

Komparasi Metode *Decision Tree* dan *Deep Learning* dalam Meramalkan Jumlah Mahasiswa *Drop Out* Berdasarkan Nilai Akademik

Dina Wahyuni Puteri*, Putu Wira Buana, I Made Sukarsa

Teknologi Informasi, Universitas Udayana

Abstrak: Akreditasi memiliki peran penting dalam proses pertimbangan setiap calon mahasiswa untuk melanjutkan pendidikan ke jenjang yang lebih tinggi. Perguruan tinggi dan program studi tentu diharapkan dapat memiliki akreditasi yang baik, guna menjamin mutu dan kualitasnya. *Drop out* menjadi salah satu permasalahan dalam menjaga dan meningkatkan mutu suatu perguruan tinggi khususnya pada program studi. Angka kasus *drop out* yang tinggi akan memicu turunnya kualitas Pendidikan di Indonesia dan berdampak pada akreditasi suatu program studi di perguruan tinggi. Upaya yang dapat dilakukan untuk mengatasi hal tersebut yaitu, dengan melakukan prediksi terkait mahasiswa yang terancam *drop out* secara dini. Penelitian ini menggunakan dua jenis metode yang berbeda yang dikomparasikan untuk menemukan hasil pemodelan terbaik. Metode yang digunakan yakni *decision tree* dengan algoritma C5.0 serta *deep learning* dengan algoritma GRU. Data peramalan yang digunakan yaitu Data Mahasiswa Program Studi Teknologi Informasi angkatan tahun 2010-2016. Metode dengan performa terbaik pada penelitian ini yaitu metode *decision tree* C5.0 yang menghasilkan nilai akurasi sebesar 95% dengan persentase kesalahan RMSE 0.13001950438859716 dan MAPE 2.26% Metode *deep learning* menunjukkan hasil yang cenderung lebih rendah dibanding *decision tree* C5.0 dengan nilai akurasi sebesar 92% dan persentase kesalahan RMSE 0.1873780487675864 MAPE 4.69%.

Kata Kunci: Akreditasi, *Decision Tree*, *Deep Learning*, *Drop Out*, GRU (*Gated Recurrent Unit*)

DOI:

<https://doi.org/10.47134/pjise.v1i2.2327>

*Correspondence: Dina Wahyuni Puteri

Email: dinawahyuni48@gmail.com

Received: 18-02-2024

Accepted: 31-03-2024

Published: 30-04-2024



Copyright: © 2024 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution-ShareAlike (CC BY SA) license (<http://creativecommons.org/licenses/by-sa/4.0/>).

Abstract: Accreditation plays a crucial role in the decision-making process for prospective students pursuing higher education. It is essential for universities and study programs to possess strong accreditation, ensuring high-quality education. Dropout rates pose challenges to maintaining and enhancing university quality, particularly in study programs. A high number of dropout cases can lead to a deterioration in Indonesia's education quality, impacting the accreditation of study programs in higher education. To address this issue, efforts can be made to predict and support students at risk of dropping out early. This research employs two methods, comparing decision tree using the C5.0 algorithm and deep learning with the GRU algorithm. Analyzing Student Data for the Information Technology Study Program from 2010-2016, the C5.0 decision tree method outperforms with 95% accuracy, 0.13001950438859716 RMSE, and 2.26% MAPE. In contrast, the deep learning method demonstrates lower performance, achieving an 92% accuracy, 0.1873780487675864 RMSE, and 4.69% MAPE.

Keywords: Accreditation, *Decision Tree*, *Deep Learning*, *Drop Out*, GRU (*Gated Recurrent Unit*)

Pendahuluan

Pendidikan memiliki peran yang sangat penting dalam memajukan serta membangun bangsa dan negara sehingga nantinya mampu menciptakan sumber daya manusia yang berkualitas. Pentingnya pendidikan di Indonesia telah termaktub dalam Pembukaan Undang-Undang Dasar 1945 yang mengamanatkan negara untuk dapat mencerdaskan kehidupan bangsa. Keberhasilan dari pendidikan suatu negara tercermin pada banyaknya kuantitas dari individu yang melanjutkan pendidikan ke jenjang yang lebih tinggi (Kemenristekdikti, 2017). Perguruan tinggi memiliki peranan dalam mempersiapkan generasi penerus bangsa yang menjadi *agents of changes* atau agen perubahan dalam mengoptimalkan pembangunan dan perubahan bangsa. Perguruan tinggi dinilai berhasil mempersiapkan generasi penerus bangsa melihat dari jumlah lulusan mahasiswa perguruan tinggi yang langsung mendapatkan pekerjaan dan mulai bersaing dalam dunia pembangunan bangsa.

