

Penerapan Algoritma *Decision Tree C4.5* Untuk Klasifikasi Mahasiswa Berpotensi *Drop out* Di Universitas Advent Indonesia

Daniel Sinaga*¹, Edwin J. Solaiman², Fergie Joanda Kaunang³

¹PT. Telekomunikasi Indonesia (Telkom), Sumatera Utara, Indonesia

^{2,3}Fakultas Teknologi Informasi, Universitas Advent Indonesia

e-mail: *¹sinagad303@gmail.com, ²edwinsolaiman@unai.edu, ³fergie.kaunang@unai.edu

Abstrak

Salah satu faktor yang menentukan kualitas perguruan tinggi adalah persentase kemampuan mahasiswa untuk menyelesaikan studi tepat waktu. Saat ini, masalah kegagalan studi siswa dan faktor - faktor penyebabnya menjadi topik yang menarik untuk diteliti. Perguruan tinggi perlu mendeteksi perilaku mahasiswa yang memiliki status "tidak diinginkan" tersebut sehingga dapat diketahui faktor - faktor penyebab kegagalannya. Berdasarkan uraian di atas, diperlukan analisa terhadap data-data mahasiswa seperti Jenis Kelamin, Umur, Agama, Tempat tinggal, IPS, Disiplin, dan Hutang, berdasarkan data mahasiswa yang dimiliki sebanyak 97 data sehingga bisa dimanfaatkan dalam pengolahan data mining. Di mana data mining digunakan untuk menggali dan mendapatkan informasi dari data dengan jumlah besar. Salah satu metode data mining adalah pengklasifikasian data. Dengan menggunakan Metode Klasifikasi dengan konsep Algoritma Decision tree C4.5 menghasilkan accuracy sebesar 90.00%, hasil dari precision adalah 87.50, dan hasil dari recall sebesar 100%. Diharapkan dapat meningkatkan keinginan Lembaga Universitas atau Perguruan tinggi untuk memberikan pikiran yang baik, pandangan, dan kebijakan baru kepada mahasiswa yang memiliki permasalahan dalam perkuliahan, dengan kata lain memaksimalkan mahasiswa dalam upaya peningkatan persentase minat kuliah mahasiswa.

Kata Kunci: *Decision Tree C4.5, Drop out, Data Mining*

Implementation of C4.5 Decision Tree Algorithm To Classify Potentially Drop out Students At Universitas Advent Indonesia

Abstract

One of the factors that determine the quality of higher education is the percentage of students' ability to complete their studies on time. At present, the problem of student failure and the factors causing it to be an interesting topic to research. Higher education institutions need to detect the behavior of students who have an "undesirable" status so that the factors causing their failure can be identified. Based on the description above, it is necessary to analyze student data such as Gender, Age, Religion, Residence, Social Studies, Discipline, and Debt, based on student data that is as much as 98 data so that it can be used in data mining processing. Where data mining is used to dig and get information from large amounts of data. One of the data mining methods is data classification. By using the classification method with the concept of the C4.5 Decision Tree Algorithm, it produces an accuracy of 90.00%, the result of precision is 87.50, and the result of the recall is 100%. It is hoped that it can increase the desire of the University or Higher Education Institution to provide good thoughts, views, and new policies to students who have problems in lectures, in other words maximizing students in an effort to increase the percentage of student interest in college.

Keywords: *Decision Tree C4.5, Drop out, Data Mining*

1. Pendahuluan

Perguruan Tinggi merupakan penyelenggara pendidikan akademik bagi mahasiswa. Lima lembaga perguruan tinggi di antaranya adalah universitas, institut, sekolah tinggi, akademi dan politeknik. Berdasarkan data yang diperoleh dari Pusat Statistik Pendidikan Badan Penelitian dan Pengembangan Departemen Pendidikan Nasional Republik Indonesia menyebutkan bahwa jumlah lembaga penyelenggara perguruan tinggi mengalami peningkatan setiap tahunnya [1].

