

**PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN  
*SUPPORT VECTOR MACHINE (SVM)* PADA ANALISIS  
SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER  
TERHADAP PT ESTEH INDONESIA MAKMUR**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**SITI SUHARYANTI**

**19.11.3064**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN  
*SUPPORT VECTOR MACHINE (SVM)* PADA ANALISIS  
SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER  
TERHADAP PT ESTEH INDONESIA MAKMUR**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**SITI SUHARYANTI**

**19.11.3064**

Kepada

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

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE (SVM)* PADA ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP PT ESTEH INDONESIA MAKMUR**

yang disusun dan diajukan oleh

**Siti Suharyanti**

**19.11.3064**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 18 Oktober 2023

Dosen Pembimbing,



Anna Baita, M.Kom

NIK. 190302290

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE (SVM)* PADA ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP PT ESTEH INDONESIA MAKMUR**

yang disusun dan diajukan oleh

**Siti Suharyanti**

**19.11.3064**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 18 Oktober 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Anna Baita, M.Kom**  
**NIK. 190302290**

**Dina Maulina, M.Kom**  
**NIK. 190302250**

**Nuri Cahyona, M.Kom**  
**NIK. 190302278**

**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 18 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 : Siti Suharyanti  
NIM : 19.11.3064

Menyatakan bahwa Skripsi dengan judul berikut:

**Perbandingan Algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) Pada Analisis Sentimen Pengguna Media Sosial Twitter Terhadap PT Esteh Indonesia Makmur**

Dosen Pembimbing : Anna Baita, 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, 18 Oktober 2023

Yang Menyatakan,



Siti Suharyanti

## HALAMAN PERSEMBAHAN

Puji dan syukur peneliti panjatkan kepada Allah SWT yang telah melimpahkan rahmat, hidayah, dan karunia-Nya sehingga peneliti dapat menyelesaikan skripsi ini. Oleh karena itu, peneliti ingin mempersembahkan skripsi ini untuk:

1. Diri sendiri yang mau belajar hal baru, berjuang melawan rasa malas, tidak menyerah, tetap fokus pada progres sendiri, dan bisa melawan tekanan yang ada sehingga skripsi ini dapat selesai.
2. Orang tua peneliti, Bapak Suyono dan Ibu Rina Apriyanti yang telah memberikan dukungan baik moral maupun finansial dan selalu mendoakan peneliti.
3. Adik peneliti, Lila Ramadhani yang sudah menjadi teman, penghibur, dan mendengarkan keluh kesah peneliti selama mengerjakan skripsi.
4. Keluarga peneliti, yang selalu memberikan dukungan dan doa untuk peneliti.
5. Teman SMA peneliti, Wafa, Dewi, dan Yogie yang selalu memberikan semangat, memotivasi, mengajak peneliti main, dan mengobrol sehingga dapat mengurangi rasa bosan dan lelah peneliti saat mengerjakan skripsi.
6. Teman kuliah peneliti, Fadillah dan Nita yang telah menjadi teman seperjuangan peneliti selama 4 tahun, menjadi penghibur, memberikan semangat, memotivasi, dan selalu membantu peneliti.
7. Max Verstappen, yang selalu menang dan memberikan hiburan kepada peneliti agar tidak bosan.

## KATA PENGANTAR

Puji dan syukur peneliti panjatkan kepada Allah SWT yang telah melimpahkan rahmat, hidayah, dan karunia-Nya sehingga peneliti dapat menyelesaikan skripsi yang berjudul “Perbandingan Algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* Pada Analisis Sentimen Pengguna Media Sosial Twitter Terhadap PT Esteh Indonesia Makmur” untuk memenuhi salah satu syarat mencapai derajat Sarjana Program Studi Informatika di Universitas Amikom Yogyakarta.

