

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN PENGGUNA  
PROVIDER TELKOMSEL PADA MEDIA SOSIAL TWITTER  
MENGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER**

**SKRIPSI**



disusun oleh

**Wahyu Prihanto**

**18.21.1132**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2019**

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN PENGGUNA  
PROVIDER TELKOMSEL PADA MEDIA SOSIAL TWITTER  
MENGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Wahyu Prihanto**

**18.21.1132**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2019**

**PERSETUJUAN**

**SKRIPSI**

**ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN  
PENGGUNA PROVIDER TELKOMSEL PADA MEDIA SOSIAL  
TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES  
CLASSIFIER**

yang dipersiapkan dan disusun oleh

**Wahyu Prihanto**

**18.21.1132**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 26 Maret 2019

**Dosen Pembimbing,**



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

## PENGESAHAN

### SKRIPSI

#### ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN PENGGUNA PROVIDER TELKOMSEL PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER

Yang dipersiapkan dan disusun oleh

**Wahyu Prihanto**

**18.21.1132**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 21 November 2019

#### Susunan Dewan Penguji

**Nama Penguji**

**Tanda Tangan**

**Anggit Dwi Hartanto, M.Kom**  
NIK. 190302163



**Wiwi Widayani, M.Kom**  
NIK. 190302272



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



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
tanggal 29 November 2019



**DEKAN FAKULTAS ILMU KOMPUTER**

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

## PERNYATAAN

Saya yang bertanda tangan 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 atau kelompok lain untuk memperoleh gelar akademis di suatu Institusi Pendidikan, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain atau kelompok lain, kecuali yang secara tertulis sebagai acuan dalam naskah ini dan disebutkan dalam daftar pustaka.

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

Yogyakarta, 26 November 2019



Wahyu Prihanto

NIM. 18.21.1132

## **MOTTO**

“Lakukan dan kerjakan sesuai dengan apa yang kamu inginkan”

“Manusia hanya bisa merencanakan dan berusaha, Allah yang menentukan”

“Merendahkan maka kamu akan ditinggikan, dan meninggikan maka kamu akan direndahkan”



## PERSEMBAHAN

Dengan rasa penuh rasa syukur karya ini penulis persembahkan untuk :

- Allah SWT, yang memberikan hidup dan memegang matinya setiap makhluk tanpa-Nya tulisan ini tiada bermakna.
- Bapak dan Ibu saya tercinta, yang selalu memberikan kasih sayang, do'a, semangat, pengorbanan, dukungan, dan kesabaran.
- Situs [stackoverflow.com](https://stackoverflow.com) dan [github](https://github.com) yang senantiasa hadir membantu setiap masalah dalam penelitian ini.
- Ibu Krisnawati, S.Si, MT sebagai pembimbing yang selalu memberikan arahan, dukungan, dan ilmu didalam menjalankan penelitian ini.
- Teman-teman kelas 18 IF Transfer yang sudah menjadi teman yang baik selama ini.
- Teman-teman dekat yang selalu memberikan support dan dukungan.
- Wahyu Prihanto atau diri saya sendiri. Saya mengapresiasi diri saya sendiri atas selesainya penelitian ini.

## KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala limpahan rahmat dan hidayah-Nya serta shalawat serta salam penulis curahkan kepada junjungan kita Nabi Muhammad SAW sehingga penulis dapat menyelesaikan skripsi dengan judul “ANALISIS SENTIMEN TERHADAP TINGKAT KEPUASAN PENGGUNA PROVIDER TELKOMSEL PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER” dengan baik.

Skripsi ini disusun guna memenuhi salah satu persyaratan untuk mendapatkan gelar Sarjana pada jurusan Informatika di Universitas Amikom Yogyakarta.

Dalam penyusunan skripsi ini, penulis sangat menyadari masih banyak kekurangan. Selain itu, terselesainya penyusunan skripsi ini adalah berkat bantuan dan bimbingan dari berbagai pihak baik secara langsung maupun tidak langsung. Oleh karena itu penulis ingin menyampaikan rasa terimakasih kepada :

1. Allah SWT atas segala berkah, hidayah bimbingan dan keridhoan ilmunya sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik.
2. Nabi Agung Muhammad SAW yang selalu menjadi panutan dan suri tauladan.
3. Kedua orang tua dan penulis sebagai sosok luar biasa yang memberikan segala dukungan dan semangat baik moril maupun material.