Permasalahan yang sering terjadi dalam mengoptimalkan pendidikan jenjang tinggi yakni permasalahan terkait dengan *drop out*. *Drop out* merupakan permasalahan yang memerlukan perhatian khusus, dikarenakan tingkat *drop out* mahasiswa yang tinggi akan dapat berpengaruh pada kualitas dari sebuah perguruan tinggi dan juga pendidikan di Indonesia. *Drop out* menjadi permasalahan yang sering berkaitan pada proses studi mahasiswa di beberapa kondisi. Dilansir dari Data Nasional pada Statistik Pendidikan Tinggi Tahun 2018 jumlah mahasiswa putus kuliah di Bali mencapai angka 2.539 dengan mahasiswa terdaftar sebanyak 143.046 dan persentase 2% (Kemenristekdikti, 2018). Selanjutnya di tahun 2019 Data Nasional Statistik Pendidikan Tinggi menjabarkan jika angka putus kuliah di Bali mencapai 6.609 dari jumlah mahasiswa yang terdaftar sebanyak 142.573 dan persentase 5% (Kemenristekdikti, 2019). Data tersebut menunjukkan jika terjadi peningkatan yang cukup signifikan dari angka putus kuliah mahasiswa di Bali. Hal tersebut tentu menjadi permasalahan yang harus diperhatikan lebih jauh lagi, karena akan memberikan dampak bagi mahasiswa maupun juga pihak perguruan tinggi. Kasus *drop out* pada perguruan tinggi tentu berpengaruh pada akreditasi serta kualitas perguruan tinggi (Sari et al., 2021). Akreditasi merupakan hal yang penting untuk menjaga dan meningkatkan reputasi dan kualitas yang dimiliki perguruan tinggi sehingga dapat bersaing di dunia pendidikan dalam melahirkan mahasiswa berprestasi penerus bangsa. Setiap perguruan tinggi tentu memiliki banyak fakultas yang terdiri dari beberapa program studi.

Penelitian sebelumnya yang berkaitan dengan kasus *drop out* mahasiswa dilakukan oleh Anita et al. (2020). Penelitian ini menggunakan algoritma C5.0 dengan hasil akhir penelitian menunjukkan bahwa pada data mahasiswa Angkatan 2015 dengan jumlah 125 data (88 data latih dan 37 data uji), menghasilkan persentase nilai *recall* sebesar 72,22%, nilai *precision* sebesar 92,85%, nilai *accuracy* sebesar 83,78%, dan nilai *error* sebesar 16,21% (Anita et al., 2020).

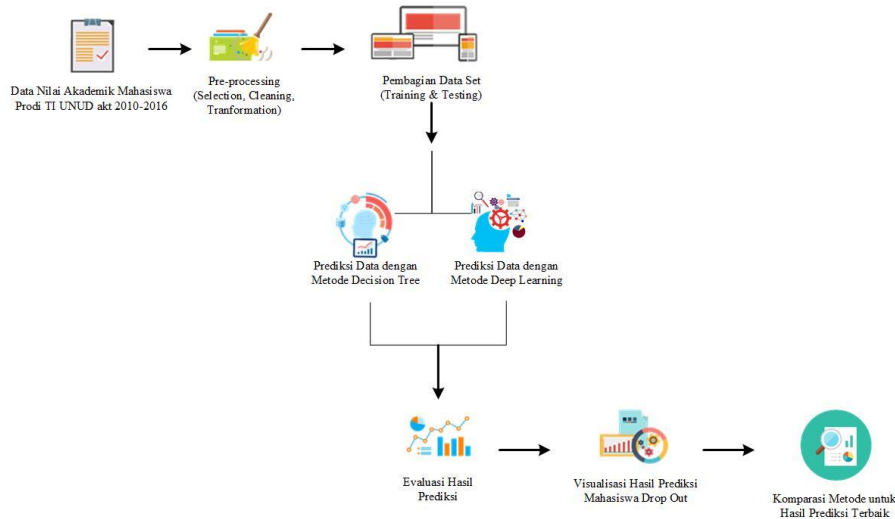
Penelitian lainnya yang menggunakan algoritma C5.0 dalam melakukan prediksi terkait dengan kelulusan mahasiswa juga telah diteliti oleh Fevi Ariska et al. (2022) dengan judul "*Student Graduation Predictions using Comparison of C5.0 Algorithm with Linear Regression*". Data yang digunakan berasal dari Universitas Prima Indonesia khususnya mahasiswa Fakultas Teknologi dan Ilmu Komputer angkatan 2017-2020 dengan jumlah 715

data. Prediksi terbaik pada penelitian tersebut yakni menggunakan algoritma C5.0, dikarenakan hasil prediksinya mencapai persentase 96,85% untuk *training* dan 93,72% untuk *testing*, sedangkan *linear regression* hanya menghasilkan hasil prediksi sebesar 33,31% untuk *training* dan 40,30 untuk *testing* (Ariska et al., 2022).

