Jurnal ICT : Information Communication & Technology
Vol. 23, N0.1, Juli 2023, pp. 101-104
p-ISSN: 2302-0261, e-ISSN: 2303-3363

Klasifikasi Mahasiswa Non-Aktif Menggunakan Grid Search Optimization Pada Algoritma Decision Tree

Prima Dina Atika[1], Ahmad Fathurrozi[2*], Robertoi[3]
[123]Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya
Email: [1] prima.dina@dsn.ubharajaya.ac.id, [2]fathur@dsn.ubharajaya.ac.id, [3] robertohutapea01@gmail.com

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 15 Juli 2023
Direvisi, 21 Juli 2023
Diiterima, 25 Juli 2023

Key words:
non-active students
Decision Tree
Grid Search Optimization

ABSTRAK

Abstract- *The number of non-active students will affect the performance of the study program. Because with so many non-active students if not handled properly it will cause students to graduate late and what is more worrying is that students drop out (DO). This is not good if the study program wants superior accreditation. Therefore, research was conducted to optimize the Decision Tree classification technique using Grid Search Optimization (GSO). The purpose of the GSO is to improve the accuracy of the classification results. The stages used are the Cross-Industry Standard Process for Data Mining (CRISP-DM) with the stages of business understanding, data understanding, data preparation, modeling, and evaluation. The results of the accuracy of applying the decision tree algorithm are 98.2% accuracy by applying grid search optimization to get the best parameters with criteria = "gini" and max_depth = 64, increasing accuracy to 98.4%.*

Kata Kunci:

Mahasiswa Non-Aktif,
Decision Tree,
Grid Search Optimization.

Abstrak- Banyaknya mahasiswa non aktif akan mempengaruhi kinerja program studi. Karena dengan banyaknya mahasiswa non aktif jika tidak ditangani dengan baik akan menyebabkan mahasiswa lulus tidak waktu dan yang lebih dikhawatirkan adalah mahasisawa menjadi drop out (DO). Hal ini tidak bagus jika program studi ingin akreditasi unggul. Oleh karena itu, dilakukan penelitian untuk mengoptimalkan teknik klasifikasi Decision Tree menggunakan Grid Search Optimization (GSO). Tujuan dari GSO adalah untuk meningkatkan akurasi hasil klasifikasi. Tahapan yang digunakan adalah *Cross-Industry Standard Process for Data Mining (CRISP-DM)* dengan tahapan *business understanding*, *data understanding*, *data prepation*, *modelling*, dan *evaluation*. Hasil akurasi penerapan algoritma decision tree akurasi 98,2% dengan menerapkan optimasi pencarian grid mendapatkan parameter terbaik dengan kriteria = "gini" dan max_depth = 64, meningkatkan akurasi menjadi 98,4%.

Penulis Korespondensi:

Prima Dina Atika
Program Studi Informatika,
Fakultas Ilmu Komputer
Universitas Bhayangkara Jakarta Raya
Jl. Raya Perjuangan Bekasi Utara, Kota Bekasi, Jawa Barat 17121, Indonesia.
Email: prima.dina@dsn.ubharajaya.ac.id

Jurnal ICT : Information Communication & Technology
Vol. 23, N0.1, Juli 2023, pp. 101-104
p-ISSN: 2302-0261, e-ISSN: 2303-3363

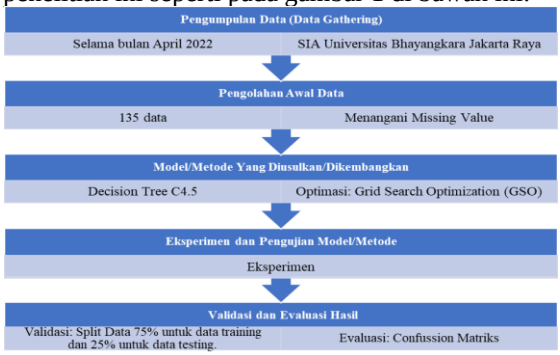
1. Pendahuluan

Pembelajaran mesin (ML) adalah teknik yang banyak digunakan untuk mengekstraksi pengetahuan dari data [1]. ML adalah area penelitian di persimpangan statistik, kecerdasan buatan, dan ilmu komputer[2]. ML juga disebut sebagai analisis prediktif atau pembelajaran statistik[3]. Konsep ML memiliki empat kategori dalam pengembangannya: pembelajaran terawasi, pembelajaran tak terawasi, pembelajaran semi terawasi dan pembelajaran penguatan. Pembelajaran terbimbing sering digunakan untuk mengklasifikasikan data ketika kumpulan data yang digunakan harus memiliki pengenal data, terutama pada variabel target. Decision tree (DT) adalah salah satu algoritma klasifikasi.