Perguruan tinggi diharapkan menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan sumber daya manusia yang berilmu, cakap dan kreatif. Semakin bertambah jumlah perguruan tinggi maka semakin meningkat pula jumlah sumber daya manusia berkualitas yang dihasilkan perguruan tinggi. Salah satu faktor yang menentukan kualitas perguruan tinggi adalah persentase kemampuan mahasiswa untuk menyelesaikan studi tepat waktu [2]. Saat ini, masalah kegagalan studi siswa dan faktor - faktor penyebabnya menjadi topik yang menarik untuk diteliti. Dalam penelitian yang dilakukan oleh [3] diperoleh 17 aturan yang dapat digunakan untuk menentukan mahasiswa yang berpotensi untuk *drop out*. Muhamad et al. [4], menggunakan algoritma C4.5 untuk mengklasifikasikan siswa-siswa yang berpotensi *drop out*. Penelitian tersebut menghasilkan 9 aturan untuk klasifikasi potensi siswa *drop out* dengan hasil evaluasi kinerja model yakni sebesar 96.15%.

Sebagian besar masalah akademik yang terjadi di perguruan tinggi mengakibatkan mahasiswa mengundurkan diri dan *drop out*. Penentuan keputusan *drop out* oleh sebuah perguruan tinggi bukan hal yang mudah dilakukan karena harus melihatkan berbagai variabel atau kriteria akademis yang telah ditetapkan oleh perguruan tinggi secara jelas dan diketahui mahasiswa sejak awal menjadi mahasiswa. Oleh karena itu diperlukan analisis seperti atribut apa saja yang mempengaruhi mahasiswa untuk melakukan *drop out* serta klasifikasi mahasiswa yang berpotensi *drop out*. Tujuan dari penelitian adalah membangun sebuah model untuk mengklasifikasikan mahasiswa yang berpotensi *drop out* dan bisa digunakan oleh Universitas Advent Indonesia untuk dapat membantu mahasiswa dan instansi universitas untuk mengetahui hal-hal yang bisa mempengaruhi mahasiswa dalam pendidikan.

2. Metode Penelitian

Penelitian ini dilakukan dengan studi pustaka, dan analisis masalah lalu mengambil data dari database UNAI, yaitu data yang dikumpulkan selama penelitian mengenai fakta-fakta dari objek yang diteliti. Subyek yang diteliti adalah mahasiswa angkatan 2017-2018. Agar lebih mudah dipahami, peneliti menjabarkan tahapan-tahapan yang dilakukan dalam penelitian ini dengan menggunakan bagan alur yang dapat dilihat pada Gambar 1.

