

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KUALITAS
DAN PELAYANAN JNE MENGGUNAKAN ALGORITMA
RANDOM FOREST**

SKRIPSI



disusun oleh

M Jembar Ashofa

15.11.9333

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KUALITAS
DAN PELAYANAN JNE MENGGUNAKAN ALGORITMA
RANDOM FOREST**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Sistem Informasi



disusun oleh

M Jembar Ashofa

15.11.9333

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

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
KUALITAS DAN PELAYANAN JNE MENGGUNAKAN
ALGORITMA RANDOM FOREST**

yang dipersiapkan dan disusun oleh

M Jembar Ashofa

15.11.9333

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 6 Februari 2019

Dosen Pembimbing,



Windha Mega P.D. M.Kom
NIK. 190302185

PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
KUALITAS DAN PELAYANAN JNE MENGGUNAKAN
ALGORITMA RANDOM FOREST**

yang dipersiapkan dan disusun oleh
M Jembar Ashofa

15.11.9333

telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Agustus 2019

Susunan Dewan Penguji

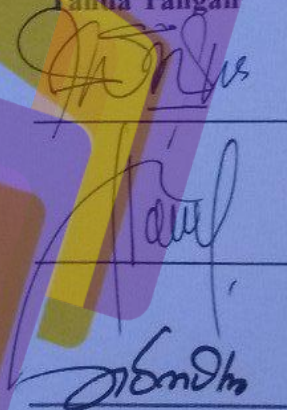
Nama Penguji

Wiwi Widayani, M.Kom
NIK. 190302272

Mulia Sulistiono, M.Kom
NIK. 190302248

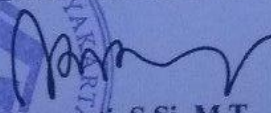
Windha Mega P.D, M.Kom
NIK. 190302185

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 5 September 2019

DEKAN FAKULTAS ILMU KOMPUTER


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



PERNYATAAN

Saya yang bertandatangan 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 untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau 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 tanggungjawab saya pribadi.

Yogyakarta, 27 Agustus 2019



M Jembar Ashofa
NIM. 15.11.9333

MOTTO

” Proses itu berantakan ”

- Austin Kleon

” Hard work betrays none, dreams betray many ”

- Hikigaya Hachiman from Oregairu

“ Ada kejahatan yang lebih kejam daripada membakar buku. Salah satunya adalah tidak membacanya. ”

- Joseph Brodsky

“ Siapapun bisa menangani kemenangan. Hanya yang kuat yang tahan kekalahan ”

- Adolf Hitler

“ Jika diri Anda dan musuh Anda mulai sombong, maka Anda menempatkan diri Anda dalam bahaya. ”

- Sun Tzu

“ Kebenaran suatu hal tidaklah ditentukan oleh berapa banyaknya orang yang mempercayainya. ”

- Ahmad Dahlan

PERSEMBAHAN

Alhamdulillah segala puji syukur atas kehadiran Allah SWT yang telah memberikan limpahan rahmat dan karunia-Nya sehingga penelitian ini dapat dilakukan dan diselesaikan dengan sebaik-baiknya. Saya juga ucapkan terimakasih untuk dukungan dan bantuan semua pihak yang membantu selesainya penelitian ini. Ucapan terimakasih saya persembahkan kepada :

1. Kedua orang tua dan semua keluarga yang senantiasa memberikan dukungan dan do'a kepada saya.
2. Ibu Windha Mega, M.Kom selaku dosen pembimbing yang telah banyak memberikan bimbingan, pelajaran, serta ilmu yang sangat bermanfaat.
3. Teman-teman kelas 15-S1IF-12 yang menjadi teman seperjuangan dari awal perkuliahan hingga saat ini.
4. Sahabat-sahabat rantau, yang banyak memberikan dukungan, kritik dan saran yang membantu proses penulisan skripsi ini.
5. Sobat Rodi, yang telah berjuang bersama dalam proses penulisan skripsi dari awal sampai selesai.
6. Sahabat kontrakan, yang sudah menemani dan bersenang-senang bersama penulis selama perkuliahan

KATA PENGANTAR

Assalamualaikum Wr.Wb.