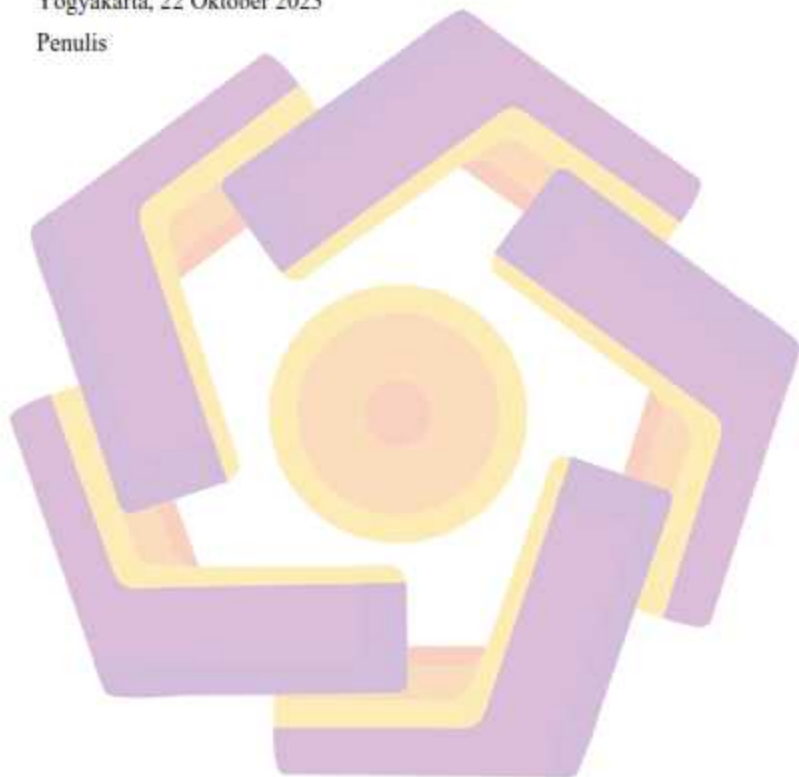
Selama mengerjakan skripsi ini, peneliti mendapat bantuan, masukan, dan dukungan dari berbagai pihak. Untuk itu peneliti ingin menyampaikan ucapan terima kasih kepada:

1. Allah SWT atas berkat dan rahmat-Nya peneliti dapat menyelesaikan pembuatan skripsi ini.
2. Bapak Hanif Al-Fatta, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Windha Mega Pradnya Duhita, M.Kom., selaku Ketua Program Studi SI Informatika.
4. Ibu Anna Baita, M.Kom., selaku Dosen Pembimbing yang telah memberikan bimbingan, ilmu, saran, dan arahan kepada peneliti demi terselesainya skripsi ini.
5. Bapak Arif Akbarul Huda, S.Si., M.Eng., selaku Dosen Wali yang telah memberikan bimbingan dan membantu peneliti selama masa studi.
6. Seluruh dosen dan staf akademik Fakultas Ilmu Komputer yang telah memberikan ilmu dan bantuannya selama proses perkuliahan.
7. Bapak, Ibu, Adik, dan keluarga yang telah memberikan dukungan moral maupun finansial.
8. Teman dekat peneliti saat SMA dan kuliah yang selalu memberikan bantuan, dukungan, dan motivasi kepada peneliti.
9. Seluruh pihak terkait yang telah membantu peneliti yang tidak dapat disebutkan satu per satu.

Peneliti menyadari sepenuhnya bahwa skripsi ini masih jauh dari sempurna, untuk itu semua jenis saran, kritik dan masukan yang bersifat membangun sangat peneliti harapkan. Semoga penelitian ini dapat memberikan manfaat dan wawasan tambahan bagi para pembaca dan khususnya bagi peneliti sendiri.

Yogyakarta, 22 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.....	viii
DAFTAR TABEL .....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN .....	xii
DAFTAR ISTILAH .....	xiii
INTISARI .....	xiv
ABSTRACT .....	xv
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	4
1.5.1 Manfaat Teoritis.....	4
1.5.2 Manfaat Praktis .....	4
1.6 Sistematika Penulisan .....	4
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>6</b>
2.1 Studi Literatur.....	6
2.2 Dasar Teori .....	14
2.2.1 PT Esteh Indonesia Makmur.....	14
2.2.2 Twitter.....	14
2.2.3 Data Mining .....	14
2.2.4 Analisis Sentimen .....	15
2.2.5 Pre-Processing.....	16
2.2.5.1 Cleansing.....	16
2.2.5.2 Case Folding .....	17
2.2.5.3 Normalization .....	17
2.2.5.4 Tokenizing .....	17
2.2.5.5 Stopword Removal .....	17
2.2.5.6 Stemming .....	18
2.2.6 TF-IDF .....	18
2.2.7 Oversampling SMOTE.....	19
2.2.8 Hyperparameter Tuning.....	19
2.2.9 Algoritma Naïve Bayes .....	22
2.2.9.1 Multinomial Naïve Bayes .....	23
2.2.10 Algoritma Support Vector Machine (SVM) .....	23

