



**PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN
EXTREME GRADIENT BOOSTING (XGBOOST) PADA
ANALISA SENTIMEN TERHADAP KEPUASAN KINERJA
TIMNAS SEPAK BOLA INDONESIA**

SKRIPSI

HAFIEF MAULANA AZIZ

41519010163

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA JAKARTA**

2022

HALAMAN PENYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Hafief Maulana Aziz

NIM : 41519010163

Program Studi : Teknik Informatika

Judul Laporan Skripsi : Perbandingan Algoritma *Naïve Bayes* dan Extreme Gradient Boosting (XGBoost) Pada Analisa Sentimen Terhadap Kepuasan Kinerja Timnas Sepak Bola Indonesia

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

UNIVERSITAS
MERCU BUANA

Jakarta, 23 Desember 2022



Hafief Maulana Aziz

HALAMAN PENGESAHAN

Saya yang bertanda tangan di bawah ini:

Nama : Hafief Maulana Aziz
NIM : 41519010163
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Perbandingan Algoritma *Naïve Bayes* dan Extreme Gradient Boosting (XGBoost) Pada Analisa Sentimen Terhadap Kepuasan Kinerja Timnas Sepak Bola Indonesia

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Eugenius Kau Suni, ST, MT
NIDN : 0312018204
Ketua Penguji : Sabar Rudiarto, M.Kom
NIDN : 0309036902
Penguji 1 : Umniy Salamah, S.Kom, MMSI
NIDN : 0306098104
Penguji 2 : Yaya Sudarya Triana, Ph.D.
NIDN : 0016016404



Jakarta, 24 Januari 2023

Mengetahui,

Koordinator Tugas



Wawan Gunawan, S.Kom., M.T.

Ketua Program Studi



Ir. Emil R. Kaburuan, Ph.D., IP

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Hafief Maulana Aziz

NIM : 41519010163

Program Studi : Teknik Informatika

Judul Tugas Akhir : Perbandingan Algoritma *Naïve Bayes* dan Extreme Gradient Boosting (XGBoost) Pada Analisa Sentimen Terhadap Kepuasan Kinerja Timnas Sepak Bola Indonesia

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS
MERCU BUANA

Jakarta, 23 Desember 2022



Hafief Maulana Aziz

ABSTRAK

Nama : Hafief Maulana Aziz
NIM : 41519010163
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Perbandingan Algoritma *Naïve Bayes* dan Extreme Gradient Boosting (XGBoost) Pada Analisa Sentimen Terhadap Kepuasan Kinerja Timnas Sepak Bola Indonesia
Pembimbing : Eugenius Kau Suni, ST, MT

Sepak Bola menjadi olahraga nomor satu di dunia dan tidak terkecuali di Indonesia sendiri, dan penilaian terhadap sebuah tim Sepak Bola sangatlah penting dilakukan untuk menjadi parameter utama dalam melakukan evaluasi untuk kedepannya. PSSI selaku induk tertinggi Sepak Bola Indonesia hanya mengevaluasi kinerja tim dari tingkat prestasinya dan pengamatan dari orang-orang terpilih PSSI yaitu Anggota Exco PSSI. Kurangnya keterlibatan teknologi informasi seperti Machine Learning dalam penilaian kinerja tim Sepak Bola menjadi permasalahan sendiri bagi PSSI. Sejauh ini pendukung Sepak Bola Indonesia belum dilibatkan dalam penilaian kinerja Tim Nasional Indonesia. Padahal penilaian pendukung Sepak Bola bisa menjadi alternatif parameter dalam mengevaluasi kinerja Tim Nasional Indonesia. Jadi dilakukannya penelitian ini adalah untuk mencari metode alternatif dalam melakukan evaluasi terhadap Timnas Sepak Bola Indonesia dengan pemanfaatan teknologi *Machine Learning*. Untuk mencari algoritma *Machine Learning* yang terbaik dilakukan perbandingan dari algoritma yang dipakai. Dalam penelitian ini dilakukan perbandingan algoritma *Naïve Bayes* dan *Extreme Gradient Boosting (XGBoost)* pada analisis sentimen cuitan netizen Indonesia terhadap kinerja Timnas Sepak Bola Indonesia. Dengan menggunakan beberapa metode *Natural Language Processing (NLP)* untuk teks processing dan fitur ekstrasi untuk pembobotan kata, didapat hasil yang cukup baik dari kedua buah model algoritma yang dipakai dengan nilai rata-rata akurasi mencapai $\pm 85\%$, dimana algoritma *XGBoost* memiliki akurasi yang lebih baik dengan nilai rata-rata akurasi 85,48% ketimbang *Naïve Bayes* yang memiliki nilai rata-rata akurasi 84,20%.

