

**ANALISIS SENTIMEN PERSEPSI PENGGUNA JNE MENGGUNAKAN  
ALGORITMA NAÏVE BAYES CLASSIFIER**

**TUGAS AKHIR**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata – 1  
Pada Program Studi Teknik Industri Fakultas Teknologi Industri



Disusun oleh:

Adelia Febriyanti (16522259)

JURUSAN TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2020



FAKULTAS  
TEKNIK INDUSTRI

Gedung K11, Mas Mansur  
Jl. Kaligung Km 14,5 Yogyakarta  
Telp. (0274) 895287, 898444 ext.2511;  
Fax. (0274) 895007

**SURAT KETERANGAN PENELITIAN**

Nomor : 209/A/Ka.Lab SIOP/FTI-UII/IV/2020

*Assalamu'alaikum Warahmatullahi Wabarakatuh*

Kami yang bertanda tangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut :

Nama : Adelia Febriyanti  
No. Mhs : 16522259  
Dosen Pembimbing : Annisa Uswatun Khasanah, S.T., M.Sc

Telah selesai melaksanakan penelitian yang berjudul " Analisis Sentimen Persepsi Pengguna JNE Menggunakan Algoritma *Naive Bayes Classifier*" di Laboratorium Statistik Industri dan Optimasi (SIOP), Prodi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 19 September sampai dengan tanggal 31 Desember 2019

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

*Wassalamu'alaikum Warahmatullahi Wabarakatuh*

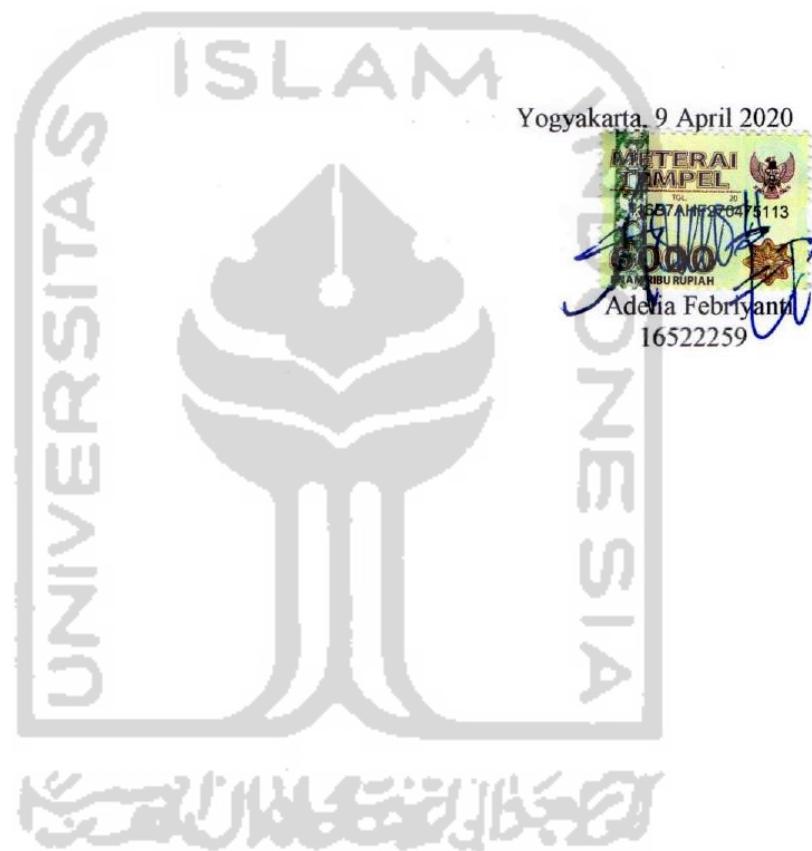
Dikeluarkan : di Yogyakarta  
Tanggal : 06 April 2020

Mengetahui,  
Kepala Lab. Statistik Industri dan Optimasi

Annisa Uswatun Khasanah, S.T., M.Sc

### PERNYATAAN KEASLIAN

Demi Allah, saya akui karya ini adalah hasil kerja saya sendiri kecuali kutipan dan ringkasan yang setiap salah satunya telah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah dalam karya tulis dan hak kekayaan intelektual maka saya bersedia ijazah yang saya terima untuk ditarik kembali oleh Universitas Islam Indonesia.





**LEMBAR PENGESAHAN PENGUJI TUGAS AKHIR**

**ANALISIS SENTIMEN PERSEPSI PENGGUNA JNE MENGGUNAKAN ALGORITMA  
NAÏVE BAYES CLASSIFIER**

Disusun oleh:

Nama : Adelia Febriyanti  
NIM : 16522259

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknik Industri

Yogyakarta, 27 April 2020

**Tim Penguji**

Annisa Uswatun K., S.T., M.B.A., M.Sc.

Ketua

Dr. Drs. Imam Djati Widodo, M.Eng.Sc.

Anggota I

Andrie Pasca Hendradewa, S.T., M.T.

Anggota II

Ir. Ali Parkhan, M.T.

Anggota III

*Dam*

*Imam*

*Ali*

*Parkhan*

Mengetahui



## **HALAMAN PERSEMPAHAN**

*Dengan Nama Allah yang Maha Pengasih dan Maha Penyayang.  
Alhamdulillahirobbil'alamin. Puji syukur tiada henti atas kehadirat Allah SWT atas  
nikmat yang telah diberikan. Skripsi ini saya persesembahkan kepada:*

*Bapak Pujadi dan Ibu Dwi Cahyani selaku kedua orang tua saya yang tiada henti  
mendo'akan, memberikan semangat, dan mendidik saya sejak kecil hingga saat ini.  
Terima kasih atas pengorbanan dan segala hal yang tidak bisa diungkapkan dengan  
kata, yang tidak akan pernah bisa dibalas oleh apapun.  
Teruntuk adik, dan keluarga besar saya, terima kasih atas semangat, dukungan, dan  
motivasi yang selalu diberikan hingga pada tahap ini. Serta sahabat dan teman-teman  
saya yang senantiasa hadir mendengarkan, membantu dan menemani hari-hari saya  
selama masa perkuliahan.*



## HALAMAN MOTTO


**فِي أَيِّ الَّأَيْ رَبِّكُمَا تُكَذِّبَانِ**

“Maka nikmat Tuhanmu yang manakah yang kamu dustakan?. ” (QS. Ar Rahman: 13)



“Maka sesungguhnya sesudah kesulitan itu ada kemudahan. Sesungguhnya sesudah kesulitan itu ada kemudahan.” (QS. Al-Insyirah:5-6)

## KATA PENGANTAR

### **Assalamu'alaikum Warahmatullaahi Wabarakatuh**

Alhamdulillah, segala puji dan syukur bagi Allah SWT atas berkah, rahmat, dan nikmat yang diberikan kepada hamba-Nya. Tidak lupa shalawat serta salam kepada Nabi besar Muhammad SAW beserta keluarga dan sahabat yang telah membawa ajaran kebenaran sehingga kini keluar dari jalan kegelapan menuju jalan terang benderang dalam rangka meraih ridho Allah SWT.

Atas rahmat dan nikmat Allah SWT, tugas akhir yang berjudul "**ANALISIS SENTIMEN PERSEPSI PENGGUNA JNE MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER**" dapat diselesaikan dengan baik. Dalam penelitian tugas akhir ini penulis mendapat pengetahuan, bimbingan, dukungan, dan saran dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. Ir. Hari Purnomo, M.T. selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Bapak Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph. D. selaku Ketua Jurusan Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia.
3. Bapak Dr. Taufiq Immawan, S.T., M.M. selaku Ketua Program Studi Teknik Industri Universitas Islam Indonesia.
4. Ibu Annisa Uswatun Khasanah, S.T., M.Sc. selaku dosen pembimbing I dan Bapak Ir. Ali Parkhan, M.T. selaku dosen pembimbing II Tugas Akhir program studi Teknik Industri Universitas Islam Indonesia.
5. Bapak Pujadi dan Ibu Dwi Cahyani selaku orang tua saya serta adik tersayang Muhammad Rizky Firmansyah dan Muhammad Rafi yang selalu memberikan kasih sayang dan *support* kepada saya hingga saat ini.
6. Kepala Laboratorium Statistika Industri dan Optimasi, Laboran, serta asisten Statistika Industri dan Optimasi (SIOP) angkatan 2016-2017. Terutama asisten angkatan 2016 antara lain Karina, Palmy, Ahen, Naufal, dan Ahmed yang selalu mendengarkan semua cerita dan keluh kesah.

7. Sahabat saya Rizka Yolanda Wiyono, yang senantiasa dengan senang hati berbagi ilmu dan mendengarkan segala keluh kesah dan Nina Muid terimakasih atas dukungan dan semangat yang diberikan kepada penulis
8. Sahabat semasa kuliah Hera Sharafina, Dhianitya Yogiari, Sinta Nofita, Dewi Ayu Ningrum, Theresa Kintan, dan Galih Pribadi Wicaksana terimakasih atas dukungan yang tiada hentinya diberikan.
9. Teman-teman serta sahabat Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia khususnya angkatan 2016

Penulis menyadari dalam menyusun dan menulis laporan Tugas Akhir ini masih terdapat kekurangan. Oleh karena itu, penulis mengharap kritik dan saran bersifat membangun untuk lebih baiknya Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat bagi semua pihak.

Wassalamu'alaikum Warohmatullahi Wabarakatuh

Yogyakarta, 9 April 2020



A blue ink signature in cursive script, appearing to read "Adelia Febriyanti".

(Adelia Febriyanti)

## ABSTRAK

Teknologi turut ikut serta dalam mengubah gaya hidup manusia secara fundamental, salah satunya dalam aspek bisnis. Salah satu bisnis yang berkembang dengan adanya teknologi adalah industri logistik. Perkembangan jasa logistik juga dibuktikan melalui banyaknya jumlah perusahaan yang bergerak di bidang jasa pengiriman barang, salah satunya adalah PT. Tiki Jalur Nugraha Ekakurir (JNE), yang telah berdiri selama 29 tahun dan memiliki jaringan yang luas di seluruh kota di Indonesia, dengan titik layanan 1.500 lokasi. JNE memiliki aplikasi bernama my JNE pada website *Google Play*, yang mendapat ulasan lebih dari 86.000 komentar dan hingga Desember 2019 hanya mendapat rating sebesar 2,4 bintang dari total keseluruhan penilaian 5 bintang. Hal ini mendorong dilakukannya penelitian terkait JNE dengan menggunakan data ulasan pengguna dari *Google Play*. Ulasan yang digunakan pada penelitian ini berjumlah 1.876 yang diklasifikasikan menjadi kelas sentimen positif dan negatif menggunakan algoritma *Naïve Bayes Classifier* serta dilakukan asosiasi kata. Klasifikasi dengan *naïve bayes classifier* dengan perbandingan data latih 90% dan data uji 10% menghasilkan akurasi yang optimum yaitu sebesar 85.87%. Selanjutnya untuk asosiasi teks didapatkan informasi bahwa pengguna JNE membicarakan “kirim”, “paket”, “kurir”, “bagus”, “aplikasi”, “cepat”, “layan”, “terima”, “bantu”, dan “bintang”. Sedangkan pada kelas sentimen negatif pengguna sering membicarakan mengenai “kirim”, “paket”, “kurir”, “kecewa”, “layan”, “jasa”, “buruk”, “aplikasi”, “parah”, dan “lambat”.

Kata Kunci: Analisis Sentimen, Asosiasi Kata, Diagram *Fishbone*, JNE, *Google Play*, *Naïve Bayes Classifier*,

## DAFTAR ISI

HALAMAN JUDUL .....	i
SURAT KETERANGAN PENELITIAN .....	ii
PERNYATAAN KEASLIAN .....	iii
LEMBAR PENGESAHAN PEMBIMBING .....	iv
LEMBAR PENGESAHAN PENGUJI TUGAS AKHIR.....	v
HALAMAN PERSEMBAHAN .....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR .....	viii
ABSTRAK.....	x
DAFTAR ISI .....	xi
DAFTAR TABEL .....	xiv
DAFTAR GAMBAR.....	xv
BAB I.....	1
PENDAHULUAN .....	1
1.1    Latar Belakang .....	1
1.2    Rumusan Masalah.....	5
1.3    Batasan Masalah .....	5
1.4    Tujuan Penelitian .....	5
1.5    Manfaat Penelitian .....	6
1.6    Sistematika Penulisan .....	6
BAB II .....	8
TINJAUAN PUSTAKA.....	8
2.1    Kajian Induktif .....	8
2.2    Kajian Deduktif.....	16
2.2.1.    Jasa Pengiriman .....	16
2.2.2.    JNE .....	17
2.2.3.    Alur Pengiriman JNE.....	19
2.2.4. <i>Google Play</i> .....	20
2.2.5. <i>Online Review</i> .....	21
2.2.6. <i>Web Scraping</i> .....	21

2.2.7. <i>Machine Learning</i> .....	22
2.2.8. <i>Text Mining</i> .....	22
2.2.9. <i>Sentiment Analysis</i> .....	23
2.2.10. <i>Classification</i> .....	24
2.2.11. Fitur dan Pembobotan.....	25
2.2.12. <i>Naïve Bayes Classifier</i> .....	26
2.2.13. Asosiasi Kata .....	27
2.2.14. <i>Diagram Fishbone</i> .....	28
BAB III .....	29
METODE PENELITIAN .....	29
3.1. Objek dan Subjek Penelitian.....	29
3.2. Populasi dan Sampel Penelitian .....	29
3.3. Pengumpulan Data.....	29
3.4. Metode Analisis Data.....	31
3.5. Alur Penelitian .....	32
BAB IV .....	37
PENGUMPULAN DAN PENGOLAHAN DATA .....	37
4.1. Pengumpulan Data.....	37
4.2. Pengolahan Data .....	40
4.2.1. Analisis Deskriptif.....	40
4.2.2. <i>Pre-processing Data</i> .....	42
4.2.3. Representasi Model .....	45
4.2.4. Pembagian Data Latih dan Data Uji .....	48
4.2.5. Klasifikasi dengan Metode <i>Naïve Bayes Classifier</i> .....	50
4.2.6. Visualisasi dan Asosiasi .....	56
4.2.7. Interpretasi Hasil.....	68
BAB V .....	72
PEMBAHASAN .....	72
5.1. Gambaran Umum Persepsi Pengguna JNE.....	72
5.2. Hasil Klasifikasi dan Asosiasi Metode <i>Naïve Bayes Classifier</i> .....	73
5.3. Rencana Pemecahan Masalah Berdasarkan Diagram <i>Fishbone</i> .....	78
BAB VI .....	85
KESIMPULAN DAN SARAN .....	85
6.1. Kesimpulan .....	85

6.2. Saran .....	86
DAFTAR PUSTAKA.....	87
LAMPIRAN .....	92



## DAFTAR TABEL

Tabel 1 1 Top Brand Index Jasa Kurir.....	3
Tabel 2.1 Kelebihan <i>Naïve Bayes Classifier</i> Untuk <i>Sentiment Analysis</i> .....	13
Tabel 2.2 Status Pengiriman .....	19
Tabel 3.1 Variabel Penelitian.....	31
Tabel 4 1 Menghapus Angka Pada Data.....	42
Tabel 4.2 Menghapus Tanda Baca.....	43
Tabel 4.3 Menghapus Singkatan.....	43
Tabel 4.4 Menyeragamkan Huruf .....	43
Tabel 4.5 Tokenizing .....	43
Tabel 4.6 Filtering.....	44
Tabel 4.7 Stemming.....	45
Tabel 4.8 Perhitungan Skor Sentimen .....	45
Tabel 4.9 Pelabelan Jumlah Kelas Sentimen .....	46
Tabel 4.10 Pelabelan Kelas Sentimen Pada Kalimat .....	47
Tabel 4.11 Pembagian Data Latih dan Data Uji 1 .....	49
Tabel 4.12 Pembagian Data Latih dan Data Uji 2 .....	49
Tabel 4.13 Pembagian Data Latih dan Data Uji 3 .....	49
Tabel 4.14 Pembagian Data Latih dan Data Uji 4 .....	50
Tabel 4.15 Pembagian Data Latih dan Data Uji 5 .....	50
Tabel 4.16 Hasil klasifikasi 1 .....	51
Tabel 4.17 Hasil Klasifikasi 2.....	52
Tabel 4.18 Hasil Klasifikasi 3.....	52
Tabel 4.19 Hasil Klasifikasi 4.....	53
Tabel 4.20 Hasil Klasifikasi 5.....	54
Tabel 4.21 Permasalahan pada JNE.....	69
Tabel 4.22 Pengelompokan Permasalahan Berdasarkan 6 Aspek .....	70
Tabel 5.1 Rencana Pemecahan Masalah.....	78

## DAFTAR GAMBAR

Gambar 1.1 Penetrasi Pengguna Internet 2018.....	2
Gambar 2.1 Logo JNE .....	17
Gambar 2 2 Proses Pengiriman JNE.....	19
Gambar 2.3 Klasifikasi menggunakan Model Classifier .....	24
Gambar 3.1 Alur Penelitian .....	32
Gambar 4.1 Ekstensi Data Scraper .....	37
Gambar 4.2 Website my JNE .....	38
Gambar 4.3 Pembuatan Query.....	38
Gambar 4.4 Kolom Query .....	39
Gambar 4.5 Membuat Recipe .....	39
Gambar 4.6 Import Data .....	40
Gambar 4.7 Hasil <i>scraping data</i> .....	40
Gambar 4.8 Grafik Jumlah Ulasan .....	41
Gambar 4.9 Persentase Rating Ulasan.....	41
Gambar 4.10 Diagram Lingkaran Jumlah Ulasan .....	47
Gambar 4.11 Persentase Pelabelan Kelas Sentimen .....	48
Gambar 4.12 Grafik Perbandingan Hasil Akurasi .....	55
Gambar 4.13 Grafik Hasil Klasifikasi Naïve Bayes Classifier.....	56
Gambar 4.14 Barplot Visualisasi Sentimen Positif .....	57
Gambar 4.15 Wordcloud Visualisasi Sentimen Positif .....	57
Gambar 4.16 Asosiasi Kata Ulasan Positif.....	58
Gambar 4.17 Barplot Visualisasi Ulasan Negatif.....	60
Gambar 4.18 Wordcloud Visualisasi Ulasan Negatif.....	61
Gambar 4.19 Asosiasi Kata Ulasan Negatif .....	62
Gambar 4.20 Barplot Visualisasi Ulasan Positif dan Negatif.....	64
Gambar 4.21 Wordcloud ulasan keseluruhan .....	65
Gambar 4.22 Asosiasi kata yang sering muncul.....	65
Gambar 4.23 Interpretasi Diagram Fishbone Ulasan Negatif.....	68

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Teknologi informasi semakin berkembang pesat. Perkembangan teknologi membuat segala sesuatunya menjadi lebih mudah. Manusia dituntut untuk senantiasa mengembangkan teknologi, mengingat kebutuhan terkait data dan informasi semakin meningkat setiap tahunnya. Tuntutan ini dilakukan untuk menunjang proses pengolahan data dan informasi agar dapat dilakukan secara lebih efektif dan efisien (Josi et al., 2014). Perkembangan teknologi yang begitu pesat membawa peradaban manusia menuju ke era serba digital. Era digital tentunya membawa dampak yang positif, salah satunya adalah perkembangan teknologi internet dengan pengguna yang terus bertambah setiap tahunnya.

Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) merupakan salah satu lembaga yang rutin melakukan survei setiap tahunnya. Berdasarkan survei pada Maret-April 2019 menunjukkan bahwa jumlah penetrasi pengguna internet di Indonesia pada tahun 2018 tercatat sebesar 64,8% seperti yang dapat dilihat pada Gambar 1.1. Jumlah penetrasi tersebut naik sebanyak 10,12% dibanding dengan tahun sebelumnya. Terjadi kenaikan pengguna sebanyak 27 juta orang, yang sebelumnya berjumlah 143,26 menjadi 171,17 juta pengguna internet, dari total penduduk Indonesia sebanyak 246,16 juta orang. Pengguna internet akan terus meningkat setiap tahunnya seiring dengan perkembangan infrastruktur jaringan di berbagai daerah Indonesia (APJII, 2018).



Gambar 1.1 Penetrasi Pengguna Internet 2018

Teknologi turut ikut serta dalam mengubah gaya hidup manusia secara fundamental, salah satunya dalam aspek bisnis. Perusahaan memanfaatkan adanya teknologi informasi untuk menarik minat pelanggan. Peningkatan jumlah pengguna internet membuat berbagai perusahaan berlomba-lomba dalam memasarkan produknya baik barang maupun jasa melalui media internet. Berdasarkan survey yang dilakukan oleh APJII (2018) diketahui bahwa berjualan *online* termasuk ke dalam 10 besar alasan mengapa responden di Indonesia menggunakan internet. Sedangkan membeli secara *online* termasuk dalam 15 besar alasan utama penggunaan internet. Oleh karena itu terdapat banyak situs jual-beli secara *online* yang dikenal juga dengan sebutan *e-commerce*.

Pertumbuhan *e-commerce* ini tentunya ditopang oleh industri logistik, yaitu perusahaan ekspedisi. Kehadiran jasa ekspedisi memudahkan *e-commerce* mulai dari pergudangan hingga distribusi sehingga dapat lebih berfokus pada penjualan tanpa harus memikirkan pengiriman barang. Pada tahun 2017, kapitalisasi jasa ekspedisi menunjukkan bahwa salah satu industri jasa terbesar di Indonesia adalah pada bidang logistik (Chrisbiyanto, 2018). Selain itu, Badan Pusat Statistik memaparkan bahwa pertumbuhan ekonomi di Indonesia sepanjang tahun 2017 mencapai 5,19% dengan sektor transportasi dan pergudangan menjadi salah satu yang paling tinggi. Perusahaan ekspedisi akan mengirimkan pesanan dari penjual agar sampai ke tangan konsumen. Dalam pengiriman barang ini pun segala sesuatunya terkontrol melalui teknologi informasi, mulai dari

penyerahan barang ke petugas ekspedisi, *update* status terkini lokasi barang, hingga barang akhirnya diterima oleh konsumen.

Perkembangan jasa logistik juga dibuktikan melalui banyaknya jumlah perusahaan yang bergerak di bidang jasa pengiriman barang. Salah satu perusahaan jasa pengiriman dan logistik yang terkenal dan paling banyak digunakan adalah PT. Tiki Jalur Nugraha Ekakurir (JNE). Berdasarkan data *Top Brand* Indonesia terkait dengan jasa kurir yang paling banyak digunakan pada tahun 2015 sampai dengan 2019 menunjukkan bahwa JNE selalu berada pada urutan teratas seperti yang dapat dilihat pada Tabel 1.1 :

Tabel 1.1 *Top Brand Index Jasa Kurir*

Sumber : (*Top Brand Award* Indonesia, 2020)

BRAND	TOP BRAND INDONESIA				
	2015	2016	2017	2018	2019
JNE	43.5%	47.6%	49.4%	45.0%	26.4%
J&T	-	-	-	13.9%	20.3%
Tiki	36.2%	35.7%	34.7%	13.6%	12.6%
Pos Indonesia	6.7%	9.6%	8.4%	11.6%	5.4%
DHL	2.1%	1.3%	1.3%	3.5%	3.8%

JNE menjalin kerjasama dengan berbagai *e-commerce* di Indonesia, salah satunya adalah tokopedia dan shopee. JNE dipilih oleh masyarakat dibanding perusahaan lainnya bukan tanpa alasan. JNE yang telah berdiri selama 29 tahun ini memiliki jaringan yang luas di seluruh kota di Indonesia, dengan titik layanan 1.500 lokasi dan jumlah karyawan lebih dari 40.000 orang. Sistem komunikasi JNE telah terhubung secara *online* dengan lebih dari 150 lokasi dengan sistem yang efektif dan efisien untuk mengetahui status dari barang yang dikirim.

JNE memiliki aplikasi bernama My JNE yang dirilis pada tanggal 27 Januari 2016 di *Google Play*. Hingga saat ini, aplikasi my JNE telah di *download* oleh lebih dari satu juta pengguna. Adapun my JNE telah mendapat ulasan sebanyak lebih dari 86.000 komentar yang mencakup bukan hanya mengenai aplikasi namun juga pelayanan JNE secara keseluruhan. Ulasan ini mencakup keluhan yang bersifat negatif dan juga saran yang positif. Ulasan pengguna merupakan salah satu media yang efektif dan efisien untuk

menemukan informasi terhadap citra dari suatu perusahaan. Hal ini dikarenakan konsumen melihat perspektif rekomendasi *word-of-mouth* (opini) pengguna lain sebelum menggunakan suatu produk maupun jasa (Fanani, 2017). Adapun *rating* yang didapatkan oleh JNE di *Google Play* hingga Desember 2019 hanya sebesar 2,4 bintang dari total keseluruhan penilaian 5 bintang.

Berdasarkan data yang dihimpun dari StatCounter (2020) pada bulan Desember 2018 – Desember 2019, *Operating System Market Share Indonesia* menunjukkan bahwa mayoritas penduduk di Indonesia merupakan pengguna Android dengan persentase sebesar 50,61%. Android memiliki berbagai layanan di dalamnya, salah satunya adalah *Google Play*. *Google Play* merupakan layanan digital milik *Google* berupa toko untuk memasarkan produk berupa aplikasi, permainan, hingga buku. *Google Play* dapat diakses melalui *playstore* (aplikasi android), *website* dan *Google TV* (Karch, 2019). *Google Play* memiliki beragam fitur, salah satunya adalah pengguna dapat memberikan ulasan terkait suatu aplikasi. Berdasarkan ulasan tersebut dapat diketahui faktor apa saja yang mempengaruhi dan perlu diperbaiki, oleh perusahaan. Ulasan dapat dianalisis menggunakan teknik *Text Mining*. *Text Mining* adalah teknik pencarian informasi berdasarkan sumber data berupa teks. Salah satu teknik dari *Text Mining* adalah analisis sentimen yang mana dapat digunakan untuk proses klasifikasi opini ke dalam kelompok positif, negatif, maupun netral. Sehingga perusahaan dapat melakukan perbaikan berdasarkan sentimen negatif yang ditemukan.

Berdasarkan uraian diatas, perlu dilakukan analisis lebih lanjut terkait ulasan pengguna JNE di situs *Google Play* untuk mengetahui opini pengguna. Dilakukan pengambilan data dengan teknik *scraping* dikarenakan efektif dan efisien baik dari segi biaya, waktu maupun tenaga. Kemudian dilakukan pengklasifikasian ulasan dalam bentuk positif dan negatif menggunakan algoritma *Naïve bayes Classifier*. Menurut Wati (2016) *Naïve bayes* merupakan metode paling sederhana dari pengklasifikasian probabilitas, memiliki tingkat akurasi yang sangat tinggi ketika diaplikasikan pada *database* dengan *bigdata*. Selain itu model *naïve bayes* juga memiliki tingkat kesalahan yang sangat minimum dibandingkan dengan algoritma klasifikasi lainnya (Liu, et al., 2016).

Setelah proses klasifikasi, pada setiap sentimen positif dan negatif dilakukan ekstraksi dan eksplorasi informasi menggunakan analisis deskriptif dan asosiasi antar kata untuk dapat menemukan kata apa yang sering muncul pada ulasan pengguna JNE.

Selanjutnya dilakukan analisis menggunakan diagram *fishbone* pada sentimen negatif untuk mengetahui faktor sebab-akibat, sehingga JNE dapat mengetahui apa saja yang perlu dilakukan untuk memperbaiki pelayanannya. Penelitian ini diharapkan mampu menyajikan klasifikasi analisis sentimen yang baik dan tepat sehingga didapatkan informasi yang berguna terkait langkah yang perlu dilakukan oleh JNE dalam memperbaiki pelayanannya dan mengembangkan perusahaannya.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka permasalahan yang dari penelitian ini dapat dirumuskan sebagai berikut :

1. Bagaimana gambaran umum persepsi pengguna JNE berdasarkan ulasan pada aplikasi My JNE di website *Google Play*?
2. Bagaimana performa algoritma *Naïve bayes Classifier* dalam mengklasifikasikan persepsi pengguna JNE menjadi kelas positif dan negatif?
3. Informasi apa yang didapatkan berdasarkan klasifikasi dan asosiasi yang telah dilakukan?
4. Faktor apa saja yang perlu dilakukan untuk memperbaiki hasil ulasan negatif yang didapat?