Upaya yang dapat dilakukan untuk mencegah dan meminimalisir terjadinya kasus serupa yakni dengan melakukan identifikasi sejak dini untuk menekan angka *drop out* melalui proyeksi ke depan berupa peramalan potensi mahasiswa yang terancam *drop out* dengan menggunakan indikator dari nilai akademik mahasiswa. Penelitian ini menggunakan metode *decision tree* C5.0 dan metode *deep learning Gated Recurrent Unit* (GRU). Kedua algoritma tersebut dikomparasikan untuk menguji tingkat akurasi prediksi yang lebih baik, dengan demikian nantinya pada proses interpretasi hasil peramalan didapatkan akurasi dari hasil peramalan mahasiswa *drop out* yang lebih baik.

Metode

Metode penelitian yang digunakan pada penelitian ini dapat digambarkan pada gambaran umum. Gambaran umum penelitian mengacu pada proses yang ada dalam data *mining*, seperti data *collection* dimana dilakukan pengumpulan data, *pre-processing* berfungsi untuk menjalankan proses *cleaning* dan *transforming* data, pembagian *dataset* dari hasil *pre-processing* dan yang terakhir analisis data dengan menggunakan metode yang dipilih (Mandalapu et al., 2023). Gambaran umum yang telah dirancang dapat dilihat pada Gambar 1 berikut.



Gambar 1. Gambaran Umum Penelitian

Gambar 1 merupakan gambaran umum pada penelitian yang dilakukan. Gambaran umum meliputi beberapa tahapan, tahap pertama yakni pengumpulan Data Nilai Akademik Mahasiswa Program Studi Teknologi Informasi Fakultas Teknik Universitas Udayana angkatan 2010-2016 yang telah dikumpulkan dalam bentuk format *form spreadsheet* pada Microsoft Excel. Data tersebut bersumber dari *history* nilai akademik yang diperoleh di USDI (Unit Sumber Daya Informasi). Tahap selanjutnya yaitu *pre-processing* data yang meliputi proses data *selection*, *cleaning*, dan *transformation*. Setelah *pre-processing* maka data akan dibagi menjadi data *training* dan data *testing* untuk kemudian dianalisis menggunakan model *decision tree* C5.0 dan model *deep learning Gated Recurrent Unit* (GRU). Berikutnya,

hasil prediksi akan dievaluasi untuk melihat performa dari masing-masing algoritma yang digunakan. Setelah dievaluasi, hasil akan divisualisasikan untuk memudahkan membaca hasil prediksi, dan tahap terakhir yaitu mengkomparasikan kedua algoritma yang digunakan untuk menilai algoritma mana yang cocok dalam melakukan prediksi mahasiswa *drop out* berdasarkan dengan data serta atribut yang digunakan.

Hasil dan Pembahasan

A. Pre-processing

Pre-processing merupakan tahapan yang harus dilakukan sebelum masuk ke dalam proses utama analisis data, yang meliputi serangkaian proses di dalamnya (Nyoman et al., 2022). Data yang diperoleh akan dikategorikan sebagai data kotor yang kemudian dimodifikasi, dihapus, dan disesuaikan dengan kebutuhan analisis untuk menghindari data yang bersifat ambigu (Budilaksana et al., 2021). Tahapan dalam *pre-processing* meliputi data *selection*, data *cleaning* dan data *transformation*.

1. Data Selection

Tahap pertama yang dilakukan yaitu data *selection*, dimana proses ini mencakup pemilihan atribut dan mengategorikan atribut jenis kelamin, IPK, SKS total, UKT, dan keaktifan terakhir. Data *selection* merupakan tahapan yang dilakukan dalam data *mining* yang digunakan untuk melakukan seleksi terhadap sekumpulan data operasional yang akan digunakan untuk proses data *mining* (Mardi, 2017). Hasil dari pemilihan atribut dapat dilihat pada Gambar 2 berikut.

	tahun	jenis_kelamin	ipk	sks_total	ukt	semester	keaktifan_terakhir	status
0	2016	Laki-laki	2.76	123.0	7000000.0	12	Keluar (Lulus)	1
1	2016	Laki-laki	NaN	NaN	5000000.0	2	Non-aktif (Belum bayar)	0
2	2016	Laki-laki	2.93	131.0	5000000.0	13	Non-aktif (Belum bayar)	0
3	2016	Laki-laki	2.91	100.0	8000000.0	13	Cuti	0
4	2016	Perempuan	2.55	38.0	5000000.0	2	Non-aktif (Belum bayar)	0

Gambar 2. Hasil Pemilihan Atribut

Gambar 2 merupakan hasil dari pemilihan atribut, atribut yang digunakan untuk analisis meliputi tahun, jenis kelamin, IPK, SKS total, UKT, semester dan keaktifan terakhir. Tahapan selanjutnya yaitu melakukan proses kategori atribut, hal tersebut dilakukan karena terlalu banyak variasi data yang ada. Sehingga untuk memudahkan proses prediksi dalam menemukan pola serupa maka perlu dilakukan proses pengelompokan atribut data. Hasil dari pengelompokan atribut dapat dilihat pada Tabel berikut.