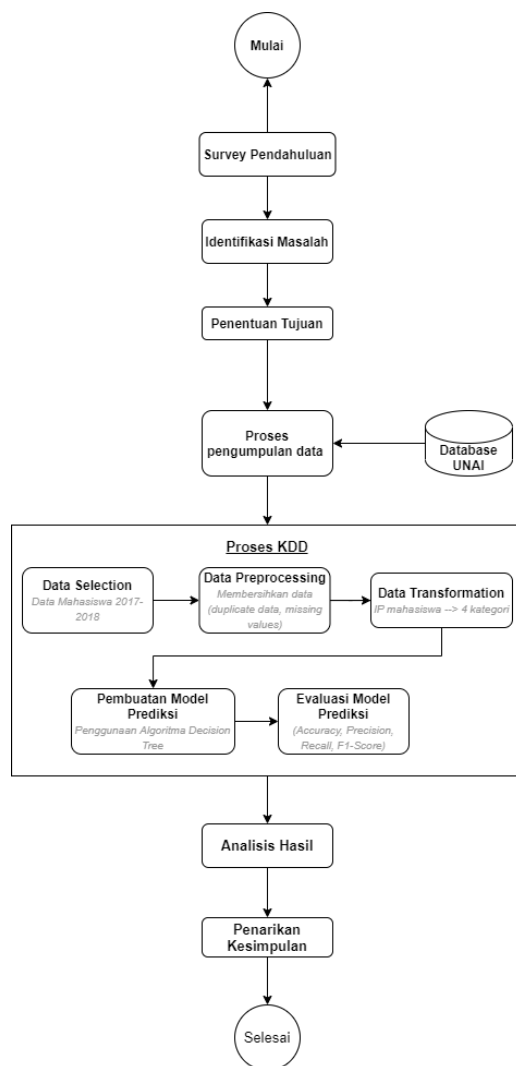
Proses KDD

Data mining adalah proses yang digunakan untuk mengidentifikasi pola yang valid, baru, berpotensi berguna, dan pada akhirnya dapat dipahami dalam data. *Data mining*, yang juga dikenal dengan Knowledge Discovery in Databases (KDD), juga dapat didefinisikan sebagai suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting [5]. KDD sendiri memiliki beberapa tahapan yang perlu dilakukan untuk memperoleh hasil berupa pola, pengetahuan ataupun informasi tertentu dari sebuah database. Tahapan-tahapan tersebut yakni *Selection, Preprocessing, Transformation, Data Mining, Interpretation and Evaluation, Data Visualization* [6]. Peneliti menggunakan tahapan-tahapan yang sama dalam melakukan penelitian ini untuk memperoleh hasil akhir dari data yang telah dikumpulkan.

Data Selection

Data selection merupakan proses pemilihan atau seleksi data dari sekumpulan data operasional yang tersimpan dalam data base. Data hasil seleksi kemudian akan digunakan untuk proses data mining, data disimpan dalam suatu berkas, yang terpisah dari data base operasional. Penulis telah melakukan

wawancara terhadap pihak Biro Administrasi Akademik UNAI untuk menentukan beberapa data yang diperlukan untuk penelitian sesuai izin dari pihak bersangkutan. Dikarenakan beberapa data bersifat pribadi dan tidak dapat digunakan oleh penulis. Data yang dapat digunakan tersebut yaitu mahasiswa Fakultas Teknologi Informasi, jurusan Teknik Informatika dan Sistem Informasi angkatan 2017-2018 yang telah melewati masa studi empat semester yaitu sebanyak 98 data.



Gambar 1 Bagan Alur Penelitian

Data Preprocessing

Data preprocessing merupakan proses pembersihan data, menghilangkan *noise* dan data yang tidak konsisten. Pada umumnya data yang diperoleh dari suatu basis data memiliki bagian-bagian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga salah ketik [7]. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan atribut *data mining* yang dimiliki. Data-data yang tidak relevan itu dibuang oleh karena dapat mengurangi mutu atau akurasi dari hasil *data mining* nantinya. Pembersihan data juga akan mempengaruhi kinerja dari sistem *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya. Data yang penulis peroleh dari Biro Administrasi Akademik UNAI juga terdapat 1 baris data yang kosong seperti tidak tercatatnya data IP semester yang ada pada semester tertentu sehingga penulis tidak memasukkan data tersebut karena dianggap dapat

mempengaruhi proses pengolahan data. Selain itu saat memindahkan data penulis juga memperbaiki kesalahan pengetikan yang terjadi saat memasukan data ke dalam sistem basis data Biro Administrasi Akademik UNAI. Jumlah total data yang digunakan dalam penelitian ini adalah 97 data.

