

**IMPROVISASI ALGORITMA NAIVE BAYES DENGAN ADABOOST
UNTUK DATASET ECOLI**

SKRIPSI



disusun oleh

Agung Triyadi

17.11.1065

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**IMPROVISASI ALGORITMA NAIVE BAYES DENGAN ADABOOST
UNTUK DATASET ECOLI**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Agung Triyadi

17.11.1065

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

IMPROVISASI ALGORITMA NAIVE BAYES DENGAN ADABOOST UNTUK DATASET ECOLI

yang dipersiapkan dan disusun oleh

Agung Triyadi
17.11.1065

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 9 Februari 2021

Dosen Pembimbing,



Hastari Utama, M.Cs
NIK. 190302230

PENGESAHAN

SKRIPSI

IMPROVISASI ALGORITMA NAIVE BAYES DENGAN ADABOOST UNTUK DATASET ECOLI

yang dipersiapkan dan disusun oleh

Agung Triyadi

17.11.1065

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Februari 2021

Susunan Dewan Penguji

Nama Penguji

Donni Prabowo, M.Kom.
NIK. 190302253

Haryoko, S.Kom, M.Cs,
NIK. 190302286

Hastari Utama, M.Cs
NIK. 190302230

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Februari 2021

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis suatu institusi Pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat menjadi tanggungjawab saya pribadi

Yogyakarta, 10 Februari 2021



Agung Triyadi
NIM. 17.11.1065

MOTTO

“Hidup itu bukan ajang untuk balapan. Setiap orang punya waktunya masing-masing. Jangan pernah membandingkan hidupmu dengan yang lain. Tidak pernah ada perbandingan antara Matahari dan Bulan. Mereka akan bersinar pada waktunya masing-masing, jadi tunggulah waktumu dan terus berjuang disertai doa untuk meraih mimpimu.”



PERSEMBAHAN

Segala puji dan syukur atas kehadiran Allah SWT atas berkat, rahmat dan hidayah-Nya sehingga penulis mampu menyelesaikan skripsi dengan baik. Penulis juga mengucapkan terimakasih kepada pihak-pihak yang telah berkontribusi baik secara langsung maupun tidak langsung, baik dalam penelitian maupun dalam penyusunan naskah. Skripsi dipersembahkan kepada:

1. Orang tua dan seluruh keluarga penulis yang senantiasa memberikan dukungan dan doa sehingga dapat menempuh Pendidikan sarjana dan menyelesaikan skripsi ini dengan baik.
2. Bapak Hastari Utama, M.Cs selaku pembimbing yang telah dengan sabar dan tulus membimbing dan memberikan ilmu kepada penulis.
3. Teman-teman UKM AMIKOM Computer Club yang telah banyak bekerja sama dalam masa berorganisasi, serta berbagi pengetahuan dan ilmu.
4. Teman-teman S1 Informatika 03 yang telah banyak bekerja sama dalam menempuh Pendidikan sarjana.
5. Semua pihak keluarga besar Universitas Amikom Yogyakarta yang tidak dapat disebutkan satu-persatu.

KATA PENGANTAR

Segala puji dan syukur atas kehadiran Allah SWT atas berkat, rahmat dan hidayah-Nya sehingga penulis mampu menyelesaikan skripsi dengan judul “Improvisasi Algoritma *Naïve Bayes* Dengan *AdaBoost* Untuk *Dataset Ecoli*” sebagai syarat untuk menyelesaikan Program Sarjana (S1) pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta, walaupun penulis menyadari masih banyak sekali kekurangan yang itu semua tidak lepas dari keterbatasan penulis.

Dalam penyusunan skripsi banyak hambatan dan rintangan yang dihadapi namun pada akhirnya dapat dilewati karena bantuan, saran, dan bimbingan dari berbagai pihak baik secara moral maupun spiritual. Pada kesempatan ini penulis ingin mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. M. Suyanto, MM, selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta dan Ketua Program Studi S1 Sistem Informasi.
3. Bapak Hastari Utama, M.Cs selaku Dosen Pembimbing skripsi.
4. Seluruh jajaran dosen Program Studi S1 Informatika Fakultas Ilmu Komputer yang telah memberikan bekal ilmu pengetahuan selama perkuliahan.