## 1.3 Batasan Masalah

Batasan masalah yang ditentukan dalam penelitian ini adalah sebagai berikut:

1. Data yang akan diklasifikasi hanya data ulasan pengguna JNE pada *Google Play*
2. Ulasan pengguna mengenai JNE pada website *Google Play* yang diambil hanya pada periode setelah aplikasi my JNE di update yaitu setelah 19 September 2019 sampai dengan 31 Desember 2019.
3. Ulasan yang diambil adalah ulasan dengan menggunakan bahasa Indonesia.

## 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Mengetahui gambaran umum persepsi pengguna JNE berdasarkan ulasan pada aplikasi My JNE di website *Google Play*.

2. Mengetahui performa algoritma *Naïve bayes* dalam mengklasifikasikan persepsi pengguna JNE menjadi kelas positif dan negatif.
3. Memperoleh informasi penting mengenai berdasarkan klasifikasi dan asosiasi yang telah dilakukan.
4. Mengetahui faktor yang perlu dilakukan untuk memperbaiki hasil ulasan negatif yang didapat.

## 1.5 Manfaat Penelitian

Manfaat penelitian yang dilakukan adalah sebagai berikut :

1. Mengetahui gambaran umum mengenai persepsi pengguna JNE berdasarkan *website Google Play*.
2. Dapat memudahkan pihak yang memiliki kepentingan untuk melihat persepsi pengguna dalam bentuk sentimen positif dan negatif serta mengetahui faktor sebab-akibat terjadinya, sehingga dapat fokus melakukan evaluasi serta perbaikan ke arah yang lebih baik.

## 1.6 Sistematika Penulisan

Sistematika penulisan dibuat untuk memberikan gambaran secara umum mengenai penelitian yang akan dilakukan. Berikut merupakan sistematika penulisan penelitian

### BAB I PENDAHULUAN

Membuat kajian singkat tentang latar belakang permasalahan, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penelitian

### BAB II KAJIAN LITERATUR

Berisi konsep dan prinsip dasar untuk memecahkan masalah penelitian. Disamping itu juga untuk memuat uraian tentang hasil penelitian yang pernah dilakukan sebelumnya oleh peneliti lain yang memiliki hubungannya dengan penelitian yang dilakukan.

### BAB III METODE PENELITIAN

Mengandung uraian pemaparan tentang populasi dan sampel penelitian, jenis dan sumber data, variabel penelitian, metode pengumpulan data, metode analisis data, dan diagram alir.

**BAB IV****PENGUMPULAN DAN PENGOLAHAN DATA**

Berisi tentang data yang diperoleh selama penelitian dan bagaimana menganalisa data tersebut. Hasil pengolahan data ditampilkan baik dalam bentuk tabel maupun grafik. Pada sub bab ini merupakan acuan untuk pembahasan hasil yang akan ditulis pada sub bab pembahasan hasil.

**BAB V****PEMBAHASAN**

Melakukan pembahasan hasil yang diperoleh dalam penelitian, dan kesesuaian hasil dengan tujuan penelitian sehingga dapat menghasilkan sebuah rekomendasi.

**BAB VI****PENUTUP**

Berisi tentang kesimpulan terhadap analisis yang dibuat dan rekomendasi atau saran-saran atas hasil yang dicapai dan permasalahan yang ditemukan selama penelitian.

**DAFTAR PUSTAKA****LAMPIRAN**

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Kajian Induktif

Penelitian yang dilakukan oleh Jumeilah (2018) dengan judul “Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE dengan *Naïve bayes*”, menggunakan metode klasifikasi dan algoritma *naïve bayes*. Penelitian ini dilakukan dengan menggunakan Rstudio. Hasil penelitian menunjukkan bahwa klasifikasi komentar pengguna JNE pada twitter sejak tanggal 14 Februari 2018 - 19 Februari 2018 dengan algoritma *naïve bayes* mampu menghasilkan nilai akurasi yang baik yaitu sebesar 85%, *precision* 78% dan *recall* 67%.

Raksanagara et al., (2016) dalam penelitiannya yang berjudul “Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode *Naïve bayes*”, menggunakan algoritma *naïve bayes classifier*. Hasil penelitian ini menunjukkan bahwa hasil analisis sentimen pada twitter akun ekspedisi JNE menghasilkan akurasi mencapai 83%.

“*Sarcasm Detection in Microblogs Using Naïve bayes and Fuzzy Clustering Technology in Society*” merupakan penelitian yang dilakukan oleh Mukherjee dan Bala (2017), dalam penelitiannya menggunakan *naïve bayes* sebagai algoritma yang mana tergolong dalam *Supervised Learning*, dan juga membandingkannya dengan algoritma FCM (*Fuzzy c-means clustering*) yang tergolong dalam *Unsupervised Learning*. Penelitian ditujukan dengan menganalisis sarkasme dari 2000 *tweet*. Hasil penelitian menunjukkan bahwa *fuzzy c-means clustering* tidak seefektif algoritma *naïve bayes* dalam mendeteksi sarkasme, dikarenakan *naïve bayes* lebih cocok untuk *dataset* berjumlah kecil. Adapun akurasi yang didapatkan adalah sebesar 65%.

Penelitian yang dilakukan oleh Dey et al., (2016) mengenai “*Sentiment Analysis of Review Datasets using Naïve bayes and K-NN Classifier*” bertujuan untuk mengetahui sentimen terkait ulasan film dan juga hotel. Digunakan algoritma *naïve bayes* dan juga *k-nearest neighbour*. Berdasarkan hasil penelitian diketahui bahwa *naïve bayes* menghasilkan akurasi yang lebih tinggi dari KNN pada kategori ulasan film yaitu lebih dari 80%. Sedangkan untuk akurasi *review* hotel diketahui bahwa kedua algoritma tersebut

memiliki perbedaan yang cukup tipis namun *naïve bayes* tetap memiliki akurasi yang lebih tinggi.

“*An Approach of Semantic Web Service Classification Based on Naïve Bayes*” merupakan penelitian yang ditulis oleh Liu et al., (2016). Penelitian tersebut dilakukan dengan menggunakan klasifikasi dan algoritma *naïve bayes*. Penelitian dilakukan dengan tujuan untuk mengklasifikasikan layanan web semantik agar dapat membuat pengguna dengan mudah menemukan layanan yang dibutuhkan dengan cepat dan akurat. Hasil penelitian menunjukkan bahwa tingkat akurasi yang dihasilkan mencapai lebih dari 90%.

Penelitian berjudul “*Textual Dissection of Live Twitter Reviews Using Naïve bayes*” ditulis oleh Kunal et al., pada tahun 2018 dilakukan dengan menggunakan algoritma *naïve bayes* dengan *tweepy* dan *textblob* sebagai pustaka *python* untuk mengklasifikasikan ulasan berupa *tweet*. Sumber data yang digunakan adalah *live dataset twitter* untuk mengklasifikasikan *tweet* positif dan negatif terkait *personality* beberapa tokoh yang sedang menjadi tren. Dalam penelitian ini penulis memperkuat alasan penggunaan *naïve bayes* dengan melakukan perbandingan antara algoritma *naïve bayes classifier* dan juga *decision tree* dengan menggunakan dataset “*Titanic*”, pengujian dilakukan menggunakan tools *rapidminer*. Hasilnya menunjukkan bahwa akurasi yang diperoleh *naïve bayes* adalah lebih tinggi yakni 92,58% dibanding *decision tree* dengan akurasi hanya sebesar 79,04%.

“*A Sentiment Analysis Visualization System For The Property Industry*” merupakan penelitian yang ditulis oleh Mahadzir et al., (2018). Penelitian ini menggunakan algoritma *naïve bayes classifier* dan *software R* sebagai tool untuk mengetahui sentimen publik terkait properti dengan studi kasus PR1MA (Perumahan Rakyat 1Malaysia) yang merupakan proyek perumahan yang dibangun pemerintah dengan harga terjangkau. Sumber data dari penelitian ini adalah twitter dengan tweet yang dibatasi yakni hanya yang berbahasa Melayu dan bahasa Inggris dengan rentang waktu Januari – April 2018. Hasil penelitian menunjukkan bahwa melalui algoritma *naïve bayes* akurasi yang didapatkan sebesar 90% dengan presisi 93% dan *recall* 88%.

Penelitian yang dilakukan oleh Sánchez-Franco et al. pada tahun 2019 dengan judul “*A Naïve bayes Strategy For Classifying Customer Satisfaction: A Study Based On Online Reviews Of Hospitality Services*”, dilakukan menggunakan metode klasifikasi dan algoritma *naïve bayes classifier* untuk menganalisis kepuasan pelanggan Hotel di Las

Vegas berdasarkan 46.172 ulasan pada situs Yelp. Algoritma *naïve bayes* mengklasifikasikan sentimen pengguna dengan presisi dan *recall* yang tinggi yaitu sebesar 89% (*True Prediction*) serta 81% (*True Prediction*). Hasil penelitian ini lebih akurat dibanding hasil statistik sebelumnya dengan sampel yang terbatas, sehingga dapat dilakukan peningkatan pelayanan baik dari segi staff, dan lain sebagainya.

Penelitian yang berjudul “*Sentiment Analysis to determine Accomodation, Shopping, and Culinary Location on Foursquare in Kupang City*” yang ditulis oleh Aliandu pada tahun 2015 menggunakan algoritma *Naïve Bayes* untuk dapat menentukan sentimen dari akomodasi (hotel), tempat belanja, dan restoran (kuliner) di Kupang, Nusa Tenggara, Indonesia berdasarkan *review* pada situs yang bernama *Foursquare*. Pada *Foursquare* terdapat tips yang mana pengguna dapat memberikan komentar terkait baik/buruknya suatu lokasi. Berdasarkan hasil penelitian diketahui bahwa metode yang memiliki kecepatan dan akurasi tinggi yang cocok untuk bigdata adalah *naïve bayes*. Dengan hasil akurasi sebesar 77,48% dengan menggunakan fitur TF-IDF sedangkan untuk *testing set* sebesar 66,22%. Untuk kategori hotel diketahui bahwa Swiss Bellin dan Aston merupakan hotel terbaik. Untuk restoran adalah KFC yang terletak di Flobamora Mall, Palapa Kafe dan Restoran, dan Dapur Nekamese. Sedangkan untuk tempat berbelanja terbaik adalah Flobamora Mall.

“*Comparative Analysis between Naïve bayes Algorithm and Decision tree to Solve WSD Using Empirical Approach*” merupakan penelitian yang dilakukan oleh Al-bayaty et al., (2016). Penelitian ini bertujuan untuk mengatasi disambiguasi makna dari suatu kata (WSD) dengan menggunakan *e-quran dataset*. Penelitian ini menggunakan algoritma *naïve bayes*, *decision tree*, dengan *wordnet* serta *senseval-3*. Hasil penelitian menunjukkan bahwa *naïve bayes* memiliki akurasi yang lebih tinggi dibanding dengan *decision tree* yaitu sebesar 62,86%. Sedangkan *decision tree* menghasilkan akurasi sebesar 45,14%.

Penelitian yang dilakukan oleh Bilal et al., (2015) mengenai “*Sentiment Classification of Roman Urdu Opinions Using Naïve bayesian, Decision tree and KNN Classification Techniques*” dilakukan dengan membandingkan tiga algoritma yaitu *decision tree*, *naïve bayes*, dan K-NN. Penelitian ini didasarkan pada opini yang menggunakan bahasa *Roman-Urdu* dengan sumber berupa blog terkait efek penggunaan facebook yang kemudian diekstraksi menggunakan *software Easy Web Extractor*. Sedangkan *tool* pengolahan data yang digunakan adalah WEKA. Berdasarkan hasil

penelitian diketahui bahwa *Naïve Bayes* memiliki akurasi terbaik dengan akurasi data set training sebesar 97,33% dan *testing set* sebesar 97,50%.

“*A Web Based Tool for Arabic Sentiment Analysis*” merupakan penelitian yang dilakukan oleh El-Masri et al., (2017). Penelitian ini dilakukan untuk mengetahui analisis sentimen pada *tweet* dengan teks Arab terkait dengan trending topik yang sedang terjadi yaitu politik menggunakan perbandingan algoritma pada *machine learning* yaitu *naïve bayes* dan *support vector machine*. *Dataset* yang digunakan berupa *tweet* dengan jumlah sebanyak 8000 yang dipilih secara acak. Penelitian dilakukan dengan menggunakan *software R*, dengan *packages* termasuk didalamnya Rweka, Shiny, dan twitteR. Berdasarkan penelitian yang ada diketahui bahwa *naïve bayes* memiliki akurasi tertinggi yaitu sebesar 70% dibanding dengan *Support Vector Machine* (SVM) dengan akurasi hanya sebesar 34%.

Ramadhani (2015) dalam Tugas Akhirnya terkait “Analisis Sentimen Menggunakan Metode *Naïve Bayes Classifier* dengan Model Dokumen Bernoulli dan *Support Vector Machine*” melakukan perbandingan antara metode *Naïve Bayes* dan *Support Vector Machine*. Dilakukan klasifikasi dengan menggunakan fungsi kernel dan proporsi dengan data training sebesar 5.412 dan data testing sebesar 2.701. Hasil dari perbandingan menunjukkan bahwa nilai akurasi dengan menggunakan metode *Naïve Bayes Classifier* (NBC) memiliki akurasi yang lebih tinggi dibanding dengan SVM yaitu sebesar 62,63%. Adapun opini positif didapatkan sebanyak 44,465% sedangkan opini negatif 55,534%.

Barfian et al., pada tahun 2017 dalam penelitiannya terkait “*Twitter Pornography Multilingual Content Identification Based on Machine Learning*” membandingkan tiga metode dari *machine learning*, yaitu *Decision Tree*, *Naïve Bayes*, dan *Support Vector Machines*. *Dataset* yang digunakan berupa *dataset* dari *twitter* menggunakan plugin API AYLIEN dengan pengolahan menggunakan *tools RapidMiner Studio* dengan keyword “*pornography*”. Data yang digunakan adalah 200 *tweets* dengan bahasa Indonesia, 200 *tweet* dengan bahasa Inggris, dan 200 *tweets* dengan bahasa Indonesia-Inggris. Pada *dataset tweet* berbahasa Indonesia, metode yang memiliki akurasi tertinggi adalah *Naïve Bayes*. Sedangkan untuk dataset berbahasa Inggris dan Indonesia-Inggris, SVM memperoleh akurasi tertinggi yaitu 83,33% dan juga 72,77%.

Dalam penelitian yang dilakukan oleh Ibrohim dan Budi, pada tahun 2018 terkait “*A Dataset and Preliminaries Study for Abusive Language Detection In Indonesian Social Media*” dilakukan pendekstrian *abusive language* (bahasa kasar) di Indonesia pada sosial media seperti Line, Facebook, Twitter, dan sosial media lainnya. Data yang digunakan berupa tweet berbahasa Indonesia yang diuji coba dengan membandingkan algoritma SVM (*Support Vector Machine*), *Decision Tree*, *Random Forest*, dan juga *Naïve Bayes*, dengan fitur dan *word n-gram*. Berdasarkan hasil penelitian diketahui bahwa hasil akurasi tertinggi adalah *Naïve Bayes Classifier* dengan kombinasi *word unigram* dan fitur *bigram*, sebesar 70,06%.

Dalam penelitian yang ditulis oleh Wongso et al., pada tahun 2017 yang berjudul *News Article Text Classification in Indonesian Language* dilakukan analisis sentimen dengan pengambilan data menggunakan teknik data *crawling* melalui situs berita Indonesia yaitu [www.cnnindonesia.com](http://www.cnnindonesia.com). Dilakukan perbandingan antara tiga metode *machine learning*, yaitu *Support Vector Machine*, *Multivariate Bernoulli Naïve Bayes*, dan *Multinomial Naïve Bayes*, sebagai alat untuk mengklasifikasikan. Selain itu juga melakukan perbandingan antara fitur seleksi TF-IDF dan algoritma SVD. Data yang digunakan adalah sebanyak 5000 dokumen dengan masing-masing tiap kategori ekonomi, kesehatan, olahraga, politik, dan teknologi berjumlah 1000 dokumen. Berdasarkan data tersebut kemudian 80% dari total dataset menjadi data training, sedangkan 20% dari dataset digunakan sebagai data testing. Hasil penelitian menunjukkan bahwa kombinasi dari TF-IDF dan algoritma *Multinomial Naïve Bayes* memperoleh hasil tertinggi yaitu 98,4%.

Dalam penelitian yang dilakukan oleh Jain dan Kumar pada tahun 2015 dengan judul “*An Effective Approach to Track Levels of Influenza-A (H1N1) Pandemic in India Using Twitter*” digunakan dataset yang bersumber dari twitter terkait dengan informasi kesadaran masyarakat, langkah pencegahan, dan lain sebagainya mengenai virus H1N1 di India. Dilakukan perbandingan pada algoritma machine learning dengan menggunakan *Naïve Bayes* (NB), *Support Vector Machine* (SVM), *Random Forest*, dan juga *Decision Tree*. *Dataset twitter* yang digunakan berjumlah 21040 dengan periode batas *tweet* yang diambil adalah sejak 15 Februari 2015 - 15 Maret 2015. Hasil penelitian menunjukkan bahwa *F-Measure* pada SVM dan *Naïve Bayes* mencapai 0,77 atau 77% dengan recall dari naïve bayes lebih tinggi yaitu 86%.

Penelitian yang dilakukan oleh Azalia et al., pada tahun 2019 dengan judul *Name Indexing in Indonesian Translation of Hadith Using Named Entity Recognition with Naïve Bayes Classifier* dilakukan ekstraksi informasi dalam rangka memudahkan pencarian informasi dalam hadits. Dilakukan pengindeksan nama dalam hadits terjemahan bahasa Indonesia. Penelitian ini dilakukan dengan menggunakan *naïve bayes* sebagai algoritma, yang mana digunakan berbagai macam fitur, yakni *unigram*, dan *tag POS*. Penelitian dilakukan dengan melakukan ekstraksi 258 nama dari 13870 token data berdasarkan 100 hadits berbahasa Indonesia. *Naïve bayes* selaku algoritma dikombinasikan dengan berbagai fitur sehingga didapatkan hasil akhir dengan akurasi tertinggi sebesar 82,63% dari skor F1.

Kajian Induktif berisi tentang penelitian – penelitian terdahulu yang terkait dengan penelitian yang telah dilakukan. Berdasarkan penelitian terdahulu dapat diketahui beberapa kesimpulan terkait algoritma *naïve bayes* dalam tabel berikut :

Tabel 2.1 Kelebihan *Naïve Bayes Classifier* Untuk *Sentiment Analysis*

<b>Penulis</b>	<b>Metode</b>	<b>Hasil Penelitian</b>
Jumeilah (2018)	<i>Naïve Classifier</i>	<i>bayes</i> Hasil klasifikasi komentar pada twitter dari tanggal 14 Februari 2018 - 19 Februari 2018 menunjukkan bahwa <i>naïve bayes</i> mampu mengklasifikasikan dengan baik dengan rata-rata akurasi sebesar 85%.
Raksanagara et al., (2016)	<i>Naïve Classifier</i>	<i>bayes</i> Akurasi <i>naïve bayes</i> pada data uji negatif memperoleh ketepatan 76% sedangkan untuk data uji positif sebesar 83%.
Mukherjee dan Bala (2017)	<i>Naïve Classifier, Clustering</i>	<i>bayes</i> Analisis berupa kata yang bersifat sarkasme dilakukan menggunakan 2000 tweet, didapatkan hasil bahwa <i>Naïve bayes</i> lebih efektif dalam mendekripsi sarkasme dibanding dengan fuzzy clustering. Adapun hasil akurasi sebesar 65%.
Dey et al., (2016)	<i>Naïve bayes, K-NN</i>	Mengklasifikasikan ulasan film dan juga

<b>Penulis</b>	<b>Metode</b>	<b>Hasil Penelitian</b>
		hotel, dengan hasil yang didapatkan adalah akurasi <i>naïve bayes</i> lebih tinggi dibanding dengan K-NN yaitu lebih dari 80%
Liu et al., (2016)	<i>Naïve bayes</i> , <i>Classification</i>	Mengklasifikasikan web semantik dengan hasil akurasi pada data training menunjukkan hasil lebih dari 90%.
Kunal et al., (2018)	<i>Naïve bayes</i> , <i>Decision tree</i>	Mengklasifikasikan sentimen tweet positif dan negatif terkait personality tokoh yang sedang trend, dengan hasil akurasi pada <i>naïve bayes</i> didapatkan sebesar 92,58% sedangkan <i>decision tree</i> sebesar 79,04%
Mahadzir et al., (2018)	<i>Naïve bayes</i>	Mengklasifikasikan sentimen masyarakat terkait properti berupa PR1MA berdasarkan twitter, hasil akurasi menggunakan metode <i>Naïve bayes</i> didapatkan sebesar 90%
Sánchez-Franco et al., (2019)	<i>Naïve bayes</i> , <i>Classifier</i>	Mengklasifikasikan sentimen pelanggan hotel, didapatkan presisi sebesar 89% dan recall 81%
Aliandu (2015)	<i>Naïve bayes</i> , <i>Classifier</i>	Analisis sentimen pada akomodasi, tempat belanja, dan lokasi kuliner di Kupang melalui <i>foursquare</i> , akurasi yang dihasilkan sebesar 77,48%
Al-Bayaty et al., (2016)	<i>Naïve bayes</i> , <i>Classifier</i> , <i>Decision tree</i>	Membandingkan dua algoritma untuk mengatasi disambiguasi pada kata (WSD) di dataset e-quran, dihasilkan bahwa <i>naïve bayes</i> memiliki akurasi lebih tinggi yaitu sebesar 62, 86% sedangkan <i>decision tree</i> 45,14%

<b>Penulis</b>	<b>Metode</b>	<b>Hasil Penelitian</b>
Bilal <i>et al.</i> , (2016)	<i>Naïve Decision KNN</i>	<i>bayes, tree,</i> Klasifikasi sentimen pada blog dengan bahasa Roman-Urdi dengan metode <i>Naïve bayes</i> menghasilkan akurasi data set training sebesar 97,33% dan testing set sebesar 97,50%.
El-Masri <i>et al.</i> , (2017)	<i>Naïve Support Machine</i>	<i>bayes, Vector</i> Analisis sentimen pada tweet berbahasa arab, hasil dari <i>Naïve bayes</i> menghasilkan akurasi yang lebih tinggi yaitu sebesar 70% sedangkan SVM hanya sebesar 34%.
Ramadhani (2015)	<i>Naïve Classifier</i>	<i>bayes</i> Nilai akurasi paling tinggi adalah dengan dan metode <i>Naïve bayes Classifier</i> (NBC)
	<i>Support Vector Machine</i>	dengan akurasi sebesar 62,6295% dengan proporsi opini untuk kelas positif sebesar 44,465% dan opini kelas negatif sebesar 55,5349%
Barfian <i>et al.</i> , (2017)	<i>Decision Tree, Naïve Bayes, Support Vector Machine</i>	Akurasi tertinggi pada dataset berbahasa Indonesia didapatkan melalui metode <i>Naïve Bayes</i> . Sedangkan untuk dataset berbahasa Inggris dan Indonesia-Inggris melalui metode SVM
Ibrohim dan Budi (2018)	<i>Support Vector Machine, Decision Tree, Random Forest, Naïve Bayes</i>	Berdasarkan hasil perbandingan empat metode diketahui hasil akurasi tertinggi adalah <i>Naïve Bayes Classifier</i> dengan <i>Forest</i> , dan <i>Naïve Bayes</i> kombinasi <i>word unigram</i> dan fitur <i>bigram</i> , sebesar 70,06%.
Wongso <i>et al.</i> , (2017)	<i>Support Vector Machine, Multivariate Bernoulli</i>	Berdasarkan perbandingan dari tiga metode yang dilakukan didapatkan hasil penelitian bahwa kombinasi dari TF-IDF <i>Naïve Bayes</i> dan algoritma <i>Multinomial Naïve Bayes</i>

<b>Penulis</b>	<b>Metode</b>	<b>Hasil Penelitian</b>
	<i>Bayes</i> , <i>Multinomial Naïve Bayes</i>	dan memperoleh hasil tertinggi yaitu 98,4%.
Jain dan Kumar, (2015)	<i>Support Vector</i> metode tersebut diketahui bahwa <i>F-Machine</i> (SVM), <i>Measure</i> pada SVM dan <i>Naïve Bayes Random Forest</i> , mencapai 0,77 atau 77% dengan recall dan <i>Decision Tree</i> dari naïve bayes lebih tinggi yaitu 86%.	Berdasarkan perbandingan dari empat metode tersebut diketahui bahwa <i>F-Machine</i> (SVM), <i>Measure</i> pada SVM dan <i>Naïve Bayes Random Forest</i> , mencapai 0,77 atau 77% dengan recall dan <i>Decision Tree</i> dari naïve bayes lebih tinggi yaitu 86%.
Azalia et al., (2019)	<i>Naïve Bayes</i>	Dengan menggunakan algoritma <i>naïve bayes</i> didapatkan akurasi yang tinggi yaitu sebesar 82,63%.

Berdasarkan kajian empiris, diketahui bahwa dengan menggunakan algoritma *Naïve bayes* diketahui bahwa hasil penelitian menunjukkan kinerja klasifikasi dengan metode tersebut memiliki nilai akurasi yang cukup menjanjikan. Namun dengan adanya penelitian ini diharapkan dapat menjadikan gambaran dan bahan evaluasi kinerja dari perusahaan di bidang jasa ekspedisi, khususnya JNE.

## 2.2 Kajian Deduktif

### 2.2.1. Jasa Pengiriman

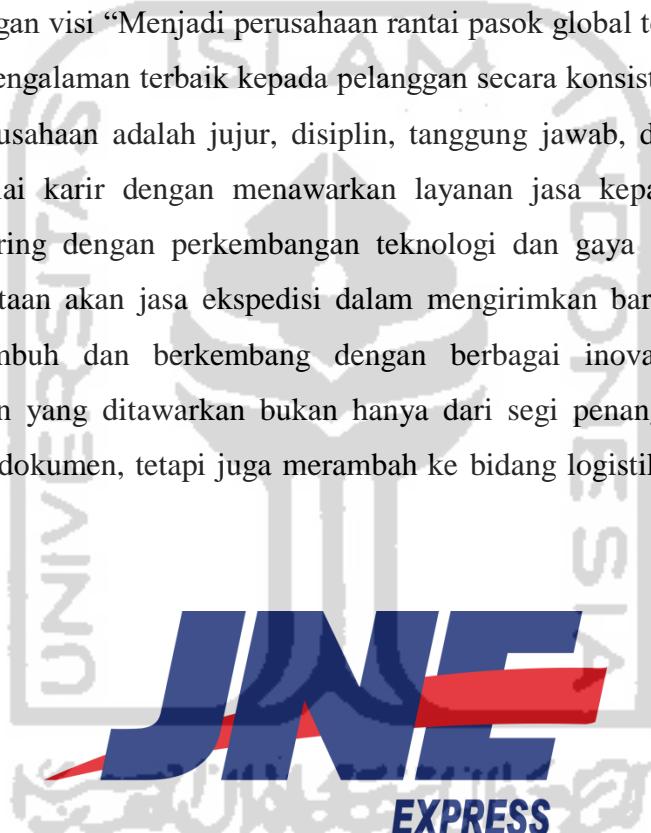
Jasa ekspedisi atau yang biasa disebut juga jasa kurir atau jasa pengiriman menurut PER-178/PJ/2006 mendefinisikan Jasa Pengiriman yaitu yang mengacu pada Keputusan Menteri Perhubungan No.KM/10 Tahun 1988 tentang Jasa Pengurusan Transportasi. Berdasarkan SK Menhub tersebut, yang dimaksud dengan Jasa Pengiriman Barang (Jasa *Freight Forwarding*) adalah : “Usaha yang ditujukan untuk mewakili kepentingan Pemilik Barang, untuk mengurus semua kegiatan yang diperlukan bagi terlaksananya pengiriman dan penerimaan barang melalui transportasi darat, laut dan udara yang dapat mencakup kegiatan penerimaan, penyimpanan, sortasi, pengepakan, penandaan pengukuran, penimbangan, pengurusan penyelesaian dokumen, penerbitan dokumen angkutan, klaim asuransi, atas pengiriman barang serta penyelesaian tagihan dan biaya-biaya lainnya terkait

dengan pengiriman barang-barang tersebut sampai dengan diterimanya barang oleh yang berhak menerimanya”.