2.2.11	<i>Confusion Matrix</i> .....	26
2.2.12	<i>Wordclouds</i> .....	31
<b>BAB III METODE PENELITIAN</b> .....		32
3.1	Objek Penelitian .....	32
3.2	Alur Penelitian .....	32
3.2.1	<i>Scraping Data</i> .....	33
3.2.2	<i>Labeling</i> .....	34
3.2.3	<i>Pre-processing Data</i> .....	35
3.2.3.1	<i>Cleansing</i> .....	36
3.2.3.2	<i>Case Folding</i> .....	37
3.2.3.3	<i>Normalization</i> .....	38
3.2.3.4	<i>Tokenizing</i> .....	39
3.2.3.5	<i>Stopword Removal</i> .....	40
3.2.3.6	<i>Stemming</i> .....	41
3.2.4	Ekstraksi Fitur TF-IDF .....	42
3.2.5	<i>Splitting Data</i> .....	42
3.2.6	<i>Oversampling SMOTE</i> .....	43
3.2.7	Klasifikasi Model <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> .....	44
3.2.8	Evaluasi .....	44
3.3	Alat dan Bahan .....	45
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....		46
4.1	Visualisasi Data <i>Labeling</i> .....	46
4.2	Visualisasi Data <i>Train</i> .....	46
4.3	Model <i>Naïve Bayes</i> .....	47
4.3.1	<i>Naïve Bayes</i> Tanpa <i>Hyperparameter Tuning</i> .....	47
4.3.2	<i>Naïve Bayes</i> Dengan <i>Hyperparameter Tuning</i> .....	49
4.4	Model <i>Support Vector Machine</i> .....	52
4.4.1	<i>Support Vector Machine</i> Tanpa <i>Hyperparameter Tuning</i> .....	52
4.4.2	<i>Support Vector Machine</i> Dengan <i>Hyperparameter Tuning</i> .....	54
4.5	Perbandingan Performa Model .....	57
4.6	<i>Wordclouds</i> .....	60
<b>BAB V PENUTUP</b> .....		64
5.1	Kesimpulan .....	64
5.2	Saran .....	64
<b>REFERENSI</b> .....		65

## DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian .....	9
Tabel 3.1 Contoh data <i>scraping</i> .....	33
Tabel 3.2 Contoh proses <i>labeling</i> .....	34
Tabel 3.3 Contoh proses <i>cleansing</i> .....	36
Tabel 3.4 Contoh proses <i>case folding</i> .....	37
Tabel 3.5 Contoh kamus kata .....	38
Tabel 3.6 Contoh proses <i>normalization</i> .....	38
Tabel 3.7 Contoh proses <i>tokenizing</i> .....	39
Tabel 3.8 Contoh proses <i>stop-word removal</i> .....	40
Tabel 3.9 Contoh proses <i>stemming</i> .....	41
Tabel 3.10 Hasil <i>splitting</i> data .....	42
Tabel 3.11 Hasil <i>oversampling</i> SMOTE .....	43
Tabel 4.1 Skenario <i>naïve bayes</i> tanpa <i>hyperparameter tuning</i> .....	47
Tabel 4.2 Hasil pengujian <i>naïve bayes</i> tanpa <i>hyperparameter tuning</i> .....	48
Tabel 4.3 Skenario <i>naïve bayes</i> dengan <i>hyperparameter tuning</i> .....	49
Tabel 4.4 Hasil pengujian <i>naïve bayes</i> dengan <i>hyperparameter tuning</i> .....	50
Tabel 4.5 Skenario SVM tanpa <i>hyperparameter tuning</i> .....	52
Tabel 4.6 Hasil pengujian SVM tanpa <i>hyperparameter tuning</i> .....	53
Tabel 4.7 Skenario SVM dengan <i>hyperparameter tuning</i> .....	54
Tabel 4.8 Hasil pengujian SVM dengan <i>hyperparameter tuning</i> .....	55
Tabel 4.9 Persentase kata pada <i>wordclouds</i> sentimen positif .....	61
Tabel 4.10 Persentase kata pada <i>wordclouds</i> sentimen negatif .....	62