Tabel 1. Hasil Pengelompokan Atribut IPK

IPK	Kategori
<2,50	Tidak Memuaskan
2,50 - 2,75	Cukup
2,76 - 3,00	Memuaskan
3,01 - 3,50	Sangat Memuaskan
>3,50	Dengan Pujian

Tabel 2. Hasil Pengelompokan Atribut SKS Total

SKS Total	Kategori
<144	Tidak Memenuhi
144	Memenuhi

Tabel 3. Hasil Pengelompokan Atribut UKT

UKT	Kategori
<1.000.000	Kurang dari Rp1.000.000
1.000.000-5.000.000	Rp1.000.000 – Rp5.000.000
>5.000.000	Lebih dari Rp5.000.000

2. Data Cleaning

Data cleaning merupakan proses menghapus data *missing value*. *Data missing value* adalah data yang bernilai kosong, dimana data tersebut tidak dapat digunakan dalam penelitian. Contoh dari data *missing value* dapat dilihat pada Gambar 3 berikut.

	tahun	jenis_kelamin	ipk	sks_total	ukt	semester	keaktifan_terakhir	status
0	2016	Laki-laki	2.76	123.0	7000000.0	12	Keluar (Lulus)	1
1	2016	Laki-laki	NaN	NaN	5000000.0	2	Non-aktif (Belum bayar)	0
2	2016	Laki-laki	2.93	131.0	5000000.0	13	Non-aktif (Belum bayar)	0
3	2016	Laki-laki	2.91	100.0	8000000.0	13	Cuti	0
4	2016	Perempuan	2.55	38.0	5000000.0	2	Non-aktif (Belum bayar)	0

Gambar 3. Contoh Data *Missing Value*

Gambar 3 merupakan contoh dari data *missing value*. *Data missing value* yang ditampilkan dari Gambar 3 yaitu terletak pada atribut IPK dan SKS Total tepatnya pada baris kedua dimana data tersebut bernilai NaN. Jumlah dari *missing value* pada data penelitian yang digunakan dapat dilihat pada Tabel 4 berikut.

Tabel 4. Jumlah Data *Missing Value*

Atribut	Jumlah Data <i>Missing Value</i>
jenis_kelamin	11
ipk	26
sks total	1
ukt	10
semester	0
keaktifan terakhir	0
status	0

3. Data Transformation

Data *transformation* merupakan proses mengubah data ke dalam bentuk yang sama. Data penelitian yang digunakan terdiri dari data *categorical* dan data *numerical* dengan demikian diperlukannya proses transformasi untuk menjadikan data dalam bentuk yang sama dengan skala tertentu. Tujuan dari proses transformasi yaitu untuk memudahkan dalam mempelajari dan memahami pola data, sehingga bisa membantu dalam pembangunan model prediksi yang optimal (Mandalapu et al., 2023). Proses data *transformation* dilakukan dengan menggunakan *library* LabelEncoder. Hasil dari data *transformation* dapat dilihat pada Gambar 4 berikut.

tahun	jenis_kelamin	ipk	sks_total	ukt	semester	keaktifan_terakhir	status	jenis_kelamin_n	ipk_n	sks_total_n	ukt_n	semester_n	keaktifan_terakhir_n	status_n	
0	2016	Laki-laki	2.76	123.0	7000000.0	12	Keluar (Lulus)	1	0	2.0	1	2.0	1.0	4	1
2	2016	Laki-laki	2.93	131.0	5000000.0	13	Non-aktif (Belum bayar)	0	0	2.0	1	1.0	2.0	7	0
3	2016	Laki-laki	2.91	100.0	8000000.0	13	Cuti	0	0	2.0	1	2.0	2.0	1	0
4	2016	Perempuan	2.55	38.0	5000000.0	2	Non-aktif (Belum bayar)	0	1	1.0	1	1.0	0.0	7	0
5	2016	Laki-laki	3.34	127.0	1000000.0	13	Cuti	0	0	3.0	1	1.0	2.0	1	0
6	2016	Laki-laki	3.45	138.0	7000000.0	13	Non-aktif (Belum bayar)	0	0	3.0	1	2.0	2.0	7	0
7	2016	Laki-laki	3.55	144.0	5000000.0	9	Keluar (Lulus)	1	0	4.0	0	1.0	1.0	4	1
8	2016	Laki-laki	3.24	135.0	5000000.0	13	Non-aktif (Belum bayar)	0	0	3.0	1	1.0	2.0	7	0

Gambar 4. Hasil Transformasi Data

Gambar 4 merupakan proses transformasi data pada setiap atribut. Gambar tersebut menunjukkan jika terdapat 6 kolom baru yang merupakan hasil transformasi data, dimana artinya dari hasil pelabelan data kemudian dibuatkan kolom baru untuk menyimpan hasil transformasi data. Penjabaran transformasi dari masing-masing atribut dapat dilihat pada Tabel berikut.