Pohon keputusan adalah algoritma yang menggabungkan perhitungan entropi dan data Gini untuk membentuk pohon keputusan. Pohon keputusan tersebut kemudian digunakan untuk mengklasifikasikan data baru yang belum memiliki pengenal (kelas target) [4][5]. Masalah klasifikasi terdiri dari kombinasi variabel dalam jumlah yang sangat besar (dimensi tinggi) yang mempengaruhi akurasi dan kualitas model klasifikasi yang dihasilkan dan dapat diselesaikan dengan menggunakan Grid Search Optimization (GSO).
Grid Search Optimization (GSO) adalah teknik optimisasi untuk algoritma validasi silang yang secara signifikan meningkatkan akurasi model klasifikasi [6][7]. Metode ini menguji kombinasi dan validasi secara individual, kemudian memilih kombinasi yang memberikan model terbaik untuk memprediksi kinerja [8]. Beberapa penelitian dengan Grid Search Optimization telah banyak digunakan, misalnya untuk meningkatkan akurasi Support Vector Regression (SVR) [9] dan menghilangkan optimasi hyperparameter dari algoritma klasifikasi [8][10].
framework/ platform/ model/ persamaan yang digunakan dijelaskan dengan baik dan detail.

2. Metode Penelitian

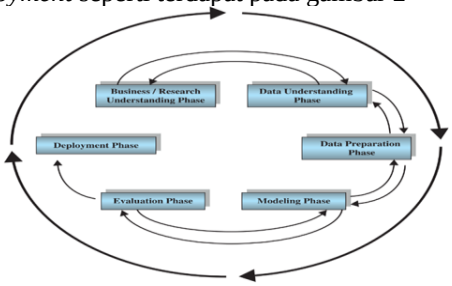
Metode penelitian yang digunakan dalam penelitian ini seperti pada gambar 1 di bawah ini:



Gambar 1 Metode Penelitian

3. Hasil dan Pembahasan

Tahapan pembahasan penelitian ini menggunakan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [11] yang teridiri dari a. *Business Understanding, b. Data Understanding, c. Data Preparation, d. Modeling, e. Evaluation, f. Deployment* seperti terdapat pada gambar 2



Gambar 2 *Cross-Industry Standard Process for Data Mining (CRISP-DM)* Sumber: [12]

3.1 Business Understanding

Permasalahan yang ada adalah banyaknya mahasiswa dengan status non aktif sehingga mempengaruhi kinerja program studi. Karena dengan banyaknya mahasiswa non aktif jika tidak ditangani dengan baik akan menyebabkan mahasiswa lulus tidak waktu dan yang lebih dikhawatirkan adalah mahasisawa menjadi *drop out (DO).* Hal ini tidak bagus jika program studi ingin akreditasi unggul, selain itu juga ingin mengetaui apa penyebab mahasiswa non-aktif. Untuk hal tersebut maka perlu diklasifikasi mahasiswa dengan status non aktif agar bisa ditangani dengan baik.

3.2. Data Understanding

Data terdiri dari kolom yaitu NPM, NIK, LokasiDaerah, SKS_Progress, IPK, Semester, Tahun_Lahir, Gender, Status, Pendidikan, Pekerjaan_Orang_Tua, Jenis_Kelas, Non-Aktif/Aktif. Sejumlah 135 record seperti pada tabel 1 di bawah ini

Table 1 Data Mahasiswa

No	Nama Kolom	Tipe Data	Deskripsi
1	NPM	Integer	NPM adalah Nomor Pokok Mahasiswa
2	NIK	Integer	NIK adalah Nomor Induk Kependudukan
3	Lokasi Daerah	String	Lokasi Daerah adalah tempat daerah

No	Nama Kolom	Tipe Data	Deskripsi
			mahasiswa tinggal
4	SKS_progress	Integer	SKS_progress adalah sks yang sudah dijalankan
5	IPK	Float	IPK adalah Indeks Prestasi Kumulatif
6	Semester	Integer	Semester adalah proses kegiatan dalam suatu jenjang dalam perkuliahan
7	Tahun_Lahir	Integer	Tahun Lahir adalah waktu yang dipakai untuk menentukan umur
8	Gender	String	Gender adalah prilaku pada hubungan social pada mahasiswa
9	Status	String	Status disini adalah sudah Menikah atau belum Menikah
10	Pendidikan	String	Pendidikan adalah tingkat Pendidikan akhir sebelum masuk perkuliahan
11	Pekerjaan Orang Tua	String	Pekerjaan orang tua adalah metrik untuk melihat pekerjaan orang tua mahasiswa
12	Jenis Kelas	String	Jenis Kelas adalah keterangan untuk melihat kelas mahasiswa
13	Non_Aktif/Aktif	String	Labeling untuk mahasiswa

3.3 *Data Preparation*

Pada tahap data preparation ini dilakukan menghapus NIK karena sudah terwakili dengan NPM. Kemudian mengganti nama kolom seperti apda tabel 2 di bawah ini:

Tabel 2 Data Mahasiswa1

No	Nama Awal Kolom	Nama Baru Kolom
1	NPM	npm
2	Lokasi Daerah	Lokasi
3	SKS yang sudah dijalankan	sks
4	IPK	ipk
5	Semester	semester
6	Tahun Lahir	tahun_lahir
7	Gender	gender
8	Status	status_kawin
9	Pendidikan	pendidikan
10	Pekerjaan Orangtua	orang_tua
11	Jenis Kelas	kelas
12	Status	label

