

**IMPLEMENTASI ALGORITMA C5.0 UNTUK  
PENENTUAN PRESTASI AKADEMIK SISWA**

**SKRIPSI**



Oleh:

**SALSABILA CITRA ESANANDA**

**NPM. 17081010029**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2021**

# **IMPLEMENTASI ALGORITMA C5.0 UNTUK PENENTUAN PRESTASI AKADEMIK SISWA**

## **SKRIPSI**

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar Sarjana  
Komputer Program Studi Informatika



Oleh:

**SALSABILA CITRA ESANANDA**

**NPM. 17081010029**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2021**

# LEMBAR PENGESAHAN

## LEMBAR PENGESAHAN

Judul : IMPLEMENTASI ALGORITMA C5.0 UNTUK PENENTUAN  
PRESTASI AKADEMIK SISWA  
Oleh : SALSABILA CITRA ESANANDA  
NPM : 17081010029

Telah Diseminarkan Dalam Ujian Skripsi  
Pada Hari Senin, 7 Juni 2021

Mengetahui,

Dosen Pembimbing

Dosen Penguji

1.

1.

  
Budi Nugroho, S.Kom, M.Kom  
NPT. 3 6009 05 0205 1

  
Eka Prakarsa Mulyawati, S.T, M.Kom  
NIP. 19860525 201803 1 001

2.

2.

  
Fetty Tri Anggarany, S.Kom, M.Kom  
NPT. 3 8202 06 0208 1

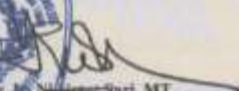
  
Christia, S.Kom, M.T  
NPT. 3 8210 10 0296 1

Menyetujui,

Dekan  
Fakultas Ilmu Komputer

Koordinator Program Studi  
Informatika,



  
Dr. Jati Nugroho Sari, MT  
NIP. 19620731 199203 2 001

  
Budi Nugroho, S.Kom, M.Kom  
NPT. 3 6009 05 0205 1

# SURAT PERNYATAAN ANTI PLAGIAT

## SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Program Studi Informatika UPN "Veteran" Jawa Timur, yang bertandatangan di bawah ini :

Nama : SALSABILA CITRA ESANANDA.

NPM : 17081010029

Menyatakan bahwa Judul Skripsi/ Tugas Akhir yang Saya ajukan dan akan dikerjakan, yang berjudul:

**"IMPLEMENTASI ALGORITMA C5.0 UNTUK PENENTUAN PRESTASI  
AKADEMIK SISWA"**

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk dan atau *software* yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN "Veteran" Jawa Timur maupun di institusi pendidikan lain. Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka Saya siap menerima segala konsekuensinya.

Surabaya, 07 Juni 2021

Hormat Saya,



**SALSABILA CITRA ESANANDA**

**NPM. 17081010029**

# IMPLEMENTASI ALGORITMA C5.0 UNTUK PENENTUAN PRESTASI AKADEMIK SISWA

**Nama Mahasiswa** : Salsabila Citra Esananda  
**NPM** : 17081010029  
**Program Studi** : Informatika  
**Dosen Pembimbing** : Budi Nugroho, S. Kom, M.Kom  
Fetty Tri Anggraeny, S.Kom, M.Kom

## ABSTRAK

Pendidikan merupakan aspek yang sangat penting di dunia yang mempengaruhi kelangsungan hidup manusia. Seiring berjalannya waktu semakin banyak perkembangan ilmu pengetahuan yang berdampak terhadap kualitas prestasi siswa. Pada saat ini, di era digital banyak kegiatan yang melibatkan penggunaan teknologi informasi dalam pengolahan data termasuk dalam menentukan prestasi akademik siswa. Hal ini membuat lebih efektif dan efisien dalam *monitoring* perkembangan siswa dibandingkan dengan cara manual yang rentan akan *human error* mengingat data yang begitu banyak.

Algoritma C5.0 merupakan salah satu metode klasifikasi yang dapat diterapkan pada kasus penentuan prestasi siswa dengan menerapkan struktur pohon dalam mengambil keputusan. Algoritma ini terdapat 2 proses pengembangan dari algoritma sebelumnya, yaitu fase *pruning* dan *boosting* di akhir proses algoritma C5.0.