Jasa pengiriman saat ini semakin berkembang pesat. Persaingan antar perusahaan jasa kurir pun semakin ketat. Terbukti bahwa saat ini terdapat berbagai perusahaan jasa kurir di Indonesia, antara lain : POS Indonesia, JNE Express, J&T Express, TIKI (Titipan Kilat), SiCepat. Pandu Logistics, dan lain sebagainya.

### **2.2.2. JNE**

PT Tiki Jalur Nugraha Ekakurir (JNE) merupakan perusahaan yang didirikan pada 26 November 1990. Dengan visi “Menjadi perusahaan rantai pasok global terdepan di dunia”, dan misi “Memberi pengalaman terbaik kepada pelanggan secara konsisten”. Adapun nilai yang dianut oleh perusahaan adalah jujur, disiplin, tanggung jawab, dan visioner. Pada awalnya JNE memulai karir dengan menawarkan layanan jasa kepabeanan atau impor barang/dokumen. Seiring dengan perkembangan teknologi dan gaya hidup masyarakat menyebabkan permintaan akan jasa ekspedisi dalam mengirimkan barang/dokumen pun meningkat. JNE tumbuh dan berkembang dengan berbagai inovasi layanan yang ditawarkan. Pelayanan yang ditawarkan bukan hanya dari segi penanganan pengiriman paket yang kecil dan dokumen, tetapi juga merambah ke bidang logistik, transportasi dan distribusi.



Gambar 2.1 Logo JNE

Berbagai peluang membuat JNE terus melebarkan sayap dengan memperluas jaringan ke berbagai daerah di Indonesia, tidak hanya di kota besar saja. Kini JNE memiliki karyawan lebih dari 40.000 orang dengan titik layanan mencapai lebih dari 6.000 lokasi. Serta lebih dari 150 lokasi JNE telah terkoneksi dengan sistem informasi online yang efektif dan efisien dalam melacak status dari pengiriman barang maupun dokumen. JNE tidak hanya mengandalkan sumber daya manusia, tetapi juga menggunakan teknologi

dalam mengembangkan inovasi, seperti mesin x-ray, GPS, hingga alat komunikasi satelit. Selain itu pada tahun 2016 JNE merilis program *7Magnificent*, aplikasi myJNE, dan *JNE Eco Courier*. JNE menjadi salah satu dari lima perusahaan jasa pengiriman yang paling dicari pada 2017-2018 berdasarkan data yang dihimpun dari *Google Trend*.

JNE menawarkan berbagai produk dan layanan yang terdiri dari tiga jenis yaitu *JNE Express*, *JNE Logistic*, dan *JNE Freight*. *JNE Express* melayani pengiriman paket dengan tujuan dalam negeri melalui lebih dari 1.500 titik layanan yang tersebar di seluruh Indonesia dengan memanfaatkan moda transportasi yang disesuaikan dengan kebutuhan pelanggan. Pelanggan dapat memilih layanan sesuai dengan kebutuhan terkait berapa lama waktu yang diinginkan agar paket tersebut sampai. Berikut ini merupakan layanan yang ditawarkan oleh *JNE Express* :

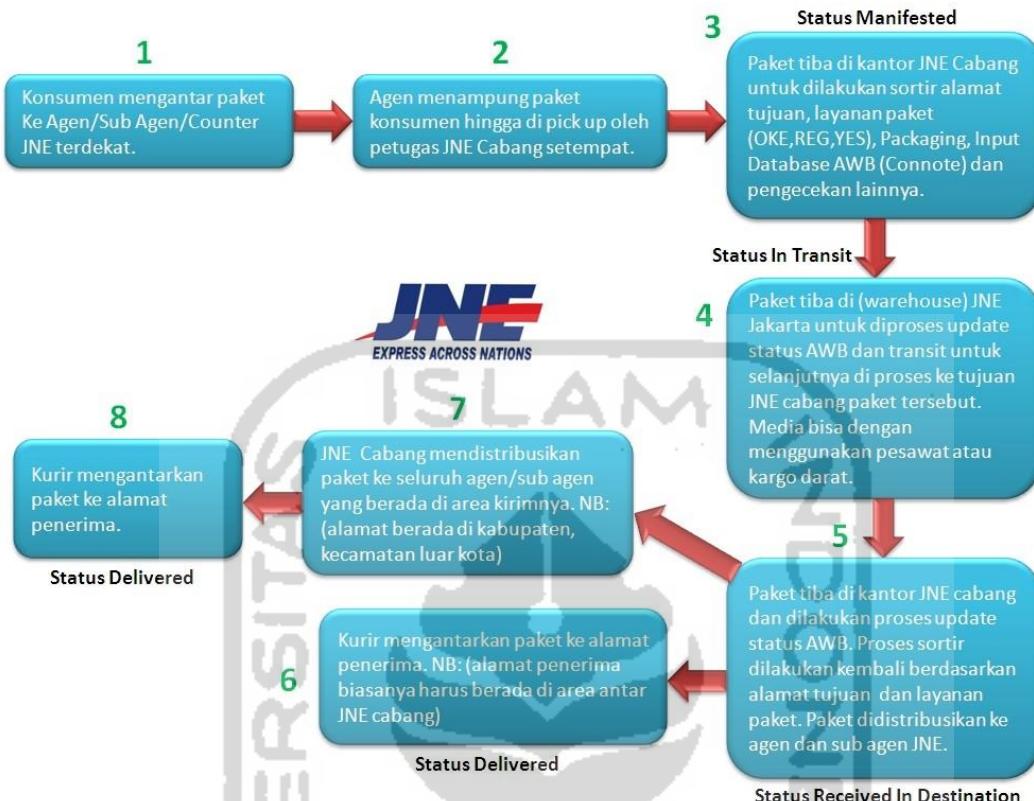
1. *Super Speed (SS)*
2. Pesanan Oleh-oleh Nusantara (PESONA)
3. Yakin Esok Sampai (YES)
4. Reguler (REG)
5. Ongkos Kirim Ekonomis (OKE)

Selain produk dan layanan diatas, terdapat juga layanan lainnya yang masih termasuk dalam *JNE Express*, yaitu *JNE Loyalty Card (JLC)*, *Money Remittance*, *Jemput Asi Seketika (JESIKA)*, *JNE Pick-Up Point (JNE PIPO)*, *@box*, *Diplomat*, *JNE Online Payment (JOP)*, *JNE Online Booking (JOB)*, *JNE Trucking (JTR)*, *Pop Box*, dan lain sebagainya. Selain *JNE Express*, terdapat juga *JNE Freight* dan *JNE Logistic* yang mana melayani layanan baik angkutan darat, laut, udara, dan juga pergudangan.

*JNE* merupakan perusahaan yang menawarkan kecepatan, kehandalan, dan komitmen tanggung jawab dalam melayani pelanggan. Kehandalan dan komitmen *JNE* ini terbukti dengan diraihnya berbagai bentuk penghargaan serta sertifikasi ISO 9001:2008 atas sistem manajemen mutu. Mulai dari tahun 1998 sampai dengan 2019 terdapat lebih dari 100 penghargaan yang diraih baik itu nasional maupun internasional. Salah satu penghargaan yang diraih adalah menjadi *1st Millennial's Choice in Courier Services Category*. Selain itu juga pada September 2019 *JNE* berhasil meraih *Oracle Award 2019* dengan Kategori *Cloud Platform Innovation* di Moscone Center, Amerika Serikat.

### 2.2.3. Alur Pengiriman JNE

JNE dalam melaksanakan proses pengiriman barang tentunya memiliki beberapa tahapan, sebagai berikut :



Gambar 2 2 Proses Pengiriman JNE

(Sumber : Dion Barus, 2019)

Berdasarkan Gambar 2.2, terdapat berbagai tahapan yang mana pengguna JNE dapat melacak paket yang dikirim melalui resi. Resi merupakan kode unik berupa nomor yang berguna untuk melacak dan mengidentifikasi keberadaan paket. Pada saat melacak nomor resi, akan terdapat tampilan status berupa beberapa istilah yang dapat dilihat pada Tabel 2.2 :

Tabel 2.2 Status Pengiriman

Status	Keterangan
<i>Shipment Received at JNE Officer</i>	Paket diterima dari pengirim atau barang baru didaftarkan di kantor JNE
<i>Shipment Picked up by JNE Courier</i>	Paket diambil oleh kurir JNE untuk dikirim ke pusat sorting

Status	Keterangan
<i>Received at Sorting Center</i>	Paket diterima digudang
<i>Departed From Transit</i>	Paket sedang dikirim untuk dipindahkan (transit) ke kota selanjutnya
<i>Shipment Forwarded to Destination</i>	Paket diteruskan ke cabang atau agen JNE yang terdekat dengan kota tujuan
<i>Received at Inbound Station</i>	Paket diterima pada cabang atau agen JNE terdekat
<i>Misroute</i>	Salah rute pengiriman
<i>Received at Origin Gateway</i>	Paket sudah diterima gudang yang berada di kota pengirim, dan nantinya akan dikirim ke pusat gudang yang berada di kota tujuan.
<i>Received at Warehouse</i>	Paket sudah sampai di gudang kota tujuan
<i>With Delivery Courier</i>	Paket akan dikirim oleh kurir ke alamat tujuan
<i>Delivered</i>	Paket telah diterima

#### 2.2.4. Google Play

*Google play* merupakan layanan digital milik *Google* yang didalamnya terdapat layanan berupa toko untuk memasarkan produk berupa aplikasi, permainan, hingga buku. *Google Play* dapat diakses melalui *playstore* (aplikasi android), *website* dan *Google TV*(Karch, 2019). *Google play* awalnya bernama *Android market*, yang dirilis perdana pada 22 Oktober 2008 dan kemudian menjadi *Google Play* yang dikenalkan pada Maret 2012 untuk menggantikan *Android Market* dan Layanan Musik *Google*.

Berdasarkan data dari Teknologi.id diketahui bahwa pada 2018 aplikasi yang tersedia di *Google Play* telah mencapai lebih dari 3,6 juta. Setelah beberapa bulan, terdapat lebih dari sepuluh ribu komentar tekstual dari aplikasi baru yang diluncurkan di *Google Play* (Liu et al., 2016). *Google play* memiliki beragam fitur yang mana salah satunya adalah pengguna dapat memberikan *review* terkait aplikasi dan lain sebagainya. Oleh karena ini *Google Play* menjadi salah satu sumber informasi terkait komentar pengguna untuk dapat dilakukan analisis sentimen.

### **2.2.5. *Online Review***

Keluhan dan saran merupakan salah satu dari empat metode yang dapat digunakan untuk mengukur kepuasan pelanggan (Kotler & Amstrong, 1996). Keluhan dan saran dapat disampaikan melalui berbagai media, baik melalui kuesioner tertulis maupun komentar yang terkoneksi dengan internet (*online review*). Menurut Mudambi & Schuff (2010) pada jurnal “*What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com*”, *Online Review* diartikan sebagai hasil evaluasi produk yang diunduh dalam situs perusahaan atau situs pihak ketiga.

Dalam penelitian ini, *online review* merupakan ulasan atau komentar terkait evaluasi dan apa yang dirasakan terkait aplikasi, film, buku, dan berbagai hal lainnya yang disampaikan melalui perangkat (komputer, handphone, dll) dan terkoneksi dengan internet pada suatu situs baik milik perusahaan maupun pihak orang ketiga.

### **2.2.6. *Web Scraping***

*Website* merupakan aplikasi yang didalamnya terdapat berbagai dokumen multimedia (teks, gambar, animasi, video) yang menggunakan protokol HTTP (*Hypertext Transfer Protocol*) dan untuk mengaksesnya digunakan perangkat lunak yang disebut *browser*. Sedangkan *Web Scraping* adalah proses pengambilan sebuah dokumen semiterstruktur dari internet, umumnya berupa halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain (Turland, 2010). Terdapat langkah yang perlu dilakukan untuk melakukan *web scraping* (Josi, et al., 2014) yaitu sebagai berikut :

1. *Create Scraping Template* : Pembuat program mempelajari dokumen HTML dari *website* yang akan diambil informasinya dari tag HTML yang mengapit informasi yang akan diambil
2. *Explore Site Navigation* : Pembuat program mempelajari teknik navigasi pada *website* yang akan diambil informasinya untuk ditirukan pada aplikasi *web scraper* yang akan dibuat
3. *Automate Navigation and Extraction* : Berdasarkan informasi yang didapatkan dari langkah 1 dan 2 diatas, aplikasi *web scraper* dibuat agar pengambilan informasi dapat dilakukan secara otomatis dari *website* yang ditentukan.

4. *Extracted Data and Package History* : Informasi yang didapat dari langkah 3 disimpan dalam tabel atau tabel-tabel *database*.

### **2.2.7. Machine Learning**

*Machine learning* merupakan pendekatan dalam AI (*Artificial Intelligence*) yang mencoba menirukan bagaimana manusia sebagai makhluk yang cerdas berproses dalam belajar dan melakukan generalisasi, sehingga pada *Machine learning* hal-hal tertentu dapat dijalankan secara otomatis (Tanaka & Okutomi, 2014). *Machine learning* adalah bidang yang mempelajari pengembangan algoritma komputer untuk mengubah data menjadi aksi yang cerdas atau secara singkat dapat juga diartikan sebagai proses mengubah data menjadi informasi (Suyanto, 2017).

Sedangkan Mohri et al., (2012) mendefinisikan *Machine learning* sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat. Pengalaman pada hal ini diartikan sebagai informasi yang sebelumnya telah tersedia dan dijadikan sebagai data pembelajaran. Dalam pembelajaran *Machine learning*, terdapat beberapa skenario, antara lain sebagai berikut:

1. *Supervised Learning* : pembelajaran menggunakan input berupa data training yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.
2. *Unsupervised Learning* : pembelajaran menggunakan input data training yang tidak diberi label. Setelah itu mencoba untuk mengelompokan data berdasarkan karakteristik-karakteristik yang ditemui.
3. *Reinforcement Learning* : data training dan data testing saling dicampur. Untuk mengumpulkan informasi pembelajaran secara aktif dengan berinteraksi ke lingkungan sehingga mendapatkan balasan untuk setiap aksi dari pembelajaran.

### **2.2.8. Text Mining**

#### **1. Pengertian Text Mining**

*Text Mining* merupakan teknik penambangan data yang digunakan untuk menganalisis dan mengklasifikasikan dokumen atau data berupa teks guna menemukan pola data dan informasi yang bermanfaat untuk tujuan tertentu sehingga dapat dilakukan analisis keterhubungan antar dokumen tersebut (Aditya, 2015). Sumber data yang digunakan pada *Text Mining* adalah kumpulan teks yang

memiliki format yang tidak terstruktur atau semi terstruktur. Adapun tugas khusus dari *Text Mining* antara lain yaitu pengkategorisasian teks (*text categorization*) dan pengelompokan teks (*text clustering*).

## **2. *Text Preprocessing***

*Text Mining* memiliki sumber data berupa dokumen yang tidak terstruktur. Berdasarkan ketidakakuratan struktur data tersebut maka diperlukan tahap text preprocessing untuk mempersiapkan teks agar menjadi lebih terstruktur. Proses yang dilakukan pada *Text Preprocessing* dilakukan dengan tahapan sebagai berikut:

- a. *Cleaning* : *Cleaning* merupakan tahap yang dilakukan untuk membersihkan dokumen dari kata-kata yang tidak diperlukan dan tidak dibutuhkan untuk mengurangi *noise*. Selain pembersihan kata, dilakukan juga penghilangan tanda baca seperti titik (.), koma (,), dan tanda baca lainnya (~!@#\$%^&\*()\_-+=[]{};”.>,<). Selain itu dilakukan pengubahan huruf dalam dokumen menjadi huruf kecil semua (*lower-case*). Pengubahan menjadi huruf kecil ini biasa disebut juga *case-folding*.
- b. *Filtering* : *Filtering* atau biasa juga disebut *stopword removal* merupakan penghapusan kata-kata yang kurang bermakna atau kurang penting seperti “dan”, “atau”, “saya”, dan lain sebagainya (Putri, 2016).
- c. *Tokenizing* : *Tokenizing* merupakan tahap pemotongan kata atau memisahkan teks menjadi kata atau potongan kalimat yang biasa disebut juga token. Kata dalam dokumen yang dipisahkan adalah kata pada kalimat yang terpisah oleh spasi.
- d. *Stemming* : *Stemming* merupakan tahap mengubah kata menjadi kata dasar (*root*) dari setiap kata hasil proses *filtering* yang mengandung imbuhan.

### **2.2.9. *Sentiment Analysis***

Analisis Sentimen atau *opinion mining* merupakan studi komputasi mengenai pendapat, perilaku, dan emosi seseorang terhadap entitas yang mana entitas tersebut dapat mewakili individu, kejadian atau topik (Kristiyanti, 2015). Melalui analisis sentimen dapat diketahui opini publik terkait suatu hal baik berupa produk, jasa, aplikasi, film, musik, dan lain sebagainya melalui review pengguna pada berbagai situs. Analisis sentimen digunakan

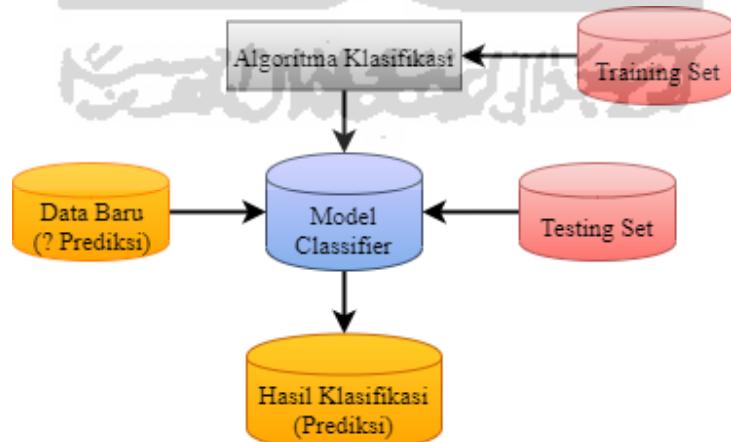
untuk melihat kecenderungan apakah pendapat atau opini pada suatu masalah cenderung termasuk dalam opini positif atau opini negatif (Rozi, Pramono, & Dahlan, 2012).

### 2.2.10. Classification

Klasifikasi merupakan teknik analisis data yang digunakan untuk menghasilkan model prediksi untuk mendeskripsikan label atau kelas data (Han & Kamber, 2006). Sedangkan menurut Prasetyo, (2012) klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Klasifikasi tergolong dalam *supervised method*.

Menurut Han & Kamber, (2006) *classification* memiliki dua tahap proses. Tahap pertama adalah membangun suatu model yang berdasarkan serangkaian *data class* yang disebut *learned model*. Model tersebut dibangun dengan menganalisa *record database*. Setiap *record* diasumsikan menjadi *predefined class* yang ditentukan oleh suatu atribut yang disebut *class label* atribut. Akibat terdapat *class label* maka tahap ini juga dikenal dengan *supervised learning*. Berbeda dengan *unsupervised learning* atau dikenal dengan *clustering* yang tidak memerlukan *class label*. Tahap pertama ini juga disebut sebagai tahap pembelajaran. Sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis data training. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan  $y = f(x)$  dimana  $y$  adalah kelas hasil prediksi dan  $x$  adalah record yang ingin diprediksi class-nya.

Berikut merupakan bagan proses klasifikasi data sampel menggunakan model *classifier* untuk mendapatkan hasil prediksi :



Gambar 2.3 Klasifikasi menggunakan *Model Classifier*

Sumber : (Prasetyo, 2006)

Beberapa persiapan yang dilakukan untuk mendapatkan hasil klasifikasi yang baik di antaranya adalah (Han & Kamber, 2006):

1. Pembersihan Data

Pembersihan data ini dilakukan untuk mengurangi kecacatan data di dalam data pelatihan, beberapa metode yang digunakan di antaranya dengan teknik *smoothing* untuk menghilangkan *noise data*, melengkapi data yang hilang dan sebagainya.

2. Analisis Relevansi

Dari beberapa atribut yang akan digunakan untuk proses klasifikasi mungkin saja terdapat atribut yang sangat berhubungan kuat satu sama lain, kedua atribut ini memiliki kemiripan sehingga menyebabkan proses klasifikasi menjadi tidak optimal, maka salah satu dari atribut ini dapat dibuang.

Hasil klasifikasi dan prediksi dapat dievaluasi menggunakan beberapa kriteria (Han & Kamber, 2006) :

1. Akurasi

Akurasi digunakan untuk mengetahui kemampuan model klasifikasi untuk dapat memberikan ketepatan hasil prediksi.

2. Kecepatan

Mengetahui kecepatan iterasi untuk mendapatkan model klasifikasi dan iterasi mendapatkan hasil prediksi.

### 2.2.11. Fitur dan Pembobotan

Pembobotan merupakan metode untuk mengubah input data menjadi suatu fitur vektor. Metode pembobotan yang umum digunakan adalah *bag-of-feature*. Sebagai contoh terdapat sederetan fitur seperti pada vektor  $\{f_1, f_2, \dots, f_n\}$  dimana yaitu sekumpulan fitur-fitur sebanyak  $n$  yang sudah ditentukan sebelumnya. Misalkan kata “puas” maka fitur vektor dari data adalah vektor.

- a. *Term Presence*

*Term Presence* (TP) ialah metode pembobotan pada suatu dokumen teks yang melihat keberadaan daftar kata-kata (*term*) atau fitur yang ada pada corpus terhadap suatu dokumen. Jika suatu fitur yang ada pada daftar fitur acuan terdapat pada dokumen yang sedang diboboti maka nilai fitur tersebut pada *feature vector* akan diberi nilai 1 dan

tidak menghiraukan jumlah kemunculan fitur tersebut. Jika fitur tersebut tidak ada pada dokumen maka diberi nilai 0 pada *feature space* (O'Keefe & Koprinska, 2009)

b. *Term Frequency*

*Term Frequency* (TF) memiliki kesamaan dengan TP yang telah dijelaskan sebelumnya, tapi yang membedakan adalah TF menghitung jumlah kemunculan fitur acuan pada suatu dokumen bukan hanya keberadaan fitur tersebut (O'Keefe & Koprinska, 2009).

c. *Term Frequency – Inverse Document Frequency* (TF-IDF)

*Term Frequency – Inverse Document Frequency* (TF-IDF) merupakan algoritma pembobotan tersusun dari dua nilai yang berasal dari dua algoritma dengan pembobotan yang berbeda, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Output* dengan fitur/term tersebut dengan kata yang sering muncul pada dokumen akan menghasilkan nilai TF-IDF yang tinggi. Sementara, fitur yang sering muncul pada dokumen akan menghasilkan nilai rendah. Dengan menggunakan metode ini terms atau fitur-fitur yang penting akan memiliki nilai tinggi dan sebaliknya untuk *term* atau fitur yang tidak penting akan memiliki nilai yang rendah (O'Keefe & Koprinska, 2009).

### 2.2.12. *Naïve Bayes Classifier*

*Naive Bayes Classifier* merupakan teknik klasifikasi berdasarkan Teorema *Bayes* dengan asumsi independensi di antara para prediktor. *Naive Bayes Classifier* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema *Bayes*. *Naïve bayes* merupakan metode paling sederhana dari pengklasifikasian probabilitas, memiliki tingkat akurasi yang sangat tinggi ketika diaplikasikan pada *database* dengan *big data* (Wati, 2016). Selain itu model *naïve bayes* juga memiliki tingkat kesalahan yang sangat minimum dibandingkan dengan algoritma klasifikasi lainnya (Liu et al., 2016)

Algoritma *naïve bayes* hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai variabel *independent*, maka hanya variansi dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan

klasifikasi, bukan keseluruhan dari matriks kovarians. Rumus Bayes dalam (Muslehatin dkk, 2017) secara umum adalah sebagai berikut:

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \quad (2.1)$$

dengan :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$  = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$  = Probabilitas hipotesis H (prior prob.)

$P(X|H)$  = Probabilitas X berdasarkan kondisi tersebut

$P(X)$  = Probabilitas dari X

Adapun aturan *Bayes* adalah sebagai berikut :

Jika  $P(h_1|x) < P(h_2|x)$ , maka x diklasifikasikan sebagai h2. Pernyataan  $P(h_1|x)$  mengindikasikan probabilitas hipotesis h1 berdasarkan kondisi x terjadi, begitu pula dengan h2. Sehingga didapat klasifikasi dari x sesuai dengan probabilitas terbesar diantara probabilitas x terhadap semua kelas.

### 2.2.13. Asosiasi Kata

Istilah korelasi sering digunakan untuk menyatakan hubungan dua atau lebih variabel yang sifatnya kuantitatif, sedangkan istilah asosiasi sering dimaknai keeratan hubungan antara dua atau lebih variabel yang sifatnya kualitatif (Ulwan, 2016). Tujuan analisis korelasi adalah untuk mencari hubungan variabel bebas (X) dengan variabel terikat (Y), dengan ketentuan data memiliki syarat-syarat tertentu (Fadlisyah, 2014).

Asosiasi merupakan proses pencarian hubungan antar elemen data. Pada penelitian ini digunakan pendekatan asosiasi untuk menemukan hubungan antar ulasan pada masing-masing klasifikasi ulasan positif dan ulasan negatif pengguna sehingga mendapatkan informasi yang dapat dijadikan bahan rujukan dalam pengembangan serta untuk meningkatkan kualitas perusahaan.

#### 2.2.14. *Diagram Fishbone*

*Diagram fishbone* atau dikenal dengan “Diagram Ishikawa” dikembangkan oleh De. Kaoru Ishikawa sekitar tahun 1943. *Diagram fishbone* umumnya digunakan untuk mengidentifikasi permasalahan dan menentukan penyebab dari permasalahan tersebut (Fritz, 2016). Diagram ini berguna untuk menganalisis dan menemukan faktor-faktor yang berpengaruh atau efek secara signifikan di dalam menentukan karakteristik kualitas output kerja. Efek ini bisa bernilai "baik" dan bisa bernilai "buruk". Jadi dengan diketahui sebab dari efek yang terjadi, diharapkan hasil dari proses produksi bisa diperbaiki dengan mengubah faktor terkontrol dari suatu proses. Diagram ini juga berguna untuk mengidentifikasi akar penyebab potensi dari suatu masalah. Diagram sebab akibat memfokuskan pada penekanan masalah atau gejala yang merupakan akar penyebab masalah. Diagram sebab akibat juga menampilkan penyebab masalah dengan cara menghubungkan penyebab-penyebab menjadi satu (Fauziah, 2009).

## BAB III

### METODE PENELITIAN

#### 3.1. Objek dan Subjek Penelitian

Objek penelitian ini adalah salah satu jasa kurir terbesar di Indonesia yaitu PT. Tiki Jalur Nugraha Ekakurir atau yang biasa disebut JNE. Adapun subjek penelitian ini adalah persepsi pengguna JNE terhadap pelayanan JNE melalui *review* pada *database website Google Play* yang hingga 31 Desember 2019 telah mencapai sekitar 87.589 komentar sejak awal pertama kali aplikasi My JNE dirilis dan akan terus bertambah jumlahnya seiring berjalannya waktu.

#### 3.2. Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah semua data *review* atau ulasan pengguna PT. Tiki Jalur Nugraha Ekakurir (JNE) dari *database website Google Play* dengan aplikasinya yang bernama My JNE dengan total 87.589 komentar. Sedangkan sampel diambil dari suatu populasi yang mana pada penelitian ini digunakan sampel ulasan pengguna JNE sejak adanya *update* aplikasi My JNE versi terbaru yaitu 2.3.12, pada tanggal 19 September 2019 sampai dengan 31 Desember 2019. Dengan jumlah sampel sebanyak 1.876 komentar.