5. Semua pihak keluarga besar Universitas Amikom Yogyakarta yang tidak dapat disebutkan satu-persatu.

Penulis menyadari dalam penyusunan skripsi ini masih banyak kekurangan serta jauh dari kata sempurna. Maka dari itu kritik dan saran yang bersifat membangun sangat diperlukan.

Akhir kata semoga penyusunan skripsi ini dapat bermanfaat bagi pembaca dalam menambah wawasan dan pengetahuan khususnya dalam bidang informatika.

Yogyakarta, 10 Februari 2021

Agung Triyadi

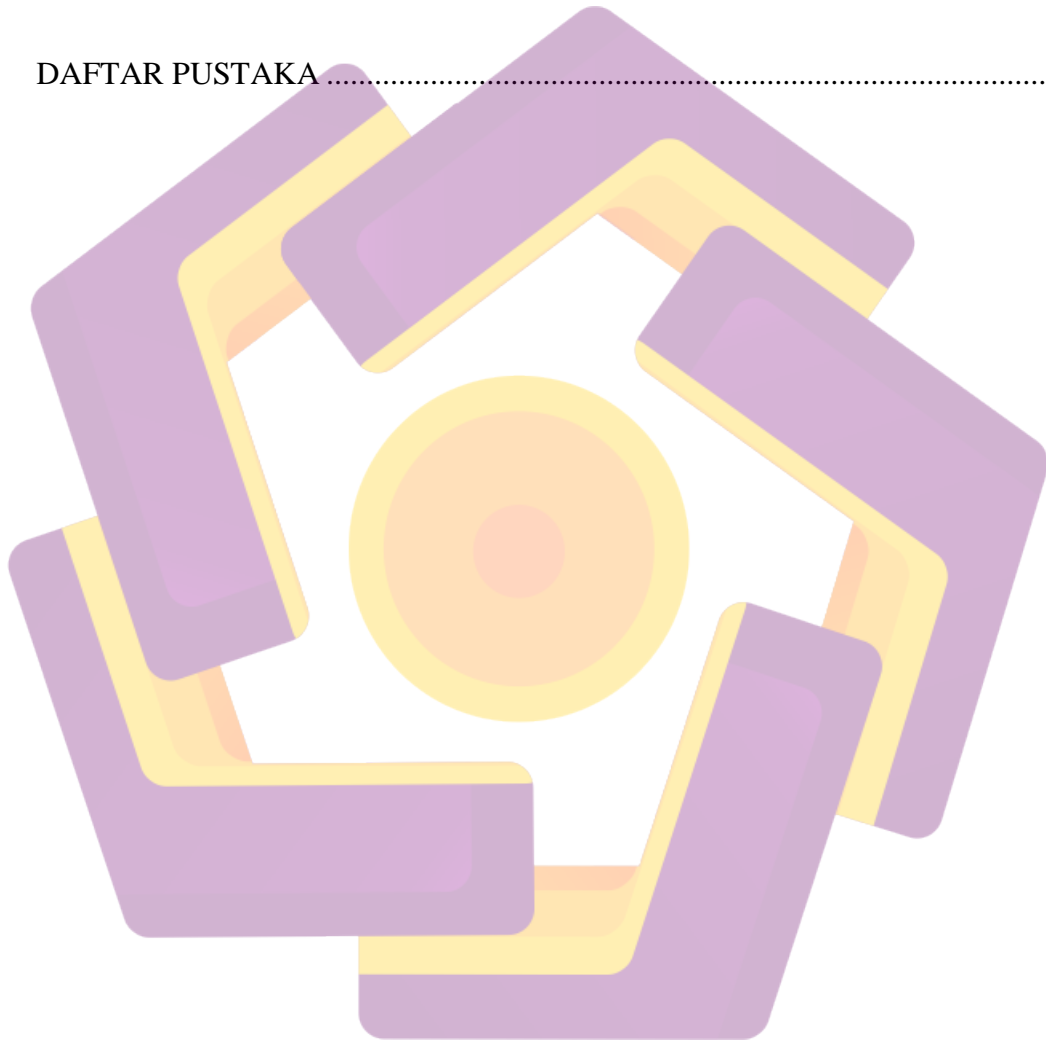
DAFTAR ISI

HALAMAN SAMBUL	i
HALAMAN JUDUL.....	ii
PERSETUJUAN	iii
PENGESAHAN	iv
PERNYATAAN.....	v
MOTTO.....	vi
PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xv
INTISARI.....	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4

