

**PENGARUH TEKNIK ENSEMBLE TERHADAP AKURASI
PADA ALGORITMA KLASIFIKASI**

SKRIPSI



disusun oleh

Alvin Aditya Rahman

16.11.0579

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

**PENGARUH TEKNIK ENSEMBLE TERHADAP AKURASI PADA
ALGORITMA KLASIFIKASI**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Alvin Aditya Rahman

16.11.0579

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

PERSETUJUAN

SKRIPSI

PENGARUH TEKNIK ENSEMBLE TERHADAP AKURASI PADA ALGORITMA KLASIFIKASI

yang dipersiapkan dan disusun oleh

Alvin Aditya Rahman

16.11.00579

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 17 Juli 2020

Dosen Pembimbing,

Mardhiya Hayaty, S.T., M.Kom

NIK. 190302108

PENGESAHAN
SKRIPSI
PENGARUH TEKNIK ENSEMBLE TERHADAP AKURASI PADA
ALGORITMA KLASIFIKASI

yang dipersiapkan dan disusun oleh

Alvin Aditya Rahman

16.11.0579

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Agustus 2020
Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dr. Andi Sunyoto, M.Kom
NIK. 190302052

Majid Rahardi, S.Kom, M.Eng
NIK. 190302393

Mardhiya Hayaty, S.T., M.Kom
NIK. 190302108

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Agustus 2020

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.

NIK. 190302038

PERNYATAAN

Saya yang bertanda tangan di bawah ini menyatakan, skripsi ini merupakan karya saya sendiri (ASLI) dan isi didalam skripsi tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu Instansi Pendidikan dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis / diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 21 Agustus 2020