#### 3.3. Pengumpulan Data

##### 3.3.1. Metode Pengumpulan Data

Metode yang digunakan untuk proses pengumpulan data dari penelitian ini adalah sebagai berikut :

###### 1. *Web Scraping*

Data pada penelitian ini adalah data sekunder yang diambil secara online dengan teknik *scraping data* menggunakan ekstensi *Scraper* bawaan yang ada di *Google Chrome* yaitu *Data Scraper*. Melalui teknik *scraper* ini didapatkan informasi dari *website* secara otomatis tanpa perlu menyalin secara manual.

## 2. Studi Pustaka

Studi pustaka dilakukan dengan menggunakan beberapa kajian literatur, buku, maupun referensi jurnal yang sekiranya berkaitan dengan tujuan, rumusan, batasan, dan metode penelitian.

### 3.3.2. Jenis dan Sumber Data

Jenis data yang digunakan pada penelitian ini adalah data sekunder. Pada penelitian ini data sekunder berperan sebagai *main data* dalam proses pengolahan untuk mengetahui persepsi pengguna terhadap pelayanan yang diberikan oleh perusahaan JNE. Adapun sumber data yang digunakan dalam penelitian ini adalah *dataset reviews* terhadap kepuasan pengguna JNE terhadap pelayanan yang diberikan di *website Google Play*, yang mana peneliti mengambil secara *online* dengan *link* berikut ini :

<https://play.google.com/store/apps/details?id=com.indivara.jneone>.

Pengambilan data secara *online* ini dinamakan *web scrapping* atau *scrapping data*. Adapun *scraping* data pada penelitian ini dilakukan dengan menggunakan *Data Scraper*. *Google Chrome* memiliki berbagai jenis *extension* (ekstensi) yang mana salah satunya adalah *Data Scraper* yang berfungsi sebagai *web scraper* untuk mengambil data atau konten yang terdapat pada situs *online* secara gratis. *Data Scraper* mengekstraksi atau mengambil data dari halaman web HTML dan mengimpornya ke dalam *spreadsheet Microsoft Excel*. Data hasil ekstraksi ini yang kemudian akan diolah untuk dapat mengetahui persepsi pengguna JNE.

### 3.3.3. Variabel dan Definisi Operasional Variabel

Variabel penelitian ini digunakan sebagai batasan yang diperhatikan dalam memilih *review* atau ulasan pengguna JNE di *website Google Play*. Adapun variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

Tabel 3.1 Variabel Penelitian

Variabel	Definisi Operasional Variabel
Waktu ( <i>Date</i> )	Waktu berupa tanggal, bulan, dan tahun saat pengguna JNE mengisi komentar pada website <i>Google Play</i> . Variabel waktu perlu diperhatikan, mengingat <i>review</i> yang diambil adalah berupa sampel dengan waktu tertentu yang telah ditentukan sebelumnya.
Ulasan ( <i>Review</i> )	Isi komentar pengguna terhadap pelayanan JNE di website <i>Google Play</i> .
<i>Rating</i>	<i>Rating</i> yang diberikan pengguna terkait pelayanan JNE di website <i>Google Play</i> berupa bintang 1 sampai dengan 5

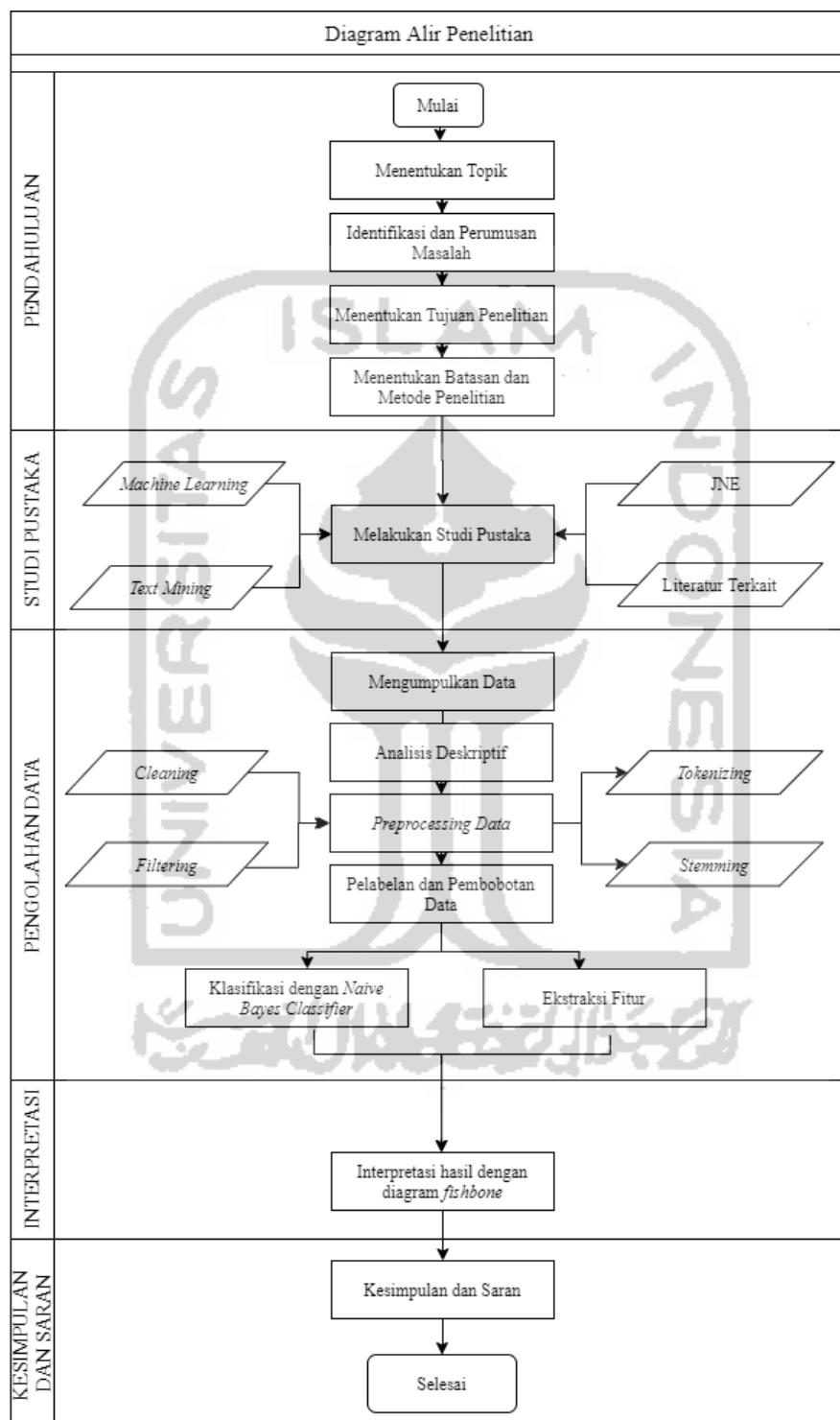
### 3.4. Metode Analisis Data

Berikut ini merupakan metode analisis data yang digunakan dalam penelitian ini, antara lain :

1. Metode analisis deskriptif, yang digunakan untuk menggambarkan secara umum *review* pengguna JNE melalui aplikasi my JNE pada situs website *Google Play*.
2. Metode *Machine learning* yaitu *Naïve bayes Classifier* yang digunakan untuk melakukan klasifikasi *review* yang berbentuk kelas positif maupun negatif.
3. *Wordcloud*, digunakan untuk melakukan visualisasi kata yang paling banyak digunakan atau paling sering muncul pada komentar pengguna di aplikasi my JNE.
4. *Association*, digunakan untuk mengidentifikasi dan membentuk pola kata yang berasosiasi dengan kata lainnya untuk mendapatkan informasi yang dianggap penting.
5. Diagram *fishbone* (sebab-akibat), digunakan untuk mengidentifikasi *review* dengan kelas negatif.

### 3.5. Alur Penelitian

Berikut merupakan alur dari penelitian ini :



Gambar 3.1 Alur Penelitian

Berdasarkan diagram alir penelitian, langkah-langkah pada penelitian ini terdiri atas sebagai berikut :

1. Mulai
2. Menentukan Topik

Langkah pertama yang perlu dilakukan adalah menentukan topik penelitian yang akan dilakukan. Adapun topik dari penelitian ini adalah analisis sentimen pada salah satu perusahaan jasa pengiriman terbesar di Indonesia yaitu PT. Tiki Jalur Nugraha Ekakurir atau yang biasa disebut JNE.

3. Identifikasi Perumusan Masalah

Setelah mengetahui topik dari penelitian, selanjutnya adalah melakukan identifikasi dan perumusan masalah. Pada tahap ini dilakukan identifikasi masalah namun bukan dengan pengamatan secara langsung, tetapi dengan melihat kondisi yang ada sekarang ini melalui *website* seperti *Google Play*. Mengingat JNE merupakan salah satu dari beberapa jasa pengiriman yang sangat sering digunakan oleh berbagai *e-commerce* dalam mendistribusikan barangnya untuk sampai ke tangan konsumen, maka perlu untuk dapat mengetahui analisis sentimen penggunanya. Kemudian dilakukan perumusan masalah yang menjadi pedoman dari penelitian ini. Adapun rumusan masalah pada penelitian ini antara lain terkait dengan bagaimana gambaran umum persepsi pengguna JNE berdasarkan ulasan pada aplikasi My JNE di *website Google Play*, bagaimana performa algoritma *Naïve bayes Classifier* dalam mengklasifikasikan persepsi pengguna JNE menjadi kelas positif dan negatif, Informasi apa yang didapatkan berdasarkan klasifikasi dan asosiasi yang telah dilakukan, dan faktor apa saja yang perlu dilakukan untuk memperbaiki hasil ulasan negatif yang didapat?

4. Menentukan Tujuan Penelitian

Pada tahap menentukan tujuan penelitian, dilakukan penentuan sekiranya hal apa saja yang ingin tercapai berdasarkan rumusan masalah. Tujuan penelitian menunjukkan hasil yang akan diperoleh dari penelitian ini. Sehingga tujuan penelitian ini adalah mengetahui gambaran umum persepsi pengguna JNE berdasarkan ulasan pada aplikasi My JNE di *website Google Play*, mengetahui performa algoritma *Naïve bayes Classifier* dalam mengklasifikasikan persepsi

pengguna JNE, memperoleh informasi berdasarkan klasifikasi dan asosiasi yang telah dilakukan, dan mengetahui faktor apa saja yang perlu dilakukan untuk memperbaiki hasil ulasan negatif yang didapat.

#### 5. Menentukan Batasan Metode Penelitian

Tentunya pada penelitian ini pun terdapat batasan yang harus dipenuhi agar penelitian tetap terarah dan tidak melebar membahas hal yang tidak perlu. Dalam penelitian ini batasan penelitian berupa rentang waktu *review* pengguna, yakni hanya komentar pada tanggal 19 September 2019 sampai dengan 31 Desember 2019 saja dengan bahasa yang digunakan adalah Bahasa Indonesia. Adapun pengambilan data hanya pada website *Google Play*. Setelah batasan, selanjutnya ditentukan metode yang sekiranya cocok untuk penelitian ini.

#### 6. Melakukan Studi Pustaka

Untuk dapat mengetahui metode yang sekiranya tepat untuk penelitian ini, maka dilakukan studi pustaka yang berkaitan dengan topik yang telah ditentukan yaitu analisis sentimen. Berikut ini beberapa studi pustaka yang menjadi referensi :

- a. *Machine learning*
- b. *Text Mining*
- c. JNE
- d. Literatur Terkait seperti mengenai *scraping data*, pengolahan data analisis sentimen menggunakan *software R*, dan lain-lain.

#### 7. Mengumpulkan Data

Pada tahap ini dilakukan pengumpulan data, namun berbeda dengan penelitian lainnya karena bukan dengan datang langsung ke perusahaan. Pengumpulan data yang akan diolah dilakukan dengan mengekstraksi data *review* secara *online* dengan teknik *scraping data* menggunakan ekstensi dari *google chrome*, yaitu *Data Scraper*. *Web scraping* dilakukan pada *review* pengguna JNE melalui aplikasi yang terdapat di *Google Play* yaitu My JNE.

#### 8. Analisis Deskriptif

Kemudian dilakukan analisis deskriptif, analisis ini digunakan untuk menggambarkan persepsi pengguna dari JNE berupa waktu *review* pengguna, *rating* yang diberikan, dan lain sebagainya atau bisa diartikan juga sebagai gambaran

secara umum *review* pengguna JNE melalui aplikasi my JNE pada situs *website Google Play*.

#### 9. *Preprocessing Data*

*Preprocessing* berguna untuk menyeleksi data dan mengubahnya menjadi data yang terstruktur dan merupakan tahapan awal untuk mengubah struktur isi dari data untuk menjadi format yang sesuai agar dapat diproses oleh *algoritma Naïve bayes Classifier*. Proses dari *preprocessing* ini dilakukan menggunakan *software R* dan *Python*, dengan proses yang dilakukan adalah berikut ini :

- a. *Cleaning* : Tahap membersihkan dokumen dari kata yang tidak diperlukan, tanda baca (~!@#\$%^&\*()\_-+=[]{};''>,<), dan mengubah huruf dalam dokumen menjadi huruf kecil semua (*lower-case*).
- b. *Filtering* : tahap menghapus kata-kata yang kurang bermakna atau tidak memiliki arti seperti kata: saya, dan, atau.
- c. *Tokenizing* : tahap pemisahan teks menjadi potongan kalimat dan kata atau yang biasa disebut token.
- d. *Stemming* : tahap pencarian kata dasar dari tiap kata pada teks

#### 10. Pelabelan dan Pembobotan Data

- a. Klasifikasi Dengan *Naïve bayes Classifier*

Metode *Machine learning* yaitu *Naïve bayes Classifier* yang digunakan untuk melakukan klasifikasi *review* yang berbentuk kelas positif maupun negatif. Adapun pada tahap ini digunakan *software R*.

- b. Visualisasi dan Asosiasi

Visualisasi dan Asosiasi dilakukan pada semua data baik yang merupakan sentimen positif maupun negatif. Visualisasi ini dilakukan dengan mengekstraksi keseluruhan informasi sehingga dapat diketahui kata yang paling sering digunakan pada komentar pengguna. Oleh karena ini bentuk dari visualisasi dan asosiasi ini nantinya akan berbentuk *wordcloud* dan *barplot*.

#### 11. Interpretasi hasil dengan diagram *Fishbone*

Setelah mengetahui visualisasi dari sentimen negatif, langkah selanjutnya adalah membuat diagram *fishbone* untuk menganalisis faktor-faktor yang menjadi masalah

di mata pengguna JNE dan mengidentifikasi hal apa saja yang perlu dilakukan dalam rangka menanggulangi permasalahan tersebut.

12. Penutup (Kesimpulan dan Saran)

Tahap terakhir dari penelitian ini adalah menarik kesimpulan dan saran berdasarkan dari hasil penelitian.

13. Selesai

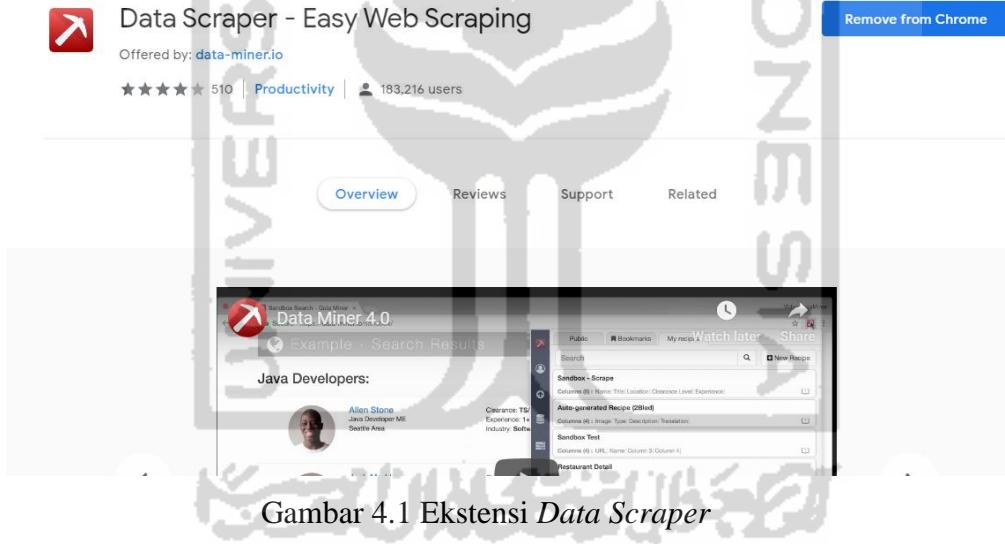


## BAB IV

### PENGUMPULAN DAN PENGOLAHAN DATA

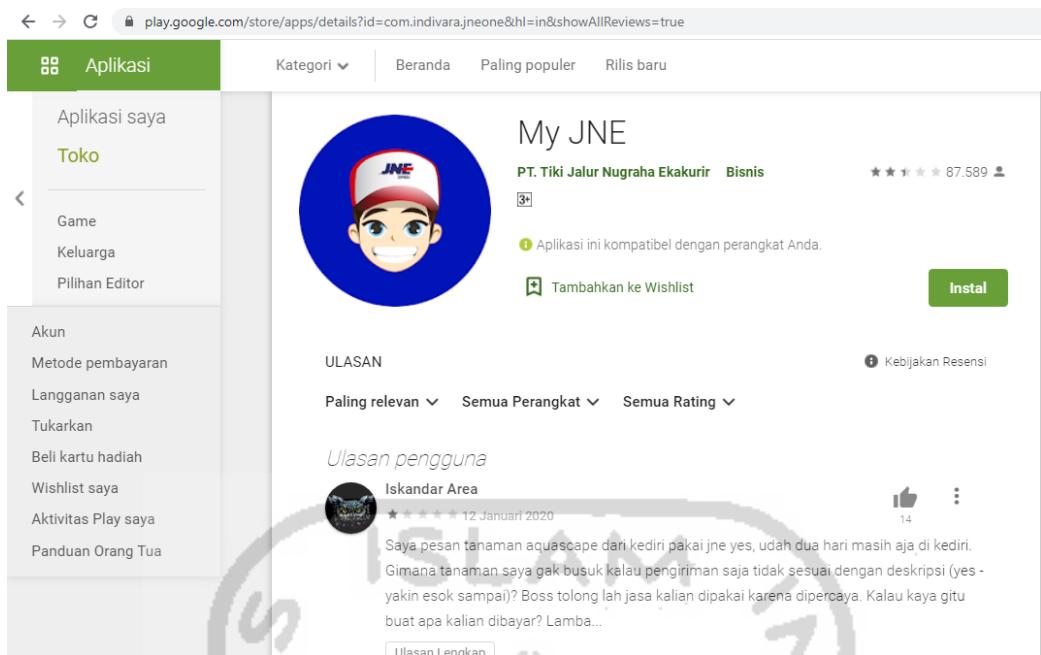
#### 4.1. Pengumpulan Data

Pada penelitian ini dilakukan pengambilan data berupa ulasan pengguna JNE melalui aplikasi my JNE pada situs *Google Play*. Pengambilan data dilakukan dengan teknik *scraping* menggunakan ekstensi dari *Google Chrome*. Terdapat beragam ekstensi dari *Google Chrome* yang dapat digunakan untuk mengambil data pada situs *online*. Salah satunya yaitu *Data Scraper*. *Data Scraper* dapat digunakan untuk mengekstraksi data pada situs secara gratis. Adapun pada penelitian ini digunakan *Data Scraper* versi 3.299.84. Data yang diekstraksi dapat diimpor menjadi *spreadsheet Microsoft Excel* maupun CSV.



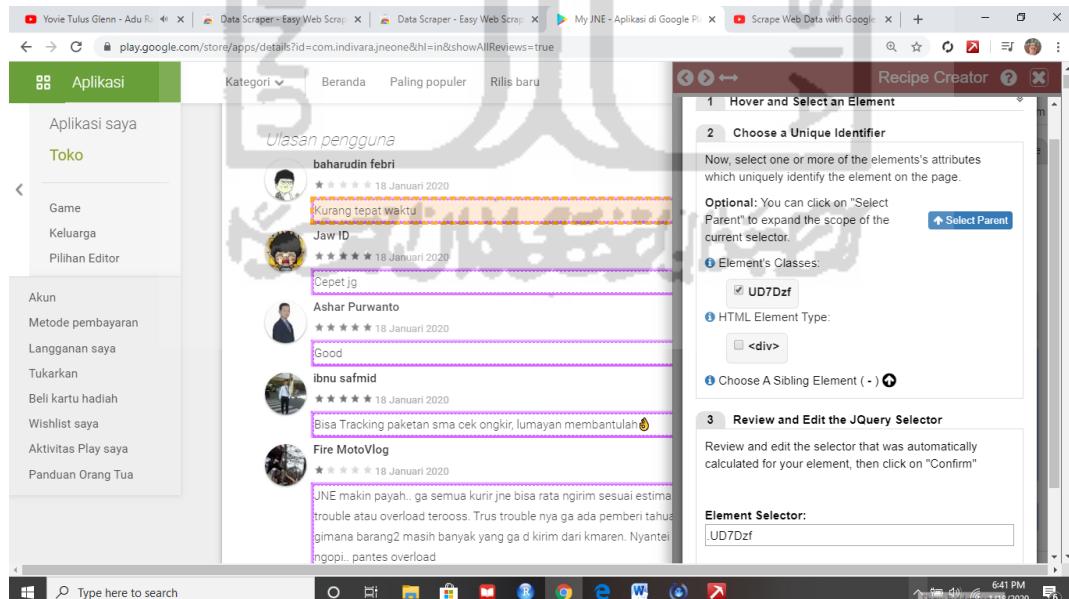
Gambar 4.1 Ekstensi *Data Scraper*

Langkah pertama yang perlu dilakukan adalah menginstall dan mengaktifkan ekstensi *Data Scraper* pada *Google Chrome*. Kemudian dapat dilakukan pengambilan data sesuai dengan alamat *domain* yang kita inginkan. Adapun pada penelitian ini data yang digunakan bersumber dari ulasan pengguna JNE dengan alamat *domain* berikut : <https://play.google.com/store/apps/details?id=com.indivara.jneone&hl=in&showAllReviews=trueadapun>.



Gambar 4.2 Website my JNE

Langkah selanjutnya adalah membuka alamat *domain* atau *website* seperti pada Gambar 4.2 selanjutnya adalah memilih data sesuai dengan batasan yang telah ditentukan. Pertama pada *start* dapat memilih sesuai dengan jenis data pada *website*. Selanjutnya klik *Cols*. Pilih sesuai dengan data yang akan kita ambil, dalam penelitian ini yaitu waktu (tanggal) dan juga komentar. Dapat dilihat untuk pembuatan setiap *query* pada *Cols* akan terdapat tampilan seperti pada Gambar 4.3 dan Gambar 4.4 :



Gambar 4.3 Pembuatan *Query*

The screenshot shows a web-based application for scraping reviews. On the left, a list of reviews is displayed with columns for Kategori, Beranda, Paling populer, and Rilis baru. The reviews are from users Muhammad Alfarizy, Zim bae, Arif widianto, Wahyu Rd, and Ivank Reey, each with a star rating and a timestamp. On the right, a configuration panel titled 'Columns' is shown. It includes a note about specifying columns for a CSV output file, a 'Cols video' icon, and a 'Find' tool to select elements. The 'Waktu' column is selected, with a selector of '.p2TkOb'. A second column is listed as 'Column 2'. Below this, a 'Find' dialog is open with 'Name: Komentar', 'Extract: Text', and 'Selector: .UD7Dzf'. Buttons for 'Add New Column' and 'Save' are at the bottom.

Gambar 4.4 Kolom *Query*

Pada penelitian ini terdapat tiga kolom yang terdiri dari waktu, komentar dan juga *rating*. Selanjutnya dapat dilakukan *scraping* dengan mengklik *Save*, mengisi *Recipe Name*, dan selanjutnya klik *Run Recipe* seperti pada Gambar 4.5 :

The screenshot shows the 'Recipe Creator' interface. At the top, there are buttons for 1 Start, 2 Rows, 3 Cols, 4 Nav, 5 Actions, 6 JS, and 7 Save. A note says 'Lastly, give a name to your recipe and then Save and Run it, to see your scraped data.' A 'Note' section says 'to run recipes with Actions or Javascript close Recipe Creator and open Data Miner to run.' Below this, a 'Recipe Name' field contains 'JNE'. A 'Recipe Description' field is empty. At the bottom, there are buttons for 'Save As New', 'Save', and a large blue 'Run Recipe' button.

Gambar 4.5 Membuat *Recipe*

Data hasil ekstraksi dapat disimpan dalam bentuk xlsx maupun csv. Adapun pada penelitian ini hasil ekstraksi data disimpan dalam file as .csv.

		Pagination:	Extracted Data:
		Not Available	New rows: 2280 Total: 2280
			<input type="button" value="Download"/> <input type="button" value="Clear"/>
Waktu	Komentar		
18 Januari 2020	Kurang tepat waktu		<input type="button" value="as .CSV"/>
18 Januari 2020	Cepet jd		<input type="button" value="as XLSX"/>
18 Januari 2020	Good		<input type="button" value="to Clipboard"/>
18 Januari 2020	Bisa Tracking paketan sma c...		

Gambar 4.6 Import Data

A	B	C
1 tanggal review		rating
2 31 Desem Sangat kecewa		1
3 31 Desem Pihak JNE nya udah bagus tepat waktu!!! Tapi kurirnya rese banget si, kalo sudah sampai knp ga telpon... Buat apa nomer		1
4 31 Desem Payah		1
5 31 Desem ini gimna oy paket gw udah 2 minggu lebih belum nyampung juga yang bener lah		1
6 31 Desem Tidak sesuai estimasinya 1-2hari tapi gak sampe2		1
7 31 Desem goooooood		5
8 31 Desem Saya terganggu dengan adanya JOB, karena penjual dapat mengirim barang kapan saja, dan memberi kode resi yg do book		1
9 31 Desem Buruk banget pelayanannya udh dekat diwiliyah gak di anter anter.. Taik emang JNE		1
10 31 Desem Semakin buruk,! D barang sudah ada no hp yg bisa di hubungi tapi kurir tidak menghubungi,! Kurirnya sangat2 buruk,! Gk r		1
11 31 Desem Karyawan JNE PEMALAS & PELIT PULSA		1
12 31 Desem Sangat kecewa sama JNE masa paketan dari Jogja di kirim ke Banjarnegara saya pantau lagi di Semarang aduh aduh Gusti		1
13 31 Desem bagus jadi aku tau udah di mana paketanku		5
14 31 Desem Kinerjanya jelek bgt,g becus kiriman saya udah dris tahun mezolitikum smpe skrg g nympe". Setiap nelfon ke jnenya sllu bi		1
15 31 Desem Jasa ekspedisi paling lambat paling lama dibanding yg lain, kalau tidak terpaksa bgt males pake jasa ini		1
16 31 Desem Mantap		5
17 31 Desem Haha JNE jatuh miskin oleh pesaing nya, jnnnn. Paket tak ada yg sampe, Mau nuntut, suruh ambil ke pusat daerah setemp		1

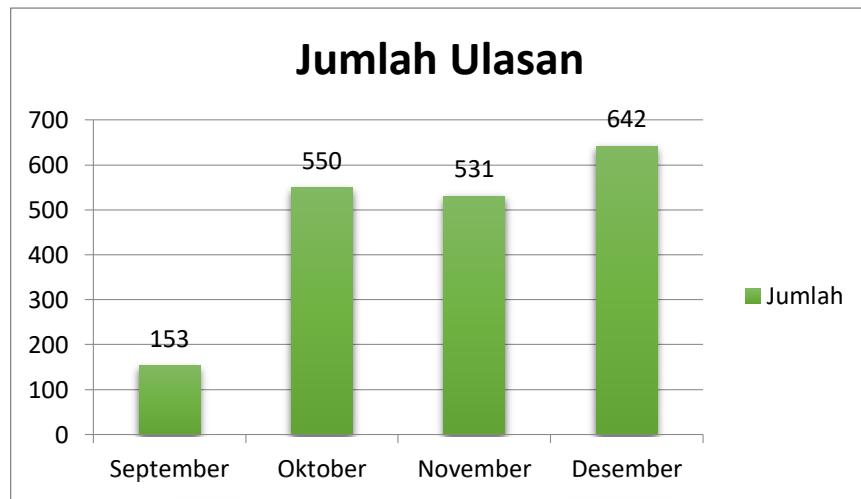
Gambar 4.7 Hasil scraping data

## 4.2. Pengolahan Data

Data yang telah diekstraksi dari *website* selanjutnya diolah agar didapatkan informasi dari data tersebut. Pengolahan data dilakukan pada ulasan pengguna JNE dengan sampel sebanyak 1.876. Terdapat dua pengolahan data yang dilakukan yang terdiri dari analisis deskriptif dan juga *pre-processing data*, yaitu sebagai berikut :

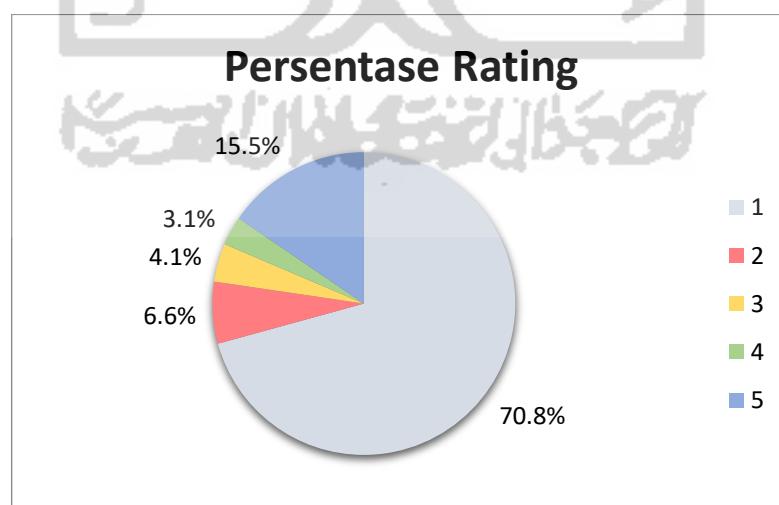
### 4.2.1. Analisis Deskriptif

Pada penelitian ini analisis deskriptif dilakukan untuk mengetahui gambaran informasi secara umum terkait pelayanan JNE berdasarkan data yang telah didapatkan dari teknik *scraping* di situs *Google Play*. Aspek yang dapat dilihat antara lain adalah jumlah ulasan pengguna berdasarkan waktu.



