

**PERBANDINGAN AKURASI METODE MACHINE  
LEARNING UNTUK SENTIMEN ANALISIS  
PEMBAYARAN DIGITAL**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Sistem Informasi



disusun oleh

**Angga Arindra Shonta**

**22.22.2481**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2023**

**PERBANDINGAN AKURASI METODE MACHINE  
LEARNING UNTUK SENTIMEN ANALISIS  
PEMBAYARAN DIGITAL**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Sistem Informasi



disusun oleh

**Angga Arindra Shonta**

**22.22.2481**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2023**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**Perbandingan Akurasi Metode Machine Learning untuk  
Sentimen Analisis Pembayaran Digital**

yang disusun dan diajukan oleh

**Angga Arindra Shonta**

**22.22.2481**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 19 Oktober 2023

Dosen Pembimbing,



**Yuli Astuti, M. Kom**

**NIK. 190302146**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**Perbandingan Akurasi Metode Machine Learning untuk  
Sentimen Analisis Pembayaran Digital**

yang disusun dan diajukan oleh

**Angga Ariendra Shonta**

**22.22.2481**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 19 Oktober 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Yoga Pristyanto, S.Kom., M.Eng**  
**NIK. 190302412**

**Irma Rofni Wulandari, S.Pd., M.Eng**  
**NIK. 190302329**

**Yuli Astuti, M.Kom**  
**NIK. 190302146**

Skrripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 19 Oktober 2023

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom.**  
**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Angga Arindra Shonta  
NIM : 22.22.2481

Menyatakan bahwa Skripsi dengan judul berikut:

### Perbandingan Akurasi Metode Machine Learning untuk Sentimen Analisis Pembayaran Digital

Dosen Pembimbing : Yuli Astuti, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Oktober 2023

Yang Menyatakan,



Angga Arindra Shonta

## HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam dan telah diselesaikan skripsi ini penulis mempersembahkannya kepada :

1. Allah SWT yang telah melimpahkan rahmat dan hidayahnya sehingga skripsi ini dapat diselesaikan.
2. Bapak Prof. Dr. M. Suyanto, MM selaku rektor Universitas Amikom Yogyakarta.
3. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Anggit Dwi Hartanto, M. Kom selaku ketua prodi Sistem Informasi Universitas Amikom Yogyakarta.
5. Ibu Yuli Astuti, M. Kom selaku dosen pembimbing yang telah memandu dan mengarahkan dalam penyusunan skripsi ini.
6. Bapak Ibu yang selalu mendukung baik moral maupun materi
7. Adik dan seluruh anggota keluarga yang selalu mendukung dan memberikan bantuan.
8. Serta semua pihak yang tidak bisa penulis sebutkan satu persatu.

## KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas kehendak-Nya maka penelitian dan penulisan skripsi dengan judul “PERBANDINGAN AKURASI METODE MACHINE LEARNING UNTUK SENTIMEN ANALISIS PEMBAYARAN DIGITAL” ini dapat diselesaikan dengan baik. Penulis berharap penelitian ini dapat bermanfaat dalam upaya mendukung perkembangan sentimen analisis pada pembayaran digital. Penulis mengharapkan masukan yang konstruktif agar dapat menyempurnakan penelitian ini. Secara khusus penulis menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. M. Suyanto, MM selaku rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Anggit Dwi Hartanto, M. Kom selaku ketua prodi Sistem Informasi Universitas Amikom Yogyakarta.
4. Ibu Yuli Astuti, M.Kom. sebagai dosen pembimbing yang telah memberikan bimbingan, masukan, dan diskusi yang intensif.
5. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu selama penulis kuliah.

Yogyakarta, 19 Oktober 2023

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL .....	x
DAFTAR GAMBAR.....	xi
INTISARI .....	xii
ABSTRACT .....	xiii
<b>BAB I</b>	
<b>PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
1.5.1 Bagi Mahasiswa .....	3
1.5.2 Bagi Pembaca .....	3
1.6 Sistematika Penulisan .....	4
<b>BAB II</b>	
<b>TINJAUAN PUSTAKA .....</b>	<b>5</b>
2.1 Studi Literatur .....	5
2.2 Sentimen Analisis .....	8
2.3 Pembayaran Digital.....	8
2.4 Twitter .....	8
2.5 Machine Learning .....	8
2.6 Crawling.....	8
2.7 API Twitter .....	9
2.8 Bahasa Python.....	9
2.9 Pra-Pemrosesan Data .....	9
2.9.1 Cleaning .....	9
2.9.2 Translation .....	10
2.9.3 Lemmatize.....	10