1.5	Manfaat Penelitian.....	5
1.6	Metode Penelitian.....	5
1.6.1	Pengumpulan Data.....	5
1.6.2	Pengolahan Data Awal	5
1.6.3	Klasifikasi.....	5
1.6.4	Boosting.....	6
1.6.5	Evaluasi	6
1.7	Sistematika Penulisan.....	6
	BAB I PENDAHULUAN	6
	BAB II LANDASAN TEORI	6
	BAB III METODE PENELITIAN.....	6
	BAB IV HASIL DAN PEMBAHASAN	7
	BAB V KESIMPULAN DAN SARAN.....	7
BAB II	LANDASAN TEORI.....	8
2.1	Tinjauan Pustaka	8
2.2	Landasan Teori.....	17
2.2.1	Data Mining.....	17
2.2.2	Preprocessing.....	17
2.2.3	Dataset Splitting	19

2.2.4	Klasifikasi.....	19
2.2.5	Naïve Bayes.....	21
2.2.6	Boosting dan Adaptive Boosting.....	24
2.2.7	Pengujian dan Evaluasi Model	28
BAB III	METODE PENELITIAN.....	34
3.1	Gambaran Umum.....	34
3.2	Alat dan Bahan.....	34
3.2.1	Alat.....	34
3.2.2	Bahan.....	35
3.3	Jalannya Penelitian.....	37
3.4	Pra Pemrosesan Data.....	41
3.4.1	Proses Transformasi Data (Normalisasi).....	41
3.5	Klasifikasi dan Boosting	44
3.6	Evaluasi.....	46
BAB IV	HASIL DAN PEMBAHASAN.....	47
4.1	Dataset.....	47
4.2	Pra Pemrosesan Data.....	49
4.2.1	Implementasi Algoritma <i>Min Max Normalization</i>	49
4.3	Implementasi Algoritma Klasifikasi Naïve Bayes Dengan AdaBoost.....	52

4.3.1	Implementasi Algoritma <i>Naïve Bayes</i> Dengan <i>Adaboost</i>	53
BAB V	KESIMPULAN DAN SARAN.....	61
5.1	Kesimpulan.....	61
5.2	Saran.....	63
DAFTAR PUSTAKA	64

