

**OPTIMASI ALGORITMA *DECISION TREE*
MENGUNAKAN *GRID SEARCH OPTIMIZATION*
UNTUK KLASIFIKASI MAHASISWA NON-AKTIF**

SKRIPSI

OLEH:

ROBERTO

201810225157



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2022**

LEMBAR PERSETUJUAN PEMBIMBING

Judul Proposal Skripsi : Optimasi Algoritma *Decision Tree*
Menggunakan *Grid Search Optimization*
Untuk Klasifikasi Mahasiswa Non-Aktif

Nama Mahasiswa : Roberto

Nomor Pokok Mahasiswa : 201810225157

Program Studi/Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 11 Juli 2022



Pembimbing I

Pembimbing II

Prima Dina Atika, S.Kom., M.Kom

NIDN: 0413066604

Ahmad Fathurrozi, S.E., M.M.S.I

NIDN: 0327117402

LEMBAR PENGESAHAN

Judul Skripsi : Optimasi Algoritma *Decision Tree* Menggunakan
Grid Search Optimization Untuk Klasifikasi
Mahasiswa Non-Aktif

Nama Mahasiswa : Roberto

Nomor Pokok Mahasiswa : 201810225157

Program Studi/Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 11 Juli 2022

Bekasi, 18 / Juli / 2022

MENGESAHKAN,

Ketua Tim Penguji : Wowon Priatna, ST, M.Ti.

NIDN : 0429118007

Penguji I : Dr. Tb. Ai Munandar, S.Kom., MT.

NIDN : 0413098403

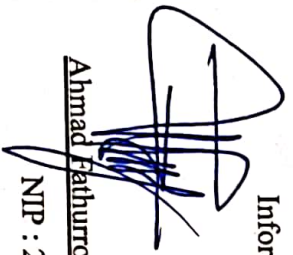
Penguji II : Prima Dina Atika, S.Kom., M.Kom.

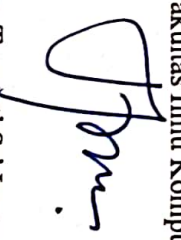
NIDN : 0311037107

MENGETAHUI,

Ketua Program Studi
Informatika

Dekan
Fakultas Ilmu Komputer


Ahmad Fathurozi, S.E., M.M.S.I.
NIP : 2012486


Dr. Dra. Tyastuti Siti Lestari, M.M.
NIP : 1408206



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Roberto
NPM : 201810225157
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Optimasi Algoritma *Decision Tree* Menggunakan *Grid Search Optimization* Untuk Klasifikasi Mahasiswa Non-Aktif

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Bekasi, 28 Juli 2022

Yang membuat pernyataan



Roberto

201810225157

ABSTRAK

Roberto. 201810225157. “Optimasi Algoritma *Decision Tree* Menggunakan *Grid Search Optimization* Untuk Klasifikasi Mahasiswa Non-Aktif”

Machine Learning (ML) merupakan teknik yang banyak digunakan untuk mengekstraksi pengetahuan dari data. ML adalah bidang penelitian di persimpangan antara statistik, kecerdasan buatan, dan ilmu komputer. ML juga dikenal sebagai analitik prediktif atau pembelajaran statistik[1]. Pada konsep klasifikasi kendala yang biasa dihadapi adalah rumit dan mahal nya proses komputasi, terutama untuk jumlah kombinasi *variable* yang sangat tinggi (dimensi tinggi). Kendala ini juga mempengaruhi akurasi dan kualitas model klasifikasi yang terbentuk. Inilah yang menjadi perhatian bagi banyak peneliti untuk diatasi dengan menawarkan berbagai pendekatan sebagai upaya untuk melakukan optimasi pada teknik klasifikasi seperti *Decision Tree* (DT). Terdapat banyak kasus yang sudah diselesaikan menggunakan algoritma DT, salah satunya terkait permasalahan klasifikasi mahasiswa non-aktif. Namun, akurasi hasil klasifikasi belum maksimal padahal batas akurasi bisa mencapai 100%. Oleh karena itu penelitian dilakukan sebagai upaya untuk mengoptimalkan teknik klasifikasi DT menggunakan *Grid Search Optimization* (GSO). Tujuannya GSO untuk menaikkan akurasi hasil klasifikasi khususnya pada kasus mahasiswa non-aktif. Hasil akurasi dari penerapan algoritma *Decision Tree* mendapatkan point sebesar 0.982 atau 98.2%, sedangkan diterapkan *Grid Search Optimization* dengan mendapatkan parameter terbaik yaitu dengan dengan criterion = ‘gini’ dan max_depth = 64. Dalam implementasi di algoritma *Decision Tree* dengan tambahan criterion = ‘gini’ dan max_depth = 64 menambah peningkatan akurasi sebesar 0.002 atau 0.2%.