Puji syukur kepada Allah SWT yang telah memberikan rahmat hidayah, karunia dan kesehatan, sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul “Pengaruh Stemming dan Stopwords Terhadap Akurasi Analisis Sentimen Menggunakan Metode Support Vector Machine”. Keberhasilan dalam menyelesaikan pembuatan laporan skripsi ini adalah berkat bantuan dan dukungan dari berbagai pihak. Maka dari itu pada kesempatan kali ini penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada:

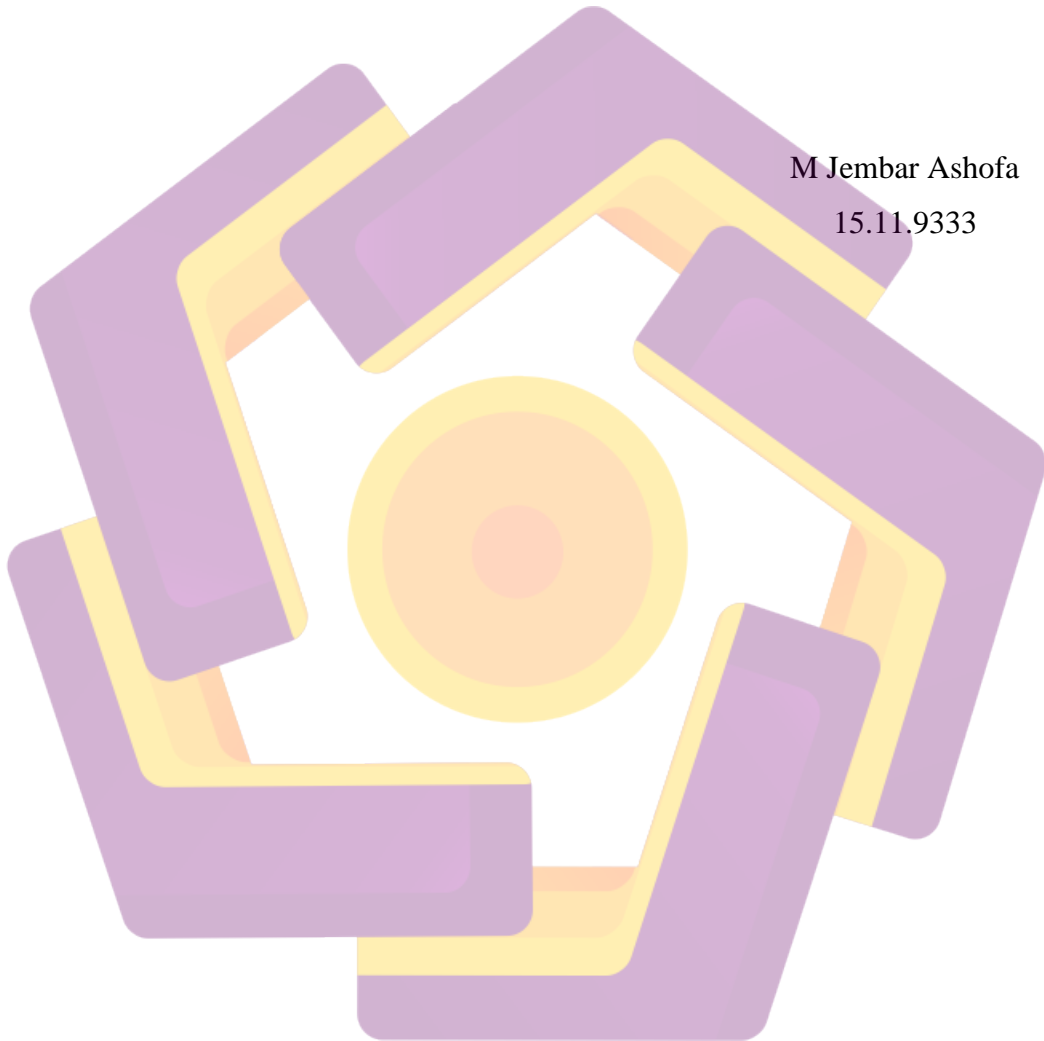
1. Bapak Prof. Dr. M. Suyanto, M.M. selaku ketua yayasan Universitas Amikom Yogyakarta
2. Ibu Krisnawati, S.Si., M.T. selaku Dekan Fakultas Ilmu Komputer dan Bapak Sudarmawan, S.T., M.T. selaku Kepala Program Studi S1 Informatika.
3. Ibu Mardhiya Hayaty, S.T, M.Kom selaku dosen pembimbing yang telah memberikan pengarahan, bimbingan dan motivasi selama proses penyusunan skripsi hingga selesai.
4. Segenap Dosen dan Karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu pengetahuan dan pengalamannya.
5. Kepada kedua orang tua penulis yang telah memberikan dukungan terbaiknya selama kuliah.
6. Kepada Teman-teman angkatan 2015 khususnya kelas 11-S1-TI12 yang telah berjuang bersama.

