/215/1/012036.
- [10] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2016.
- [11] D. G. D. Funcion, 'Predicting Student Academic Performance in Computer Organization Course: Using J48 Algorithm', *Indian J. Sci. Technol.*, vol. 11, no. 47, pp. 1–8, 2017, doi: 10.17485/ijst/2018/v11i47/130870.
- [12] M. Hemantkumar and A. Adholiya, 'Predicting Students' Performance using J48 Decision Tree', *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 4, no. March 2019, pp. 132–136, 2019.
- [13] S. Wiyono, D. S. Wibowo, and M. F. Hidayatullah, 'Comparative Study of KNN, SVM and Decision Tree Algorithm for Student 's Performance Prediction', *Int. J. Comput. Sci. Appl. Math.*, vol. 6, no. 2, pp. 50–53, 2020.
- [14] D. Y. Putri, R. Andreswari, and M. A. Hasibuan, 'Analysis of Students Graduation Target Based on Academic Data Record Using C4.5 Algorithm Case Study: Information Systems Students of Telkom University', *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1–6, 2019, doi: 10.1109/CITSM.2018.8674366.
- [15] D. Petkovic *et al.*, 'Using the random forest classifier to assess and predict student learning of Software Engineering Teamwork', *Proc. - Front. Educ. Conf. FIE*, pp. 1–7, 2016, doi: 10.1109/FIE.2016.7757406.
- [16] M. Utari, B. Warsito, and R. Kusumaningrum, 'Implementation of Data Mining for Drop-Out Prediction using Random Forest Method', *2020 8th Int. Conf. Inf. Commun. Technol. ICoICT 2020*, 2020, doi: 10.1109/ICoICT49345.2020.9166276.
- [17] E. Budiman, Haviluddin, N. Dengan, A. H. Kridalaksana, M. Wati, and Purnawansyah, 'Performance of Decision Tree C4.5 Algorithm in Student Academic Evaluation', *Int. Conf. Comput. Sci. Technol.*, vol. 488, pp. 380–389, 2018, doi: https://doi.org/10.1007/978-981-10-8276-4_36.
- [18] N. Benediktus and R. S. Oetama, 'The Decision Tree C5.0 Classification Algorithm for Predicting Student Academic Performance', *Ultim. J. Tek. Inform.*, vol. 12, no. 1, pp. 14–19, 2020, doi: 10.31937/ti.v12i1.1506.
- [19] W. Baswardono, D. Kurniadi, A. Mulyani, and D. M. Arifin, 'Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification', *J. Phys. Conf. Ser.*, vol. 1402, no. 6, 2019, doi: 10.1088/1742-6596/1402/6/066055.
- [20] A. Ishak, K. Siregar, Asfiryati, R. Ginting, and M. Afif, 'Orange Software Usage in Data Mining Classification Method on the Dataset Lenses', *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1003, no. 1, 2020, doi: 10.1088/1757-899X/1003/1/012113.
- [21] A. Ram and H. Vishwakarma, 'Diabetes Prediction using Machine learning and Data Mining Methods', *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1116, no. 1, p. 012135, 2021, doi: 10.1088/1757-899x/1116/1/012135.
- [22] G. Mariscal, Ó. Marbán, and C. Fernández, 'A survey of data mining and knowledge discovery process models and methodologies', *Knowl. Eng. Rev.*, vol. 25, no. 2, pp. 137–166, 2010, doi: 10.1017/S0269888910000032.
- [23] I. Sumitra, R. Fachrudin, and S. Supatmi, 'Forecasting the Inflation using Hybrid SARIMA-Single Exponential Smoothing for Determining Minimum Costs of Living Index in Bandung City', 2019, doi: 10.4108/eai.18-7-2019.2287710.