Tabel 5. Pelabelan pada Atribut Jenis Kelamin

jenis_kelamin	jenis_kelamin_n
Laki-laki	0
Perempuan	1

Tabel 6. Pelabelan pada Atribut IPK

ipk	ipk_n
<2.50	0
2.50-2.75	1
2.76-3.00	2
3.01-3.50	3
>3.50	4

Tabel 7. Pelabelan pada Atribut SKS Total

sks_total	sks_total_n
>144	0
<144	1

Tabel 8. Pelabelan pada Atribut UKT

ukt	ukt_n
<1.000.000	0
1.000.000-4.999.999	1
>5.000.000	2

Tabel 9. Pelabelan pada Atribut Keaktifan Terakhir

keaktifan_terakhir	keaktifan_terakhir_n
Aktif(belajar)	0
Cuti	1
Keluar(Dikeluarkan)	2
Keluar(Lainnya)	3
Keluar(Lulus)	4
Keluar(Mengundurkan diri)	5
Keluar(Pindah)	6
Non-aktif(belum bayar)	7

B. Pembagian Data *Train* dan Data *Test*

Data mahasiswa yang sebelumnya sudah ditransformasikan, kemudian dibagi menjadi 2 bagian, yaitu data *train* dan data *test*. Pembagian data ini dilakukan berdasarkan rentang tahun yakni mahasiswa angkatan tahun 2010-2014 dijadikan sebagai data *train* atau latih dan mahasiswa angkatan tahun 2015-2016 dijadikan sebagai data *test* atau data uji. Total jumlah untuk data *train* yakni 483 dan untuk data *test*-nya sebanyak 286 data. Rincian dari pembagian data dapat dilihat pada Tabel 10 berikut.

Tabel 10. Pembagian Data *Train* dan Data *Test*

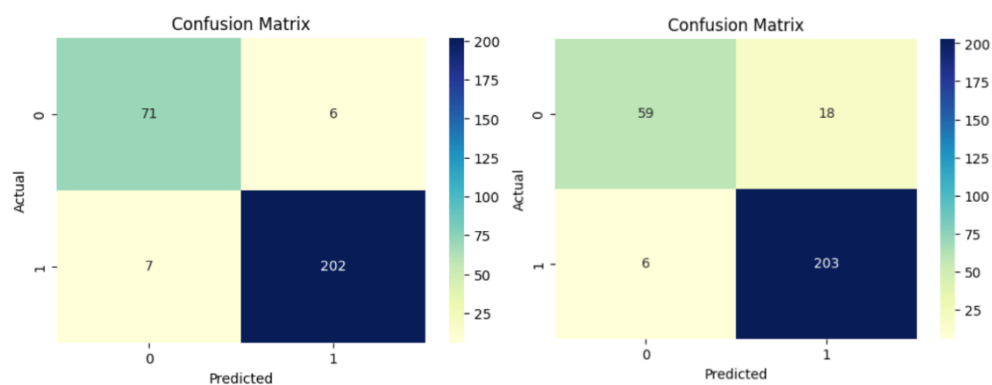
Data <i>Train</i>	Data <i>Test</i>
Data Mahasiswa Tahun 2010-2014 (483 data)	Data Mahasiswa Tahun 2015-2016 (286 data)

C. Peramalan

Peramalan merupakan suatu proses atau kegiatan yang dilakukan untuk memperkirakan hasil di masa mendatang sesuai dengan kebutuhan, yang didasari pada data yang ada sebelumnya dengan kurun waktu tertentu yang akan dianalisis menggunakan metode tertentu (Waliyansyah & Saputro, 2020) Peramalan juga diartikan sebagai suatu proses perkiraan pengukuran berdasarkan data yang relevan dengan masa lalu yang kemudian dianalisis menggunakan metode statistika di mana nantinya dapat memperbaiki peristiwa yang akan terjadi pada masa yang akan datang (Bengnga & Ishak, 2018).

Proses prediksi pada penelitian menggunakan metode *decision tree*. *Decision tree* merupakan suatu teknik dalam data *mining* yang dapat digunakan untuk melakukan proses klasifikasi, *clustering*, dan prediksi (Putra, 2017). Algoritma *decision tree* yang digunakan yaitu algoritma C.50. Algoritma C5.0 memiliki beberapa tahapan dimana nantinya semua data yang akan digunakan pada proses penelitian akan dijadikan akar dari pohon keputusan, selanjutnya atribut yang dipilih sebagai indikator dalam proses penelitian akan dijadikan sebagai pembagi pada sampel tersebut (Kastawan et al., 2018). Metode kedua yang digunakan yaitu *deep learning*. *Deep learning* merupakan suatu konsep *machine learning* yang didasarkan pada jaringan saraf tiruan atau *artificial neural networks*. *Deep learning* digunakan untuk menangani data yang berskala besar, *noisy*, dan bersifat tidak struktur (Syafa & Lestandy, 2021). *Deep learning* menawarkan teknik pembelajaran mendalam yang bisa digunakan untuk meningkatkan nilai akurasi prediksi yang optimal dengan kecepatan waktu komputasi (Menon, 2020). Algoritma *deep learning* yang digunakan yaitu *Gated Recurrent Unit* atau (GRU). Arsitektur GRU memuat dua *gate*, yang terdiri dari *reset gate* dan juga *update gate* (Zaman et al., 2019).