Penulis menyadari sepenuhnya bahwa laporan skripsi ini masih sangat jauh dari kesempurnaan, itu semua tidak lepas dari keterbatasan pengetahuan dan kemampuan dari penulis sendiri. Untuk itu, penulis mengharapkan kritik dan saran yang bersifat membangun guna mencapai kesempurnaan yang selalu penulis

harapkan sehingga dapat bermanfaat bagi penulis, serta pihak-pihak yang membutuhkan.

Yogyakarta, 27 Agustus 2019

M Jembar Ashofa
15.11.9333

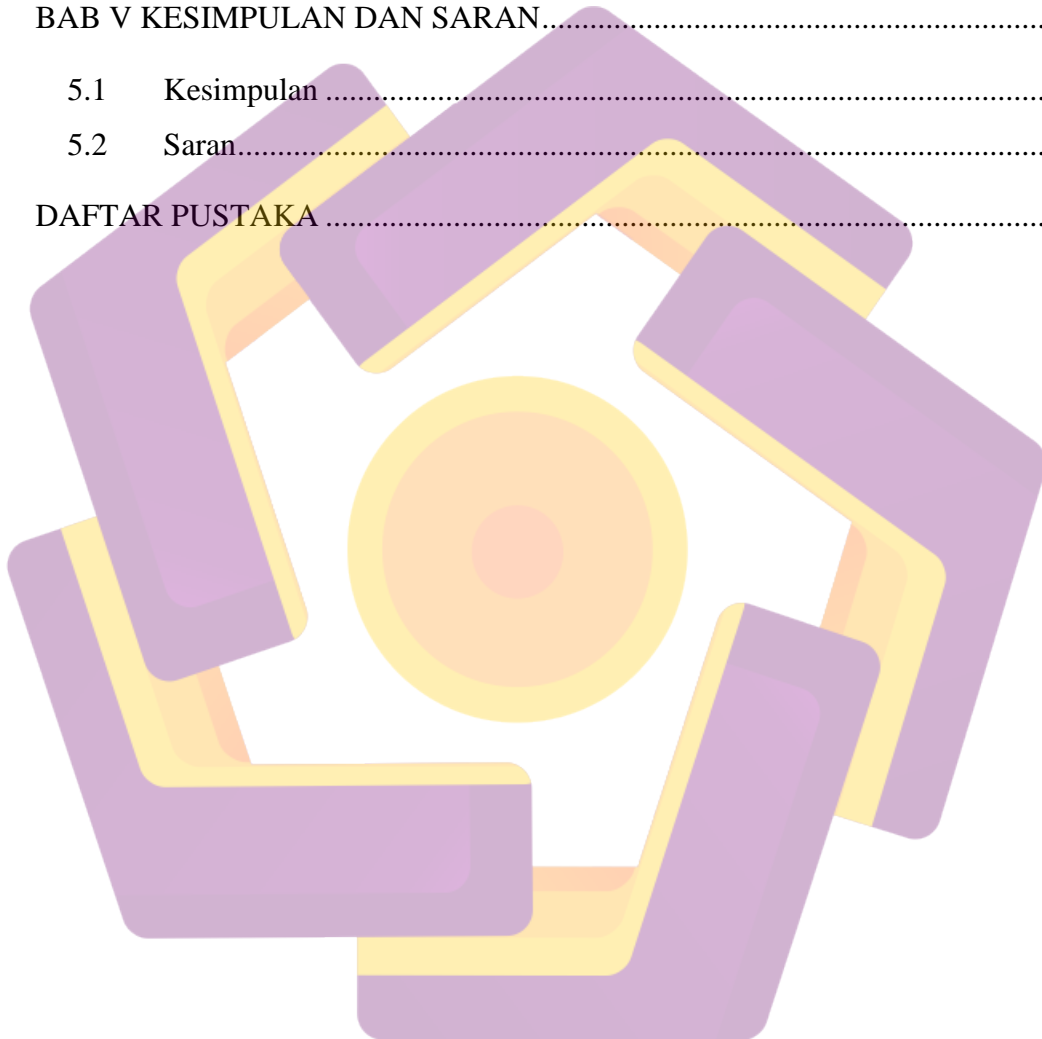


DAFTAR ISI

JUDUL	ii
PERSETUJUAN	iii
PENGESAHAN	iv
PERNYATAAN.....	v
MOTTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Maksud dan Tujuan Penelitian.....	3
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan	6
BAB II LANDASAN TEORI	8
2.1 Tinjauan Pustaka	8
2.2 Dasar Teori.....	9
2.3.1 Data Mining	9