Setelah model pohon keputusan terbentuk, dilakukan pengujian awal sebelum *boosting* menghasilkan akurasi 68,75% dan 93,15% setelah dilakukan *boosting*. Kemudian, dilakukan pengujian sistem menggunakan 144 data *testing* yang menghasilkan nilai *recall* sebesar 84,32%, *precision* sebesar 81,62%, dan nilai *accuracy* sebesar 84,03%.

**Kata Kunci:** C5.0, *Entropy*, *Information Gain*, *Pruning*, *Boosting*, Prestasi Akademik.

## **KATA PENGANTAR**

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan laporan dan penelitian skripsi dengan judul “IMPLEMENTASI ALGORITMA C5.0 UNTUK PENENTUAN PRESTASI AKADEMIK SISWA”. Penelitian ini dilakukan sebagai bentuk pertanggungjawaban penulis dalam menyelesaikan studi dan menempuh gelar Sarjana Komputer Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Selama melaksanakan penelitian dan penulisan laporan ini, penulis telah banyak menerima bimbingan dan arahan serta dukungan yang membantu hingga terselesaikannya penelitian ini.

Penulis menyadari jika dalam penulisan laporan dan penelitian skripsi ini masih jauh dari sempurna. Atas kekurangannya, penulis sangat terbuka menerima masukan kritik serta saran yang membangun ke arah yang lebih baik untuk dijadikan sebagai pembelajaran penulis di kemudian hari. Selain itu, penulis berharap semoga laporan dan penelitian skripsi ini bermanfaat bagi seluruh pihak.

Surabaya, Juni 2021

Penulis

## UCAPAN TERIMA KASIH

Ucapan terimakasih penulis sampaikan kepada Allah SWT yang selalu ada sebagai tempat mengadu atas semua masalah yang dihadapi penulis dan selalu membantu dengan memberikan kelancaran, kemudahan serta kekuatan kepada penulis dalam keadaan apapun. Selain itu, dalam melaksanakan penelitian dan penulisan laporan ini juga tidak lepas dari bimbingan, arahan, serta dukungan dari berbagai pihak hingga terselesaikannya penelitian ini. Penulis mengucapkan terimakasih kepada yang terhormat :

1. Ibu Sri Sulistyorini yang selalu memberikan semangat motivasi agar terus berjuang dan tidak menyerah, mendoakan untuk keberhasilan dan kesuksesan penulis serta alm. Ayah yang selalu mendukung penulis sampai akhir hidupnya.
2. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Dr. Ir. Ni Ketut Sari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Bapak Budi Nugroho, S.Kom, M.Kom selaku Koordinator Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur sekaligus sebagai Dosen Pembimbing I yang telah sabar dan ikhlas meluangkan waktunya dalam membimbing serta mengarahkan penulis dari awal penelitian hingga terselesaikannya penelitian.
5. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom selaku Dosen Pembimbing II yang telah sabar dan ikhlas meluangkan waktunya dalam membimbing serta

mengarahkan penulis dari awal penelitian hingga terselesaikannya penelitian.

6. Segenap Dosen Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmu serta pengalamannya selama dalam perkuliahan.
7. Adik dari penulis, Intania serta keluarga besar yang selalu mendoakan dan mendukung agar terselesaikannya penelitian skripsi ini.
8. Mohammad Yusuf Efendi yang selalu ada sebagai tempat meluapkan segala keluh kesah yang dihadapi, selalu memberikan semangat, motivasi serta *support* kepada penulis dalam keadaan apapun.
9. Yusmi Nur Aini, Sumiati Ratna Sari, Tria Nurmayasari, Tentra Olivia, Chilyatun Nisa’, Cahaya Putri Agustika, Prameswari Reksa Agami, Erlina Dyah Karisma yang selalu menemani, mendukung dan memberikan semangat dari awal perkuliahan hingga sekarang.
10. Alfiyatus, Ghanis, dan Aldha yang selalu menemani saat susah dan senang serta memberikan dukungan motivasi kepada penulis.
11. Teman-teman dari Program Studi Informatika yang bersedia menjadi bagian dari kehidupan penulis.
12. Semua pihak yang telah berperan dalam proses pengerjaan penelitian ini yang tidak dapat disebutkan satu per satu.