Data Transformation

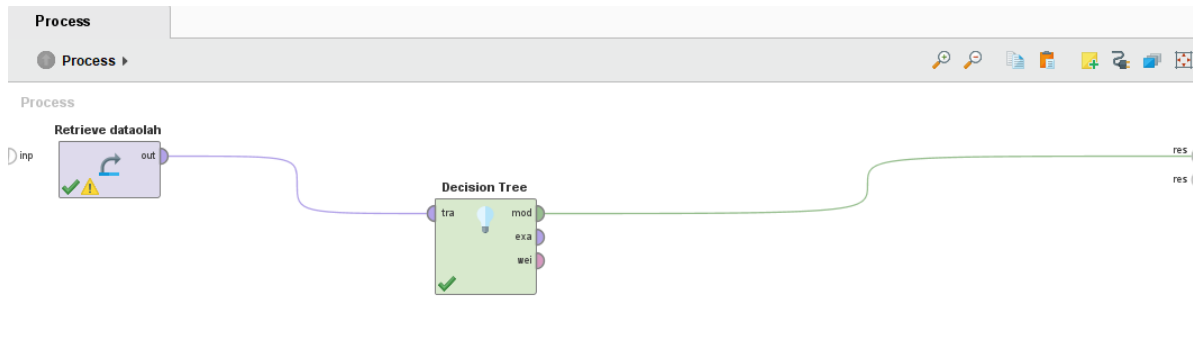
Data transformation merupakan proses mengubah data menjadi bentuk yang di perlukan [8]. Pada penelitian ini penulis menggunakan atribut Jenis kelamin, Umur, Agama, Tempat Tinggal, dan IP Semester. Peneliti kemudian menambahkan dua atribut lainnya yakni atribut Disiplin untuk melihat dari sudut pandang yang lain yakni sikap mahasiswa selama berkuliah, serta atribut Keuangan untuk mengetahui apakah mahasiswa memiliki masalah keuangan selama berkuliah. Berdasarkan buku panduan UNAI pengelompokan indeks prestasi mahasiswa terbagi menjadi 4 predikat IPS yang kemudian peneliti kelompokkan menjadi empat kategori yaitu *Under* dengan IP 2.33 – 2.75, *Memuaskan* dengan IP 2.76 – 3.00, *Sangat Memuaskan* dengan IP 3.01 – 3.50 dan *Dengan Pujian* dengan IP 3.51 – 4.00.

Tabel 1 Data setelah transformasi

Berpotensi	Jenis Kelamin	Umur	Agama	Status Tinggal	IP Semester	Disiplin	Masalah Keuangan
Ya	Laki-Laki	23	KRISTEN ADVENT	Outside	Under	Pernah	Ya
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Sangat Memuaskan	Pernah	Ya
Ya	Laki-Laki	20	KRISTEN ADVENT	Outside	Dengan Pujian	Tidak	Tidak
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Under	Pernah	Tidak
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Dengan Pujian	Tidak	Tidak
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Memuaskan	Tidak	Tidak
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Dengan Pujian	Tidak	Tidak
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Sangat Memuaskan	Pernah	Ya
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Sangat Memuaskan	Pernah	Tidak
Ya	Laki-Laki	22	KRISTEN ADVENT	Outside	Under	Pernah	Ya
Tidak	Perempuan	21	KRISTEN ADVENT	Outside	Dengan Pujian	Tidak	Tidak
Tidak	Laki-Laki	21	KRISTEN PROTESTAN	Outside	Sangat Memuaskan	Tidak	Tidak
Tidak	Laki-laki	20	KRISTEN ADVENT	Outside	Sangat Memuaskan	Pernah	Tidak
Tidak	Laki-Laki	22	KRISTEN ADVENT	Outside	Dengan Pujian	Tidak	Tidak
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Dengan Pujian	Tidak	Tidak
Ya	Laki-Laki	23	KRISTEN ADVENT	Outside	Memuaskan	Pernah	Ya
Ya	Laki-laki	23	KRISTEN ADVENT	Outside	Under	Pernah	Tidak
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Sangat Memuaskan	Tidak	Ya
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Dengan Pujian	Tidak	Tidak
Ya	Laki-Laki	21	KRISTEN PROTESTAN	Outside	Under	Pernah	Tidak
Tidak	Laki-Laki	21	KRISTEN ADVENT	Outside	Dengan Pujian	Tidak	Tidak

Pembuatan Model Menggunakan *Data Mining*

Setelah proses transformasi data dilakukan maka diperoleh data akhir yang akan digunakan dalam penelitian ini. Data tersebut yang berjumlah 97 data merupakan *data training* yang kemudian diolah dengan menggunakan perangkat lunak *Rapidminer Studio* dan algoritma yang digunakan adalah algoritma *Decision Tree C4.5* [9].