Hasil dari peramalan menggunakan metode *decision tree* dan *deep learning* dengan menggunakan *confusion matrix* dapat dilihat pada Gambar 5 berikut.

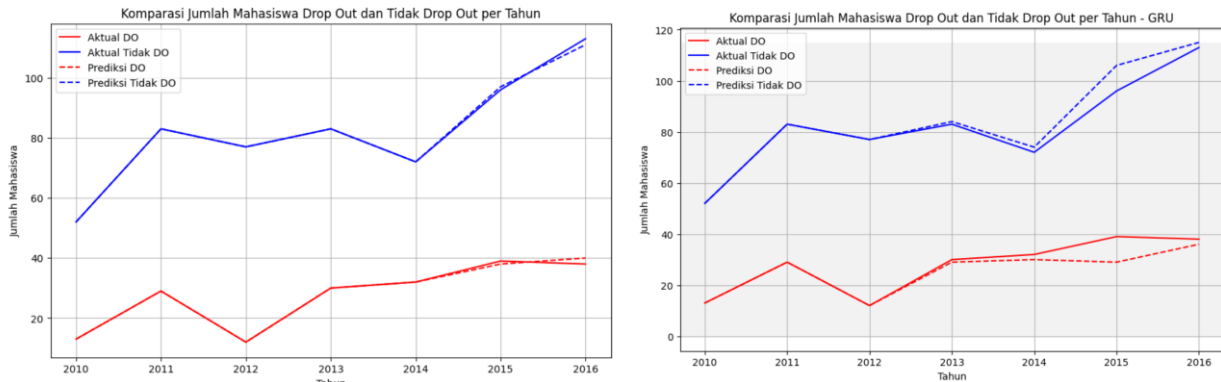


Gambar 5. Hasil *Confusion Matrix* pada *Decision Tree* C5.0 dan *Deep Learning* GRU

Gambar 5 menunjukkan hasil prediksi dari metode *decision tree* dan *deep learning* yang diuraikan menggunakan *confusion matrix*. Gambar pertama merupakan *confusion matrix* milik *decision tree* yang menghasilkan jumlah data *True Positive* (TP) sebesar 71, dimana artinya memang benar terklasifikasi *drop out*. Jumlah *False Positive* (FP) yaitu 6, dimana FP ini menjelaskan bahwa data yang terklasifikasi *drop out* ternyata tidak *drop out*. *False Negative* (FN) menunjukkan angka 7, dimana artinya data yang terklasifikasi tidak *drop out* sebenarnya adalah *drop out*. Nilai terakhir yakni *True Negative* (TN) yang memiliki jumlah data 202 yang menjelaskan jika data yang terklasifikasi tidak *drop out* memang benar tidak *drop out*. Gambar kedua merupakan *confusion matrix* milik *deep learning* yang menghasilkan jumlah data *True Positive* (TP) sebesar 59, dimana artinya memang benar terklasifikasi *drop out*. Jumlah *False Positive* (FP) yaitu 18, dimana FP ini menjelaskan bahwa data yang terklasifikasi *drop out* ternyata tidak *drop out*. *False Negative* (FN) menunjukkan angka 6, dimana artinya data yang terklasifikasi tidak *drop out* sebenarnya adalah *drop out*. Nilai terakhir yakni *True Negative* (TN) yang memiliki jumlah data 203 menjelaskan jika data yang terklasifikasi tidak *drop out* memang benar tidak *drop out*.

D. Visualisasi Data

Tujuan dari proses visualisasi yaitu agar mudah memahami hasil prediksi, sehingga informasi yang dihasilkan akan dapat digunakan sesuai dengan kebutuhan dari penelitian yang dilakukan. Visualisasi dari prediksi menggunakan *decision tree* dan *deep learning* dapat dilihat pada Gambar 6 berikut.



Gambar 6. Visualisasi Hasil Peramalan *Decision Tree* C5.0 dan *Deep Learning* GRU

Gambar 6 merupakan hasil visualisasi dari peramalan menggunakan metode *decision tree* dan *deep learning*. Gambar tersebut menjabarkan komparasi antara nilai aktual dan juga nilai prediksi di setiap tahunnya. Nilai aktual direpresentasikan dengan garis lurus, dan nilai prediksi direpresentasikan dengan garis putus-putus. Berdasarkan dengan visualisasi tersebut, terlihat jika gambar pertama yang merupakan peramalan dengan metode *decision tree* memiliki tingkat akurasi yang lebih baik, mengingat nilai aktual dan nilai prediksi tidak terlampau jauh berbeda. Pada gambar kedua milik *deep learning* terlihat jika nilai aktual dan nilai prediksi memiliki perbedaan yang cukup terlihat pada garis lurus dan garis putus-putus yang ditampilkan, khususnya pada tahun 2014-2016.