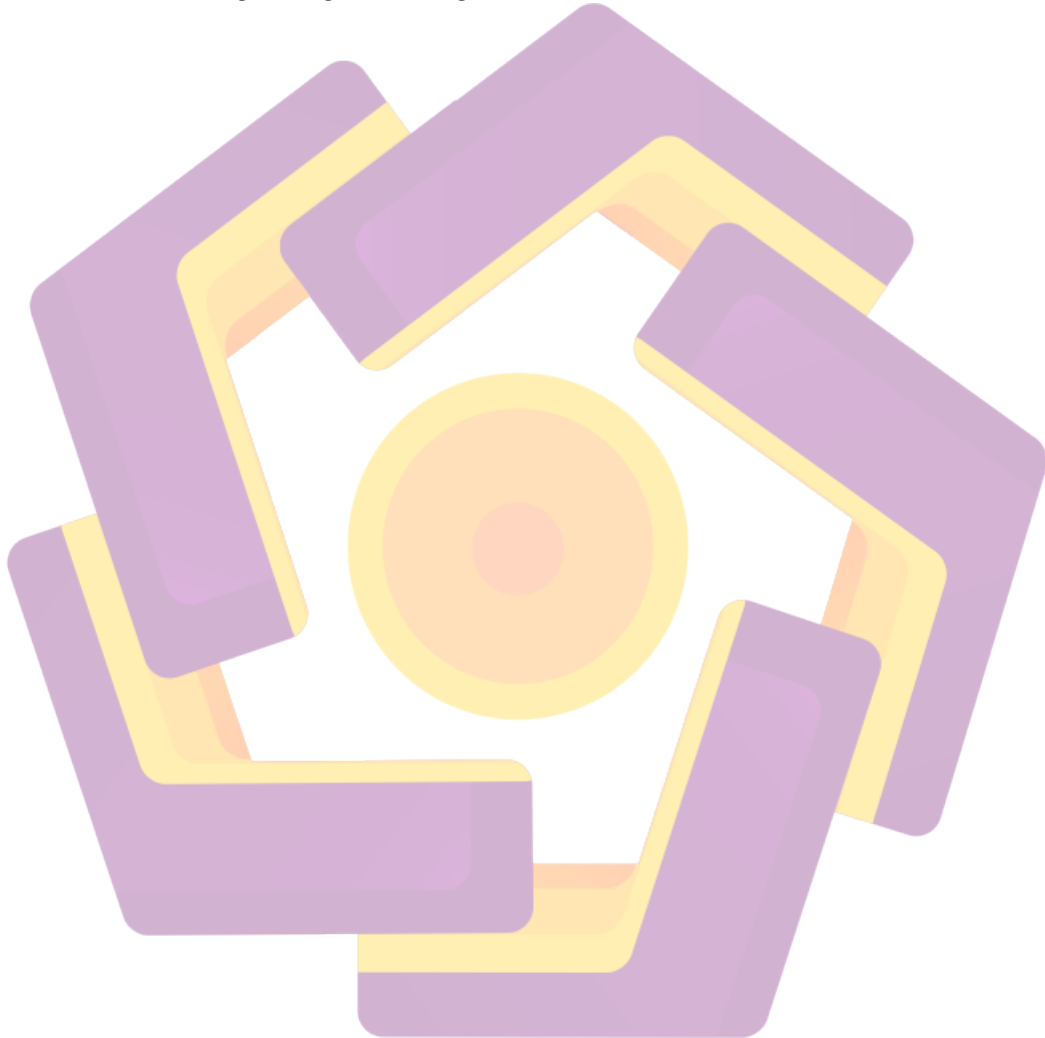


Alvin Aditya Rahman
NIM 16.11.0579

MOTTO

“Allah tidak akan membebani seseorang melainkan sesuai dengan kesanggupannya. ...” - QS. Al-Baqarah: 286