## DAFTAR GAMBAR

Gambar 2.1	<i>Hyperplane linear SVM</i>	(Sumber: <a href="https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machine-algorithm5.png">https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machine-algorithm5.png</a> )	24
Gambar 2.2	<i>Hyperplane non-linear SVM</i>	(Sumber: <a href="https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machine-algorithm9.png">https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machine-algorithm9.png</a> )	24
Gambar 2.3	<i>Confusion matrix</i>		27
Gambar 2.4	<i>Kurva ROC</i>	(Sumber: <a href="https://miro.medium.com/v2/resize:fit:828/0*mo8tN4x5jSwvncZQ">https://miro.medium.com/v2/resize:fit:828/0*mo8tN4x5jSwvncZQ</a> )	29
Gambar 2.5	<i>Kurva PR</i>	(Sumber: <a href="https://www.statology.org/wp-content/uploads/2021/09/precisionRecall2-768x612.png">https://www.statology.org/wp-content/uploads/2021/09/precisionRecall2-768x612.png</a> )	30
Gambar 2.6	<i>Learning curve</i>	(Sumber: <a href="https://scikit-learn.org/0.18/images/sphx_glr_plot_learning_curve_002.png">https://scikit-learn.org/0.18/images/sphx_glr_plot_learning_curve_002.png</a> )	30
Gambar 2.7	<i>Visualisasi wordclouds</i>	(Sumber: <a href="https://raw.githubusercontent.com/jb0hn/wordcloud-example/master/python.png">https://raw.githubusercontent.com/jb0hn/wordcloud-example/master/python.png</a> )	31
Gambar 3.1	<i>Alur Penelitian</i>		32
Gambar 3.2	<i>Alur pre-processing data</i>		36
Gambar 3.3	<i>Hasil TF-IDF</i>		42
Gambar 4.1	<i>Grafik dataset sebelum splitting</i>		46
Gambar 4.2	<i>a) Grafik data train sebelum SMOTE b) Grafik data train setelah SMOTE</i>		47
Gambar 4.3	<i>Confusion matrix naïve bayes tanpa hyperparameter tuning</i>		48
Gambar 4.4	<i>Learning curve naïve bayes tanpa hyperparameter tuning</i>		49
Gambar 4.5	<i>Confusion matrix naïve bayes dengan hyperparameter tuning</i>		51
Gambar 4.6	<i>Learning curve naïve bayes dengan hyperparameter tuning</i>		51
Gambar 4.7	<i>Confusion matrix SVM tanpa hyperparameter tuning</i>		53
Gambar 4.8	<i>Learning curve SVM tanpa hyperparameter tuning</i>		54
Gambar 4.9	<i>Confusion matrix SVM dengan hyperparameter tuning</i>		56
Gambar 4.10	<i>Learning curve SVM dengan hyperparameter tuning</i>		56
Gambar 4.11	<i>Kurva AUC-ROC semua model</i>		57
Gambar 4.12	<i>Wordclouds sentimen positif</i>		60
Gambar 4.13	<i>Wordclouds sentimen negatif</i>		62

## DAFTAR LAMBANG DAN SINGKATAN



NB	<i>Naïve Bayes</i>
SVM	<i>Support Vector Machine</i>
Warganet	Warga Internet
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
PSO	<i>Particle Swarm Optimization</i>
ASCII	<i>American Standard Code for Information Interchange</i>
NLTK	<i>Natural Language Toolkit</i>
AUC	<i>Area Under the Curve</i>
TF-IDF	<i>Term Frequency Inverse Document Frequency</i>
NLP	<i>Natural Language Processing</i>
KBBI	Kamus Besar Bahasa Indonesia
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
URL	<i>Uniform Resource Locator</i>
ROC	<i>Receiver Operating Characteristic</i>
PR	<i>Precision-Recall</i>