Terimakasih atas dukungannya selama ini, semoga Allah SWT memberi balasan atas kebaikan seluruh pihak kepada penulis.

Surabaya, Juni 2021

Penulis



## DAFTAR ISI

LEMBAR PENGESAHAN .....	i
SURAT PERNYATAAN ANTI PLAGIAT .....	i
ABSTRAK .....	i
KATA PENGANTAR .....	ii
UCAPAN TERIMA KASIH.....	iii
DAFTAR ISI.....	v
DAFTAR GAMBAR .....	viii
DAFTAR TABEL.....	x
DAFTAR KODE.....	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	4
1.5 Manfaat.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Pendahulu .....	5
2.2 Prestasi Akademik .....	7
2.3 <i>Data Mining</i> .....	8
2.4 <i>Decision Tree</i> .....	9

2.5	Algoritma C5.0.....	10
2.5.1	<i>Pruning</i> .....	17
2.5.2	<i>Boosting</i> .....	20
2.6	<i>Confusion Matrix</i> .....	36
2.7	<i>Kaggle</i> .....	38
BAB III METODE PENELITIAN.....		39
3.1	Dataset Penelitian .....	40
3.2	Seleksi Data .....	42
3.3	Pembagian Data.....	44
3.4	Pembentukan Model.....	44
3.4.1	<i>Pruning</i> .....	52
3.4.2	<i>Boosting</i> .....	54
3.5	Skenario Pengujian Sistem .....	57
BAB IV HASIL DAN PEMBAHASAN .....		58
4.1	Implementasi Data.....	58
4.2	Implementasi Desain <i>Interface</i> .....	59
4.2.1	Tampilan Awal.....	59
4.2.2	Tampilan Menu Data Latih .....	60
4.2.3	Tampilan Menu Algoritma C5.0 .....	60
4.2.4	Tampilan Menu Proses Boosting .....	65
4.2.5	Tampilan Menu Pengujian Data.....	69

4.2.6	Tampilan Menu History Pengujian .....	70
4.3	Pembahasan .....	71
4.3.1	Perhitungan Entropy.....	72
4.3.2	Perhitungan Information Gain.....	73
4.3.3	Perhitungan Tingkat Galat (Pruning).....	74
4.3.4	Perhitungan Boosting .....	74
4.4	Pengujian Sistem .....	77
4.4.1	Pengujian Awal Sistem .....	77
4.4.2	Pengujian Sistem Setelah <i>Boosting</i> .....	78
4.4.3	Perbandingan dengan Penelitian Pendahulu .....	79
BAB V KESIMPULAN DAN SARAN.....		81
5.1	Kesimpulan .....	81
5.2	Saran.....	82
DAFTAR PUSTAKA .....		83
Lampiran 1. Dataset Penelitian Sebelum Proses Diskretisasi .....		85
Lampiran 2. Dataset Penelitian Setelah Proses Diskretisasi .....		87
Lampiran 3. Data <i>Training</i> (Data Latih untuk Pembentukan Pohon Keputusan Awal).....		89
Lampiran 4. Hasil Pengujian Awal Sebelum Boosting.....		91
Lampiran 5. Hasil Pengujian Setelah <i>Boosting</i> .....		100
BIODATA PENULIS .....		109