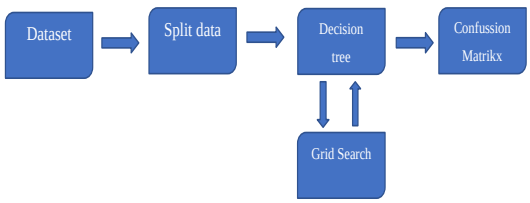
Menagani data yang kosong *(Handling Missing Value)* dengan mengisikan record null tersebut menggunakan data yang sering muncul yaitu modus. Data yang kosong dan setelah ditangani seperti pada tabel 3 di bawah ini:

Tabel 3 Data Mahasiswa setalah Handling Missing value

No	Nama Kolom	Data yang kosong	Handling Missing Value
	lokasi	7	0
	sks	0	0
	ipk	0	0
	semester	33	0
	tahun_lahir	0	0
	gender	0	0
	status_kawin	0	0
	pendidikan	3	0
	orang_tua	0	0
	kelas	0	0
	label	0	0
	lokasi	7	0
	sks	0	0

3.4 Modelling

Rancangan penelitian seperti terdapat pada gambar 4 di bawah ini:



Gambar 4 Rancangan penelitian

Dataset yang digunakan sebanyak 18.000 dengan split data 75% untuk data training dan 25%.

Modelling yang digunakan Decision Tree Classifier dan Decision Tree Classifier ditambahkan Grid Search.

Grid Search digunakan untuk mencari semua grid yang ada di data train dan data test dan untuk mencari parameter yang terbaik utnuk menentukan optimasi pada akurasi sehingga prediksi pada Decision Tree lebih sempurna.

Tahap selanjutnya dengan mencari param_grid entropy dan gini dengan max_depth [2,4,5,16,32,64] dan dipakai di variable X_res, dan y_res. Grid Search Optimization mendapatkan parameter yang terbaik di criterion 'gini' dengan max_depth = 64. Sehingga model Decision Tree dengan ditambahkan criterion 'gini' dan max_depth = 64.

3.5. Evaluation

Pemodelan dengan Decision Tree Classifier hasil akurasi 98,29%, sedangkan Decision Tree Classifier ditambahkan Grid Search hasil akurasinya adalah 98,42% Berikut tabel 4 hasil pengujian:

Tabel 4 Hasil Pengujian:

Algoritma	Akurasi	Recall	Precision
Decision Tree	98.29%	98.02%	98.58%
Decision Tree + Grid Search Optimization	98.42%	97.93%	98.93%

4. Kesimpulan

Hasil akurasi model Decision Tree Classifier adalah sebesar 98,29% sedangkan model dengan Decision Tree Classifier dan Gird Search adalah sebesar 98,42 terdapat peningkatan meski tidak significant yaitu sebesar 0.13%. Mahasiswa mempunyai sks dibawah sama dengan 88 pada semester 6 diprediksikan non-aktif. Sehingga disarankan prodi memantau dan dilakukan pembinaan untuk mahasiswa yang mempunyai sks dibawah 88 sks pada semester 6 agar tidak non aktif.

Daftar Pustaka

[1] E. Alpaydın, *Introduction to Machine Learning Second Edition*, vol. 1107. 2014. doi: 10.1007/978-1-62703-748-8_7.

[2] S. Suthaharan, *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, 36th ed. Springer US, 2015. doi: 10.1007/978-1-4899-7641-3.

[3] A. C. Muller and S. Guido, *Introduction to Machine Learning with Python*, 1 st. 1005 Gravenstein Highway North, Sebastopol, CA 95472.: O'Reilly Media, Inc., 2023. doi: 10.2174/97898151244221230101.

[4] A. Shukla, R. Tiwari, and R. Kala, *Real Life Applications of Soft Computing*. US: CRS Press, Taylor & Francis Group, LLC., 2010.

[5] Y. Zamrodah, "Data Mining Concepts and Techniques," vol. 15, n, 2016.

[6] D. M. Belete and M. D. Huchaiah, "Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results," *Int. J. Comput. Appl.*, no. September, 2021, doi: 10.1080/1206212X.2021.1974663.

[7] G. Behera and N. Nain, "GSO-CRS: grid search optimization for collaborative recommendation system," *Indian Acad. Sci.*, vol. 47, no. 158, pp. 1–12, 2019, doi: https://doi.org/10.1007/s12046-022-01924-0.

[8] A. . Wahyu Nugraha, "Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search," *J. Sist. Inf.*, vol. 11, no. 2, pp. 391–401, 2022.

[9] and S. S. L. Septiningrum, H. Yasin, "Prediksi Indeks Harga Saham Gabungan Menggunakan Support Vector Regression (Svr) Dengan Algoritma Grid Search," *Gaussian*, vol. 4, no. 2, pp. 315–321, 2015.

[10] L. Yang and A. Shami, *On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice*. 2020.

[11] Larose, "Discovering Knowledge in Data: An Introduction to Data Mining," 2005.

[12] P. Chapman *et al.*, *CRIPS DM 1.0 Step by Step Data Mining Guide*. 2000.