4. Bapak Prof. Dr. M. Suyanto, M.M., Rektor Universitas Amikom Yogyakarta, yang telah memberikan kebijaksanaan akademik kepada penulis dalam menyelesaikan tugas akhir.
5. Ibu Krisnawati, S.Si, MT selaku dosen pembimbing penulis, yang telah memberikan bimbingan, arahan dan masukan, sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
6. Seluruh staff pengajar dan karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu, bimbingan, arahan dan masukan sehingga penulis dapat memperoleh ilmu yang bermanfaat.
7. Teman – teman satu angkatan 18 IF Transfer yang telah mendukung dan memberikan semangat dalam penyelesaian tugas akhir ini.
8. Semua pihak yang telah membantu dalam pelaksanaan dan penyusunan tugas akhir ini.

Semoga segala bantuan dan dukungan yang telah diberikan bermanfaat untuk penulis maupun pembaca.

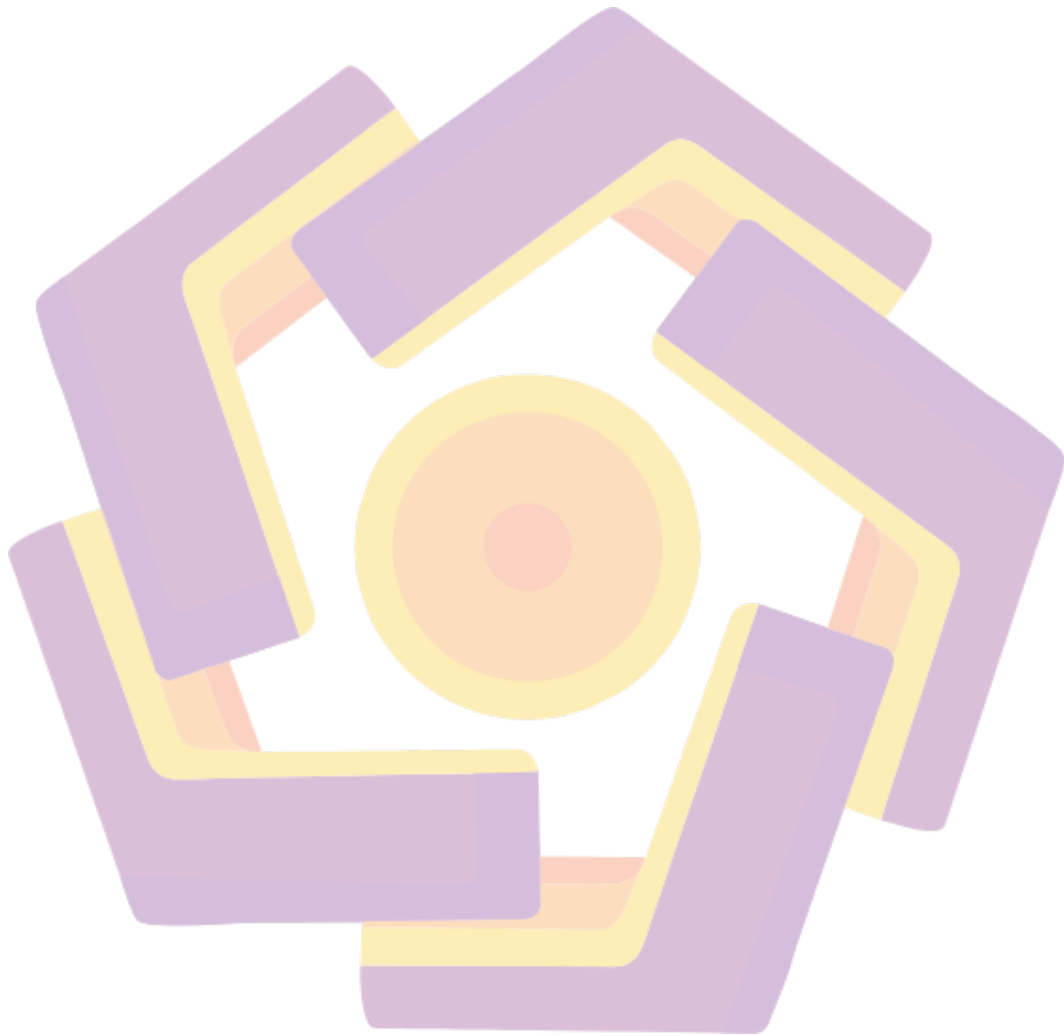
## DAFTAR ISI

HALAMAN JUDUL.....	ii
PERSETUJUAN .....	iii
PENGESAHAN .....	iv
PERNYATAAN.....	v
MOTTO .....	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR .....	xvi
INTISARI.....	xviii
<i>ABSTRACT</i> .....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penulisan .....	3
1.6 Metodologi Penelitian .....	4
1.7 Sistematika Penulisan.....	6
BAB II LANDASAN TEORI.....	8
2.1 Tinjauan Pustaka .....	8
2.2 Pengertian Data Mining.....	10