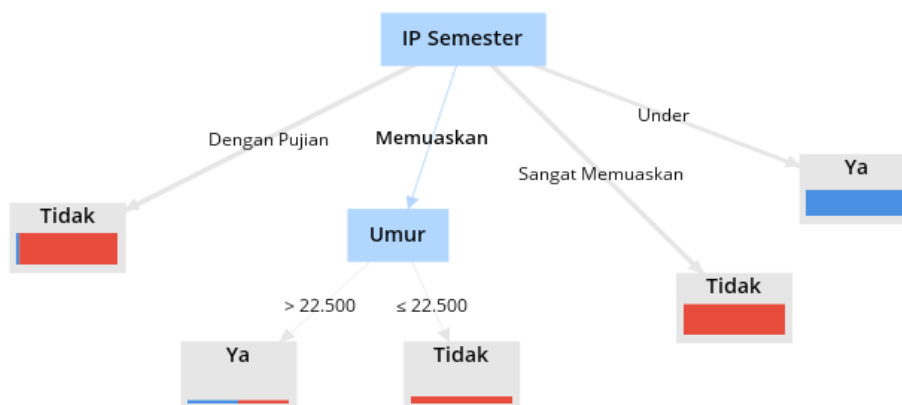
Gambar 2 Pembuatan model dan pemilihan algoritma dalam Rapidminer

RapidMiner Studio

RapidMiner adalah *platform* perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*text mining*), dan analisis prediktif (*Predictive analytics*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototype dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan [10].

3. Hasil dan Pembahasan

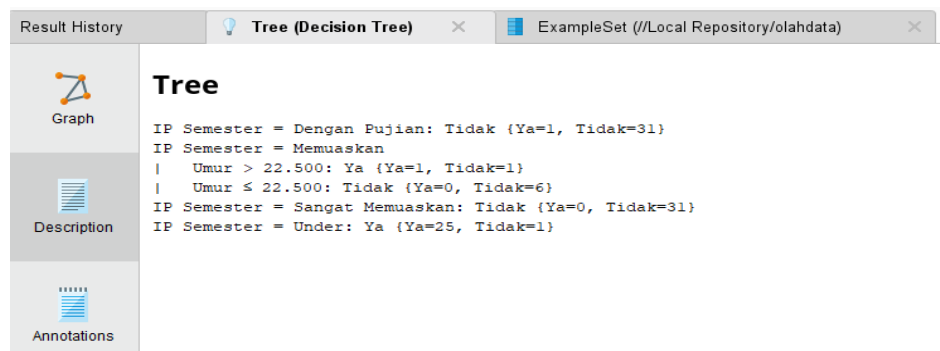
Hasil uji pohon keputusan yang dilakukan oleh algoritma *Decision Tree C4.5* di *RapidMiner Studio* menunjukkan bahwa IP Semester menjadi *root*. Gambar pohon keputusan dapat dilihat seperti pada Gambar 3. Berdasarkan hasil tersebut, atribut IP Semester merupakan hal yang sangat berpengaruh dalam klasifikasi mahasiswa berpotensi *drop out*. Dapat dilihat bahwa mahasiswa yang memiliki IP semester dengan kategori *Under* yaitu dengan IP 2.33 – 2.75 diklasifikasikan sebagai mahasiswa berpotensi *drop out* sedangkan mahasiswa dengan kategori Sangat Memuaskan dan Dengan Pujian tidak berpotensi *drop out*.