Katakunci : analisis sentimen, *Naïve Bayes*, xgboost, sepak bola, dan indonesia

ABSTRACT

Nama : Hafief Maulana Aziz
NIM : 41519010163
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Perbandingan Algoritma *Naïve Bayes* dan Extreme Gradient Boosting (XGBoost) Pada Analisa Sentimen Terhadap Kepuasan Kinerja Timnas Sepak Bola Indonesia
Pembimbing : Eugenius Kau Suni, ST, MT

Football is the number one sport in the world and Indonesia itself is no exception, and the assessment of a football team is very important to be carried out to become the main parameter in evaluating for the future. PSSI as the supreme parent of Indonesian Football only evaluates team performance from the level of achievement and observations from PSSI elected people, namely PSSI Exco Members. The lack of involvement of information technology such as Machine Learning in assessing the performance of soccer teams is a problem for PSSI. So far, Indonesian Football supporters have not been involved in evaluating the performance of the Indonesian National Team. Even though the assessment of football supporters can be an alternative parameter in evaluating the performance of the Indonesian National Team. So this research was conducted to find alternative methods in evaluating the Indonesian Football National Team by utilizing Machine Learning technology. To find the best Machine Learning algorithm, a comparison of the algorithms used is carried out. In this study, a comparison of the Naïve Bayes and Extreme Gradient Boosting (XGBoost) algorithms was carried out in the sentiment analysis of Indonesian netizens' tweets on the performance of the Indonesian Football National Team. By using several Natural Language Processing (NLP) methods for text processing and extraction features for word weighting, good results are obtained from the two algorithm models used with an average accuracy value of $\pm 85\%$, where the XGBoost algorithm has higher accuracy. better with an average accuracy value of 85.48% compared to Naïve Bayes which has an average accuracy value of 84.20%.

Keywords: *sentiment analisys, Naïve Bayes, xgboost, football, and indonesia*

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Skripsi ini. Penulisan Laporan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Kedua orang tua yang telah memberikan support serta kasih sayang hingga saya bisa menyelesaikan tugas akhir ini.
2. Bapak Prof. Dr. Ir. Andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana.
3. Bapak Yaya Sudarya Triana, M. Kom., Ph.D selaku Dekan Fakultas Ilmu Komputer dan Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
4. Bapak Ir. Emil R. Kaburuan, Ph.D., IPM selaku Ketua Program Studi
5. Bapak Eugenius Kau Suni, ST, MT selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini.
6. Bapak Sabar Rudiarto, M.Kom selaku Dosen Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
7. Ibu Umniy Salamah, S.Kom, MMSI selaku Dosen Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
8. Teman-teman Teknik Informatika 2019, sahabat dan kerabat yang selalu memberikan support dan waktu untuk selalu mengingatkan dalam menyelesaikan tugas akhir ini.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 20 Desember 2022



Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1. 1 Latar Belakang	1
1. 2 Perumusan Masalah.....	3
1. 3 Tujuan Penelitian.....	3
1. 4 Manfaat Penelitian.....	3
1. 5 Batasan Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2. 1 Penelitian Terdahulu	5
2. 2 Teori Pendukung	8
BAB III METODE PENELITIAN	10
3. 1 Jenis Penelitian.....	10
3. 2 Tahapan Penelitian	10
3. 3 Dataset	11
3. 4 Pelabelan Data.....	11
3. 5 <i>Pre-Processing</i>	12
3. 6 Penyeimbangan Data.....	14
3. 7 Ekstrasi Fitur	15
3. 8 Algoritma Klasifikasi	15
3. 9 Pengujian.....	17
BAB IV HASIL DAN PEMBAHASAN	19
4. 1 Dataset.....	19
4. 2 Analisis Hasil	22
BAB V KESIMPULAN DAN SARAN	28
5. 1 Kesimpulan.....	28

5. 2	Saran.....	28
DAFTAR PUSTAKA		29
LAMPIRAN		31



DAFTAR TABEL

TABEL 1. PENELITIAN TERKAIT	5
TABEL 2. KAMUS LEXICON.....	12
TABEL 3. CONTOH PRE-PROCESSING.....	14
TABEL 5. NAIVE BAYES - COUNT VECTORIZER (70:30)	23
TABEL 6 NAIVE BAYES - COUNT VECTORIZER (80:20)	23
TABEL 7. NAIVE BAYES - TF-IDF (70:30).....	23
TABEL 8. NAIVE BAYES - TF-IDF (80:20).....	24
TABEL 9. XGBOOST - COUNT VECTORIZER (70:30).....	24
TABEL 10. XGBOOST - COUNT VECTORIZER (80:20).....	24
TABEL 11. XGBOOST - TF-IDF (70:30)	25
TABEL 12. XGBOOST - TF-IDF (80:20)	25
TABEL 13. RATA - RATA AKURASI.....	26



DAFTAR GAMBAR

GAMBAR 1. TAHAPAN PENELITIAN.....	11
GAMBAR 2. SAMPEL RAW DATA.....	19
GAMBAR 3. SAMPEL PROCESSED DATA	20
GAMBAR 4. SAMPEL DATA UNTUK TESTING	20
GAMBAR 5. FREKUENSI KEMUNCULAN KATA PADA DATASET	21
GAMBAR 6. WORDCLOUD KATA PADA DATA POSITIF	22
GAMBAR 7. WORDCLOUD KATA PADA DATA NEGATIF.....	22
GAMBAR 8. GRAFIK BATANG NILAI KESELURUHAN ALGORITMA YANG DIPAKAI	26



DAFTAR LAMPIRAN

LAMPIRAN 1 . BIMBINGAN.....	31
LAMPIRAN 2 . LUARAN TUGAS AKHIR	32
LAMPIRAN 3 . BUKTI SUBMIT / PUBLISHED ARTIKEL ILMIAH / HKI.	33
LAMPIRAN 4 . NASKAH ARTIKEL JURNAL.....	34
LAMPIRAN 5 . CURICULUM VITAE.....	44
LAMPIRAN 6 . SP HAK CIPTA	45
LAMPIRAN 7 . SPH HAK CIPTA	46
LAMPIRAN 8 . KARTU TANDA PENDUDUK.....	47