## DAFTAR GAMBAR

<b>Gambar 2.1</b> Parent Node .....	16
<b>Gambar 2.2</b> Proses Pruning .....	20
<b>Gambar 2.3</b> Model Pohon Keputusan.....	23
<b>Gambar 2.4</b> Model Pohon Keputusan Iterasi Kedua .....	32
<b>Gambar 3.1</b> Tahapan Penelitian.....	39
<b>Gambar 3.2</b> Flowchart Algoritma C5.0 .....	45
<b>Gambar 3.3</b> Parent Node Awal Dataset Penelitian.....	51
<b>Gambar 3.4</b> Flowchart proses Pruning .....	53
<b>Gambar 3.5</b> Flowchart proses Boosting .....	55
<b>Gambar 4.1</b> Tampilan Awal Website (Menu Beranda).....	59
<b>Gambar 4.2</b> Tampilan Menu Data Latih.....	60
<b>Gambar 4.3</b> Menu Algoritma C5.0.....	61
<b>Gambar 4.4</b> Menu Perhitungan Entropy, Information Gain, dan Galat .....	61
<b>Gambar 4.5</b> Tampilan Information Gain Tertinggi dan Tingkat Galat .....	62
<b>Gambar 4.6</b> Tampilan dari keputusan Pruning dan keadaan Confidence Value .	62
<b>Gambar 4.7</b> Tampilan Menu Model Pohon Keputusan .....	63
<b>Gambar 4.8</b> Tampilan Sub Menu Boosting .....	64
<b>Gambar 4.9</b> Tampilan Menu Proses Boosting.....	65
<b>Gambar 4.10</b> Tampilan Menu Perhitungan Dataset Random.....	66
<b>Gambar 4.11</b> Menu Data Latih Akhir.....	67
<b>Gambar 4.12</b> Tampilan Sub Menu Data Latih Akhir .....	67
<b>Gambar 4.13</b> Tampilan Perhitungan Entropy, Information Gain, Galat .....	67

<b>Gambar 4.14</b> Tampilan Hasil Boosting .....	68
<b>Gambar 4.15</b> Tampilan Pohon Keputusan Akhir .....	68
<b>Gambar 4.16</b> Tampilan Pengujian Data .....	69
<b>Gambar 4.17</b> Tampilan Hasil Uji Data.....	70
<b>Gambar 4.18</b> Tampilan History Pengujian .....	71
<b>Gambar 4.19</b> Tampilan Menu History Pengujian Boosting .....	71

## DAFTAR TABEL

<b>Tabel 2.1</b> Contoh Data Rekomendasi Kinerja Pegawai .....	12
<b>Tabel 2.2</b> Pengelompokan Jumlah Kasus Data .....	13
<b>Tabel 2.3</b> Nilai Information Gain setiap Atribut .....	15
<b>Tabel 2.4</b> Hasil Pengujian Pohon Keputusan .....	24
<b>Tabel 2.5</b> Pemberian Bobot Awal .....	25
<b>Tabel 2.6</b> Pemberian Bobot Terbaru .....	27
<b>Tabel 2.7</b> Pembentukan Data Training baru .....	29
<b>Tabel 2.8</b> Pengelompokan Data Iterasi Kedua .....	30
<b>Tabel 2.9</b> Nilai Information Gain Iterasi Kedua .....	31
<b>Tabel 2.10</b> Hasil Pengujian Model Pohon Kedua .....	33
<b>Tabel 2.11</b> Pemberian Bobot Awal Iterasi Kedua .....	34
<b>Tabel 2.12</b> Tabel Confusion Matrix .....	36
<b>Tabel 2.13</b> Confusion Matrix Prediksi Kanker .....	37
<b>Tabel 3.1</b> Aturan Diskretisasi .....	43
<b>Tabel 3.2</b> Pengelompokan data setiap kelas .....	46
<b>Tabel 3.3</b> Pengelompokan Jumlah Kasus Data Training terhadap Kelas.....	47
<b>Tabel 3.4</b> Nilai Information Gain .....	50
<b>Tabel 4.1</b> Confusion Matrix Pengujian Awal.....	77
<b>Tabel 4.2</b> Tabel Confusion Matrix Boosting .....	78
<b>Tabel 4.3</b> Hasil Penelitian Pendahulu.....	80

## DAFTAR KODE

<b>Kode Program 4.1</b> Script Perhitungan nilai Pi .....	72
<b>Kode Program 4.2</b> Menghitung Entropy .....	73
<b>Kode Program 4.3</b> Perhitungan Information Gain .....	73
<b>Kode Program 4.4</b> Perhitungan Tingkat Galat .....	74
<b>Kode Program 4.5</b> Perhitungan bobot pada fase Boosting .....	75
<b>Kode Program 4.6</b> Fungsi Pembentukan Data Training Baru.....	76