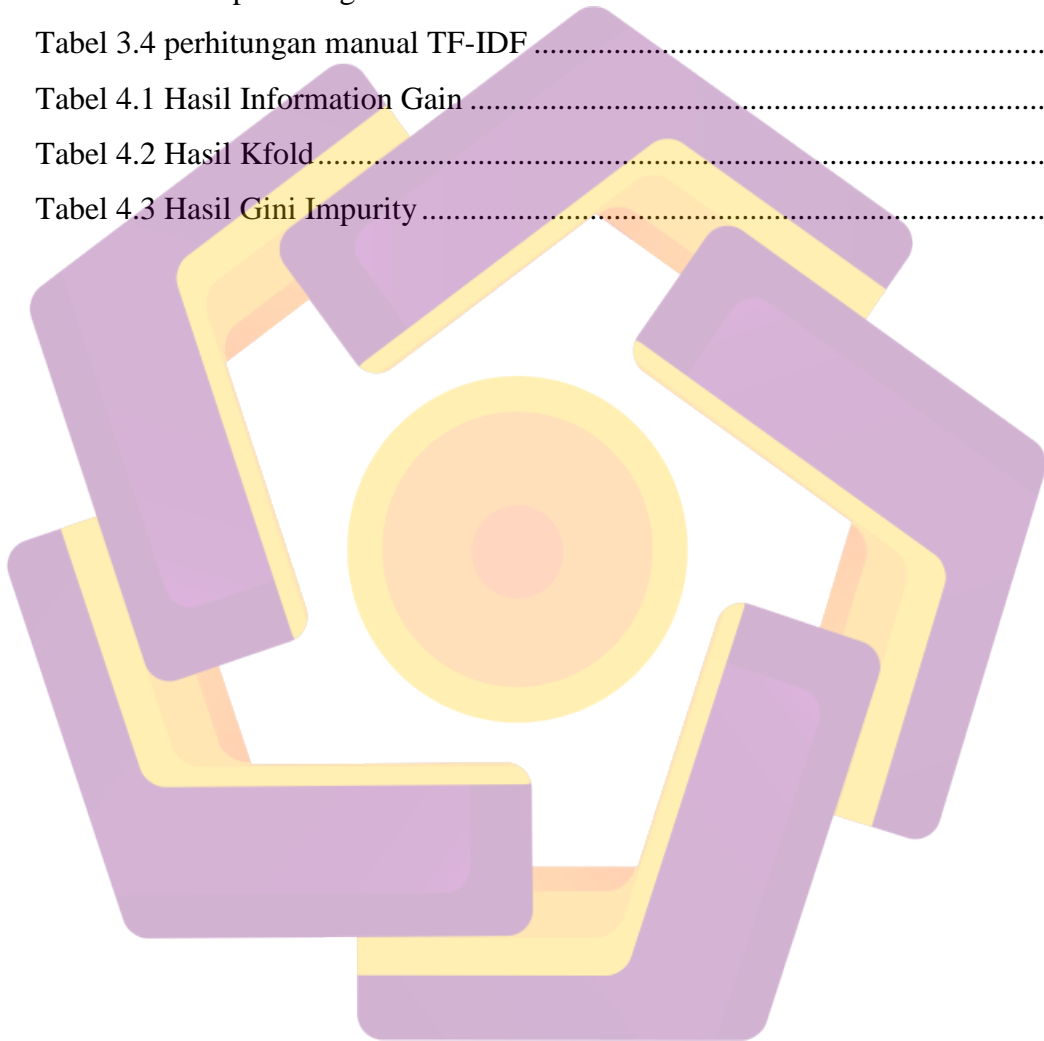
2.3.2	Supervised Learning.....	10
2.3.3	Text Mining.....	10
2.3.4	Analisis Sentimen	10
2.3.5	Preprocessing	11
2.3.6	Decision Tree	12
2.3.7	Random Forest	14
2.3.8	Evaluasi Klasifikasi.....	14
2.3.9	<i>Term Frequency-Inverse Document Frequency</i>	16
2.3.10	Flowchart	16
2.3.11	Validasi	18
BAB III METODOLOGI PENELITIAN.....		19
3.1	Alat dan Bahan.....	19
3.1.1	Alat.....	19
3.1.2	Bahan.....	20
3.2	Langkah Kerja.....	20
3.2.1	Pengumpulan Data	22
3.2.2	Preprocessing Data.....	22
3.2.3	TF-IDF	24
3.2.4	Random Forest	28
3.2.5	Evaluasi	28
3.2.6	Validasi	29
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Eksperimen.....	30
4.1.1	Pengumpulan Data	30
4.1.1.1	Registrasi ke Twitter API.....	30
4.1.1.2	Pengumpulan Data	31
4.1.2	Preprocessing Data.....	33
4.1.3	Pembagian Data	38
4.1.4	Pembobotan Kata	39
4.1.5	Pelatihan dan Pengujian Data	41

4.1.6	Evaluasi	42
4.1.7	Validasi	43
4.2	Pembahasan	45
4.2.1	Pembahasan Evaluasi	45
4.2.2	Pembahasan Validasi	47
BAB V KESIMPULAN DAN SARAN.....		49
5.1	Kesimpulan	49
5.2	Saran.....	49
DAFTAR PUSTAKA		51



DAFTAR TABEL

Tabel 2.1 Flowchart	17
Tabel 3.1 Spesifikasi Perangkat	19
Tabel 3.2 Tabel Hasil TF	25
Tabel 3.3 hasil perhitungan IDF	26
Tabel 3.4 perhitungan manual TF-IDF	27
Tabel 4.1 Hasil Information Gain	42
Tabel 4.2 Hasil Kfold	44
Tabel 4.3 Hasil Gini Impurity	45