## DAFTAR ISTILAH

C	Mengontrol penalti yang dikenakan pada kesalahan klasifikasi dalam pembentukan model
K	Jumlah tetangga terdekat yang akan digunakan dalam klasifikasi atau regresi.
Gamma	Mempengaruhi sejauh mana pengaruh dari setiap sampel dalam pembentukan margin
Alpha	Parameter yang menentukan tingkat <i>smoothing</i> pada probabilitas kelas
Coef0	Mengontrol sejauh mana pengaruh fungsi <i>non-linear</i> akan mempengaruhi hasil pemetaan data
Degree	Mengontrol derajat polinomial yang digunakan untuk transformasi data
Fit_prior	Mengontrol apakah model akan memperhitungkan probabilitas prior dari kelas-kelas yang ada dalam data pelatihan
Kernel	Fungsi matematis yang digunakan SVM untuk menghitung "skalar" produk antara pasangan data.
Instance	Data

## INTISARI

Twitter merupakan salah satu media sosial yang banyak digunakan oleh masyarakat Indonesia, lewat twitter pengguna dapat berbagi informasi, berkomunikasi dengan teman atau kerabat, dan mengetahui topik yang sedang trending atau ramai dibicarakan baik di Indonesia maupun dunia. Pada 25 September 2022 PT Esteh Indonesia Makmur melayangkan surat somasi kepada salah satu pengguna twitter dengan *username* @ghandoyy karena tweet-nya yang mengkritik salah satu produk minuman dari perusahaan tersebut. Dalam tweet-nya, ghandoyy mengkritik bahwa minuman chizu red velvet memiliki rasa yang terlalu manis, tidak lupa ia menambahkan kata-kata kasar pada tweet tersebut. Baik kritik ghandoyy maupun surat somasi dari PT Esteh Indonesia Makmur mengundang banyak reaksi dari warganet, seperti masyarakat biasa, *public figure*, tenaga kesehatan, maupun kompetitor. Dibutuhkan analisis sentimen untuk mengetahui kecenderungan respon tersebut dan mengklasifikasikannya menjadi sentimen positif dan negatif. Metode yang digunakan dalam penelitian ini adalah *Naïve Bayes* dan *Support Vector Machine*, masing-masing model akan menerapkan *hyperparameter tuning* dan tanpa *hyperparameter tuning*. *Naïve Bayes* akan menggunakan parameter *alpha* dan *fit\_prior*, sedangkan *Support Vector Machine* akan menggunakan parameter *kernel*, *C*, *gamma*, *coef0*, dan *degree*. Hasil penelitian menunjukkan bahwa SVM tanpa *hyperparameter tuning* memiliki performa yang lebih baik daripada NB, dengan akurasi 90,39%, presisi 78,91%, *recall* 59,17%, *f1-score* 62,51%, AUC-ROC 0,85, dan AUC-PR 0,42.

**Kata kunci:** Analisis Sentimen, PT Esteh Indonesia Makmur, *Naïve Bayes*, *Support Vector Machine*, *Hyperparameter Tuning*.

## ABSTRACT

Twitter is one of the social media that is widely used by Indonesians, through twitter users can share information, communicate with friends or relatives, and find out topics that are trending or being discussed both in Indonesia and the world. On September 25, 2022 PT Esteh Indonesia Makmur sent a subpoena to one of the twitter users with the username @ghandooy because of his tweet criticizing one of the company's beverage products. In his tweet, ghandooy criticized that the chizu red velvet drink had a taste that was too sweet, not forgetting that he added harsh words to the tweet. Both ghandooy's criticism and the subpoena from PT Esteh Indonesia Makmur invited many reactions from netizens, such as ordinary people, public figures, health workers, and competitors. Sentiment analysis is needed to determine the tendency of these responses and classify them into positive and negative sentiments. The methods used in this research are Naïve Bayes and Support Vector Machine, each model will apply hyperparameter tuning and without hyperparameter tuning. Naïve Bayes will use alpha and fit prior parameters, while Support Vector Machine will use kernel, C, gamma, coef0, and degree parameters. The results show that SVM without hyperparameter tuning has better performance than NB, with 90.39% accuracy, 78.91% precision, 59.17% recall, 62.51% f1-score, 0.85 AUC-ROC, and 0.42 AUC-PR.

**Keyword:** Sentiment Analysis, PT Esteh Indonesia Makmur, Naïve Bayes, Support Vector Machine, Hyperparameter Tuning.