“Imung deung teba tar ganu dala reta lero wulan” - Black Finit

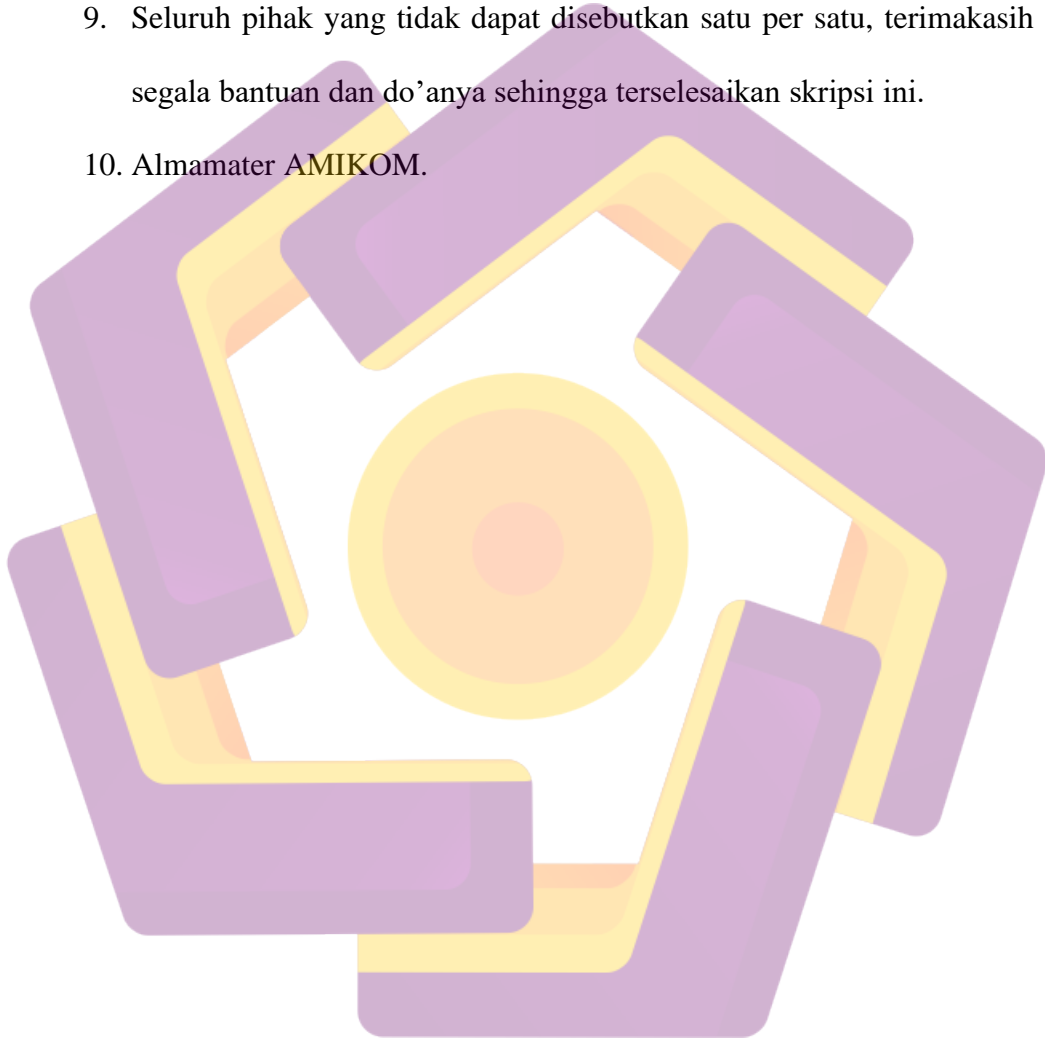


HALAMAN PERSEMBAHAN

Puji Syukur saya panjatkan kepada Allah SWT, karena berkat rahmat dan karunia-Nya sehingga saya dapat menyelesaikan skripsi ini. Tulisan sederhana ini tak lepas dari orang-orang terkasih yang selalu memberikan doa dan dukungan dalam segala bentuk. Untuk itu tulisan ini saya persembahkan untuk:

1. Kedua pahlawan terhebat yang tak lain adalah kedua orang tuaku, Bapak Hirmayanto Suhir dan Ibu Nursoimah yang tanpa lelah memberikan doa, dukungan, semangat, motivasi dan kasih sayang yang tiada ujungnya.
2. Kakakku Hilda Nuraeni Makrufah yang selalu memberikan dukungan, semangat dan selalu mengisi hari-hariku dengan canda tawa dan kasih sayangnya.
3. Ibu Mardhiya Hayaty, S.T., M.Kom selaku dosen pembimbing sekaligus dosen wali, terimakasih banyak atas bimbingannya sehingga skripsi ini dapat terselesaikan.
4. Teman-teman "Daily New Jokes" dan "Gedandangan Crew" yang selalu memberikan kata-kata sambat dan motivasi setiap hari.
5. Teman-teman kelas Informatika-09 yang selalu mendukung dan memberi bantuan selama 4 tahun ini. Terimakasih atas kebersamaan yang terjalin selama ini.
6. Kontrakan muslim dan PSU 4.2.1 yang selalu mewarnai hari-hariku, sehingga hariku lebih menyenangkan.

7. Teman-teman gabut yang selalu menemani saat tidak ada inspirasi dalam mengerjakan skripsi ini.
8. *Google, Stackoverflow*, dan forum-forum sejenisnya yang telah banyak berjasa dalam proses pengerjaan skripsi ini.
9. Seluruh pihak yang tidak dapat disebutkan satu per satu, terimakasih atas segala bantuan dan do'anya sehingga terselesaikan skripsi ini.
10. Almamater AMIKOM.



KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan Rahmat-Nya, sehingga skripsi yang disusun sebagai salah satu syarat kelulusan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta ini dapat terselesaikan dengan baik.

Pada kesempatan ini dengan segala ketulusan, keikhlasan serta kerendahan hati penulis ingin mengucapkan banyak berterima kasih yang sebesar-besarnya dan penghargaan setinggi-tingginya kepada semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini. Dalam penulisan laporan skripsi ini, peneliti banyak mendapatkan bantuan serta semangat dari berbagai pihak. Untuk itu peneliti menyampaikan rasa hormat, dan terimakasih kepada:

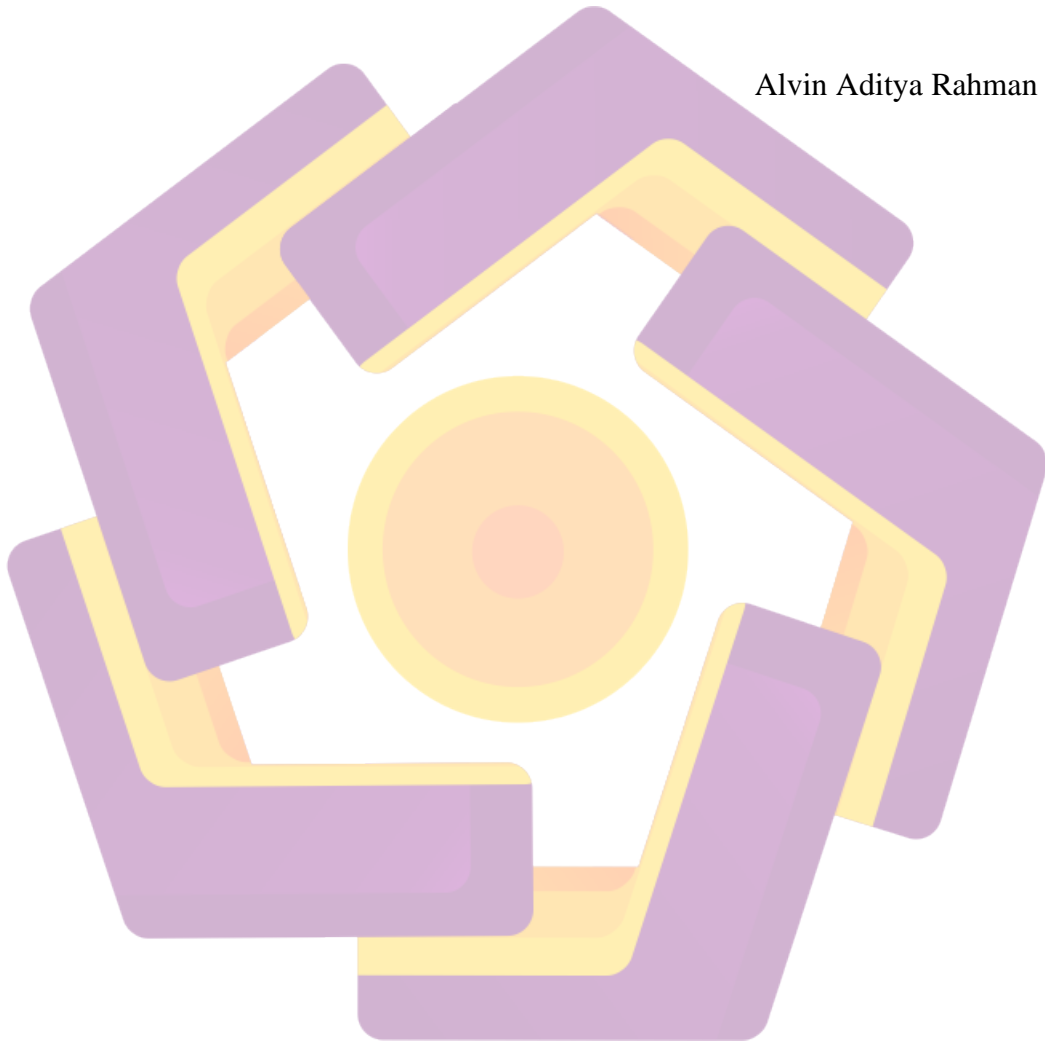
1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Sudarmawan, MT. selaku Ketua Program Studi S1 Informatika.
3. Ibu Mardhiya Hayati, S.T, M.Kom. selaku Dosen Pembimbing, yang telah memberikan pengarahan yang sangat membantu dalam proses pembuatan skripsi ini.
4. Semua pihak yang telah membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu per satu.

Peneliti juga memohon maaf kepada semua pihak jika dalam pelaksanaan dan penulisan laporan skripsi ini terdapat kesalahan atau hal yang kurang berkenan,

semua tidak lepas karena keterbatasan peneliti. Semoga karya sederhana ini dapat bermanfaat bagi kita semua. Aamiin.