DAFTAR GAMBAR

Gambar 3.1 Flowchart Klasifikasi Sentimen	21
Gambar 3.2 Flowchart Preprocessing	23
Gambar 4.1 App Detail	30
Gambar 4.2 Keys dan Token.....	31
Gambar 4.3 Tampilan Antarmuka TAGS	32
Gambar 4.4 Jumlah Data yang Didapat	32
Gambar 4.5 Dataset Penelitian.....	32
Gambar 4.6 Baca Dataset.....	33
Gambar 4.7 Kode Casefolding	33
Gambar 4.8 Hasil Case Folding	33
Gambar 4.9 Kode Cleansing	34
Gambar 4.10 Kode Slangwords	35
Gambar 4.11 Hasil Cleansing	35
Gambar 4.12 Kode Stemming.....	36
Gambar 4.13 Hasil Stemming Sastrawi	36
Gambar 4.14 Kode Stopword Remover	36
Gambar 4.15 Hasil Stopwords Removing.....	37
Gambar 4.16 Kode Filtering	37
Gambar 4.17 Hasil Filtering.....	37
Gambar 4.18 Kode Tokenizing.....	38
Gambar 4.19 Kode Pembagian Data.....	39
Gambar 4.20 Kode TF-IDF.....	40
Gambar 4.21 Hasil TF-IDF	41
Gambar 4.22 Kode Klasifikasi Random Forest	41
Gambar 4.23 Kode Confusion Matric.....	43
Gambar 4.24 Kode Kfold.....	43
Gambar 4.25 Grafik Confusion Matrix Information Gain	45
Gambar 4.26 Grafik Confusion Matrix Gini.....	46
Gambar 4.27 Grafik Perbandingan Akurasi CM Information gain dan Gini.....	47

Gambar 4.28 Perbandingan Kfold Information Gain dan Gini 48



INTISARI

JNE atau nama resminya adalah Tiki Jalur Nugraha Ekakurir adalah perusahaan yang bergerak dibidang pengiriman dan logistik. Internet saat ini menjadi cara berkomunikasi yang sangat efisien, karena berbagai lapisan masyarakat Indonesia hampir bisa menggunakan internet. Begitu juga dengan penyedia jasa, banyak yang menggunakan sosial media sebagai pelayanan terhadap pelanggannya.

Untuk mengetahui pelanggan puas atau tidak, maka harus melalui proses peninjauan atau analisis data yang telah tersedia menggunakan metode sentimen analisis. Data yang akan diambil dan diolah adalah tweet dari pelanggan jasa pengiriman *JNE* mengenai pelayanan pihak pengiriman jasa tersebut.

Data diperoleh dengan cara scrapping menggunakan alat dari pihak ketiga yaitu TAGSv6.1, lalu data yang didapat dengan scrapping tersebut akan dirapihkan dengan melakukan penghapusan simbol, ubah ke huruf kecil, hapus kata yang tidak perlu, setelah itu kalimat dibagi menjadi kata tersendiri untuk dicari bobot dari kata tersebut. Selanjutnya menggunakan algoritma random forest dengan metode information gain dan gini, data dilatih dengan label yang sudah diberi yaitu positif,netral dan negatif. Hasil akurasi yang didapat menggunakan information gain adalah sebesar 67.61% dan gini sebesar 67.33%.

Kata Kunci: *random forest, gini, analisis sentimen, data mining, entropy*

ABSTRACT

JNE or its official name is Tiki Jalur Nugraha Ekakurir is a company engaged in shipping and logistics. The internet is now a very efficient way of communicating, because various layers of Indonesian society can almost use the internet. Likewise with service providers, many use social media as a service to their customers. To find out whether customers are satisfied or not, one must go through a process of reviewing or analyzing available data using the sentiment analysis method.

The data to be retrieved and processed is a tweet from JNE shipping service customers regarding the service delivery party. Data obtained by scrapping using a tool from a third party, TAGSv6.1, then the data obtained by scrapping will be trimmed by removing the symbol, change it to lowercase, delete unnecessary words, after that the sentence is divided into separate words to find the weight from that word.

Furthermore, using a random forest algorithm with information gain and gini methods, the data is trained with labels that have been given that are positive, neutral and negative. Accuracy results obtained using information gain is 67.61% and gini is 67.33%.

Keyword: *random forest, gini, sentiment analysis, data mining, entropy*