Gambar 3 Hasil Pohon Keputusan

Adapun *rules* atau aturan yang dihasilkan dari uji pohon keputusan dapat dilihat pada Gambar 4 dimana diperoleh 5 rules atau aturan untuk klasifikasi mahasiswa berpotensi *drop out* yakni :

- Jika IP Semester mahasiswa berada pada kategori Dengan Pujian (3.51 – 4.00) maka mahasiswa tersebut tidak berpotensi *drop out*
- Jika IP Semester mahasiswa berada pada kategori Memuaskan (2.76 – 3.00) dan berumur di atas 22 tahun maka mahasiswa tersebut memiliki kemungkinan untuk *drop out*
- Jika IP Semester mahasiswa berada pada kategori Memuaskan (2.76 – 3.00) dan berumur di atas 22 tahun maka mahasiswa tersebut memiliki kemungkinan untuk *drop out*
- Jika IP Semester mahasiswa berada pada kategori Sangat Memuaskan (3.01 – 3.50) dan berumur di bawah 22 tahun maka mahasiswa tersebut tidak berpotensi *drop out*
- Jika IP Semester mahasiswa berada pada kategori *Under* (2.33 – 2.75) maka mahasiswa tersebut berpotensi untuk *drop out*



Gambar 4 Deskripsi Aturan dari Hasil Uji Pohon Keputusan

Selanjutnya peneliti melakukan proses evaluasi kinerja dari model yang telah dikembangkan. Untuk mengukur kinerja dari model yang telah dikembangkan dalam penelitian ini peneliti menggunakan *Accuracy*, *Precision*, dan *Recall*. Hasil yang diperoleh yaitu *Accuracy* 90%, *Precision* 87.5%, dan *Recall* 100%.

4. Kesimpulan

Berdasarkan hasil penelitian dan pengolahan data yang telah dilakukan mengenai Mahasiswa berpotensi untuk *drop out*, maka penulis mengambil kesimpulan bahwa algoritma *data mining Decision Tree* dapat mengklasifikasikan mahasiswa yang berpotensi *drop out* dengan hasil evaluasi yakni atribut yang sangat berpengaruh adalah IP semester. 2. Pengolahan data pada penelitian ini menggunakan algoritma *Decision Tree C4.5* dengan 97 data training menghasilkan akurasi 90.00% yang tergolong sangat tinggi. Namun, untuk meningkatkan hasil dari penelitian ini, perlu ditambahkan jumlah data serta atribut – atribut yang berkaitan dengan kegiatan mahasiswa sehingga dapat memberikan hasil dan analisis yang lebih baik lagi.

5. Daftar Pustaka

- [1] N. Makarim, "Rencana Strategis Kementerian Pendidikan dan Kebudayaan 2020-2024," Kementerian Pendidikan dan Kebudayaan, Jakarta, 2020.

- [2] Y. T. Samuel, B. Jonathan and J. Naibaho, "Predicting Timely Students Graduation Using the Decision Tree J48 Method at Universitas Advent Indonesia," *TeIKa*, vol. 9, no. 1, pp. 43-52, 2019.
- [3] A. H. Nasrullah, "Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out," *ILKOM Jurnal Ilmiah*, vol. 10, no. 2, pp. 244-250, 2018.
- [4] Muhammad, A. P. Windarto and Suhada, "PENERAPAN ALGORITMA C4.5 PADA KLASIFIKASI POTENSI SISWA DROP OUT," *Jurnal KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, 2019.
- [5] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees and A. Zanasi, *Discovering Data Mining: From Concept to Implementation*, Prentice-Hall, Inc., 1998.
- [6] A. K. Pujari, *Data Mining Techniques*, Orient Blackswan, 2013.
- [7] M. Ridwan, H. Suyono and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Jurnal EECCIS*, vol. 7, no. 1, pp. 59-64, 2013.
- [8] D. Firdaus, "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer," *Jurnal Format*, vol. 6, no. 2, pp. 91-97, 2017.
- [9] I. Melissa, "Building Data Mining Decision Tree Model for Predicting Employee Performance," *eJAICT: Journal of Applied Information, Communication, and Technology*, vol. 6, no. 2, pp. 75-86, 2019.
- [10] V. Kotu and B. Deshpande, *Predictive Analytics and Data Mining: Concepts and Practice with Rapidminer*, MA, USA: Elsevier, Inc., 2015.