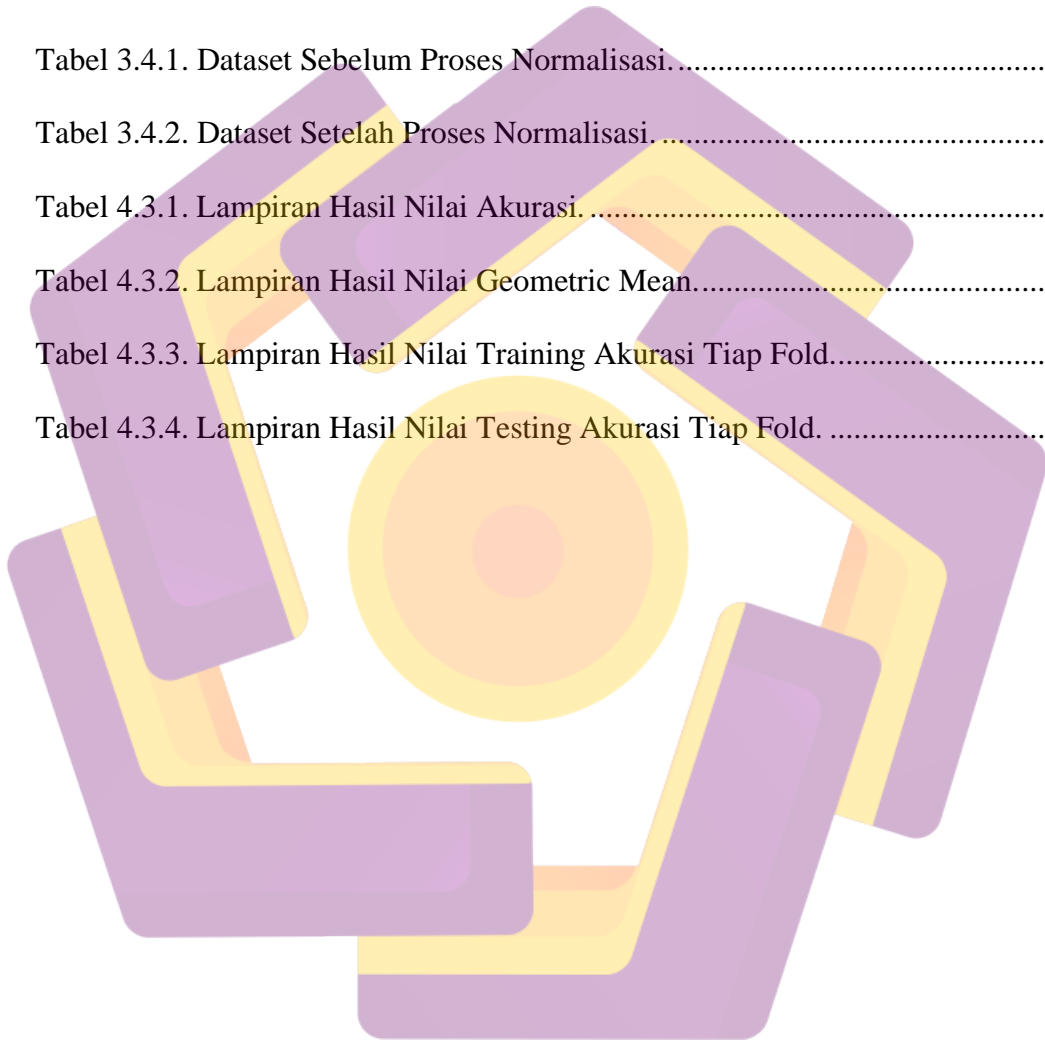


DAFTAR GAMBAR

Gambar 2.2.1. Distribusi Min Max Normalization [11][12].....	18
Gambar 2.2.2. Ilustrasi Skema Data Splitting [12].	19
Gambar 2.2.3. Bagan Proses Klasifikasi [13].	21
Gambar 2.2.4. AdaBoost sebuah algoritma boosting [11].	26
Gambar 2.2.5. Ilustrasi 10-Fold Cross Validation.	30
Gambar 3.3.1. Alur Penelitian.....	38
Gambar 3.3.2. Skenario Alur Klasifikasi Beserta Evaluasinya.	40
Gambar 3.5.1. Skenario Perbandingan Klasifikasi.	45
Gambar 4.1.1. Plot Distribusi Kelas Dataset ecoli.....	48
Gambar 4.1.2. Plot Bar Kelas Dataset ecoli.....	48
Gambar 4.2.1. Plot Distribution Dataset ecoli Sebelum Proses Min Max Normalization.....	50
Gambar 4.2.2. Plot Bar Dataset ecoli Sebelum Proses Min Max Normalization.	50
Gambar 4.2.3. Plot Distribution Dataset ecoli Setelah Proses Min Max Normalization.....	51
Gambar 4.2.4. Plot Bar Dataset ecoli Sebelum Proses Min Max Normalization.	51
Gambar 4.3.1. Nilai Akurasi Training.	53
Gambar 4.3.2. Nilai Akurasi Testing.	54
Gambar 4.3.3. Nilai Geometric Mean Training.	55
Gambar 4.3.4. Nilai Geometric Mean Testing.	56
Gambar 4.3.5. Nilai Cross Validation.....	57