Kata Kunci: *Machine Learning*, Klasifikasi, Mahasiswa Non-Aktif, *Decision Tree*, *Grid Search Optimization*.

ABSTRACT

Roberto. 201810225157. “Optimization of Decision Tree Algorithm Using Grid Search Optimization for Classification Non-Active Students.”

Machine Learning (ML) is a technique that is widely used to extract knowledge from data. ML is a research field at the intersection of statistics, artificial intelligence, and computer science. ML is also known as predictive analytics or statistical learning [1]. In the concept of classification, the constraints that are commonly encountered are complicated and expensive computational processes, especially for a very high number of variable combinations (high dimensions). This constraint also affects the accuracy and quality of the formed classification model. This is a concern for many researchers to overcome by offering various approaches in an effort to optimize classification techniques such as Decision Tree (DT). There are many cases that have been solved using the DT algorithm, one of which is related to the problem of classifying non-active students. However, the accuracy of the classification results has not been maximized even though the accuracy limit can reach 100%. Therefore, this research was conducted as an effort to optimize the DT classification technique using Grid Search Optimization (GSO). The goal of GSO is to increase the accuracy of the classification results, especially in the case of non-active students. The accuracy results from the application of the Decision Tree algorithm get 0.982, or 98.2% points, while the Grid Search Optimization is applied by getting the best parameters, namely with criterion = 'gini' and max_depth = 64. In the implementation of the Decision Tree algorithm with additional criterion = 'gini' and max_depth = 64, this adds an accuracy increase of 0.002 or 0.2%.

Keywords: Machine Learning, Non-Active Students, Classification, Decision Tree, Grid Search Optimization.

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Roberto
NPM : 201810225157
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalty-Free Right)**, atas karya ilmiah saya yang berjudul :

Optimasi Algoritma *Decision Tree* Menggunakan *Grid Search Optimization* Untuk Klasifikasi Mahasiswa Non-Aktif

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bekasi
Pada tanggal : 28 Juli 2022
Yang Menyatakan



Roberto
201810225157

KATA PENGANTAR

Puji syukur saya panjatkan kehadiran Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya saya dapat menyelesaikan Skripsi yang disusun sebagai salah satu syarat untuk mencapai gelar Sarjana Komputer (S1) pada Program Studi Informatika Universitas Bhayangkara Jakarta Raya.

Penyusunan laporan, mendapat banyak sekali bantuan, dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini ingin menyampaikan terima kasih kepada :

1. Kepada Tuhan Yang Maha Esa yang telah memberikan nikmat dan rahmat dalam melaksanakan penyusunan proposal skripsi.
2. Bapak Inspektur Jenderal Polisi (Purn) Dr. Drs. H. Bambang Karsono, S.H., M.M. Selaku Rektor Universitas Bhayangkara Jakarta Raya
3. Ibu Dr. Dra. Tyastuti Sri Lestari, M.M. Selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
4. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. Selaku Ketua Program Studi Informatika dan Selaku Dosen Pembimbing 2.
5. Ibu Prima Dina Atika, S.Kom., M.Kom. Selaku Dosen Pembimbing 1 penyusunan laporan Skripsi.
6. Ibu Dian Hartanti, S.Kom., M.M.S.I. Selaku Dosen Pembimbing Akademik.
7. Kedua Orang Tua saya yang selalu memberi semangat dan dukungan, mendoakan setiap harinya agar diberikan kesehatan dan kemudahan dalam menyelesaikan Skripsi.