Yogyakarta, 21 Agustus 2020

Alvin Aditya Rahman

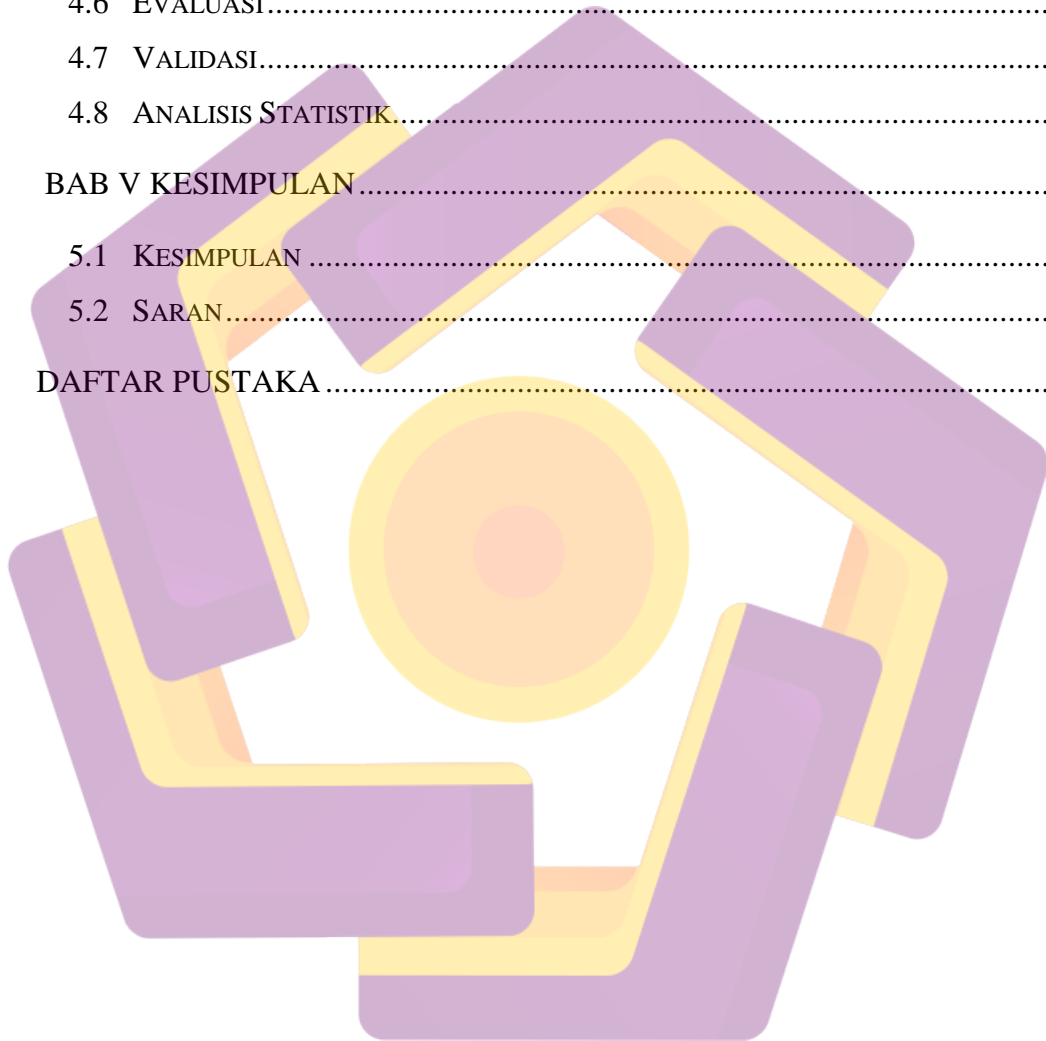


DAFTAR ISI

JUDUL	I
PERSETUJUAN.....	II
PENGESAHAN.....	III
PERNYATAAN	IV
MOTTO.....	V
HALAMAN PERSEMBAHAN.....	V
KATA PENGANTAR	VIII
DAFTAR ISI.....	X
DAFTAR TABEL	XIII
INTISARI.....	XVI
ABSTRACT	XVII
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH	2
1.3 HIPOTESIS PENELITIAN	2
1.4 BATASAN MASALAH	2
1.5 MAKSUD DAN TUJUAN PENELITIAN	3
1.6 MANFAAT PENELITIAN	3
1.7 METODE PENELITIAN	3
1.7.1 Metode Pengumpulan Dataset.....	3
1.7.2 Pre-processing	3
1.7.3 Implementasi Algoritma Klasifikasi.....	4
1.7.4 Evaluasi Hasil Klasifikasi	4
1.7.5 Uji Statistik.....	4
1.8 SISTEMATIKA PENULISAN.....	4
BAB II LANDASAN TEORI	6