Gambar 4.8 Grafik Jumlah Ulasan

Berdasarkan Gambar 4.8, dengan rentang waktu pengambilan data selama 19 September 2019 sampai dengan 31 Desember 2019 diperoleh hasil grafik yang cenderung naik. Adapun pada bulan September hanya terdapat 153 ulasan dikarenakan data diambil sesuai dengan tanggal *upgrade* versi terbaru aplikasi my JNE yaitu tanggal 19 September. Pada bulan Oktober terdapat 550 ulasan, sedangkan pada bulan November menurun menjadi 531 ulasan. Selanjutnya pada bulan Desember ulasan meningkat menjadi 642, hal ini diperkirakan terjadi karena pada bulan tersebut terdapat Hari Belanja Online Nasional (HARBOLNAS) yang meningkatkan jumlah penjualan *e-commerce* sehingga penggunaan jasa pengiriman termasuk didalamnya JNE juga turut meningkat. Selain itu dilakukan analisis terkait persentase *rating* yang dapat dilihat pada Gambar 4.9 :



Gambar 4.9 Persentase Rating Ulasan

Berdasarkan Gambar 4.9 diketahui bahwa terdapat 5 kategori rating yang berbeda, yang mana pengguna memberikan bintang dengan nilai 1 sampai dengan 5. Bintang 1 termasuk dalam kategori “tidak bagus”, bintang 2 “kurang bagus”, bintang 3 “netral”, bintang 4 “bagus” dan bintang 5 “sangat bagus”. Berdasarkan diagram lingkaran diatas dapat diketahui bahwa dari total ulasan sebanyak 1.876, sebanyak 70,8% atau sebanyak 1.328 pengguna memberikan rating 1 atau tidak bagus. Selanjutnya untuk persentase terbanyak kedua adalah 15,5% atau 290 pengguna yang mana memberikan rating 5 atau sangat bagus. Kemudian diikuti oleh rating 2 sebanyak 6,6% dengan 123 pengguna, rating 3 sebanyak 4,1% dengan 77 pengguna, dan rating 4 sebanyak 3,1% dengan 58 pengguna.

#### **4.2.2. Pre-processing Data**

*Pre-processing Data* dilakukan dengan berbagai tahapan. Adapun *pre-processing data* pada penelitian ini dilakukan dengan menggunakan *tools* yaitu R Studio dan Python berdasarkan library sastrawi. Berikut merupakan tahapan *pre-processing* yang dilakukan, antara lain sebagai berikut :

##### **1. Cleaning Data**

Pada tahap ini dilakukan pembersihan dokumen dari kata ataupun hal yang tidak diperlukan seperti *noise*, baik berupa singkatan, kata yang tidak diperlukan, tanda baca, dan lain sebagainya. Berikut merupakan contoh kalimat ulasan hasil dari *cleaning data* :

- Menghapus angka :

Tabel 4 1 Menghapus Angka Pada Data

Data Input	Data Output
Pengalaman paket jelek,kirim tgl	Pengalaman paket jelek,kirim tgl ..
<b>28.12.2019</b> sampai sekarang tgl	sampai sekarang tgl belum berjalan
<b>30</b> belum berjalan juga	juga paketnya.parah jne.
paketnya.parah jne.	

- Menghapus tanda baca

Contoh kalimat ulasan :

Tabel 4.2 Menghapus Tanda Baca

Data Input	Data Output
Pengalaman paket jelek,kirim tgl .. sampai sekarang tgl belum berjalan juga paketnya.parah jne.	Pengalaman paket jelek kirim tgl sampai sekarang tgl belum berjalan juga paketnya parah jne

- Menghapus singkatan

Tabel 4.3 Menghapus Singkatan

Data Input	Data Output
Pengalaman paket jelek kirim <b>tgl</b> sampai sekarang <b>tgl</b> belum berjalan juga paketnya parah jne	Pengalaman paket jelek kirim sampai sekarang belum berjalan juga paketnya parah jne

- Menyeragamkan huruf menjadi huruf kecil (*lower-case*)

Tabel 4.4 Menyeragamkan Huruf

Data Input	Data Output
<b>P</b> engalaman paket jelek kirim sampai sekarang belum berjalan juga paketnya parah jne	pengalaman paket jelek kirim sampai sekarang belum berjalan juga paketnya parah jne

## 2. Tokenizing

*Tokenizing* merupakan tahap memisahkan teks menjadi potongan kata yang independen dan tidak saling mempengaruhi atau disebut juga *token*. Melalui *tokenizing* proses perhitungan dari frekuensi kemunculan setiap kata dalam *corpus* dapat lebih mudah terdeteksi. Berikut merupakan contoh penerapan *tokenizing* pada kalimat yang tertera pada Tabel 4.5 :

Tabel 4.5 Tokenizing

Data Input	Data Output
pengalaman paket jelek kirim	“pengalaman” “paket” “jelek”

Data Input	Data Output
sampai sekarang belum berjalan juga paketnya parah jne	“kirim” “sampai” “sekarang” “belum” “berjalan” “juga” “paketnya” “parah” “jne”

### 3. Filtering

*Filtering* merupakan tahap yang dilakukan untuk menghilangkan kata dalam suatu dokumen, atau mengurangi kata yang terdapat didalam *corpus* yang disebut *stopwords*. Tahap *filtering* biasa juga disebut *stopword removal*. *Stopword* digunakan untuk menghilangkan kata yang tidak berpengaruh, yang apabila dihilangkan tidak mengurangi informasi yang terdapat pada suatu kalimat. Penelitian ini menggunakan *stopword* berbahasa Indonesia yang disusun oleh Bouge (2011). Adapun terdapat beberapa kata yang ditambahkan secara manual dikarenakan tidak terdapat didalam *stopword*. Contoh dari *stopwords* seperti ”saya”, ”untuk”, ”kamu”, ”yang”, dan lain sebagainya. Berikut merupakan contoh *filtering* pada kalimat ulasan seperti pada Tabel 4.6:

Tabel 4.6 *Filtering*

Data Input	Data Output
“pengalaman” “paket” “jelek”	“paket” “jelek” “kirim”
“kirim” “sampai” “sekarang”	“berjalan” “parah”
“belum” “berjalan” “juga”	
“paketnya” “parah” “jne”	

### 4. Stemming

*Stemming* merupakan tahap yang dilakukan untuk mengubah kata pada teks menjadi kata dasar. Pada tahap ini dilakukan penghapusan imbuhan pada kata hasil dari proses *filtering*, baik yang mengandung imbuhan induktif maupun deduktif. Selain berdasarkan library sastrawi pada Python, dilakukan *stemming* menggunakan *packages* katadasaR yang terdapat pada Rstudio. Adapun berikut merupakan contoh proses *stemming* pada ulasan :

Tabel 4.7 Stemming

Data Input	Data Output
“paket”	“paket”
“jelek”	“jelek”
“kirim”	“kirim”
“berjalan”	“jalan”
“parah”	“parah”

Setelah dilakukan *pre-processing*, data awal sampel yang berjumlah 1.876 berubah menjadi 1.843 dikarenakan terdapat ulasan yang hanya berupa karakter simbol maupun emotikon dan tidak berupa kata-kata oleh karena itu tidak digunakan.

#### 4.2.3. Representasi Model

Representasi model dilakukan dengan membuat model dari data yang masih berbentuk kata-kata agar dapat diolah dan dihitung. Representasi dilakukan dengan melakukan perhitungan skor kata berdasarkan kamus *lexicon* dan kemudian memberi label kelas sentimen berdasarkan skor yang ada. Skor sentimen dihitung berdasarkan jumlah kata yang terdeteksi pada kamus *lexicon*.

##### 1. Perhitungan Skor Sentimen

Perhitungan skor sentimen dilakukan, dengan contoh kalimat sebagai berikut :

Tabel 4.8 Perhitungan Skor Sentimen

Ulasan	Kata Positif	Kata Negatif
aplikasi <b>bagus</b> kirim <b>lama</b> <b>telat</b>	bagus	lama
bikin emosi <b>parah</b>		telat
		parah
Jumlah	1	3
Perhitungan	Skor = 1 – 3 = -2	

Berdasarkan kalimat *review* pada tabel 4.7 diketahui bahwa terdapat 1 kata positif dan 3 kata negatif pada kamus *lexicon*. Kata positif yang terdeteksi adalah “bagus”, sedangkan kata negatif yang terdeteksi adalah “lama”, “telat”,

dan “parah”. Adapun rumus yang digunakan untuk dapat memperoleh skor sentimen adalah sebagai berikut :

$$\text{Skor} = \text{Jumlah kata positif} - \text{Jumlah kata negatif}$$

Oleh karena itu, perhitungan berdasarkan review pada kalimat diatas :

$$\text{Skor} = 1 - 3 = -2$$

Adapun pada penelitian ini perhitungan skor dilakukan secara otomatis oleh aplikasi R yang selanjutnya memberikan label dari setiap kelas sentimen.

## 2. Pelabelan Kelas Sentimen

Tahap berikutnya setelah mendapatkan skor terhadap sentimen adalah melakukan pelabelan kelas sentimen. Pelabelan dilakukan dengan membagi data menjadi tiga kelas sentimen yaitu sentimen positif, sentimen netral, dan sentimen negatif dengan keterangan sebagai berikut :

- Sentimen negatif :  $\text{skor} < 0$
- Sentimen netral :  $\text{skor} = 0$
- Sentimen positif :  $\text{skor} > 0$

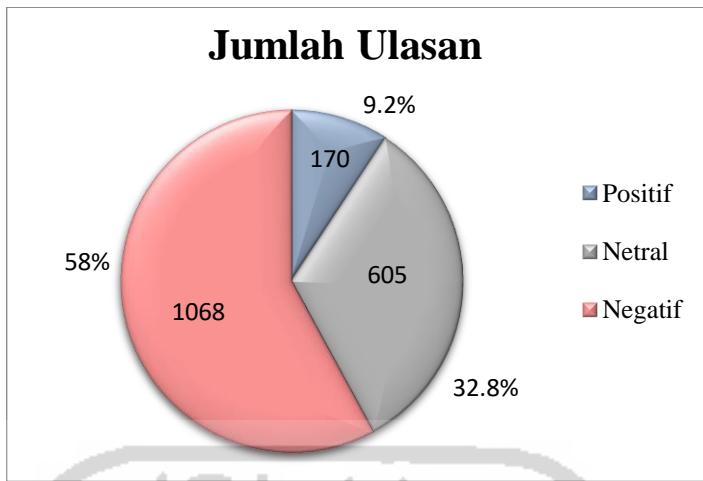
Kelas sentimen negatif merupakan pernyataan yang bersifat ketidakpuasan pelanggan, penghinaan dan segala sesuatunya yang bersifat negatif. Kelas sentimen positif merupakan pernyataan yang mengandung ungkapan pujian, rasa terimakasih, dan lain sebagainya. Adapun untuk kelas sentimen netral tidak mengandung kata yang bersifat positif maupun negatif, atau yang bersifat seimbang. Kelas sentimen netral dapat terjadi karena terdapat dua kemungkinan, yaitu :

- a. Tidak terdapat kata sentimen pada data yang terdeteksi pada kamus
- b. Jumlah skor kata positif sama dengan jumlah skor negatif.

Berikut merupakan hasil pelabelan kelas sentimen pada ulasan JNE pada *Google Play* :

Tabel 4.9 Pelabelan Jumlah Kelas Sentimen

<b>Sentimen</b>	<b>Jumlah</b>
Positif	170
Netral	605
Negatif	1068



Gambar 4.10 Diagram Lingkaran Jumlah Ulasan

Kelas netral selanjutnya akan direduksi karena peneliti hanya akan menggunakan dua pelabelan kelas sentimen yaitu sentimen positif dan sentimen negatif. Pelabelan kelas sentimen netral dilakukan secara manual oleh peneliti. Pelabelan kelas sentimen netral dilakukan berdasarkan beberapa pertimbangan (Widodo, et al., 2020), antara lain :

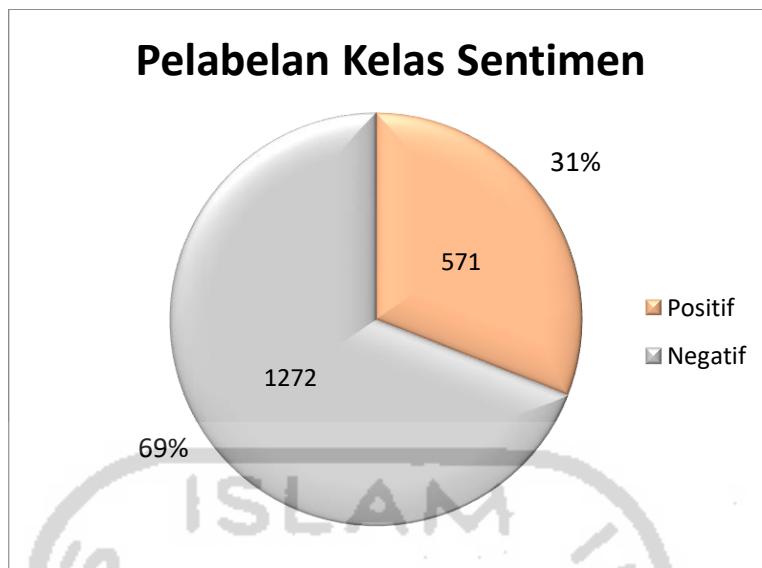
- Jika pada *review* tidak teridentifikasi kata positif atau negatif maka akan diklasifikasikan pada kelas positif
- Jika jumlah kata positif atau negatif sama akan diklasifikasikan sebagai kelas negatif dengan pertimbangan bahwa agar pihak terkait dapat melihat masalah-masalah yang terjadi pada aplikasi setelah aplikasi tersebut dirilis dan digunakan oleh masyarakat.

Adapun berikut merupakan contoh pelabelan kelas sentimen, seperti yang dapat dilihat pada Tabel 4.10 :

Tabel 4.10 Pelabelan Kelas Sentimen Pada Kalimat

Kelas Sentimen	Skor	Ulasan
Positif	2	jasa kurir <b>top terpercaya</b>
Negatif	-2	aplikasi <b>bagus</b> kirim <b>lama telat</b> bikin emosi <b>parah</b>
Netral	0	coba dulu

Hasil reduksi pelabelan kelas netral, diperoleh hasil sebagai berikut :



Gambar 4.11 Persentase Pelabelan Kelas Sentimen

Berdasarkan Gambar 4.11 diketahui bahwa hasil pelabelan kelas sentimen dengan frekuensi terbanyak adalah sentimen negatif yaitu sebesar 1.272 ulasan atau 69% sedangkan sentimen positif hanya sebesar 571 ulasan atau 31% dari total 1.843 ulasan.

#### 4.2.4. Pembagian Data Latih dan Data Uji

Data latih digunakan oleh algoritma klasifikasi untuk membentuk sebuah model *classifier*, model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada. Semakin besar data latih yang digunakan, maka akan semakin baik *machine* dalam memahami pola data. Sedangkan *testing set* atau data uji merupakan dataset yang digunakan untuk mengetahui tingkat akurasi dari model yang dihasilkan (*classifier*). Data uji digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Data yang digunakan untuk data latih dan data uji adalah data yang telah memiliki label kelas.

Adapun pada penelitian ini dilakukan lima kali percobaan dengan perbandingan antara data latih dan data uji yang berbeda, antara lain sebagai berikut :

1) Perbandingan data latih sebesar 60% dan data uji 40% :

Tabel 4.11 Pembagian Data Latih dan Data Uji 1

<b>Klasifikasi</b>	<b>Jumlah</b>	<b>Data Latih (60%)</b>	<b>Data Uji (40%)</b>
Positif	571	343	228
Negatif	1.272	763	509
Total	1.843	1.106	737

Berdasarkan perbandingan 60% : 40%, dari total 1.843 ulasan berbahasa Indonesia, digunakan 1.106 data latih dan 737 data uji.

2) Perbandingan data latih sebesar 70% dan data uji 30% :

Tabel 4.12 Pembagian Data Latih dan Data Uji 2

<b>Klasifikasi</b>	<b>Jumlah</b>	<b>Data Latih (70%)</b>	<b>Data Uji (30%)</b>
Positif	571	400	171
Negatif	1.272	890	382
Total	1.843	1.290	553

Berdasarkan perbandingan 70% : 30%, dari total 1.843 ulasan berbahasa Indonesia, digunakan 1.290 data latih dan 553 data uji.

3) Perbandingan data latih sebesar 80% dan data uji 20% :

Tabel 4.13 Pembagian Data Latih dan Data Uji 3

<b>Klasifikasi</b>	<b>Jumlah</b>	<b>Data Latih (80%)</b>	<b>Data Uji (20%)</b>
Positif	571	457	114
Negatif	1.272	1017	254
Total	1.843	1.474	368

Berdasarkan perbandingan 80% : 20%, dari total 1.843 ulasan berbahasa Indonesia, digunakan 1.474 data latih dan 368 data uji.

- 4) Perbandingan data latih sebesar 85% dan data uji 15% :

Tabel 4.14 Pembagian Data Latih dan Data Uji 4

Klasifikasi	Jumlah	Data Latih (85%)	Data Uji (15%)
Positif	571	486	85
Negatif	1.272	1081	191
Total	1.843	1.567	276

Berdasarkan perbandingan 85% : 15%, dari total 1.843 ulasan berbahasa Indonesia, digunakan 1.567 data latih dan 276 data uji.

- 5) Perbandingan data latih sebesar 90% dan data uji 10% :

Tabel 4.15 Pembagian Data Latih dan Data Uji 5

Klasifikasi	Jumlah	Data Latih (90%)	Data Uji (10%)
Positif	571	513,9 ≈ 514	57,1 ≈ 57
Negatif	1.272	1.144,8 ≈ 1.145	127,2 ≈ 127
Total	1.843	1.659	184

Berdasarkan perbandingan 90% : 10%, dari total 1.843 ulasan berbahasa Indonesia, digunakan 1.659 data latih dan 184 data uji.

#### 4.2.5. Klasifikasi dengan Metode *Naïve Bayes Classifier*

Setelah dilakukan penentuan terkait data latih dan data uji selanjutnya dilakukan klasifikasi dengan metode *Naïve Bayes Classifier*. Analisis *Naïve Bayes Classifier* dilakukan dengan membuat *confusion matrix* agar dapat diketahui nilai akurasi, *recall* dan presisi. Menurut Manning, Prabhakar, dan Hinrich (2009) *confusion matrix* merupakan salah satu *tools* penting yang digunakan pada *machine learning* untuk melakukan visualisasi dan biasanya memuat dua kategori atau lebih. *Confusion matrix* digunakan untuk mengevaluasi performansi model yang dibangun oleh algoritma klasifikasi. Pada penelitian ini *confusion matrix* dibuat dengan menggunakan aplikasi *Rstudio*. Berdasarkan pembagian data latih dan data uji yang

berbeda, didapatkan hasil *confusion matrix* dengan nilai akurasi, presisi, dan *recall* yang berbeda pula yang dapat dilihat pada tabel berikut :

- 1) Pembagian Data Latih 60% Data Uji 40%

Tabel 4.16 Hasil klasifikasi 1

Prediksi	Aktual		<i>Class Precision</i>
	Negatif	Positif	
<b>Negatif</b>	416	50	89,27%
<b>Positif</b>	93	178	65,7%
<b>Class Recall</b>	81,7%	78%	
<b>Akurasi</b>	80,6%		

Berdasarkan *confusion matrix* pada Tabel 4.20 dengan menggunakan metode *Naïve Bayes Classifier*, diketahui bahwa prediksi ulasan negatif berjumlah 466 ulasan. Dari 466 prediksi ulasan negatif, terdapat kesesuaian bahwa ulasan negatif yang termasuk aktual negatif adalah sebanyak 416 data yang telah terklasifikasi dengan benar, dan terdapat ketidaksesuaian karena kebenarannya ulasan adalah positif (aktual positif) sebanyak 14 ulasan. Sehingga pada prediksi negatif didapatkan *Class Precision* sebanyak 89,27%. Selanjutnya pada prediksi positif diketahui terdapat 271 ulasan dengan jumlah kesesuaian bahwa ulasan yang aktual positif sebanyak 178 yang telah terklasifikasi dengan benar dan terdapat ketidaksesuaian atau dengan kata lain terdapat kesalahan prediksi sebanyak 93 ulasan positif yang masuk dalam ulasan negatif. Sehingga didapatkan nilai presisi untuk kelas positif sebesar 65,7%. Selanjutnya dari hasil *confusion matrix* tersebut diperoleh tingkat akurasi sebesar 80,6%, yang artinya dari 737 data ulasan yang diuji, terdapat 594 ulasan yang diklasifikasikan secara benar oleh model *Naïve Bayes Classifier*. Sehingga *error rate* (100% - 80,6%) sebesar 19,4%.

2) Data Latih 70% Data Uji 30%

Tabel 4.17 Hasil Klasifikasi 2

Prediksi	Aktual		<i>Class Precision</i>
	Negatif	Positif	
<b>Negatif</b>	320	38	89,3%
<b>Positif</b>	60	135	69,23%
<b>Class Recall</b>	84,2%	78%	
<b>Akurasi</b>	82,3%		

Berdasarkan Tabel 4.17 dengan menggunakan metode *Naïve Bayes Classifier*, diketahui bahwa prediksi ulasan negatif berjumlah 358 ulasan. Dari 358 prediksi ulasan negatif, terdapat kesesuaian bahwa ulasan negatif yang termasuk aktual negatif adalah sebanyak 320 data yang telah terklasifikasi dengan benar, dan terdapat ketidaksesuaian karena kebenarannya ulasan adalah positif (aktual positif) sebanyak 38 ulasan. Sehingga pada prediksi negatif didapatkan *Class Precision* sebesar 89,3%. Selanjutnya pada prediksi positif diketahui terdapat 195 ulasan dengan jumlah kesesuaian bahwa ulasan yang aktual positif sebanyak 135 yang telah terklasifikasi dengan benar dan terdapat ketidaksesuaian karena kebenarannya ulasan adalah negatif (aktual negatif) sebanyak 60 ulasan. Sehingga didapatkan nilai presisi untuk kelas positif sebesar 69,23%. Selanjutnya dari hasil *confusion matrix* tersebut diperoleh tingkat akurasi sebesar 82,3%, yang artinya dari 553 data ulasan yang diuji, terdapat 455 ulasan yang diklasifikasikan secara benar oleh model *Naïve Bayes Classifier*. Sehingga *error rate* (100% - 82,3%) sebesar 17,7%. Adapun pada hasil klasifikasi dengan perbandingan data uji dan data latih sebesar 70:30 ini memiliki hasil akurasi yang lebih tinggi dibanding pada klasifikasi sebelumnya dengan perbandingan 60:40.

3) Data Latih 80% Data Uji 20%

Tabel 4.18 Hasil Klasifikasi 3

Prediksi	Aktual		<i>Class Precision</i>
	Negatif	Positif	
<b>Negatif</b>	218	26	89,3%

<b>Prediksi</b>	<b>Aktual</b>		<b>Class Precision</b>
	<b>Negatif</b>	<b>Positif</b>	
<b>Positif</b>	31	94	75,2%
<b>Class Recall</b>	87,6%	78,3%	
<b>Akurasi</b>	84,55%		

Berdasarkan Tabel 4.18 dengan menggunakan metode *Naïve Bayes Classifier*, diketahui bahwa prediksi ulasan negatif berjumlah 244 ulasan. Dari 244 prediksi ulasan negatif, terdapat kesesuaian bahwa ulasan negatif yang termasuk aktual negatif adalah sebanyak 218 data yang telah terkласifikasi dengan benar, dan terdapat ketidaksesuaian karena kebenarannya ulasan adalah positif (aktual positif) sebanyak 26 ulasan. Sehingga pada prediksi negatif didapatkan *Class Precision* sebesar 89,3%. Selanjutnya pada prediksi positif diketahui terdapat 125 ulasan dengan jumlah kesesuaian bahwa ulasan yang aktual positif sebanyak 94 yang telah terklasifikasi dengan benar dan terdapat ketidaksesuaian karena kebenarannya ulasan adalah negatif (aktual negatif) sebanyak 31 ulasan. Sehingga didapatkan nilai presisi untuk kelas positif sebesar 75,2%. Selanjutnya dari hasil *confusion matrix* tersebut diperoleh tingkat akurasi sebesar 84,55%, yang artinya dari 369 data ulasan yang diuji, terdapat 312 ulasan yang diklasifikasikan secara benar oleh model *Naïve Bayes Classifier*. Sehingga *error rate* ( $100\% - 84,55\%$ ) sebesar 15,45%. Adapun pada hasil klasifikasi dengan perbandingan data uji dan data latih sebesar 80:20 ini memiliki hasil akurasi yang lebih tinggi dibanding pada klasifikasi sebelumnya dengan perbandingan 60:40 dan 70:30.

#### 4) Data Latih 85% data Uji 15%

Tabel 4.19 Hasil Klasifikasi 4

<b>Prediksi</b>	<b>Aktual</b>		<b>Class Precision</b>
	<b>Negatif</b>	<b>Positif</b>	
<b>Negatif</b>	162	21	88,5%
<b>Positif</b>	19	74	79,6%
<b>Class Recall</b>	89,5%	77,9%	
<b>Akurasi</b>	85,5%		

Berdasarkan Tabel 4.19 dengan menggunakan metode *Naïve Bayes Classifier*, diketahui bahwa prediksi ulasan negatif berjumlah 183 ulasan. Dari 183 prediksi ulasan negatif, terdapat kesesuaian bahwa ulasan negatif yang termasuk aktual negatif adalah sebanyak 162 ulasan yang telah terkласifikasi dengan benar, dan terdapat ketidaksesuaian karena kebenarannya ulasan adalah positif (aktual positif) sebanyak 21 ulasan. Sehingga pada prediksi negatif didapatkan *Class Precision* sebanyak 88,5%. Selanjutnya pada prediksi positif diketahui terdapat 93 ulasan dengan jumlah kesesuaian bahwa ulasan yang aktual positif sebanyak 74 yang telah terklasifikasi dengan benar dan terdapat ketidaksesuaian karena kebenarannya ulasan adalah negatif (aktual negatif) sebanyak 19 ulasan atau dengan kata lain terdapat kesalahan prediksi sebanyak 19 ulasan positif yang masuk dalam ulasan negatif. Sehingga didapatkan nilai presisi untuk kelas positif sebesar 70,6%. Selanjutnya dari hasil *confusion matrix* tersebut diperoleh tingkat akurasi sebesar 85,5%, yang artinya dari 276 data ulasan yang diuji, terdapat 236 ulasan yang benar diklasifikasikan oleh model *Naïve Bayes Classifier*. Sehingga *error rate* ( $100\% - 85,5\%$ ) sebesar 14,5%.