2.3	Pengertian Analisa Sentimen.....	14
2.4	Pengertian Text Mining.....	15
2.5	Pre-Processing .....	16
2.6	Part Of Speech Tagging (POS Tagging) .....	19
2.6.1	Stanford POS Tagger .....	21
2.7	Term Frequency – Inverse Document Frequency (TF-IDF).....	22
2.8	Cosine Similarity.....	24
2.9	Naïve Bayes Classifier .....	25
2.10	Pengujian.....	26
2.11	Validasi dan Evaluasi .....	27
2.12	Flowchart.....	29
<b>BAB III PERANCANGAN .....</b>		<b>32</b>
3.1	Analisis Sistem.....	32
3.1.1	Analisis Kebutuhan Fungsional .....	32
3.1.2	Analisis Kebutuhan Non-Fungsional .....	32
3.2	Gambaran Umum Sistem .....	33
3.3	Analisis Sentimen dengan Algoritma Naïve Bayes Classifier .....	37
3.3.1	Pengumpulan Data .....	37
3.3.2	Pelabelan Tweet .....	38
3.3.3	Pre-processing.....	40
3.3.4	Training.....	59
3.4	Perancangan Antarmuka Pengguna.....	76
3.4.1	Perancangan Antarmuka Halaman Utama .....	76
3.4.2	Perancangan Antarmuka Hasil Prediksi.....	77
3.5	Perancangan Pengumpulan Data Twitter .....	77

3.6	Perancangan File Training.....	78
BAB IV IMPLEMENTASI .....		80
4.1	Deskripsi Implementasi .....	80
4.2	Implementasi Pengunduhan Data Tweet.....	80
4.3	Implementasi Pelabelan Data .....	81
4.4	Implementasi Preprocessing.....	84
4.4.1	Tweet Cleaning .....	84
4.4.2	Tokenize.....	85
4.4.3	Slang Replacement.....	85
4.4.4	POS Tagging.....	87
4.4.5	POS Filtering.....	87
4.4.6	Stemming .....	88
4.4.7	Feature List.....	88
4.5	Implementasi TF-IDF Cosine Similarity.....	90
4.6	Implementasi Training Naïve Bayes Classifier.....	92
4.7	Implementasi Antarmuka Prediksi .....	94
4.8	Implementasi Antarmuka Hasil Prediksi.....	95
4.9	Evaluasi Model.....	96
4.9.1	Penerapan POS Filtering.....	98
4.9.2	Evaluasi Model Terhadap Hasil Pelatihan .....	99
4.9.3	Evaluasi Model Terhadap Prediksi .....	103
BAB V KESIMPULAN .....		107
5.1	Kesimpulan.....	107
5.2	Saran.....	109
DAFTAR PUSTAKA .....		111

LAMPIRAN..... 1



## DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terkait Sistem Yang Dibangun.....	9
Tabel 2.2 Ilustrasi Tweet Cleaning .....	17
Tabel 2.3 Ilustrasi Tweet Case Folding dari Cleaning.....	17
Tabel 2.4 Ilustrasi Tweet Tokenization dari Case Folding .....	18
Tabel 2.5 Ilustrasi Tweet Stemming dari Tokenization.....	18
Tabel 2.6 Ilustrasi POS Tagging .....	19
Tabel 2.7 Tagset.....	20
Tabel 2.8 Confussion Matrix .....	28
Tabel 2.9 Simbol-Simbol pada Flowchart .....	30
Tabel 3.1 Tweet yang Telah Dikumpulkan dari Twitter .....	37
Tabel 3.2 Tweet Berlabel.....	39
Tabel 3.3 Tweet Cleaning .....	41
Tabel 3.4 Tokenize.....	43
Tabel 3.5 Normalisasi Kata.....	45
Tabel 3.6 Remove Stopword.....	47
Tabel 3.7 POS Tagging.....	49
Tabel 3.8 POS Filtering .....	53
Tabel 3.9 Stemming .....	55
Tabel 3.10 Feature List .....	57
Tabel 3.11 Vektor Feature.....	58
Tabel 3.12 Term Positive .....	61