2.9.4 Tokenize.....	10
2.9.5 Remove Stopwords.....	10
2.10 Pembobotan Kata & Normalisasi TF-IDF.....	11
2.11 Proses Klasifikasi.....	12
2.11.1 Logistic Regression.....	12
2.11.2 Stochastic Gradient Descent Classifier.....	12
2.11.3 Tensorflow.....	13
2.11.4 Keras.....	13
2.12 Evaluasi Model.....	13
2.13 Word Cloud.....	13
2.14 Confusion Matrix.....	14
2.15 Analisis Hasil.....	14
<b>BAB III</b>	
<b>METODE PENELITIAN.....</b>	<b>15</b>
3.1 Objek Penelitian.....	15
3.2 Alur Penelitian.....	15
3.2.1 Pengumpulan Dataset.....	16
3.2.2 Pra-Pemrosesan Data.....	17
3.2.2.1 Cleaning.....	17
3.2.2.2 Translation.....	18
3.2.2.3 Lemmatize.....	18
3.2.2.4 Tokenize.....	18
3.2.3 Master Data.....	19
3.2.4 Perhitungan Bobot dan Normalisasi Kata.....	19
3.2.5 Pelatihan dan Pengujian.....	21
3.2.6 Evaluasi Model.....	22
3.2.6.1 Confusion Matrix.....	22
3.2.6.2 Validasi Model.....	23
<b>BAB IV</b>	
<b>HASIL DAN PEMBAHASAN.....</b>	<b>24</b>
4.1 Pengumpulan Dataset.....	24
a. Library Tweepy.....	24
b. Install Library.....	24
c. Import Library.....	25
d. API Twitter.....	25
e. Crawling.....	25
f. Library Snsrape.....	26
g. Install Library.....	26

4.2	Pra-Pemrosesan Data .....	27
a.	Cleaning .....	27
b.	Menghapus Mention Username, Retweet, dan Simbol.....	27
c.	Menghapus Stock Market Tickers, Hashtags, dan Angka .....	28
d.	Menghapus Data Duplikat .....	29
e.	Translation.....	30
f.	Polarity dan Subjectivity .....	31
g.	Lemmatize.....	32
h.	Tokenize.....	33
i.	Remove Stopwords .....	34
4.3	Master Data .....	34
4.4	Perhitungan Bobot dan Normalisasi Kata .....	35
4.5	Hasil Sentimen Analisis.....	36
a.	Tensorflow .....	38
b.	Logistic Regression.....	38
c.	Stochastic Gradient Descent Classifier .....	39
4.6	Evaluasi dan Validasi.....	40
a.	Visualisasi Grafik.....	40
b.	Word Cloud.....	41
4.7	Hasil Model.....	42
<b>BAB V</b>		
<b>PENUTUP.....</b>		<b>47</b>
5.1	Kesimpulan .....	47
5.2	Saran .....	47
<b>REFERENSI.....</b>		<b>48</b>

## DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian .....	6
Tabel 3.1 Data Tweets dari Twitter .....	17
Tabel 3.2 Hasil Proses Cleaning .....	18
Tabel 3.3 Hasil Proses Translate.....	18
Tabel 3.4 Hasil Proses Lemmatize.....	18
Tabel 3.5 Hasil Proses Tokenize.....	18
Tabel 3.6 Hasil Proses Remove Stopwords.....	19
Tabel 3.7 Master Data.....	19
Tabel 3.8 Dokumen.....	20
Tabel 3.9 Kata Kunci .....	20
Tabel 3.10 Proses Perhitungan Bobot.....	21
Tabel 3.11 Hasil Perhitungan Normalisasi Bobot.....	21
Tabel 3.12 Hasil Pelatihan dan Pengujian .....	22
Tabel 3.13 Hasil Validasi Model.....	23
Tabel 4.24 Hasil Perbandingan Akurasi Metode Machine Learning.....	46

## DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian .....	16
Gambar 3.2 Confusion Matrix .....	23
Gambar 4.1 Halaman Twitter Developer .....	24
Gambar 4.2 Dataset Hasil Crawling .....	27
Gambar 4.3 Hasil Cleaning Proses Menghapus Mention Username, Retweet, dan Simbol .....	28
Gambar 4.4 Hasil Cleaning Proses Menghapus Stock Market Tickers, Hashtags, dan Angka .....	29
Gambar 4.5 Hasil Cleaning Proses Menghapus Data Duplikat .....	30
Gambar 4.6 Hasil Translate .....	31
Gambar 4.7 Hasil Polarity dan Subjectivity .....	32
Gambar 4.8 Hasil Lemmatize .....	33
Gambar 4.10 Hasil Remove Stopwords .....	34
Gambar 4.11 Master Data .....	35
Gambar 4.13 Hasil Pengujian Model Menggunakan Tensorflow (Keras) .....	37
Gambar 4.14 Hasil Prediksi Sentimen Menggunakan Tensorflow .....	38
Gambar 4.15 Hasil Prediksi Sentimen Menggunakan Logistic Regression .....	39
Gambar 4.16 Hasil Prediksi Sentimen Menggunakan Stochastic Gradient Descent Classifier .....	40
Gambar 4.17 Visualisasi Grafik Sentimen Analisis Tensorflow, Logistic Regression, dan Stochastic Gradient Descent Classifier .....	41
Gambar 4.18 Hasil Perbandingan Sentimen Analisis Pembayaran Digital .....	41
Gambar 4.19 Word Cloud Sentimen Analisis Pembayaran Digital .....	42
Gambar 4.20 Hasil Perubahan Menggunakan Label Encoder .....	43

## INTISARI

Penelitian ini membandingkan tiga metode klasifikasi yaitu Logistic Regression, Stochastic Gradient Descent Classifier, dan Tensorflow pada analisis sentimen data pembayaran digital yang diambil dari Twitter. Tujuan penelitian adalah untuk menemukan metode klasifikasi terbaik dalam analisis sentimen pada data pembayaran digital di Twitter. Metode penelitian ini menggunakan matrik evaluasi standar, termasuk akurasi, precision, recall, dan F1 score, untuk mengukur kinerja ketiga metode klasifikasi. Hasil penelitian menunjukkan bahwa metode Logistic Regression memiliki nilai akurasi 95.4%, recall 96.9%, dan F1 score tertinggi dengan nilai 90.9%, sedangkan metode Stochastic Gradient Descent Classifier memiliki nilai precision tertinggi dengan nilai 94.7%. Dari evaluasi tersebut, dapat disimpulkan bahwa metode Logistic Regression adalah metode klasifikasi terbaik untuk analisis sentimen pada data pembayaran digital di Twitter. Namun, perlu dicatat bahwa hasil ini didapatkan dari dataset yang digunakan dalam penelitian ini. Hasil analisis mungkin akan berbeda jika menggunakan dataset yang berbeda. Oleh karena itu, hasil penelitian ini sebaiknya dianggap sebagai panduan awal, dan diperlukan penelitian lanjutan untuk memvalidasi keunggulan metode Logistic Regression dalam berbagai konteks dan dataset yang berbeda.

**Kata kunci:** Analisis Sentimen, Data Pembayaran Digital, Twitter, Metode Klasifikasi, Logistic Regression, Stochastic Gradient Descent Classifier, Tensorflow.

## ABSTRACT

This study compares three classification methods, namely Logistic Regression, Stochastic Gradient Descent Classifier, and Tensorflow, on sentiment analysis of digital payment data taken from Twitter. The aim of the research is to find the best classification method for sentiment analysis of digital payment data on Twitter. The research methodology utilizes standard evaluation metrics, including accuracy, precision, recall, and F1 score, to measure the performance of these three classification methods. The research results indicate that the Logistic Regression method achieved an accuracy of 95.4%, a recall of 96.9%, and the highest F1 score of 90.9%, while the Stochastic Gradient Descent Classifier method demonstrated the highest precision rate at 94.7%. From this evaluation, it can be concluded that the Logistic Regression method is the best classification method for sentiment analysis of digital payment data on Twitter. However, it is important to note that these results are specific to the dataset used in this study. Analysis outcomes may differ when applying different datasets. Therefore, these research findings should be considered as preliminary guidance, and further research is needed to validate the effectiveness of the Logistic Regression method in various contexts and with different datasets.

**Keyword:** Sentiment Analysis, Digital Payment Data, Twitter, Classification Methods, Logistic Regression, Stochastic Gradient Descent Classifier, Tensorflow.