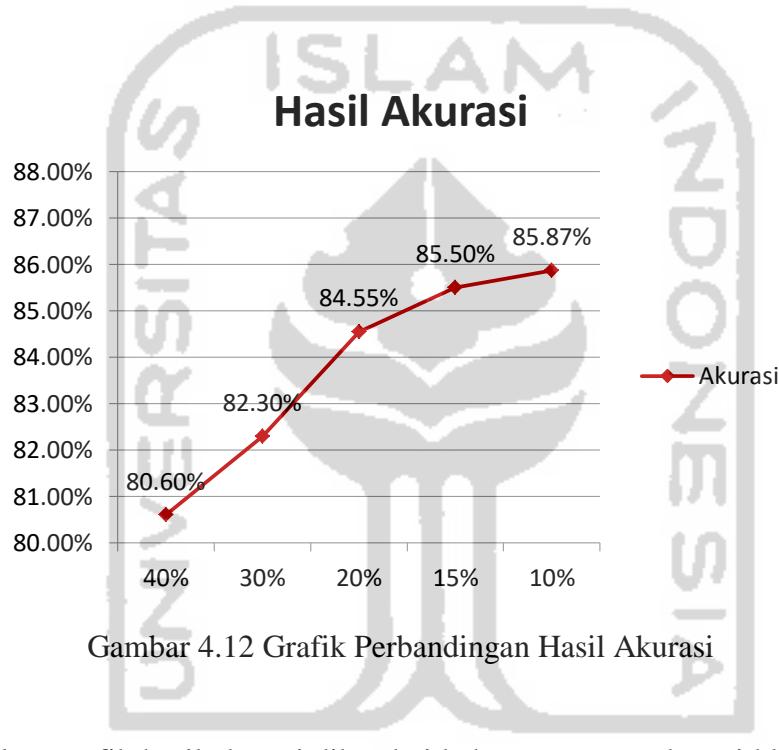
##### 5) Data Latih 90% Data Uji 10%

Tabel 4.20 Hasil Klasifikasi 5

Prediksi	Aktual		<i>Class Precision</i>
	Negatif	Positif	
<b>Negatif</b>	108	14	88,5%
<b>Positif</b>	12	50	80,6%
<b>Class Recall</b>	90%	78,1%	
<b>Akurasi</b>	85,87%		

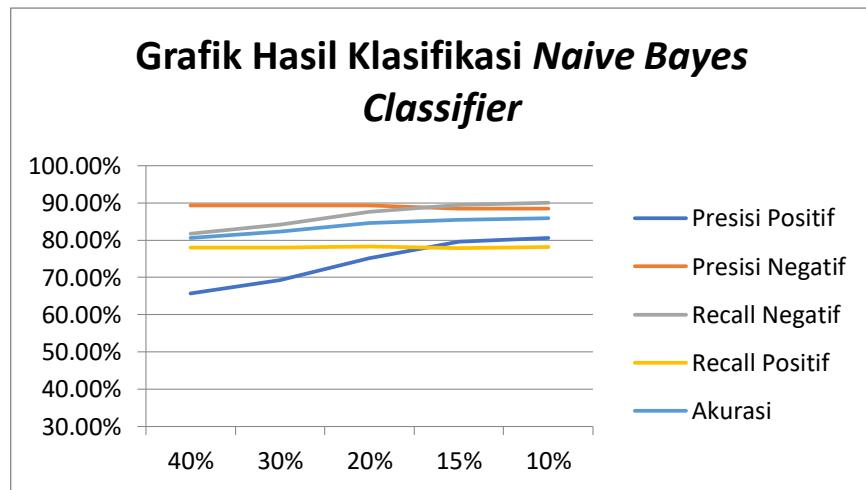
Berdasarkan Tabel 4.20 dengan menggunakan metode *Naïve Bayes Classifier*, diketahui bahwa prediksi ulasan negatif berjumlah 122 ulasan. Dari 122 prediksi ulasan negatif, terdapat kesesuaian bahwa ulasan negatif yang termasuk aktual negatif adalah sebanyak 108 data yang telah terklasifikasi dengan benar, dan terdapat ketidaksesuaian karena kebenarannya ulasan adalah positif (aktual positif) sebanyak 14 ulasan. Sehingga pada prediksi negatif didapatkan *Class Precision* sebanyak

88,5%. Selanjutnya pada prediksi positif diketahui terdapat 62 ulasan dengan jumlah kesesuaian bahwa ulasan yang aktual positif sebanyak 50 yang telah terklasifikasi dengan benar dan terdapat ketidaksesuaian karena kebenarannya ulasan adalah negatif (aktual negatif) sebanyak 12 ulasan atau dengan kata lain terdapat kesalahan prediksi sebanyak 12 ulasan positif yang masuk dalam ulasan negatif. Sehingga didapatkan nilai presisi untuk kelas positif sebesar 80,6%. Selanjutnya dari hasil *confusion matrix* tersebut diperoleh tingkat akurasi sebesar 85,87%, yang artinya dari 184 data ulasan yang diuji, terdapat 158 ulasan yang benar diklasifikasikan oleh model *Naïve Bayes Classifier*. Sehingga *error rate* ( $100\% - 85,87\%$ ) sebesar 14,13%.



Gambar 4.12 Grafik Perbandingan Hasil Akurasi

Berdasarkan grafik hasil akurasi diketahui bahwa rata-rata akurasi klasifikasi dengan Metode *Naïve Bayes Classifier* akan cenderung naik seiring dengan berkurangnya data yang digunakan dalam pengujian (Data uji). Adapun hasil maksimal didapatkan pada perbandingan data uji sebesar 10% dan data latih 90%, dengan akurasi tertinggi yaitu 85.87% dan presisi lebih dari 80% untuk kelas positif maupun negatif. Adapun secara keseluruhan nilai presisi maupun *recall* mengalami peningkatan seiring berkurangnya data uji yang digunakan seperti yang dapat dilihat pada grafik Gambar 4.13 berikut :



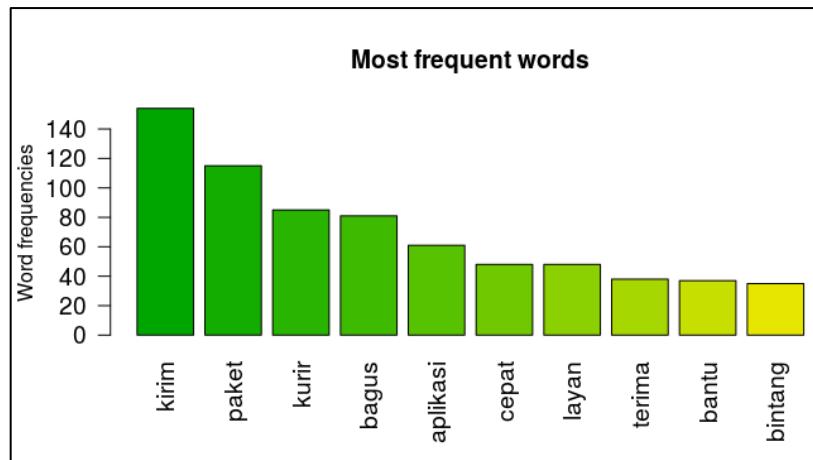
Gambar 4.13 Grafik Hasil Klasifikasi *Naïve Bayes Classifier*

#### 4.2.6. Visualisasi dan Asosiasi

Visualisasi dan asosiasi dilakukan pada keseluruhan data, baik pada sentimen positif maupun sentimen negatif. Visualisasi dan asosiasi dilakukan dengan tujuan untuk memperoleh informasi secara menyeluruh mengenai apa yang sering dibicarakan terkait topik yang dianalisis yaitu opini pengguna JNE. Selain itu dilakukan pula pencarian asosiasi antar kata yang paling sering muncul secara bersamaan, agar dapat memperkuat pencarian informasi. Berikut merupakan visualisasi dan asosiasi kata dari setiap klasifikasi kelas sentimen :

##### 4.2.6.1. Ulasan Positif

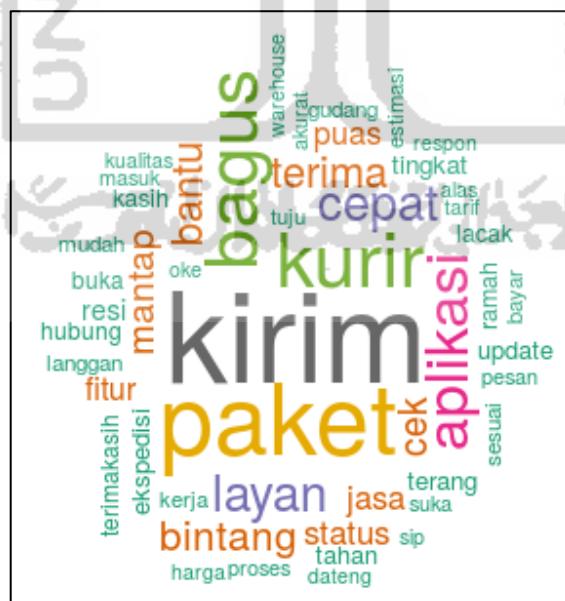
Pada visualisasi dan asosiasi ulasan positif ini digunakan data ulasan positif dari hasil pelabelan berdasarkan kamus *lexicon* maupun manual yang sebelumnya telah dilakukan. Untuk mendapatkan kata yang paling sering diulas oleh pengguna JNE, data ulasan positif diekstraksi berulang kali, yang selanjutnya kemudian diidentifikasi berdasarkan frekuensi kata pada ulasan. Berikut merupakan hasil ekstraksi informasi berupa visualisasi dalam bentuk *barplot* berdasarkan ulasan pengguna JNE dengan kelas sentimen positif :



Gambar 4.14 Barplot Visualisasi Sentimen Positif

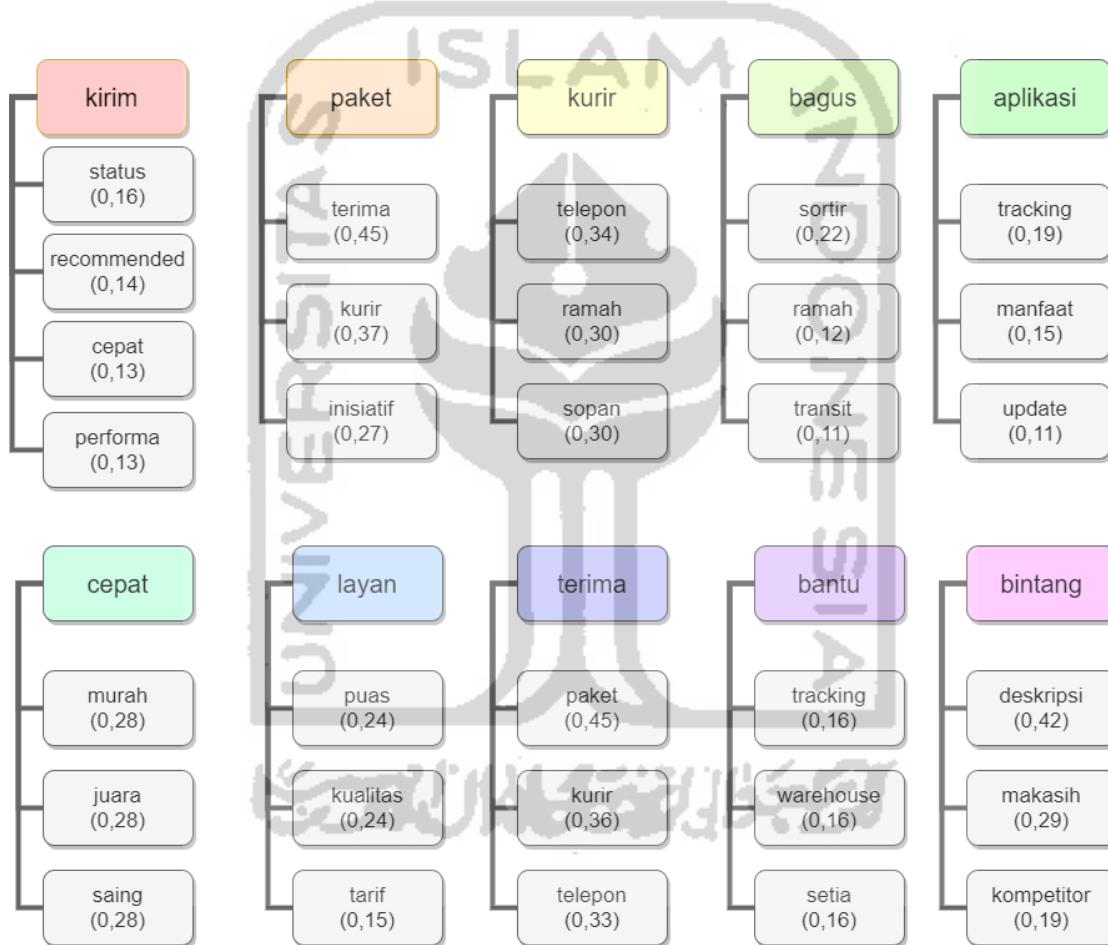
Berdasarkan Gambar 4.14 pada data ulasan kelas sentimen positif dengan jumlah ulasan sebanyak 571, didapatkan beberapa kata yang paling sering muncul, dalam hal ini peneliti memunculkan 10 kata dengan frekuensi terbanyak antara lain : kata “kirim” sebanyak 154 kali, kata “paket” sebanyak 115 kali, kata “kurir” sebanyak 85 kali, kata “bagus” sebanyak 81 kali, kata “aplikasi” sebanyak 61 kali, kata “cepat” sebanyak 48 kali, kata “layan” sebanyak 48 kali, kata “terima” sebanyak 38 kali, kata “bantu” sebanyak 37 kali, dan kata “bintang” sebanyak 35 kali.

Berdasarkan kata-kata yang muncul seperti pada gambar 4.13 selanjutnya dibuat *wordcloud* atau kumpulan kata dari ulasan yang sering dibahas oleh pengguna JNE, yaitu dapat dilihat pada gambar 4.15.



Gambar 4.15 Wordcloud Visualisasi Sentimen Positif

Berdasarkan Gambar 4.15 diketahui bahwa visualisasi dalam bentuk *wordcloud* dapat memperlihatkan dengan lebih jelas kata-kata positif yang sering muncul pada ulasan pengguna JNE. Semakin besar ukuran kata pada *wordcloud* maka semakin tinggi pula frekuensi kata tersebut, artinya pengguna sering menggunakan kata tersebut sebagai topik pembicaraan dengan penilaian positif dalam ulasan. Beberapa kata yang paling sering dibahas adalah terkait kirim, paket, kurir, bagus, aplikasi, cepat, layan, dan lain-lain. Selanjutnya, dilakukan pencarian asosiasi antar kata yang berhubungan dengan kata yang sering muncul secara bersamaan dan didapatkan hasil sebagai berikut :



Gambar 4.16 Asosiasi Kata Ulasan Positif

Berdasarkan Gambar 4.16 diperoleh asosiasi kata pada klasifikasi kelas positif. Proses ekstraksi informasi dengan asosiasi dilakukan secara berulang-ulang dengan cara menyaring kata-kata yang memiliki hubungan dengan kata lain dan didasarkan pada relevansi kata dengan topik yang diulas.

Berdasarkan asosiasi yang terkait dengan kata “kirim” diperoleh informasi bahwa dari segi pengiriman JNE telah memiliki performa yang baik dan juga cepat sehingga *recommended* dimata pengguna, serta pengguna memberikan saran agar JNE melakukan peningkatan agar status pengiriman senantiasa *update*.

Asosiasi yang terkait dengan kata “paket” menunjukkan informasi bahwa paket telah diterima oleh *customer* dengan baik dan *customer* memberikan saran agar kurir memiliki inisiatif untuk menelepon atau menghubungi terkait kabar paket yang dikirim.

Asosiasi yang terkait dengan kata “kurir” menunjukkan informasi bahwa pengguna merasa kurir telah memberikan pelayanan yang ramah dan sopan. Pengguna memberikan saran agar kurir menelepon penerima ataupun menitipkan paket melalui tetangga.

Asosiasi yang terkait dengan kata “bagus” menunjukkan informasi bahwa JNE telah melayani dengan bagus karena dilakukan sortir terlebih dahulu dan juga ramah. Pengguna juga memberikan saran terkait transit paket di *warehouse*. Pengguna berharap agar waktu transit tidak begitu lama.

Asosiasi kata yang berkaitan dengan “aplikasi” menunjukkan informasi bahwa pengguna merasa aplikasi membantu pengguna dalam melakukan *tracking*, selain itu aplikasi juga memberikan manfaat bagi pengguna. Adapun pengguna memberikan saran agar aplikasi senantiasa *update*.

Asosiasi kata yang berkaitan dengan “cepat” menunjukkan informasi bahwa selain cepat JNE juga menawarkan harga yang murah, mampu bersaing dengan kompetitor, dan juga juara bagi penggunanya.

Asosiasi kata yang berkaitan dengan “layan” menunjukkan informasi bahwa pengguna JNE merasa puas atas pelayanan yang ditawarkan oleh JNE dan merasa JNE telah memiliki kualitas yang baik. Selain itu dari segi tarif, harga yang ditawarkan sebanding dengan kualitasnya.

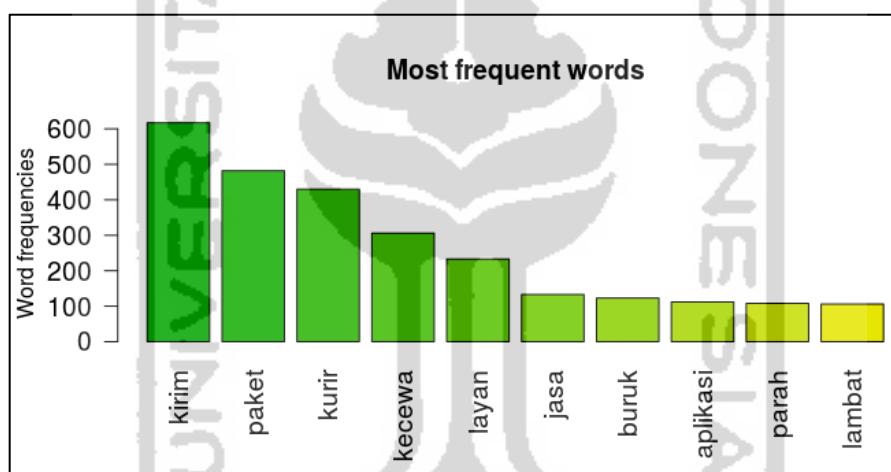
Asosiasi kata yang berkaitan dengan kata “terima” menunjukkan informasi bahwa paket diterima oleh pelanggan dan diharapkan agar kurir menelepon penerima.

Asosiasi kata yang berkaitan dengan kata “bantu” menunjukkan bahwa pengguna menilai bahwa sistem *tracking* dan *warehouse* yang ada pada JNE telah cukup baik dan pengguna akan setia menggunakan JNE.

Asosiasi kata yang berkaitan dengan kata “bintang” menunjukkan bahwa deskripsi yang diberikan oleh JNE telah cukup lengkap, pengguna berterimakasih atas pelayanan yang ditawarkan, serta JNE dianggap mampu bersaing dengan kompetitor.

#### 4.2.6.2. Ulasan Negatif

Berdasarkan total ulasan sebanyak 1.843, diketahui bahwa 69% diantaranya termasuk dalam ulasan negatif atau sebanyak 1.272 ulasan. Visualisasi dan Asosiasi Negatif dilakukan pada data ulasan negatif secara berulang-ulang hingga didapatkan informasi mengenai ulasan negatif yang sering dibicarakan oleh pengguna JNE. Hasil ekstraksi informasi berupa ulasan negatif diidentifikasi berdasarkan frekuensi kata dalam ulasan, selain itu juga didasarkan pada relevansi kata dengan topik yang mengacu pada sentimen negatif. Berikut adalah visualisasi hasil ekstraksi informasi dalam bentuk *barplot* yang didapatkan dari ulasan pengguna dengan klasifikasi negatif :



Gambar 4.17 Barplot Visualisasi Ulasan Negatif

Pada ulasan kelas sentimen negatif pada JNE, berdasarkan 1.272 ulasan diketahui bahwa terdapat kata yang paling banyak muncul. Pada hal ini divisualisasikan 10 kata dengan frekuensi tertinggi seperti yang dapat dilihat pada Gambar 4.17 Kata yang paling banyak muncul diantaranya : kata “kirim” dengan frekuensi sebanyak 617 kali, kata “paket” dengan frekuensi sebanyak 482 kali, kata “kurir” dengan frekuensi sebanyak 430 kali, kata “kecewa” dengan frekuensi sebanyak 306 kali, kata “layan” dengan frekuensi sebanyak 233 kali, kata “jasa” dengan frekuensi sebanyak 133 kali, kata “buruk” dengan frekuensi sebanyak 123 kali, kata “aplikasi” dengan frekuensi

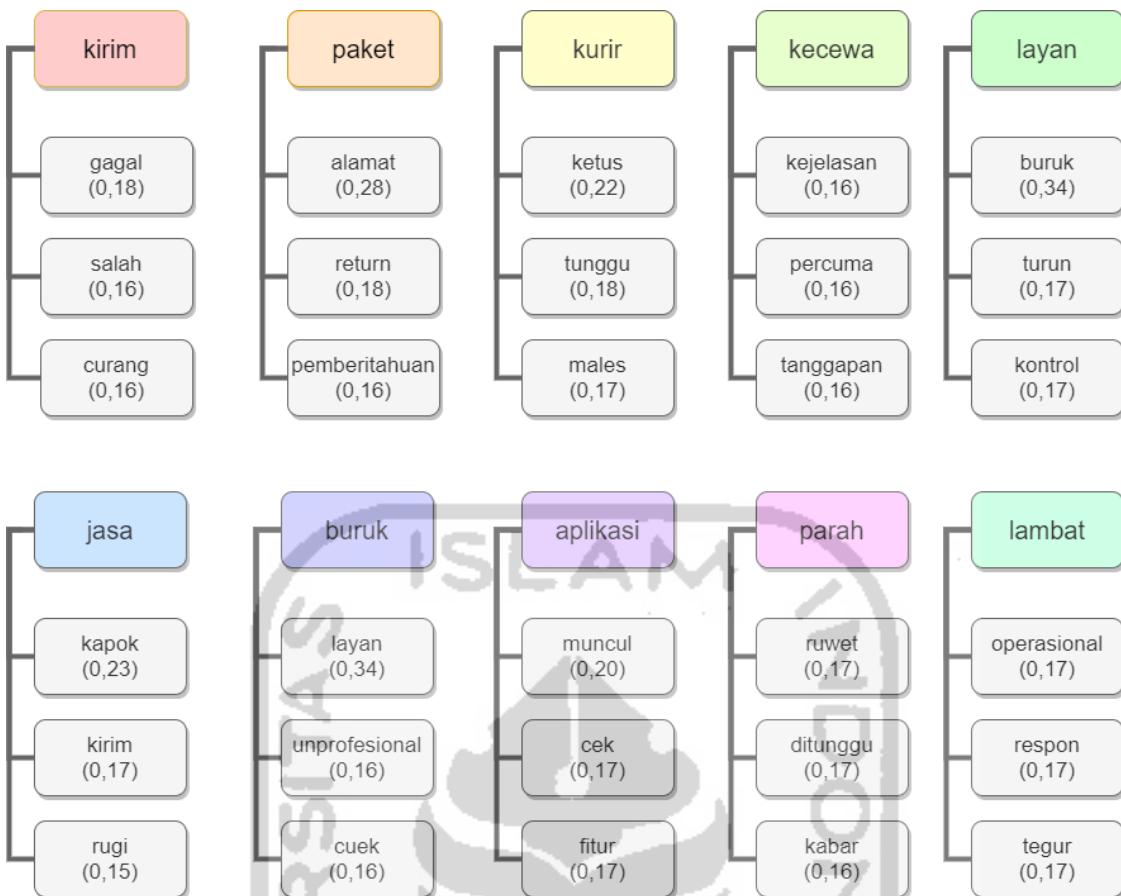
sebanyak 112 kali, kata “parah” dengan frekuensi sebanyak 108 kali, serta kata “lambat” dengan frekuensi sebanyak 106 kali.

Berdasarkan kata-kata yang muncul seperti pada gambar 4.17 selanjutnya dibuat *wordcloud* atau kumpulan kata dari ulasan negatif yang sering dibahas oleh pengguna JNE, yaitu dapat dilihat pada Gambar 4.18 :



Gambar 4.18 Wordcloud Visualisasi Ulasan Negatif

Berdasarkan Gambar 4.18 diketahui bahwa visualisasi dalam bentuk *wordcloud* dapat memperlihatkan dengan lebih jelas kata-kata negatif yang sering muncul pada ulasan pengguna JNE. Semakin besar ukuran kata pada *wordcloud* maka semakin tinggi pula frekuensi kata tersebut, artinya pengguna sering menggunakan kata tersebut sebagai topik pembicaraan dengan penilaian negatif dalam ulasan. Beberapa kata yang paling sering dibahas adalah terkait kirim, paket, kurir, kecewa, layan, dan lain-lain. Selanjutnya, dilakukan pencarian asosiasi antar kata yang berhubungan dengan kata yang sering muncul secara bersamaan dan didapatkan hasil sebagai berikut :



Gambar 4.19 Asosiasi Kata Ulasan Negatif

Berdasarkan Gambar 4.19 diperoleh asosiasi kata pada klasifikasi kelas negatif. Proses ekstraksi informasi dengan asosiasi dilakukan secara berulang-ulang dengan cara menyaring kata-kata yang memiliki hubungan dengan kata lain dan didasarkan pada relevansi kata dengan topik yang diulas.

Asosiasi kata yang berkaitan dengan “kirim” menunjukkan informasi bahwa pengguna mengalami gagal pengiriman. Gagal pengiriman dalam hal ini diartikan bahwa paket yang telah sampai di *warehouse* kemudian dikembalikan lagi kepada pengirim. Selain itu juga paket salah kirim (tersesat) dan kurir curang.

Asosiasi kata yang berkaitan dengan “paket” menunjukkan bahwa pengguna telah mencantumkan alamat yang lengkap namun kemudian paket tersesat atau salah alamat, kurir beralasan bahwa alamat belum lengkap sehingga paket terlambat datang dan juga tersesat ke rumah lain, kemudian paket di *return* atau dikembalikan lagi kepada pengirim. Selain itu pengguna merasa JNE masih tergolong kurang dalam memberikan pemberitahuan terkait informasi paket.

Asosiasi kata terkait “kurir” menunjukkan bahwa pengguna merasa kurir ketus, selain itu kurir juga dinilai malas dalam melaksanakan tugasnya, pengguna menunggu paket namun tak kunjung tiba dikarenakan kinerja kurir yang kurang baik.

Asosiasi kata terkait “kecewa” menunjukkan informasi bahwa pengguna merasa tidak mendapat kejelasan dan tanggapan terkait paket mereka sehingga merasa percuma menggunakan jasa JNE.

Asosiasi kata terkait “layan” menunjukkan informasi bahwa pengguna merasa JNE memberikan pelayanan yang buruk jika dibandingkan dengan kompetitornya. Pengguna juga merasa kinerja JNE kian menurun sehingga memberikan saran sebaiknya melakukan kontrol yang ketat baik CS maupun kurir terkait dengan pemberian layanan kepada pelanggan.