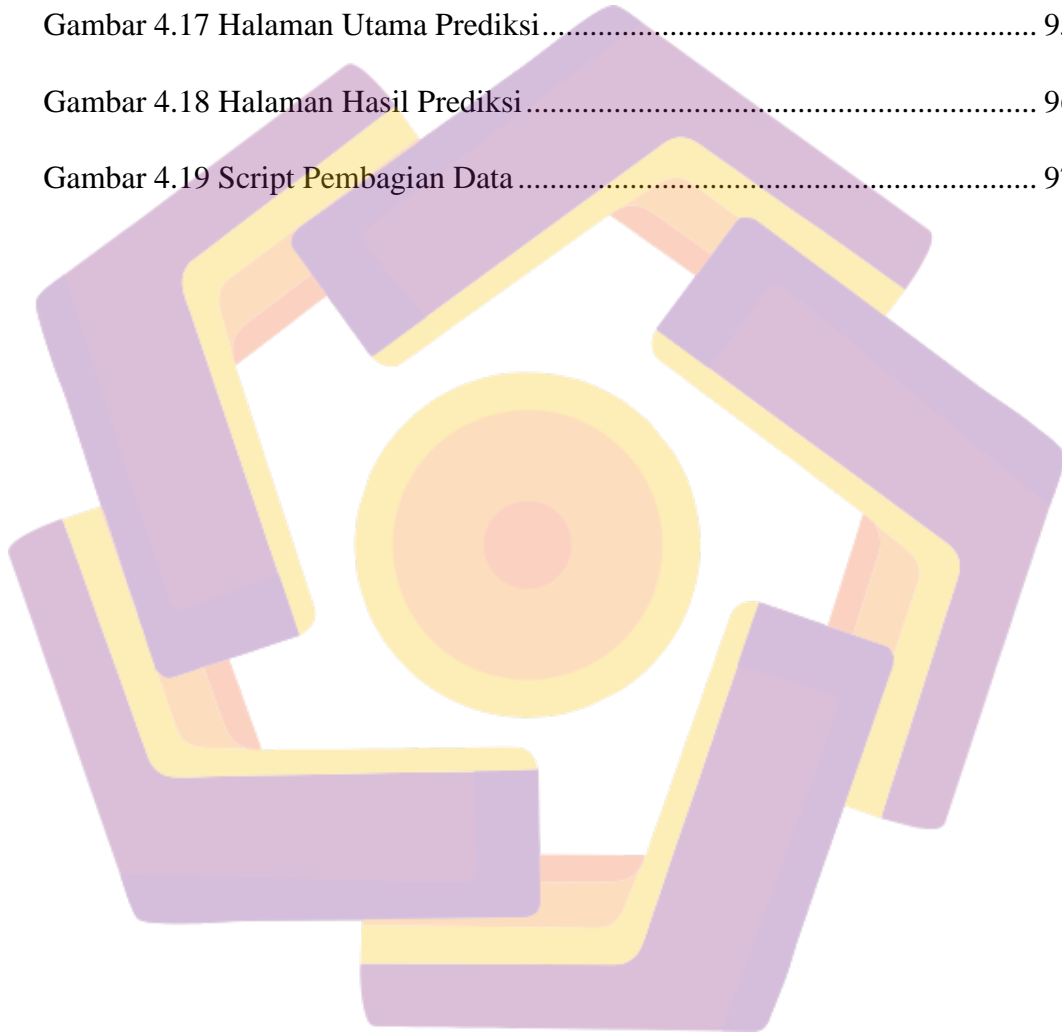
Tabel 3.13 Term Negative.....	62
Tabel 3.14 Term Neutral .....	63
Tabel 3.15 TF Tweet Ke 4 .....	65
Tabel 3.16 DF Tweet Ke 4.....	65
Tabel 3.17 IDF Tweet Ke 4.....	66
Tabel 3.18 TF-IDF Tweet Ke 4.....	67
Tabel 3.19 Cosine Similarity .....	69
Tabel 3.20 Data Latih.....	69
Tabel 3.21 Confussion Matrix.....	75
Tabel 3.22 Accuracy, Precision, dan Recall .....	75
Tabel 4.1 Perbandingan Jumlah Feature .....	98
Tabel 4.2 Pengukuran Akurasi, Precision, Recall, dan F1-Score .....	99
Tabel 4.3 Perbandingan Hasil Pelatihan POS Filter dan Tanpa POS Filter ....	103
Tabel 4.4 Evaluasi Prediksi.....	104
Tabel 4.5 Confussion Matrix Model POS Filtering .....	105
Tabel 4.6 Confussion Matrix Model Tanpa POS Filtering.....	106