Penelitian ini masih banyak kekurangan, baik dari materi maupun teknik penyajiannya. Oleh karena itu, dari penelitian ini mengharapkan kritik dan saran yang membangun.

Bekasi, 22 Juni 2022



Roberto

201810225157

DAFTAR ISI

	Halaman
LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PENGESAHAN.....	iii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vi
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI.....	vii
KATA PENGANTAR	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xiii
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	2
1.3. Rumusan Masalah	2
1.4. Batasan Masalah.....	3
1.5. Tujuan dan Manfaat Penelitian	3
1.6. Sistematika Penulisan.....	4
BAB 2 LANDASAN TEORI.....	5
2.1. Tinjauan Pustaka	5
2.2. Landasan Teori.....	6
2.2.1. <i>Data Mining</i>	6
2.2.2. Status Mahasiswa.....	9
2.2.3. <i>Decision Tree</i>	10
2.2.4. Algoritma C4.5.....	11
2.2.5. <i>Grid Search Optimization</i>	17
2.2.6. <i>Hyperparameter</i>	18

2.2.7. <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	18
2.2.8. <i>Confusion Matrix</i>	20
2.2.9. <i>Python</i>	24
BAB 3 METODOLOGI PENELITIAN	Error! Bookmark not defined.
3.1. Waktu dan Tempat Penelitian	25
3.2. Kerangka Penelitian	25
3.3. Pengumpulan Data	26
3.3.1. <i>Business Understanding</i>	26
3.3.2. <i>Data Understanding</i>	26
3.3.3. <i>Data Preparation</i>	27
BAB 4 PERANCANGAN SISTEM DAN IMPLEMENTASI	37
4.1. Hasil	37
4.1.1. <i>Dataset</i>	37
4.1.2. <i>Split Data</i>	38
4.1.3. <i>Modeling</i>	39
4.1.4. <i>Grid Search Optimization</i>	40
4.1.5. <i>Evaluation</i>	41
4.2. Pembahasan	43
4.2.1. <i>Modeling</i>	43
4.2.2. <i>Grid Search Optimization</i>	49
4.2.3. <i>Evaluation</i>	50
BAB 5 PENUTUP	53
5.1. Kesimpulan	53
5.2. Saran	53
DAFTAR PUSTAKA	54
LAMPIRAN	56

DAFTAR TABEL

	Halaman
Table 2.1 State of The Art Sumber: (Hasil Penelitian, 2022)	5
Table 2.2 Kelebihan dan Kekurangan Decision Tree Sumber: (Dios Kurniawan, 2022)	10
Table 2.3 Contoh Dataset	12
Table 2.4 Perhitungan Entropy	13
Table 2.5 Perhitungan Gini	15
Table 2.6 Contoh Dataset SMOTE	19
Table 2.7 Metric Evaluation	22
Table 2.8 Perhitungan Evaluation	23
Table 2.9 Contoh Perhitungan Confusion Matrix	23
Table 2.10 Hasil Perhitungan Confusion Matrix	24
Table 3.1 Data Understanding Sumber: (Hasil Penelitian, 2022)	26
Table 3.2 Pergantian Nama Kolom Sumber: (Hasil Penelitian, 2022)	28
Table 3.3 Missing Value Sumber: (Hasil Penelitian, 2022)	29
Table 3.4 Hasil Handle Missing Value Sumber: (Hasil Penelitian, 2022)	30
Table 4.1 Perhitungan <i>Feature</i> lokasi Sumber: (Hasil Penelitian, 2022)	43
Table 4.2 Perhitungan <i>Feature</i> ipk Sumber: (Hasil Penelitian, 2022)	44
Table 4.3 Perhitungan <i>Feature</i> ipk Sumber: (Hasil Penelitian, 2022)	45
Table 4.4 Perhitungan <i>Feature</i> tahun_lahir Sumber: (Hasil Penelitian, 2022)	47
Table 4.5 Perhitungan <i>Feature</i> gender Sumber: (Hasil Penelitian, 2022)	47
Table 4.6 Perhitungan <i>Feature</i> pendidikan Sumber: (Hasil Penelitian, 2022)	48