Asosiasi kata terkait dengan “jasa” menunjukkan informasi bahwa pengguna merasa kapok menggunakan layanan JNE. Ketika melakukan pengiriman menggunakan JNE pengguna merasa rugi karena harga yang ditawarkan dirasa tidak sesuai dengan kualitasnya.

Asosiasi kata terkait dengan “buruk” menunjukkan informasi bahwa pengguna merasa bahwa pelayanan yang diberikan oleh JNE masih tergolong buruk dan tidak profesional. Selain itu pengguna merasa bahwa kurir maupun CS terkesan cuek.

Asosiasi kata terkait dengan “aplikasi” menunjukkan informasi bahwa ketika melakukan pengecekan terhadap paket, update lokasi paket berada tidak muncul. Pengguna juga memberikan saran agar fitur pada aplikasi ditambahkan dan senantiasa *update*.

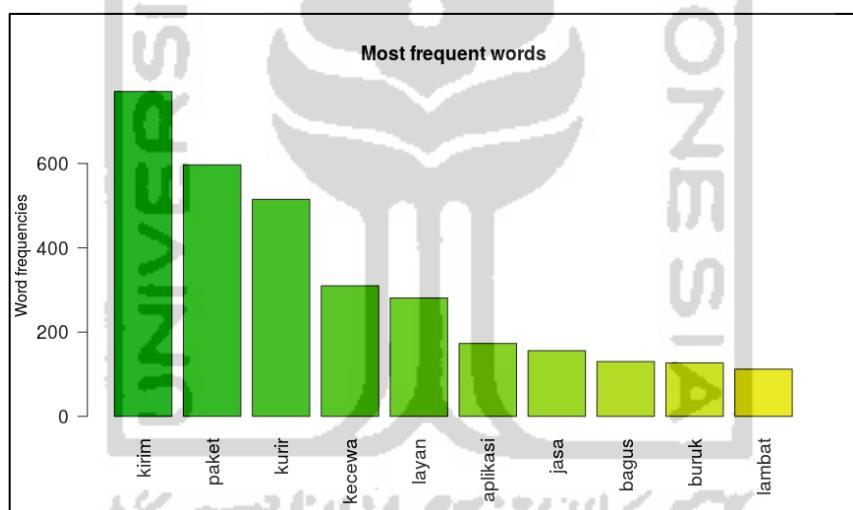
Asosiasi kata terkait dengan “parah” menunjukkan informasi bahwa pengguna mengeluhkan pelayanan yang diberikan oleh JNE terkesan ruwet. Pengguna telah menunggu kabar terkait paket namun tidak mendapatkan kabar dari JNE.

Asosiasi kata terkait dengan kata “lambat” menunjukkan informasi bahwa pengguna mengeluhkan operasional JNE yang terkesan lambat. Respon yang diberikan pun sangat minim. Pengguna telah menunggu kabar namun tidak mendapatkan kejelasan selain itu waktu penerimaan pun molor atau tidak sesuai dengan estimasi yang dijanjikan dengan yang di *claim* oleh JNE. Contohnya yaitu paket YES (Yakin Esok Sampai) namun nyatanya tidak selalu sampai dalam satu hari. Sehingga pengguna memberikan saran agar pihak-pihak yang terlibat dalam

penyaluran agar diberikan teguran atau sanksi agar tidak terjadi hal yang sama dikemudian hari.

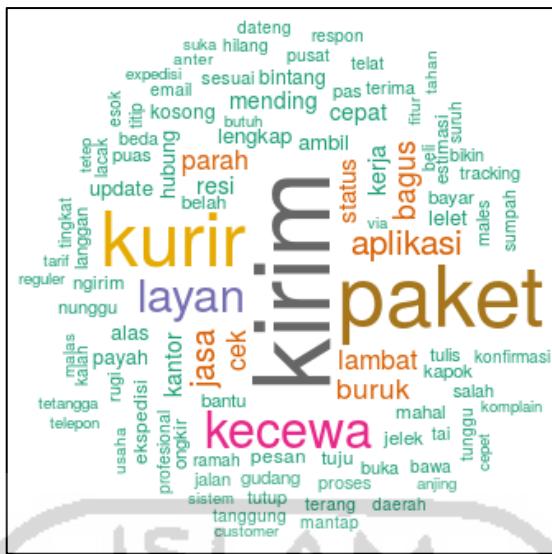
#### 4.2.6.3. Integrasi Ulasan Positif dan Negatif

Berdasarkan visualisasi dan asosiasi yang sebelumnya telah dilakukan pada ulasan positif dan negatif diketahui bahwa terdapat enam kata yang muncul baik pada ulasan positif maupun negatif. Kata tersebut terdiri atas : “kirim”, “paket”, “kurir”, “layan”, “aplikasi”, dan “jasa”. Pada penelitian ini dilakukan integrasi antara ulasan positif dan negatif untuk dapat memperjelas terkait permasalahan pokok apa yang terdapat pada keenam kata yang muncul di kedua kelas tersebut. Hasil ekstraksi informasi diidentifikasi berdasarkan frekuensi kata dalam ulasan, selain itu juga didasarkan pada relevansi kata dengan topik yang mengacu pada sentimen negatif dan juga positif. Berikut adalah visualisasi hasil ekstraksi informasi dalam bentuk *barplot* yang didapatkan dari ulasan pengguna secara keseluruhan :



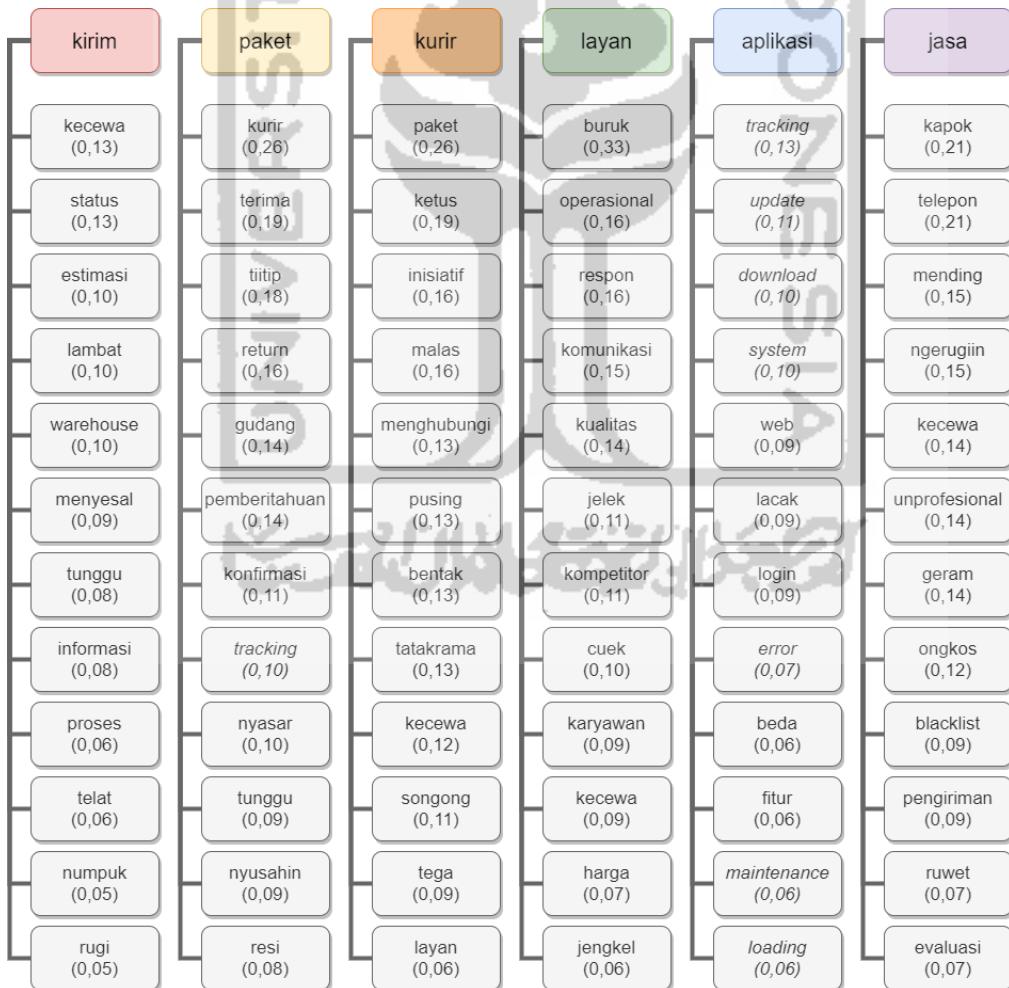
Gambar 4.20 *Barplot* Visualisasi Ulasan Positif dan Negatif

Berdasarkan ulasan kelas positif dan negatif dengan jumlah keseluruhan 1.843 ulasan diketahui bahwa pada 10 kata yang sering muncul terdapat kata “kirim” sebanyak 771 kali, “paket” sebanyak 597 kali, “kurir” sebanyak 515 kali, “kecewa” sebanyak 310 kali, “layan” sebanyak 281 kali, “aplikasi” sebanyak 173 kali, “jasa” sebanyak 156 kali, “bagus” sebanyak 130 kali, “buruk” sebanyak 127 kali, dan “lambat” sebanyak 112 kali. Berdasarkan kata-kata yang muncul barplot Gambar 4.20 selanjutnya dibuat *wordcloud* atau kumpulan kata yang sering dibahas oleh pengguna JNE, yaitu dapat dilihat pada Gambar 4.21 :



Gambar 4.21 *Wordcloud* ulasan keseluruhan

Selanjutnya untuk asosiasi kata hanya dilakukan pada enam kata berikut yang dilakukan dengan tujuan untuk menemukan pokok permasalahan dari JNE :



Gambar 4.22 Asosiasi kata yang sering muncul

Berdasarkan Gambar 4.22 diperoleh informasi sebagai berikut :

1) Kirim

Pada kata kirim diperoleh informasi bahwa pengguna JNE merasa kecewa dengan pengiriman yang dilakukan. Status pengiriman tidak *update*, selain itu estimasi pengiriman tidak sesuai dengan yang dijanjikan. JNE dinilai lambat dalam melakukan proses pengiriman. Pengguna mengeluhkan pengiriman terhambat pada *warehouse*, yang mana diketahui bahwa paket sudah terdapat di *warehouse* namun kemudian hanya dibiarkan saja menumpukan dan tidak diantar ke alamat yang dituju. Pengguna juga merasa menyesal menggunakan jasa pengiriman JNE. Pengguna menunggu informasi terkait pengiriman yang dilakukan, namun dari pihak JNE masih sangat minim dalam memberikan informasi. Pengguna menilai bahwa proses pengiriman pada JNE perlu diperbaiki karena sering terlambat dan menyebabkan pengguna rugi karena paket yang dikirim tiba tidak sesuai estimasi.

2) Paket

Kata paket berkaitan erat dengan kurir dikarenakan kurir berhubungan langsung dengan proses pengantaran paket kepada pengguna. Pengguna merasa paket tidak diterima tepat waktu. Selain itu pada resi tertera bahwa paket telah diterima namun ternyata nyasar. Pengguna JNE memberikan saran sebaiknya apabila pihak penerima sedang tidak dirumah, sebaiknya kurir menitipkan paket kepada tetangga. Pengguna juga mengeluhkan terkait paket yang *direturn* dengan alasan bahwa alamat yang tertera belum lengkap sedangkan pengguna merasa telah memberikan alamat secara rinci. Selain itu, pengguna mengeluhkan paket yang tertahan di gudang, yang mana ketika dilakukan *tracking* diketahui bahwa paket telah sampai di gudang kota tujuan namun sangat lama untuk dikirim. Pengguna yang meminta konfirmasi dari pihak gudang kemudian mendapat pemberitahuan bahwa paket tertahan di gudang demi kelengkapan laporan. Hal ini membuat penerima paket perlu menunggu dan dianggap menyusahkan karena proses yang lama.

3) Kurir

Jika sebelumnya pada kata paket berkaitan erat dengan kata kurir, begitu pula yang dapat dilihat pada kata ‘kurir’ berkaitan erat dengan paket. Berdasarkan hasil asosiasi diketahui bahwa penilaian pengguna didasarkan pada pelayanan kurir saat mengantarkan paket. Pengguna JNE menilai bahwa kurir ketus dan tidak memiliki

inisiatif untuk menelepon atau menghubungi penerima paket apabila terjadi kendala dalam pengantaran. Selain itu kurir dinilai malas dalam menjalankan tugas dan kewajibannya. Pengguna juga menilai kurir JNE kurang memiliki tatakrama yang baik, sebagai contoh terdapat pelanggan yang dibentak oleh kurir. Selain itu kurir juga dinilai songong, sehingga mengecewakan pelanggan atas pelayanan yang diberikan oleh kurir saat mengantar paket.

#### 4) Layan

Pada kata layan diperoleh informasi bahwa layanan yang diberikan oleh JNE dinilai buruk dimata pengguna. Pengguna JNE memberikan saran bahwa manajemen operasional JNE perlu diperbaiki. Selain itu respon yang diberikan pun masih dinilai terlalu lamban (*slow respon*) dan tidak ada komunikasi dari kurir atau *customer service*. Kualitas JNE juga dinilai jelek oleh pengguna. Kemudian pengguna membandingkan pelayanan antara yang diberikan oleh JNE dan kompetitor. Pelayanan yang diberikan oleh karyawan JNE dinilai cuek, sehingga pelanggan kecewa dan jengkel. Harga yang ditawarkan pun dinilai tidak sesuai dengan pelayanan yang diberikan.

#### 5) Aplikasi

Berdasarkan kata aplikasi diketahui bahwa terdapat informasi berkaitan yang erat dengan kata *tracking*. Pelanggan menggunakan aplikasi untuk melakukan *tracking* paket. Pengguna menyarankan agar aplikasi diupdate. Pada beberapa pengguna, aplikasi memiliki masalah dimana ketika pengguna telah mendownload, aplikasi tidak bisa dibuka. Selain itu ada pula pengguna yang memberikan ulasan bahwa sebaiknya pengguna lain tidak mendownload aplikasi dikarenakan aplikasi buruk. Pelanggan juga memberikan saran agar JNE memperbaiki sistemnya. Selain itu ketika dilakukan pelacakan paket menggunakan *website*, diketahui bahwa *website error*. Kemudian pengguna juga menilai fitur yang diberikan oleh JNE masih perlu ditambah atau ditingkatkan dan aplikasi JNE juga memerlukan *maintenance* secara rutin agar *loading* aplikasi tidak membutuhkan waktu yang lama seperti saat ini.

#### 6) Jasa

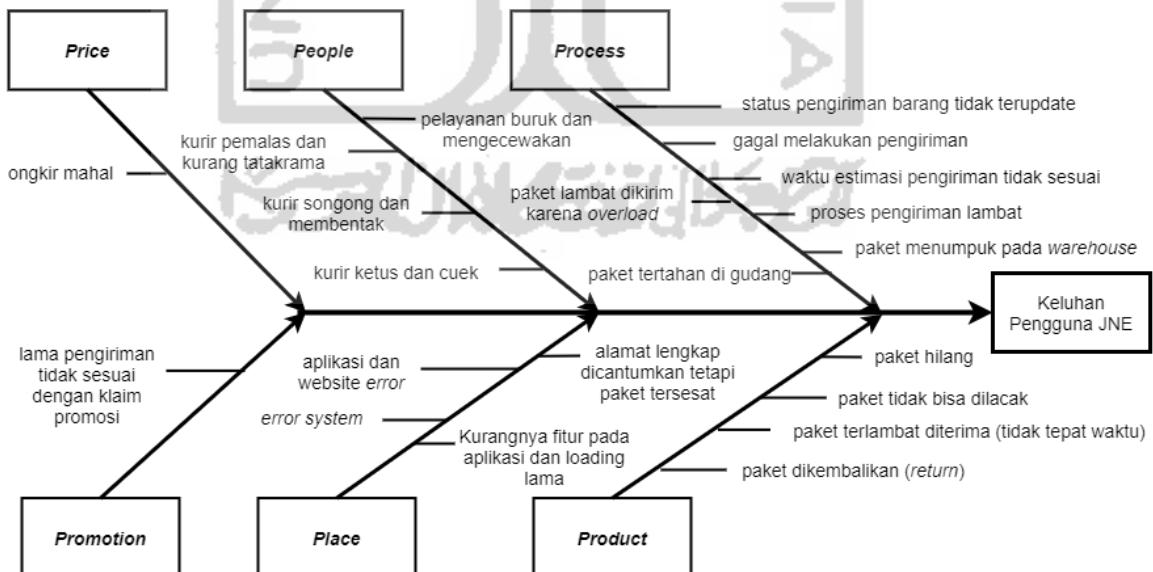
Berdasarkan kata jasa diketahui bahwa pengguna merasa kapok menggunakan jasa JNE. Pelanggan mengeluhkan bahwa pihak JNE tidak menelepon nomor telepon yang tertera dipaket (nomor penerima) sehingga tidak ada konfirmasi atau

pemberitahuan. Pelanggan membandingkan pelayanan JNE dengan kompetitor, Pelanggan berpendapat bahwa dengan menggunakan jasa JNE merugikan pelanggan dan membuat pelanggan kecewa karena dinilai tidak profesional. Pelanggan juga merasa geram dikarenakan proses pengiriman yang terkesan ruwet namun dengan ongkos kirim yang mahal tidak sesuai dengan jasa yang ditawarkan. Pelanggan memberikan saran agar jasa JNE segera dievaluasi untuk perbaikan kedepannya.

Ketika dilakukan integrasi antara ulasan positif dan negatif, diketahui bahwa kata yang muncul pada asosiasi cenderung bersifat negatif. Hal ini dikarenakan jumlah ulasan negatif lebih banyak dibanding positif. Berdasarkan asosiasi dari integrasi ulasan positif dan negatif ini selanjutnya dijadikan pertimbangan dalam interpretasi hasil dikarenakan ditemukan pokok permasalahan secara mendetail untuk melengkapi asosiasi negatif yang sebelumnya telah dilakukan. Adapun kata yang muncul dan bersifat positif disini dapat dilihat yaitu berupa saran untuk perbaikan JNE kedepannya.

#### 4.2.7. Interpretasi Hasil

Berdasarkan hasil ulasan negatif yang didapatkan dari data, maka diperoleh informasi mengenai masalah yang terjadi terkait dengan ulasan negatif pada jasa pengiriman JNE yang dianalisis dengan diagram *fishbone* sebagai berikut:



Gambar 4.23 Interpretasi Diagram *Fishbone* Ulasan Negatif

Berdasarkan Gambar 4.23 diketahui bahwa terdapat aspek *Process, People, Price, Promotion, Place*, dan *Product* yang mana didalamnya terdapat informasi terkait faktor penyebab JNE mendapat ulasan negatif. Permasalahan yang terdapat pada gambar 4.23 didasarkan pada informasi yang didapatkan melalui visualisasi dan asosiasi yang telah dilakukan sebelumnya.

Berikut merupakan informasi yang didapatkan sesuai dengan asosiasi negatif yang ada baik pada ulasan negatif maupun pada integrasi ulasan positif dan negatif sesuai dengan Gambar 4.19 dan Gambar 4.22 :

Tabel 4.21 Permasalahan pada JNE

Kata	Keterangan
Kirim	<ul style="list-style-type: none"> <li>- Status pengiriman tidak <i>update</i></li> <li>- Gagal melakukan pengiriman</li> <li>- Waktu estimasi pengiriman tidak sesuai</li> <li>- Proses pengiriman lambat</li> <li>- Pengiriman terhambat pada <i>warehouse</i> atau paket menumpuk pada <i>warehouse</i></li> </ul>
Paket	<ul style="list-style-type: none"> <li>- Paket tidak diterima tepat waktu</li> <li>- Paket hilang</li> <li>- Paket direturn</li> <li>- Paket tertahan di gudang daerah tujuan</li> <li>- Paket tidak bisa dilacak</li> <li>- Paket lambat dikirim</li> <li>- Alamat lengkap dicantumkan tetapi paket tersesat</li> <li>- Paket terlambat dikirim karena <i>overload</i></li> </ul>
Kurir	<ul style="list-style-type: none"> <li>- Kurir ketus dan cuek</li> <li>- Kurir malas</li> <li>- Kurir kurang tatakrama</li> <li>- Kurir songong dan membentak</li> </ul>
Kecewa	Berkaitan dengan jasa yang diberikan, termasuk didalamnya proses pengiriman, pelayanan yang diberikan oleh kurir, dan aplikasi
Layan	<ul style="list-style-type: none"> <li>- Pelayanan buruk dan mengecewakan</li> <li>- <i>Customer service</i> maupun kurir <i>slowrespon</i></li> </ul>

Kata	Keterangan
Jasa	<ul style="list-style-type: none"> <li>- Proses yang terkesan ruwet dibanding kompetitor</li> <li>- Ongkos kirim mahal</li> </ul>
Buruk	Berkaitan dengan pelayanan, manajemen operasional, sistem pengiriman, sumber daya manusia dan aplikasi
Aplikasi	<ul style="list-style-type: none"> <li>- Aplikasi dan <i>website error</i> saat <i>tracking</i> paket</li> <li>- Berkaitan dengan pengiriman : status pengiriman tidak <i>update</i></li> <li>- <i>Loading</i> lama</li> <li>- <i>Error System</i></li> <li>- Fitur kurang</li> </ul>
Parah	Berkaitan dengan proses yang dinilai ruwet
Lambat	Berkaitan dengan respon yang diberikan dan terlambatnya paket (tidak tepat waktu)

Berdasarkan tabel 4.21 didapatkan informasi yang kemudian dilakukan pengelompokan kembali berdasarkan 6 aspek yang digunakan yaitu *Process, People, Price, Promotion, Place*, dan *Product* yaitu sebagai berikut :

Tabel 4.22 Pengelompokan Permasalahan Berdasarkan 6 Aspek

Aspek	Permasalahan
<i>Process</i>	<p>Terkait dengan proses pengiriman :</p> <ul style="list-style-type: none"> <li>- Status pengiriman tidak update</li> <li>- Gagal melakukan pengiriman</li> <li>- Waktu estimasi pengiriman tidak sesuai</li> <li>- Proses pengiriman lambat</li> <li>- Pengiriman terhambat pada <i>warehouse</i> atau paket menumpuk pada <i>warehouse</i></li> <li>- Manajemen operasional buruk</li> <li>- Paket tertahan di gudang daerah tujuan</li> <li>- Paket lambat dikirim karena <i>overload</i></li> </ul>
<i>People</i>	<p>Berkaitan dengan kurir :</p> <ul style="list-style-type: none"> <li>- Kurir ketus dan cuek</li> <li>- Kurir malas dan kurang tatakrama</li> </ul>

Aspek	Permasalahan
	<ul style="list-style-type: none"> <li>- Kurir songong dan membentak</li> <li>Berkaitan dengan pelayanan :</li> <li>- Pelayanan buruk dan mengecewakan</li> </ul>
<i>Price</i>	Ongkos kirim mahal
<i>Promotion</i>	Lama pengiriman tidak sesuai slogan promosi
<i>Place</i>	<ul style="list-style-type: none"> <li>- Aplikasi dan <i>website error</i></li> <li>- <i>Error System</i></li> <li>- Alamat lengkap dicantumkan tetapi paket tersesat</li> <li>- Fitur kurang pada aplikasi dan <i>loading</i> aplikasi lama</li> </ul>
<i>Product</i>	<ul style="list-style-type: none"> <li>- Paket tidak diterima tepat waktu</li> <li>- Paket hilang</li> <li>- Paket direturn</li> <li>- Paket tidak bisa dilacak</li> </ul>

## BAB V

### PEMBAHASAN

#### 5.1. Gambaran Umum Persepsi Pengguna JNE

Berdasarkan data ulasan dari pengguna JNE pada 19 September sampai dengan 31 Desember 2019, didapatkan ulasan sebanyak 1.876. Pada kurun waktu tersebut diketahui bahwa ulasan terbanyak ada pada bulan Desember 2019 seperti yang dapat dilihat pada Gambar 4.8. Hal ini terjadi dikarenakan pada bulan Desember terdapat HARBOLNAS yang memicu peningkatan jumlah penjualan *e-commerce* sehingga penggunaan jasa pengiriman termasuk JNE juga turut meningkat. Kemudian dari segi *rating*, sebanyak 70,8% pengguna atau 1.328 pengguna memberikan rating 1 atau tidak bagus. Setelah dilakukan analisis deskriptif untuk melihat gambaran umum dari pengguna JNE selanjutnya dilakukan *pre-processing data*. Pada tahap *pre-processing* data ulasan yang awalnya berjumlah 1.876 kemudian berubah jumlahnya menjadi 1.843 ulasan. Banyaknya jumlah ulasan yang berkurang ini disebabkan karena adanya ulasan yang hanya berupa emotikon atau simbol sehingga hilang ketika dilakukan pembersihan data. Pada awalnya ulasan dibagi menjadi tiga kelas sentimen, yaitu positif sebanyak 170 ulasan, netral sebanyak 605 ulasan, dan negatif sebanyak 1.68 ulasan. Banyaknya jumlah sentimen netral disebabkan karena adanya kata yang tidak terdeteksi dan jumlah kata positif serta negatif yang setara. Namun sentimen netral dinilai kurang memberikan masukan dan manfaat bagi pihak JNE sehingga label netral dimasukkan ke sentimen positif atau negatif berdasarkan beberapa pertimbangan. Pelabelan kelas sentimen netral akan dilakukan secara manual, dengan ketentuan : jika pada ulasan tidak teridentifikasi kata positif atau negatif maka akan diklasifikasikan pada kelas positif dan jika jumlah kata positif atau negatif sama maka akan diklasifikasikan sebagai kelas negatif dengan pertimbangan agar pihak terkait dapat melihat masalah-masalah yang terjadi.

Pada ulasan yang mengandung pernyataan positif seperti ungkapan memuji, ungkapan terimakasih dan lain sebagainya dimasukkan ke dalam kategori sentimen positif. Adapun untuk ulasan klasifikasi yang mengandung pernyataan negatif seperti penghinaan, ketidakpuasan, dan sebagainya dimasukkan ke dalam kategori sentimen

negatif. Oleh karena itu hanya digunakan dua pelabelan yaitu sentimen positif dan negatif. Berdasarkan hasil akhir pelabelan diperoleh informasi bahwa jumlah ulasan sentimen negatif memiliki frekuensi lebih besar dibanding sentimen positif. Hal ini ditunjukkan oleh Gambar 4.11, dimana ulasan pada kelas sentimen negatif berjumlah 69% atau 1.272 ulasan. Sedangkan 31% atau 571 ulasan termasuk sentimen positif.

## 5.2. Hasil Klasifikasi dan Asosiasi Metode *Naïve Bayes Classifier*

Klasifikasi dengan metode *Naïve Bayes Classifier* diawali dengan membagi jumlah data uji dan data latih. Pada penelitian ini dilakukan perbandingan dengan persentase data latih dan data uji yang berbeda sebanyak lima kali. Pembagian data pertama diawali dengan data latih 60% (1.106 ulasan) dan data uji 40% (737 ulasan). Pada pembagian data latih dan data uji yang pertama diketahui *Class Precision* negatif 89,27% sedangkan untuk kelas positif sebesar 65,7%. Adapun akurasi yang dihasilkan sebesar 80,6%.

Pembagian data kedua dilakukan dengan data latih 70% (1.290 ulasan) dan data uji 30% (553 ulasan). Pada pembagian data latih dan data uji yang kedua dihasilkan *Class Precision* kelas negatif sebesar 89,3% sedangkan untuk kelas positif sebesar 69,23% dengan akurasi yang dihasilkan lebih tinggi dibandingkan pembagian data pertama, yaitu 82,3%.

Pembagian data ketiga dilakukan dengan data latih 80% (1.474 ulasan) dan data uji 20% (368 ulasan). Pada pembagian data latih dan data uji yang ketiga dihasilkan *Class Precision* kelas negatif sebesar 89,3% sedangkan untuk kelas positif meningkat menjadi 75,2%, dengan akurasi yang dihasilkan lebih tinggi dibandingkan pembagian data pertama dan kedua, yaitu 84,55%.