E. Komparasi Metode *Decision Tree* dan *Deep Learning*

Metode *decision tree* dan *deep learning* telah diramalkan kemudian diuji kinerja modelnya. Setiap metode memiliki nilai evaluasi yang berbeda. Model terbaik dinilai berdasarkan dari hasil evaluasi klasifikasi yang dapat dilihat pada Tabel 11 berikut.

Tabel 11. Komparasi Hasil Pengujian Performa Algoritma

Metode	Akurasi	Precision	Recall	F1-Score	RMSE	MAPE
<i>Decision Tree</i> C5.0	95%	94%	94%	94%	0.13001950438859716	2.26%
<i>Deep Learning</i> GRU	92%	91%	87%	89%	0.1873780487675864	4.69%

Tabel 11 menunjukkan perbandingan dari setiap hasil evaluasi masing-masing metode, baik itu untuk *decision tree* C5.0 maupun *deep learning* GRU. Hasil pengukuran kinerja dari kedua algoritma menunjukkan bahwa performa dari algoritma *decision tree* C5.0 memiliki nilai yang lebih besar dibandingkan *deep learning* GRU. *Decision tree* C5.0 memiliki

akurasi sebesar 95%, *precision* 94%, *recall* 94% hingga *f1-score* 94%. *Deep learning* GRU memiliki akurasi 92%, *precision* 91%, *recall* 87%, dan *f1-score* 89%. Nilai RMSE dan MAPE yang dimiliki oleh *decision tree* C5.0 juga lebih rendah dibandingkan *deep learning* GRU, yang artinya semakin rendah nilai dari RMSE dan MAPE maka semakin baik pula kinerja model dalam melakukan prediksi.

Kesimpulan

Kesimpulan yang dapat diambil berdasarkan dengan rumusan masalah serta hasil yang diperoleh dari penelitian yakni pemodelan menggunakan *decision tree* C5.0 pada kasus *drop out* mahasiswa TI angkatan 2010-2016 menghasilkan nilai akurasi sebesar 95%. Evaluasi dari kinerja *decision tree* C5.0 diuraikan dalam nilai *precision* 94%, *recall* 94% hingga *f1-score* 94%. Akurasi dari pemodelan menggunakan *deep learning* GRU yakni sebesar akurasi 92%, *precision* 91%, *recall* 87%, dan *f1-score* 89%. Data penelitian memiliki kecocokan pada pendekatan *decision tree*, *deep learning* juga berhasil melakukan proses prediksi namun nilai yang dihasilkan lebih rendah. *Deep learning* cenderung cocok dengan tipe data yang kompleks dan bervolume besar. Performa yang dihasilkan pada pemodelan *decision tree* dan *deep learning* memperlihatkan hasil jika *decision tree* memiliki kinerja yang baik dalam melakukan prediksi *drop out* mahasiswa, dibuktikan dengan hasil akurasi dan evaluasi. Hasil komparasi tentu saja menunjukkan apabila *decision tree* C5.0 merupakan metode yang lebih baik dalam melakukan prediksi mahasiswa *drop out*.

Penelitian yang dilakukan memiliki kelemahan pada data yang digunakan yang kurang bervariasi, sehingga proses penentuan faktor-faktor belum bisa diidentifikasi dengan maksimal. Diharapkan kedepannya, data yang digunakan memiliki atribut yang lebih beragam, untuk membantu dalam proses identifikasi pola yang kompleks, yang tentunya akan berpengaruh pada kualitas hasil prediksi. Atribut yang digunakan juga harus sudah diteliti pengaruhnya terhadap analisis prediksi yang dilakukan. Penggunaan metode lain sangat dianjurkan untuk bisa melihat hasil perbandingan dengan lebih baik lagi, agar nantinya bisa didapat kinerja metode lain yang memiliki hasil yang lebih optimal.

Daftar Pustaka

- Anita, N., Intan, P. Y., & Yulia, P. E. (2020). Prediksi Ketepatan Waktu Kelulusan. *Jurnal Informatika Dan Sistem Informasi (JIFoSI)*, 1(3), 817–825.
- Ariska, F., Sihombing, V., & Irmayani, I. (2022). Student Graduation Predictions Using Comparison of C5.0 Algorithm With Linear Regression. *Sinkron*, 7(1), 256–266. <https://doi.org/10.33395/sinkron.v7i1.11261>
- Bengnga, A., & Ishak, R. (2018). Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo. *ILKOM Jurnal Ilmiah*, 10(2), 136–143. <https://doi.org/10.33096/ilkom.v10i2.274.136-143>
- Bhat, I., Umadevi, V., Jagadeesh, N., & ... (2023). Tender Coconut Classification using Decision Tree and Deep Learning Technique. ... *Processing and* <https://ieeexplore.ieee.org/abstract/document/10117353/>