DAFTAR TABEL

Tabel 2.1.1. Tabel Perbandingan Tinjauan Pustaka.....	12
Tabel 2.2.1. Tabel Confusion Matrix Dua Kelas.....	31
Tabel 3.2.1. Dataset ecoli.....	36
Tabel 3.4.1. Dataset Sebelum Proses Normalisasi.....	42
Tabel 3.4.2. Dataset Setelah Proses Normalisasi.....	44
Tabel 4.3.1. Lampiran Hasil Nilai Akurasi.....	58
Tabel 4.3.2. Lampiran Hasil Nilai Geometric Mean.....	59
Tabel 4.3.3. Lampiran Hasil Nilai Training Akurasi Tiap Fold.....	59
Tabel 4.3.4. Lampiran Hasil Nilai Testing Akurasi Tiap Fold.....	60



INTISARI

Permasalahan klasifikasi yang sering dijumpai adalah rendahnya tingkat akurasi sebuah *learning algorithms* yang termasuk dalam *weak classifiers* adalah sebuah permasalahan serius dalam *machine learning* karena tingkat akurasi dari sebuah *weak classifiers* yang rendah, hal tersebut berdampak pada hasil yang kurang akurat sehingga diperlukan sebuah tambahan *learning algorithms* pada *weak classifiers* menjadi sebuah *strong classifiers*.

Dalam penelitian ini akan dilakukan sebuah implementasi penggunaan *learning algorithms* yaitu *Naïve Bayes classifier* yang mana *learning algorithms* tersebut masih masuk kedalam *weak classifiers*, selanjutnya akan dibandingkan dengan *learning boosting algorithms* yaitu *Adaptive Boosting* atau biasa disebut dengan *AdaBoost*. *AdaBoost* digunakan untuk meningkatkan nilai *learning algorithms* serta mampu beradaptasi dengan *weak classifiers* secara baik dan meningkatkan akurasi dari *Naïve Bayes Classifier*.

Dari skenario tersebut diperoleh peningkatan nilai yang membuat *AdaBoost* cukup handal dalam meningkatkan sebuah nilai dari *weak classifier*, nilai yang diperoleh *Naïve Bayes* saja tidak di *boosting* pada *dataset ecoli* adalah Akurasi (*train*: 76,39%, *test*: 75,51%), *Geometric Mean* (*train*: 84,65%, *test*: 82,85%). Sedangkan untuk *Naïve Bayes* dengan *AdaBoost* pada *dataset ecoli* adalah Akurasi (*train*: 87,08%, *test*: 84,75%), *Geometric Mean* (*train*: 71,80%, *test*: 60,94%).

Kata Kunci : *Boosting, Naïve Bayes Classifier, Adaptive Boosting, Weak Classifier, Strong Classifier, Learning Algorithms.*

ABSTRACT

A classification problem that is often encountered is the low level of accuracy of learning algorithms that are included in the weak learners, which is a serious problem in machine learning because the level of accuracy of weak learners is low, it results in inaccurate results so that additional learning algorithms are needed on weak learners becomes a strong learners.

In this study, an implementation of the use of learning algorithms, namely the Naïve Bayes classifier, will be carried out in which learning algorithms are still included in the weak learners, and will then be juxtaposed with learning boosting algorithms, namely Adaptive Boosting or commonly known as AdaBoost. AdaBoost is used to increase the value of learning algorithms and is able to adapt well to weak learners and increase the accuracy of the Naïve Bayes Classifier.

From this scenario, an increase in value was obtained that made AdaBoost reliable in increasing a value from a weak classifier, the value obtained by Naïve Bayes alone was not boosting on the ecoli dataset, namely Accuracy (train: 76,39%, test: 75,51%), Geometric Mean (train: 84,65%, test: 82,85%). Whereas for Naïve Bayes with AdaBoost on the ecoli dataset, Accuracy (train: 87,08%, test: 84,75%), Geometric Mean (train: 71,80%, test: 60,94%).

Keywords: *Boosting, Naïve Bayes Classifier, Adaptive Boosting, Weak Classifier, Strong Classifier, Learning Algorithms.*