2.1	TINJAUAN PUSTAKA	6
2.2	DASAR TEORI.....	9
2.2.1	Data Mining	9
2.2.2	Preprocessing Data	9
2.2.3	Klasifikasi	9
2.2.3.1	Naïve Bayes.....	10
2.2.3.2	Support Vector Machine	11
2.2.4	Imbalanced Data.....	16
2.2.5	Ensemble Learning	17
2.2.5.1	Bagging	18
2.2.5.2	Boosting	19
2.2.6	Evaluasi.....	21
2.2.7	Validasi	22
2.2.8	Uji Statistik.....	23
BAB III METODE PENELITIAN		26
3.1	ALAT DAN BAHAN.....	26
3.1.1	Alat Penelitian	26
3.1.2	Bahan Penelitian.....	26
3.2	ALUR PENELITIAN.....	27
3.2.1	Pengumpulan Dataset	28
3.2.2	Preprocessing	29
3.2.3	Implementasi Algoritma	29
3.2.4	Evaluasi.....	29
3.2.5	Validasi	30
3.2.6	Analisis Statistik.....	30
BAB IV ANALISIS DAN PEMBAHASAN		31
4.1	PENGUMPULAN DATASET	31
4.2	PENDEFINISIAN DATASET	32
4.3	PREPROCESSING	32
4.4	DATA SPLITTING	34

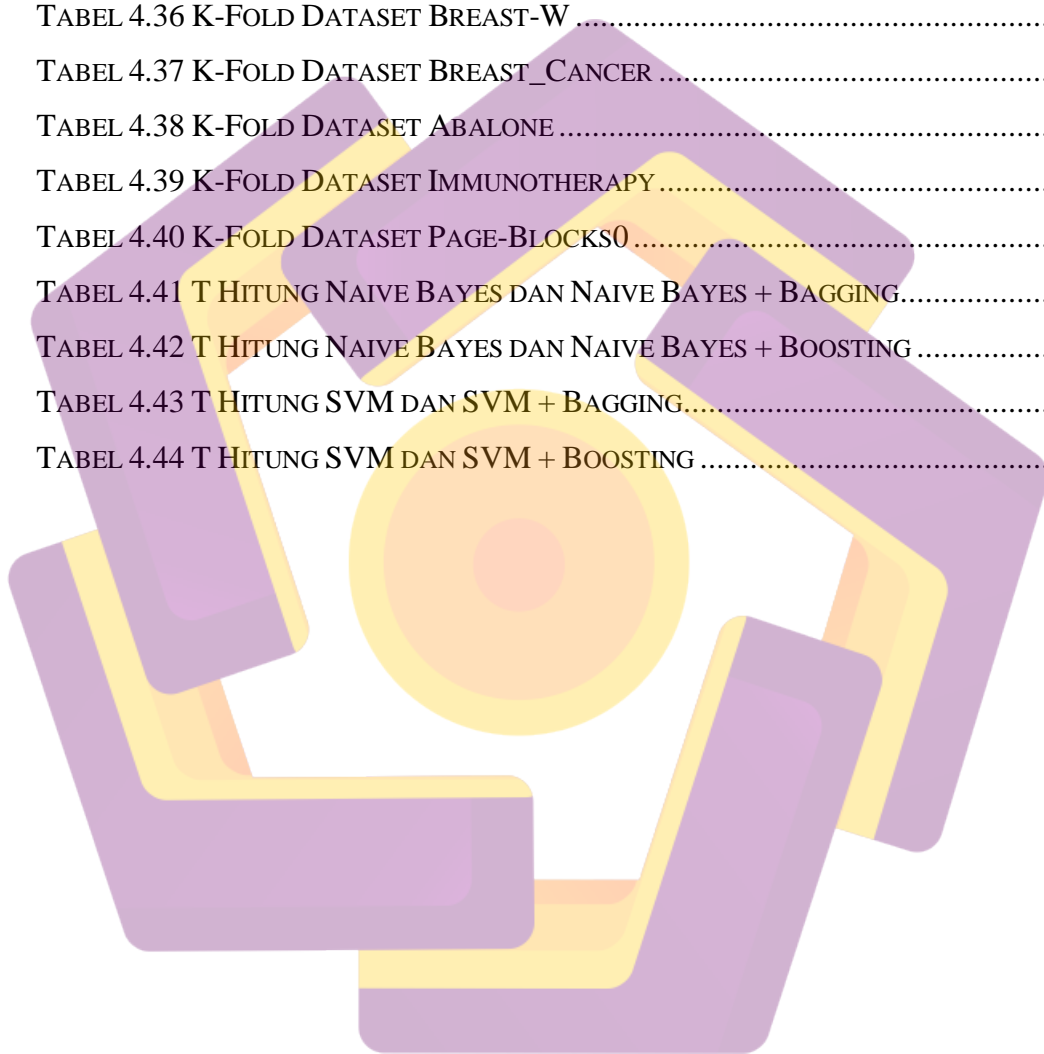
4.5	IMPLEMENTASI ALGORITMA	35
4.5.1	Naïve Bayes	35
4.5.2	Support Vector Machine	38
4.5.3	Bagging	40
4.5.4	Boosting (Adaboost).....	43
4.6	EVALUASI.....	48
4.7	VALIDASI.....	54
4.8	ANALISIS STATISTIK.....	59
BAB V KESIMPULAN		66
5.1	KESIMPULAN	66
5.2	SARAN.....	66
DAFTAR PUSTAKA		67