## DAFTAR GAMBAR

Gambar 2.1 Ilustrasi K-Fold Cross Validation .....	28
Gambar 2.2 Contoh Flowchart.....	31
Gambar 3.1 Gambaran Umum Sistem .....	34
Gambar 3.2 Preprocessing .....	40
Gambar 3.3 Proses Training.....	59
Gambar 3.4 Rancangan Halaman Utama Prediksi.....	76
Gambar 3.5 Rancangan Halaman Hasil Prediksi .....	77
Gambar 3.6 Script Pengunduhan Data Twitter .....	78
Gambar 3.7 Data Tweet Berlabel.....	79
Gambar 4.1 Script Pengumpulan Data Twitter .....	81
Gambar 4.2 Hasil Pengumpulan Data Twitter .....	81
Gambar 4.3 Kuisisioner Validasi Data Training .....	82
Gambar 4.4 Responer Kuisisioner .....	83
Gambar 4.5 Data Tweet Berlabel.....	83
Gambar 4.6 Script Tweet Cleaning.....	84
Gambar 4.7 Script Tweet Tokenize .....	85
Gambar 4.8 List Slang.....	86
Gambar 4.9 Script Slang Replacement .....	86
Gambar 4.10 Script POS Tagging.....	87
Gambar 4.11 Script POS Filtering .....	88
Gambar 4.12 Script Stemming.....	88



Gambar 4.13 Script Pembuatan Feature List .....	89
Gambar 4.14 Script Penyimpanan Feature List .....	90
Gambar 4.15 Script TF-IDF dan Cosine Similarity .....	91
Gambar 4.16 Script Training Naïve Bayes Classifier .....	93
Gambar 4.17 Halaman Utama Prediksi.....	95
Gambar 4.18 Halaman Hasil Prediksi.....	96
Gambar 4.19 Script Pembagian Data .....	97



## INTISARI

Media sosial Twitter merupakan microblog yang memungkinkan penggunanya untuk menuliskan Tweet tentang berbagai topik dan membahas isu-isu yang sedang terjadi. Tweet pada Twitter dapat berbentuk teks, foto, maupun video. Tetapi pada penelitian ini hanya akan menggunakan cuitan yang berbentuk teks saja. Tweet tersebut nantinya akan digunakan untuk membuat analisis sentiment terhadap tingkat kepuasan pengguna provider Telkomsel menggunakan algoritma Naïve Bayes Classifier(NBC).

Naïve Bayes Classifier(NBC) adalah algoritma yang menggunakan metode probabilitas dan statistic yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Ciri utama dari Naïve Bayes ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi/kejadian. Keuntungan menggunakan Naïve Bayes adalah bahwa metode ini hanya membutuhkan data pelatihan yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

Algoritma ini digunakan sebagai pengklasifikasi Tweet ke dalam tiga sentiment yaitu positif, negative, netral. Penelitian ini berfokus pada pengujian akurasi algoritma Naïve Bayes Classifier tanpa menggunakan metode POS Filtering dan menggunakan metode POS Filtering. Akurasi yang dihasilkan oleh algoritma Naïve Bayes Classifier tanpa menggunakan metode POS Filtering sebesar 90% sedangkan menggunakan metode POS Filtering sebesar 87%. Kesimpulan penelitian ini bahwa akurasi algoritma Naïve Bayes Classifier dapat mencapai 90% tanpa menggunakan metode POS Filtering.

**Kata Kunci:** Twitter, Klasifikasi, Naïve Bayes Classifier, POS Filtering.

## **ABSTRACT**

*Social media Twitter is a microblog that allows users to write Tweets about various topics and discuss current issues. Tweets on Twitter can be in the form of text, photos or videos. But this research will only use text in the form of text. The tweet will later be used to make sentiment analysis of the level of user satisfaction of Telkomsel providers using the Naïve Bayes Classifier (NBC) algorithm.*

*Naïve Bayes Classifier (NBC) is an algorithm that uses probability and statistical methods proposed by British scientist Thomas Bayes. The main characteristic of Naïve Bayes is a very strong assumption of independence from each condition event. The advantage of using Naïve Bayes is that this method only requires a small amount of training data to determine the estimated parameters needed in the classification process.*

*This algorithm is used to classify Tweets into three sentiments, namely positive, negative, neutral. This research focuses on testing the accuracy of the Naïve Bayes Classifier algorithm without using the POS Filtering method and using the POS Filtering method. The accuracy produced by the Naïve Bayes Classifier algorithm without using the POS Filtering method by 90% while using the POS Filtering method by 87%. The conclusion of this study is that the accuracy of the Naïve Bayes Classifier algorithm can reach 90% without using the POS Filtering method.*

**Keywords:** *Twitter, Classification, Naïve Bayes Classifier, POS Filtering.*