- Budilaksana, D., Sukarsa, M., Agung, A., Agung, K., & Wiranatha, C. (2021). Implementing kNearest Neighbor Methods to Predict Car Prices. *JURNAL ILMIAH MERPATI*, 9(1), 58–71.
- Chen, Y., Chen, W., Pal, S. C., Saha, A., & ... (2022). Evaluation efficiency of hybrid deep learning algorithms with neural network decision tree and boosting methods for predicting groundwater potential. *Geocarto* <https://doi.org/10.1080/10106049.2021.1920635>
- Eom, J. Y., Jang, S. U., & Jeon, W. S. (2023). Wi-Sniffer: Wifi-based intruder detection system using deep learning and decision tree. *2023 IEEE 97th Vehicular* <https://ieeexplore.ieee.org/abstract/document/10199724/>
- Grégoire, G., Fortin, J., Ebtehaj, I., & Bonakdari, H. (2023). Forecasting Pesticide Use on Golf Courses by Integration of Deep Learning and Decision Tree Techniques. *Agriculture*. <https://www.mdpi.com/2077-0472/13/6/1163>
- Kastawan, P. W., Wiharta, D. M., & Sudarma, M. (2018). Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil. *Majalah Ilmiah Teknologi Elektro*, 17(3), 371. <https://doi.org/10.24843/mite.2018.v17i03.p11>
- Kataria, S., & Nafis, M. T. (2019). Internet banking fraud detection using deep learning based on decision tree and multilayer perceptron. *2019 6th International Conference on* <https://ieeexplore.ieee.org/abstract/document/8991389/>
- Kemenristekdikti. (2017). *Laporan Kinerja 2017*.
- Kemenristekdikti, P. (2018). *Statistik Pendidikan Tinggi 2018*.
- Kemenristekdikti, P. (2019). *Statistik Pendidikan Tinggi Tahun 2019*.
- Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. *IEEE Access*, 11, 60153–60170. <https://doi.org/10.1109/ACCESS.2023.3286344>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Menon, S. P. (2020). Deep learning for prediction of amyotrophic lateral sclerosis using stacked auto encoders. In *Int. J. Big Data Management* (Vol. 1, Issue 2).
- Nyoman, I., Bayu, K., Made, I., Dwi Suarjaya, A., & Wira Buana, P. (2022). Classification of Indonesian Population's Level Happiness on Twitter Data Using N-Gram, Naïve Bayes, and Big Data Technology. *International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)*, 12(5).
- Putra, A. (2017). Solusi Prediksi Mahasiswa Drop Out Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma. *Jurnal SIMETRIS*, 84(1), 177–184. <https://doi.org/10.1103/PhysRevE.94.033206>
- Romeo, L., Devanna, R. P., Marani, R., & ... (2023). Scale-invariant semantic segmentation of natural RGB-D images combining decision tree and deep learning models. ... *Sensing and* <https://doi.org/10.1117/12.2677371.short>

- Sari, E. Y., Kusriani, K., & Sunyoto, A. (2021). Analisis Akurasi Jaringan Syaraf Tiruan Dengan Backpropagation Untuk Prediksi Mahasiswa Dropout. *Creative Information Technology Journal*, 6(2), 85. <https://doi.org/10.24076/citec.2019v6i2.235>
- Syafa, L., & Lestandy, M. (2021). Penerapan Deep Learning untuk Prediksi Kasus Aktif Covid-19. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(1), 453–457.
- Tandon, P., Ghanavati, S., Cheetirala, S. N., & ... (2023). A Hybrid Decision Tree and Deep Learning Approach Combining Medical Imaging and Electronic Medical Records to Predict Intubation Among Hospitalized *JMIR Formative* <https://formative.jmir.org/2023/1/e46905/>
- Waliyansyah, R. R., & Saputro, N. D. (2020). Forecasting New Student Candidates Using the Random Forest Method. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 11(1), 44. <https://doi.org/10.24843/lkjiti.2020.v11.i01.p05>
- Yan, J., Li, C., & Liu, Y. (2021). Insecurity early warning for large scale hybrid AC/DC grids based on decision tree and semi-supervised deep learning. *IEEE Transactions on Power Systems*. <https://ieeexplore.ieee.org/abstract/document/9399284/>
- Zaman, L., Sumpeno, S., & Hariadi, M. (2019). Analisis Kinerja LSTM dan GRU sebagai Model Generatif untuk Tari Remo. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 8(2), 142. <https://doi.org/10.22146/jnteti.v8i2.503>
- Zhao, L., Han, W., Niu, P., Lu, Y., Zhang, F., & ... (2023). Using nomogram, decision tree, and deep learning models to predict lymph node metastasis in patients with early gastric cancer: a multi-cohort study. *American Journal of* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9906085/>
- Zhao, L., Han, W., Niu, P., Lu, Y., Zhang, F., Jiao, F., Zhou, X., & ... (2022). Nomogram, Decision Tree and Deep Learning Models to Predict Lymph Node Metastasis of Patients with Early Gastric Cancer: A Multi-Cohort Study. *papers.ssrn.com*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4185033