DAFTAR TABEL

TABEL 2.1 HASIL PERBANDINGAN PENELITIAN.....	7
TABEL 2.2 CONFUSION MATRIX.....	22
TABEL 4.3 KETERANGAN DATASET.....	31
TABEL 4.4 TABEL PENGELOMPOKAN KELAS 0.....	35
TABEL 4.5 TABEL PENGELOMPOKAN KELAS 1.....	36
TABEL 4.6 SAMPEL DATASET SVM.....	38
TABEL 4.7 TABEL MATRIKS KERNEL LINEAR.....	38
TABEL 4.8 TABEL DATASET PROSES BAGGING.....	41
TABEL 4.9 BOOTSTRAP 1.....	41
TABEL 4.10 BOOTSTRAP 2.....	41
TABEL 4.11 BOOTSTRAP 3.....	41
TABEL 4.12 HASIL PREDIKSI BOOTSTRAP 1.....	42
TABEL 4.13 HASIL PREDIKSI BOOTSTRAP 2.....	42
TABEL 4.14 HASIL PREDIKSI BOOTSTRAP 3.....	42
TABEL 4.15 TABEL HASIL BAGGING.....	42
TABEL 4.16 DATA TRAINING ADABOOST.....	43
TABEL 4.17 PROBABILITAS NILAI BOBOT AWAL.....	44
TABEL 4.18 NILAI PREDIKSI UNTUK $H_1(x)$	44
TABEL 4.19 HASIL UPDATE DISTRIBUSI $H_1(x)$	45
TABEL 4.20 NILAI PREDIKSI UNTUK $H_2(x)$	45
TABEL 4.21 HASIL UPDATE DISTRIBUSI $H_2(x)$	46
TABEL 4.22 NILAI PREDIKSI UNTUK $H_3(x)$	46
TABEL 4.23 HASIL UPDATE DISTRIBUSI $H_3(x)$	47
TABEL 4.24 CONFUSION MATRIX DATASET DIABETES.....	48
TABEL 4.25 HASIL CONFUSION MATRIX.....	50
TABEL 4.26 PERBANDINGAN AKURASI.....	50
TABEL 4.27 PERBANDINGAN SENSITIVITAS.....	51
TABEL 4.28 PERBANDINGAN PRESISI.....	52
TABEL 4.29 PERBANDINGAN SPECIFICITY.....	52
TABEL 4.30 PERBANDINGAN ERROR RATE.....	53

TABEL 4.31 K-FOLD DATASET DIABETES	54
TABEL 4.32 K-FOLD DATASET GLASS0.....	55
TABEL 4.33 K-FOLD DATASET WISCONSIN.....	55
TABEL 4.34 K-FOLD DATASET POKER-9_vs_7	56
TABEL 4.35 K-FOLD DATASET HTRU2.....	56
TABEL 4.36 K-FOLD DATASET BREAST-W	57
TABEL 4.37 K-FOLD DATASET BREAST_CANCER	57
TABEL 4.38 K-FOLD DATASET ABALONE	58
TABEL 4.39 K-FOLD DATASET IMMUNOTHERAPY	58
TABEL 4.40 K-FOLD DATASET PAGE-BLOCKS0	59
TABEL 4.41 T HITUNG NAIVE BAYES DAN NAIVE BAYES + BAGGING.....	59
TABEL 4.42 T HITUNG NAIVE BAYES DAN NAIVE BAYES + BOOSTING	61
TABEL 4.43 T HITUNG SVM DAN SVM + BAGGING.....	62
TABEL 4.44 T HITUNG SVM DAN SVM + BOOSTING	63