Table 4.7 Perhitungan <i>Feature</i> orang_tua Sumber: (Hasil Penelitian, 2022)	48
Table 4.8 Perhitungan <i>Feature</i> kelas Sumber: (Hasil Penelitian, 2022).....	48
Table 4.9 Total Gain dan Gini Sumber: (Hasil Penelitian, 2022)	49
Table 4.10 Hasil Evaluasi Sebelum dan Sesudah Optimasi Sumber: (Hasil Penelitian, 2022).....	51



DAFTAR GAMBAR

	Halaman
Gambar 2.1 Cross-Industry Standard Process for Data Mining (CRISP-DM) Sumber: (Larose, 2014).....	7
Gambar 2.2 Contoh Decision Tree Sumber: (Nong Ye, 2014).....	17
Gambar 3.1 Kerangka Penelitian Sumber: (Hasil Penelitian, 2022)	25
Gambar 3.2 Melihat Outlier Semester Sumber: (Hasil Penelitian, 2022).....	31
Gambar 3.3 Melihat Variable Target Label Sumber: (Hasil Penelitian, 2022).....	31
Gambar 3.4 Hasil Handling Outlier Sumber: (Hasil Penelitian, 2022)	32
Gambar 3.5 Melihat Data Categorical Sumber: (Hasil Penelitian, 2022).....	32
Gambar 3.6 Hasil Handling Data Categorical Sumber: (Hasil Penelitian, 2022) .	33
Gambar 3.7 Hasil Drop Kolom Sumber: (Hasil Penelitian, 2022).....	34
Gambar 3.8 Hasil Kolom Variable y Sumber: (Hasil Penelitian, 2022).....	34
Gambar 3.9 Melihat <i>Dataset</i> Sebelum <i>Oversampling Data</i> Sumber: (Hasil Penelitian, 2022).....	35
Gambar 3.10 Hasil <i>Resample</i> Sumber: (Hasil Penelitian, 2022).....	36
Gambar 4.1 Rancangan <i>Modeling</i> Sumber: (Hasil Penelitian, 2022).....	37
Gambar 4.2 <i>Dataset</i> Sumber: (Hasil Penelitian, 2022).....	38
Gambar 4.3 <i>Feature Selection</i> Sumber: (Hasil Penelitian, 2022)	38
Gambar 4.4 Ukuran <i>Train-Test Split Data</i> Sumber: (Hasil Penelitian, 2022)	39
Gambar 4.5 Hasil Akurasi <i>Train</i> dan <i>Test</i> Sumber: (Hasil Penelitian, 2022).....	40
Gambar 4.6 Hasil dari <i>Grid Search</i> Sumber: (Hasil Penelitian, 2022).....	40

Gambar 4.7 Hasil <i>Decision Tree</i> Dengan <i>Parameter</i> Sumber: (Hasil Penelitian, 2022)	41
Gambar 4.8 <i>Confusion Matrix Model</i> Sumber: (Hasil Penelitian, 2022).....	42
Gambar 4.9 Hasil Pencarian <i>Grid Search Optimization</i> Sumber: (Hasil Penelitian, 2022)	49
Gambar 4.10 Hasil <i>Decision</i> Dengan <i>Hyperparameter</i> Terbaik Sumber: Hasil Penelitian (2022).....	50
Gambar 4.11 <i>Visualization Tree</i> Sumber: (Hasil Penelitian, 2022)	52



DAFTAR LAMPIRAN

	Halaman
Lampiran I.....	57
Lampiran II.....	61
Lampiran III.....	62
Lampiran IV.....	63