DAFTAR GAMBAR

GAMBAR 2.1 ILUSTRASI METODE SVM.....	12
GAMBAR 2.2 MAXIMUM MARGIN HYPERPLANE.....	13
GAMBAR 2.3 CLASS OVERLAPPING	17
GAMBAR 2.4 SMALL DISJUNCTS	17
GAMBAR 2.5 ILUSTRASI K-FOLD CROSS VALIDATION	23
GAMBAR 3.6 ALUR PENELITIAN	27
GAMBAR 4.7 SCRIPT PENDEFINISIAN DATASET.....	32
GAMBAR 4.8 DATASET DENGAN DATA KOSONG.....	32
GAMBAR 4.9 SCRIPT DROP CLASS TARGET.....	33
GAMBAR 4.10 SCRIPT HANDLING MISSING DATA.....	33
GAMBAR 4.11 HASIL PEMBERSIHAN DATA KOSONG.....	34
GAMBAR 4.12 SCRIPT DATA SPLITTING.....	34
GAMBAR 4.13 SCRIPT NAIVE BAYES	37
GAMBAR 4.14 SCRIPT SUPPORT VECTOR MACHINE	40
GAMBAR 4.15 SCRIPT BAGGING DENGAN NAIVE BAYES	43
GAMBAR 4.16 SCRIPT BAGGING DENGAN SVM	43
GAMBAR 4.17 SCRIPT ADABOOST DENGAN NAIVE BAYES.....	47
GAMBAR 4.18 SCRIPT ADABOOST DENGAN SVM.....	48
GAMBAR 4.19 SCRIPT CONFUSION MATRIX	48
GAMBAR 4.20 SCRIPT K-FOLD CROSS VALIDATION.....	54

INTISARI

Klasifikasi dengan imbalanced dataset adalah topik penelitian populer dalam machine learning dan data mining, karena hampir semua dataset memiliki distribusi kelas yang miring atau biasa disebut imbalanced data. Namun saat ini masih sedikit penelitian tentang penanganan imbalance data terutama menggunakan teknik ensemble. Sehingga diperlukan penelitian lebih lanjut untuk menentukan metode terbaik.

Pada penelitian ini, peneliti mencoba menganalisa pengaruh teknik ensemble terhadap akurasi pada algoritma klasifikasi dengan 10 dataset imbalance yang berbeda. Algoritma klasifikasi yang digunakan pada penelitian ini yaitu naive bayes dan support vector machine. Adapun teknik ensemble yang digunakan pada penelitian ini yaitu bagging dan boosting. Proses analisis dilakukan dengan membandingkan akurasi algoritma saat sebelum dan sesudah diterapkan teknik ensemble.

Dengan langkah penelitian yang sudah dilakukan, didapatkan hasil perbandingan akurasi algoritma. Dari semua dataset yang digunakan terdapat peningkatan akurasi pada enam dataset saat menerapkan boosting pada algoritma naive bayes. Setelah didapatkan hasil perbandingan akurasi dilanjutkan uji signifikansi dengan hasil H_0 diterima dan H_1 ditolak yang artinya tidak ada peningkatan yang signifikan saat diterapkan teknik ensemble terhadap akurasi algoritma klasifikasi naive bayes dan support vector machine.

Kata Kunci: data mining, klasifikasi, naive bayes, support vector machine, imbalance dataset, teknik ensemble, bagging, boosting.

ABSTRACT

Classification with an imbalanced dataset is a popular research topic in machine learning and data mining because almost all datasets have a sloping class distribution or so-called imbalanced data. However, there is still little research on handling imbalanced data, especially using ensemble techniques. So that further research is needed to decide the best method.

In this study, the researcher tried to analyze the effect of the ensemble technique on the accuracy of the classification algorithm with 10 different imbalance datasets. The classification algorithm used in this study is a naïve bayes and support vector machine. The ensemble technique used in this research is bagging and boosting. The analysis process is carried out by comparing the accuracy of the algorithm before and after the ensemble technique is applied.

With the research steps that have been carried out, the results of the algorithm accuracy comparison are got. Of all the datasets used, there is an increase in the accuracy of the six datasets when applying boosting to the Naive Bayes algorithm. After obtaining the results of the comparison of the accuracy, the significance test was continued with the results of H_0 being accepted and H_1 was rejected, so there was no significant increase when the ensemble technique applied to the accuracy of the naïve Bayes classification algorithm and the support vector machine.

Keywords: *data mining, classification, naïve bayes, support vector machine, imbalance dataset, ensemble technique, bagging, boosting.*