Pembagian data keempat dilakukan dengan data latih 85% (1.567 ulasan) dan data uji 15% (276 ulasan). Pada pembagian data latih dan data uji yang keempat dihasilkan *Class Precision* kelas negatif sebesar 88,5% sedangkan untuk kelas positif meningkat menjadi 79,6%, dengan akurasi yang dihasilkan lebih tinggi dibandingkan tiga perbandingan data sebelumnya, yaitu 85,5%.

Pembagian data terakhir dilakukan dengan data latih 90% (1.659 ulasan) dan data uji 10% (184 ulasan). Pada pembagian data latih dan data uji yang terakhir dihasilkan *Class Precision* kelas negatif sebesar 88,5% sedangkan untuk kelas positif meningkat hingga lebih dari 80% yaitu 80,6%, dengan akurasi yang dihasilkan lebih tinggi dibandingkan semua pembagian data sebelumnya, yaitu 85,87%.

Berdasarkan hasil penerapan metode *Naïve Bayes Classifier*, hasil klasifikasi menunjukkan bahwa pada data uji dan data latih yang berbeda dihasilkan akurasi yang cukup baik karena telah lebih dari 80%. Adapun dari lima perbedaan pembagian data yang telah dilakukan kemudian dilakukan perbandingan sehingga diketahui bahwa yang terbaik adalah pembagian data latih 90% dan data uji 10%. Hal ini dibuktikan melalui nilai akurasi tertinggi dibanding dengan perbandingan lainnya yaitu 85,87% dengan *Class Precision* diatas 80%, untuk kelas negatif 88,5% dan positif 80,6%. Kemudian berdasarkan hasil klasifikasi *naïve bayes* pada grafik Gambar 4.13 diketahui bahwa selain akurasi, *precision* dan *recall* juga secara garis besar meningkat diiringi dengan berkurangnya data uji yang digunakan.

Pendeteksian sentimen pada penelitian ini menggunakan kamus *lexicon* atau *bag of words*. Pada *bag of words* sistem mempelajari kata-kata yang terdapat dalam suatu sumber tanpa adanya pengkategorian kata berdasarkan struktur penyusun kalimat, seperti kata benda, kata kerja, kata sifat, dan lain sebagainya. Hal inilah yang menjadi pemicu terjadinya kesalahan pada sistem pelabelan yang disebabkan oleh kerancuan padanan kata. Sebagai contoh seperti kata “tidak bagus”, padanan kata ini dapat berubah menjadi kata “bagus”, yang diketahui bahwa keduanya saling bertolak belakang. Oleh karena itu penting dilakukannya asosiasi kalimat, tidak cukup dengan kata saja, contohnya menggunakan *semantic parsing*.

Berdasarkan visualisasi yang telah dilakukan pada ulasan kelas sentimen positif dengan jumlah ulasan sebanyak 571, didapatkan beberapa kata yang paling sering muncul, dalam hal ini 10 kata dengan frekuensi terbanyak antara lain : kata “kirim” sebanyak 154 kali, kata “paket” sebanyak 115 kali, kata “kurir” sebanyak 85 kali, kata “bagus” sebanyak 81 kali, kata “aplikasi” sebanyak 61 kali, kata “cepat” sebanyak 48 kali, kata “layan” sebanyak 48 kali, kata “terima” sebanyak 38 kali, kata “bantu” sebanyak 37 kali, dan kata “bintang” sebanyak 35 kali.

Asosiasi kata pada kelas sentimen positif dilakukan dengan melakukan ekstraksi informasi secara berulang dan didasarkan pada relevansi kata dengan topik yang ada. Pada Gambar 4.16 didapatkan informasi bahwa asosiasi kata dilakukan pada kata “kirim”, “paket”, “kurir”, “bagus”, “aplikasi”, “cepat”, “layan”, “terima”, “bantu”, dan “bintang”.

Berdasarkan asosiasi kata diketahui bahwa pengguna merasa dari segi pengiriman JNE telah memiliki performa yang baik dan juga cepat sehingga *recommended* dimata pengguna, serta pengguna memberikan saran agar JNE melakukan peningkatan agar status pengiriman senantiasa *update*. Adapun paket diterima oleh pelanggan dengan baik dan pelanggan memberikan saran agar kurir memiliki inisiatif untuk menelepon atau menghubungi terkait kabar paket yang dikirim. Kurir dinilai baik karena memberikan pelayanan yang ramah dan sopan. Pengguna memberikan saran agar kurir menelepon penerima ataupun menitipkan paket melalui tetangga. JNE dianggap telah melayani dengan bagus karena dilakukan sortir terlebih dahulu dan juga ramah. Pengguna juga memberikan saran terkait transit paket di *warehouse*. Pengguna berharap agar waktu transit tidak begitu lama. Adapun pengguna menilai bahwa aplikasi membantu pengguna dalam melakukan *tracking*, selain itu aplikasi juga memberikan manfaat bagi pengguna. Adapun pengguna memberikan saran agar aplikasi senantiasa *update*. Selain cepat JNE juga dianggap menawarkan harga yang murah, mampu bersaing dengan kompetitor, dan juga juara bagi penggunanya. Pengguna berterimakasih atas pelayanan yang ditawarkan dan mendukung JNE dengan memberikan berbagai saran untuk perbaikan JNE kedepannya.

Sedangkan berdasarkan 1.272 ulasan negatif atau 69% dari 1.843 ulasan diketahui bahwa terdapat kata yang paling banyak muncul. Adapun 10 kata dengan frekuensi tertinggi seperti yang dapat dilihat pada Gambar 4.17. Kata yang paling banyak muncul diantaranya : kata “kirim” dengan frekuensi sebanyak 617 kali, kata “paket” dengan frekuensi sebanyak 482 kali, kata “kurir” dengan frekuensi sebanyak 430 kali, kata “kecewa” dengan frekuensi sebanyak 306 kali, kata “layan” dengan frekuensi sebanyak 233 kali, kata “jasa” dengan frekuensi sebanyak 133 kali, kata “buruk” dengan frekuensi sebanyak 123 kali, kata “aplikasi” dengan frekuensi sebanyak 112 kali, kata “parah” dengan frekuensi sebanyak 108 kali, serta kata “lambat” dengan frekuensi sebanyak 106 kali.

Asosiasi kata pada kelas sentimen negatif dilakukan dengan melakukan ekstraksi informasi secara berulang dan didasarkan pada relevansi kata dengan topik yang ada. Pada Gambar 4.19 didapatkan informasi bahwa asosiasi kata dilakukan pada kata “kirim”, “paket”, “kurir”, “kecewa”, “layan”, “jasa”, “buruk”, “aplikasi”, “parah”, dan “lambat”.

Pengguna menilai telah memberikan alamat secara rinci namun mengalami gagal pengiriman. Selain itu pengguna juga menduga gagal pengiriman ini terjadi dikarenakan kinerja kurir yang tidak baik. Pengguna menilai kurir ketus dan pemalas. Pengguna merasa jengkel dikarenakan paket yang tidak kunjung tiba dan bahkan ada yang hilang ataupun tersesat ke rumah lain. Kemudian paket di *return* atau dikembalikan lagi kepada pengirim. Selain itu pengguna merasa JNE masih tergolong kurang dalam memberikan pemberitahuan terkait informasi paket. Pengguna berharap pihak JNE segera menindaklanjuti. Pengguna merasa kecewa karena tidak mendapat kejelasan dan tanggapan terkait paket mereka sehingga merasa percuma menggunakan jasa JNE. Pengguna membandingkan kecepatan atau laju dari kemampuan JNE dengan kompetitornya dalam mengantarkan barang, yang mana sangat disayangkan kinerja JNE dalam mengirim barang termasuk lama dan tidak sesuai estimasi. Pengguna beropini bahwa pelayanan yang diberikan oleh JNE masih tergolong buruk bila dibandingkan dengan kompetitornya. Aplikasi JNE disarankan untuk ditambahkan fitur dan senantiasa di *update* karena saat melakukan *tracking* paket tidak dapat dilacak. Pengguna mengeluhkan pelayanan yang diberikan oleh JNE terkesan ruwet. Pengguna telah menunggu kabar namun tidak mendapatkan kejelasan selain itu waktu penerimaan pun molor atau tidak sesuai dengan estimasi yang dijanjikan dengan yang di *claim* oleh JNE. Contohnya yaitu paket YES (Yakin Esok Sampai) namun nyatanya tidak selalu sampai dalam satu hari. Sehingga pengguna memberikan saran bahwa baik CS maupun kurir perlu dikontrol kembali dalam memberikan pelayanan kepada pelanggan.

Integrasi ulasan positif dan negatif dilakukan pada keseluruhan ulasan (positif dan negatif) yaitu 1.843 ulasan. Integrasi ini bertujuan untuk menemukan pokok permasalahan dari JNE dan apa yang paling sering dibicarakan secara garis besar keseluruhan ulasan. Berdasarkan visualisasi yang dilakukan diketahui bahwa terdapat 10 kata yang sering muncul seperti yang dapat dilihat pada Gambar 4.20 terdapat kata

“kirim” sebanyak 771 kali, “paket” sebanyak 597 kali, “kurir” sebanyak 515 kali, “kecewa” sebanyak 310 kali, “layan” sebanyak 281 kali, “aplikasi” sebanyak 173 kali, “jasa” sebanyak 156 kali, “bagus” sebanyak 130 kali, “buruk” sebanyak 127 kali, dan “lambat” sebanyak 112 kali.

Berdasarkan visualisasi yang telah dilakukan diketahui bahwa terdapat enam kata yang selalu muncul baik pada ulasan positif maupun negatif, antara lain : kata “kirim”, “paket”, “kurir”, “layan”, “aplikasi”, dan “jasa”. Pada keenam kata tersebut kemudian dilakukan asosiasi menggunakan *software* Rstudio dan juga dianalisis dengan melihat data ulasan hasil *scraping*.

Berdasarkan hasil asosiasi pada Gambar 4.22 diketahui bahwa sebelumnya pada ulasan positif pengguna menyarankan agar status pengiriman senantiasa *update*, sedangkan pada ulasan negatif pengguna mengeluhkan status pengiriman yang tidak update. Hal ini menjadi salah satu contoh terkait integrasi yang berkaitan dengan pengiriman yang dilakukan oleh JNE. Selain itu pengguna mengeluhan proses pengiriman yang dinilai lambat dan tidak sesuai dengan waktu estimasi yang dijanjikan (paket tidak sampai tepat waktu). Salah satu titik permasalahan adalah pengiriman terhambat yang disebabkan adanya paket menumpuk pada *warehouse*. Selain penumpukan paket, diketahui juga terdapat permasalahan terkait paket yang tertahan digudang. Adapun ketika dilakukan *tracking* diketahui bahwa paket telah sampai di gudang kota tujuan namun membutuhkan waktu yang sangat lama untuk dikirim. Berdasarkan salah satu ulasan diketahui bahwa terdapat pengguna yang meminta konformasi dari pihak gudang namun kemudian diberitahukan bahwa paket masih tertahan demi kelengkapan laporan. Selain itu penumpukan paket disebabkan oleh berbagai hal, salah satunya pengguna menilai hal ini disebabkan karena kinerja kurir yang kurang baik. Pengguna menilai kurir malas dalam mengantarkan paket sehingga paket menumpuk. Selain itu pengguna juga mengkritik pelayanan yang diberikan oleh kurir terkesan ketus dan cuek sehingga kurang dari segi tatakrama. Pelanggan juga memberikan saran agar kurir memiliki inisiatif untuk menelepon ketika akan mengantarkan paket atau memberitahu informasi terkait *update* terkini dari paket apabila terjadi kendala dalam pengantaran sehingga pelanggan mendapat kejelasan dan tidak menunggu tanpa kepastian. Dari segi aplikasi pun pengguna menyarankan agar senantiasa dilakukan *update*, penambahan fitur, dan *maintenance* secara berkala. Selain

itu terkait dengan ongkos kirim, dikarenakan paket yang sering terlambat datang, hilang, maupun tersesat, pengguna merasa ongkos kirim yang ditetapkan tergolong mahal dikarenakan tidak sesuai dengan pelayanan yang diberikan. Pelanggan juga memberikan saran agar jasa JNE segera dievaluasi untuk perbaikan kedepannya.

### **5.3. Rencana Pemecahan Masalah Berdasarkan Diagram *Fishbone***

Berdasarkan diagram *fishbone* pada Gambar 4.23 diketahui informasi yang menjadi penyebab mengapa JNE mendapat ulasan negatif dari penggunanya. Terdapat 6 faktor yaitu *Process*, *People*, *Price*, *Promotion*, *Place*, dan *Product*. Langkah selanjutnya ialah menentukan rencana pemecahan masalah dari permasalahan yang ada. Adapun rencana pemecahan masalah dapat dilihat pada tabel berikut :

Tabel 5.1 Rencana Pemecahan Masalah

<b>Aspek</b>	<b>Permasalahan</b>	<b>Rencana Pemecahan Masalah</b>
<i>Process</i>	<p>Terkait dengan proses pengiriman :</p> <ul style="list-style-type: none"> <li>- Status pengiriman tidak update</li> <li>- Gagal melakukan pengiriman</li> <li>- Waktu estimasi pengiriman tidak sesuai</li> <li>- Proses pengiriman lambat</li> <li>- Pengiriman terhambat pada <i>warehouse</i> atau paket menumpuk pada <i>warehouse</i></li> <li>- Paket tertahan digudang di daerah tujuan</li> <li>- Paket lambat dikirim karena <i>overload</i></li> </ul>	<p>Proses pengiriman merupakan masalah pokok yang menyebabkan terjadinya keterlambatan paket sampai ke tangan konsumen. Pada hal ini paket terhambat pada <i>warehouse</i> dan adanya paket yang menumpuk di gudang karena lamanya waktu <i>transit</i> perlu di analisis lebih lanjut oleh pihak JNE. Pemecahan masalah yang dapat dilakukan antara lain :</p> <ol style="list-style-type: none"> <li>1. Paket yang tertahan di gudang daerah tujuan dalam waktu yang lama mengindikasikan adanya paket yang tercecer atau luput saat di sortir petugas gudang. Oleh karena itu perlu dilakukan</li> </ol>

<b>Aspek</b>	<b>Permasalahan</b>	<b>Rencana Pemecahan Masalah</b>
		<p>pengelompokan paket sejak awal agar tidak ada yang tercecer dan memperbaiki tata letak gudang yang diatur berdasarkan jenis paket dan alamat tujuan</p> <ol style="list-style-type: none"> <li>2. Menandai paket sesuai dengan kode tertentu sesuai alamat tujuan agar mudah diidentifikasi</li> <li>3. Paket yang telah diberi kode tertentu sesuai alamat dikirim dari gudang ke masing-masing <i>Drop Point</i> yang tersebar di beberapa wilayah sehingga kurir hanya mengantarkan barang sesuai dengan wilayah <i>Drop Point</i> nya saja</li> <li>4. Apabila terdapat masalah di gudang berkaitan dengan terhambatnya pengiriman via udara atau kargo, pihak JNE sebaiknya menghubungi <i>customer</i> agar <i>customer</i> mengetahui kendala yang terjadi dan dapat mengerti alasan mengapa paket tiba tidak sesuai estimasi</li> <li>5. Meminimalisir adanya proses yang tidak diperlukan untuk mempercepat waktu <i>transit</i></li> </ol>

<b>Aspek</b>	<b>Permasalahan</b>	<b>Rencana Pemecahan Masalah</b>
		<p>6. Memberikan target harian minimum yang wajib dipenuhi untuk pengantaran paket</p> <p>7. Pihak <i>developer</i> segera memperbaiki sistem cek resi baik pada <i>website</i> maupun aplikasi my JNE agar status pengiriman senantiasa <i>update</i></p> <p>8. Memperbaiki dan memperketat sistem agar tidak terjadi lagi gagal pengiriman dikarenakan alamat yang dinilai tidak lengkap. Saat diawal pengirim mendaftarkan paketnya, paket wajib mencantumkan alamat lengkap, jika tidak maka tidak akan diproses</p>
<i>People</i>	<p>Berkaitan dengan kurir :</p> <ul style="list-style-type: none"> <li>- Kurir ketus dan cuek</li> <li>- Kurir malas dan kurang tatakrama</li> <li>- Kurir songong dan membentak</li> </ul> <p>Berkaitan dengan pelayanan :</p> <ul style="list-style-type: none"> <li>- Pelayanan buruk dan mengecewakan</li> </ul>	<p>Terkait dengan <i>people</i> atau SDM dari JNE, dapat dilakukan dengan:</p> <ol style="list-style-type: none"> <li>1. Melakukan <i>training</i> bagi kurir dan <i>customer service</i> terkait pelayanan pelanggan</li> <li>2. Membuat SOP bagi <i>customer service</i> maupun kurir untuk melayani pertanyaan pelanggan dengan baik.</li> <li>3. Menambahkan fitur penilaian pada aplikasi berupa pemberian rating, kritik dan saran bagi kurir yang</li> </ol>

<b>Aspek</b>	<b>Permasalahan</b>	<b>Rencana Pemecahan Masalah</b>
		<p>mengantarkan paket agar pelanggan dapat memberikan penilaian terhadap kinerja kurir.</p> <p>4. Menambahkan fitur informasi terkait kurir yang mengantarkan paket yang didalamnya terdapat nomor telepon atau akses chat kurir secara langsung agar ketika dilakukan <i>tracking</i> paket akan muncul nomor telepon kurir yang sedang mengurus paket dan pelanggan dapat langsung menghubungi kurir bila ada yang ingin ditanyakan</p> <p>5. Menyediakan <i>database</i> berupa nomor telepon dari tiap agen JNE di aplikasi untuk mempermudah pelanggan dalam mencari informasi keberadaan paket bila paket belum berada di tangan kurir</p> <p>6. Memperketat kontrol atau pengawasan</p> <p>7. Melaksanakan evaluasi kinerja salah satunya yaitu berdasarkan penilaian pelanggan pada aplikasi</p>

<b>Aspek</b>	<b>Permasalahan</b>	<b>Rencana Pemecahan Masalah</b>
		8. Memberikan hukuman tegas bagi <i>customer service</i> maupun kurir yang lalai dan tidak memberikan pelayanan dengan baik.
<i>Price</i>	Ongkos kirim mahal	Melakukan pengkajian kembali terhadap ongkos kirim yang dilakukan dengan membandingkan harga dengan kompetitor dan memberikan potongan ongkos kirim bagi pelanggan yang kecewa akibat paket tidak datang tepat waktu
<i>Promotion</i>	Lama pengiriman tidak sesuai klaim promosi	Pihak JNE harus lebih konsisten dalam pelaksanaan pengiriman disesuaikan dengan klaim promosi yang ada dan memberikan informasi kepada pelanggan apabila terdapat kendala yang menyebabkan paket tidak tiba tepat pada waktunya
<i>Place</i>	<ul style="list-style-type: none"> <li>- Aplikasi dan <i>website error</i></li> <li>- <i>Error System</i></li> <li>- Alamat lengkap dicantumkan tetapi paket tersesat</li> <li>- Kurangnya fitur pada aplikasi dan <i>loading</i> lama</li> </ul>	<p>Adapun terkait aplikasi dan alamat paket, rencana pemecahan masalahnya adalah sebagai berikut:</p> <ol style="list-style-type: none"> <li>1. Melakukan update <i>sistem</i> secara berkala baik pada aplikasi maupun pada <i>website</i> dan melakukan pengujian terhadap aplikasi hingga tidak terdeteksi <i>error</i></li> </ol>

<b>Aspek</b>	<b>Permasalahan</b>	<b>Rencana Pemecahan Masalah</b>
		<p>2. Melakukan penambahan fitur yang diperlukan untuk menunjang kinerja JNE dan melakukan <i>maintenance</i> terhadap aplikasi maupun website setidaknya setahun sekali</p> <p>3. Menerapkan peraturan bahwa pengirim wajib mencantumkan alamat lengkap, jika tidak maka tidak akan diproses</p> <p>4. Membuat SOP terkait pengantaran paket, seperti kurir menghubungi penerima dan mengkonfirmasi alamat yang dituju.</p>
<i>Product</i>	<ul style="list-style-type: none"> <li>- Paket tidak diterima tepat waktu</li> <li>- Paket hilang</li> <li>- Paket direturn</li> <li>- Paket tidak bisa dilacak</li> </ul>	<p>Terkait paket, berikut rencana pemecahan masalah yang dapat dilakukan :</p> <ol style="list-style-type: none"> <li>1. Memberikan informasi dan penjelasan kepada pelanggan apabila pada proses pengiriman terdapat kendala yang menyebabkan paket tidak datang tepat waktu</li> <li>2. Melakukan perbaikan tata letak gudang seperti yang sudah dijelaskan dibagian proses untuk meminimalisir terjadinya paket hilang atau</li> </ol>

<b>Aspek</b>	<b>Permasalahan</b>	<b>Rencana Pemecahan Masalah</b>
		<p>tercecer</p> <p>3. Membuat SOP dan memberikan tindakan tegas bagi kurir yang dengan sengaja lalai dalam melaksanakan tugas</p> <p>4. Membuat SOP terkait dengan pengantaran paket, seperti kurir menghubungi penerima dan mengkonfirmasi alamat yang dituju.</p> <p>5. Memberikan hukuman tegas kepada kurir agar tidak lalai ataupun membawa lari paket pelanggan.</p> <p>6. Terkait dengan paket yang tidak bisa dilacak berhubungan dengan aplikasi dan website yang mana perlu dilakukan <i>update</i> secara berkala terkait keberadaan dari setiap paket yang ada</p>

## BAB VI

### KESIMPULAN DAN SARAN

#### 6.1. Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan di bab sebelumnya, diperoleh beberapa kesimpulan yakni sebagai berikut :

1. Pada kurun waktu 19 September – 31 Desember 2019 didapatkan ulasan pengguna JNE pada website *Google Play* sebanyak 1.876 ulasan. Adapun pengguna paling banyak memberikan rating 1 dengan persentase 70,8% atau sebanyak 1.328 pengguna. Sedangkan pada pelabelan kelas sentimen setelah dilakukan *pre-processing* diketahui ulasan yang dapat diolah sebanyak 1.843 yang kemudian diperoleh jumlah ulasan negatif sebanyak 1.272 ulasan (69%) dan ulasan positif sebanyak 571 ulasan (31%).
2. Penerapan metode *Naïve Bayes Classifier* dilakukan dengan menggunakan lima jenis perbandingan data latih dan data uji. Didapatkan hasil bahwa perbandingan data latih dan data uji sebesar 90% : 10% menghasilkan tingkat akurasi, *precision*, dan *recall* paling optimal dengan akurasi sebesar 85,87%
3. Berdasarkan hasil klasifikasi dan asosiasi yang dilakukan, didapatkan informasi bahwa secara umum pengguna membahas enam hal yakni kirim, paket, kurir, layan, aplikasi, dan jasa. Pada kelas sentimen positif pengguna banyak membahas mengenai kata “kirim”, “paket”, “kurir”, “bagus”, “aplikasi”, “cepat”, “layan”, “terima”, “bantu”, dan “bintang”. Sedangkan pada kelas sentimen negatif pengguna sering membicarakan mengenai “kirim”, “paket”, “kurir”, “kecewa”, “layan”, “jasa”, “buruk”, “aplikasi”, “parah”, dan “lambat”.
4. Berdasarkan diagram *fishbone* didapatkan permasalahan yang didasarkan pada enam faktor yaitu *Process*, *People*, *Price*, *Promotion*, *Place*, dan *Product*. Dari keenam faktor tersebut diperoleh permasalahan sebanyak 21, antara lain : status pengiriman barang tidak *update*, paket tidak bisa dilacak, gagal melakukan pengiriman, waktu estimasi pengiriman tidak sesuai, proses pengiriman lambat, pengiriman terhambat pada *warehouse*, paket tertahan di gudang di daerah tujuan, paket lambat dikirim karena *overload*, pelayanan buruk dan mengecewakan, kurir

ketus dan cuek, kurir pemalas dan kurang tatakrama, kurir songong dan membentak, ongkir mahal, lama pengiriman tidak sesuai dengan klaim promosi, alamat lengkap dicantumkan tetapi paket tersesat, *error system*, aplikasi dan *website error*, kurangnya fitur pada aplikasi dan loading lama, paket tidak diterima tepat waktu, paket hilang, paket direturn, paket tidak bisa dilacak.

## 6.2. Saran

Berdasarkan beberapa kesimpulan diatas, adapun saran yang dapat diberikan adalah sebagai berikut :

### 1. Saran bagi pihak JNE

Bagi pihak JNE, hasil ekstraksi informasi dari ulasan-ulasan yang telah diberikan oleh pengguna khususnya ulasan yang berbentuk negatif dapat dijadikan bahan evaluasi khususnya dalam rangka peningkatan kepuasan pengguna, peningkatan kualitas dari SDM dalam memberikan pelayanan semaksimal mungkin, serta untuk pengembangan pembaharuan aplikasi dan bisnis selanjutnya.

### 2. Saran untuk penelitian selanjutnya

Peneliti selanjutnya dapat mengembangkan penelitian dengan menggunakan berbagai bahasa dikarenakan penelitian ini masih dibatasi oleh ulasan berbahasa Indonesia. Kemudian penelitian ini baru menggunakan satu algoritma yaitu *Naïve Bayes Classifier*, sehingga penelitian selanjutnya dapat dilakukan menggunakan pendekatan *machine learning* lainnya sebagai pembanding performa algoritma *Naïve Bayes Classifier* untuk mengklasifikasikan ulasan. Selain itu pada penelitian ini hanya menganalisis satu jasa pengiriman, untuk penelitian selanjutnya dapat membandingkan lebih dari satu jasa pengiriman. Serta pelabelan kelas sentimen yang digunakan dalam penelitian ini masih terbatas menggunakan kamus *lexicon* atau *bag of words*, sehingga kata-kata negasi belum dapat teridentifikasi dengan baik. Untuk penelitian selanjutnya diharapkan dapat melakukan penanganan khusus terhadap kata negasi agar hasil yang didapatkan lebih akurat, salah satunya menggunakan sistem pelabelan yang memiliki tingkatan lebih tinggi seperti *semantic parsing* yang dapat menerjemahkan dalam bentuk kalimat.

## DAFTAR PUSTAKA

- Aditya. (2015). Penggunaan Web Crawler Untuk Menghimpun Tweet Dengan Metode Pre-Processing Text Mining. *Jurnal Infotel*, 7.
- Al-Bayaty, B. F. Z., & Joshi, S. (2016). Comparative Analysis between Naïve Bayes Algorithm and Decision Tree to Solve WSD Using Empirical Approach. *Lecture Notes on Software Engineering*, 4(1), 82–86. <https://doi.org/10.7763/LNSE.2016.V4.228>
- Aliandu, P. (2015). Sentiment Analysis to Determine Accommodation, Shopping and Culinary Location on Foursquare in Kupang City. *Procedia Computer Science*, 72, 300–305. <https://doi.org/10.1016/J.PROCS.2015.12.144>
- APJII. (2018). Hasil Survei Penetrasi dan Perilaku Pengguna Internet Indonesia 2018. Retrieved December 10, 2019, from <https://www.apjii.or.id/content/read/39/410/Hasil-Survei-Penetrasi-dan-Perilaku-Pengguna-Internet-Indonesia-2018>
- Award, T. B. (2020). Top Brand Index. Retrieved from [https://www.topbrandaward.com/top-brand-index/?tbi\\_find=jne](https://www.topbrandaward.com/top-brand-index/?tbi_find=jne)
- Azalia, F. Y., Bijaksana, M. A., & Huda, A. F. (2019). Name Indexing in Indonesian Translation of Hadith using Named Entity Recognition with Naïve Bayes Classifier. *Procedia Computer Science*, 157, 142–149. <https://doi.org/10.1016/J.PROCS.2019.08.151>
- Barfian, E., Iswanto, B. H., & Isa, S. M. (2017). Twitter Pornography Multilingual Content Identification Based on Machine Learning. *Procedia Computer Science*, 116, 129–136. <https://doi.org/10.1016/J.PROCS.2017.10.024>
- Barus, D. (2019). Info Lengkap: Proses Pengiriman Barang JNE. Retrieved August 5, 2020, from <https://dionbarus.com/info-lengkap-proses-pengiriman-barang-jne/>
- Bilal, M., Israr, H., Shahid, M., & Khan, A. (2016). Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. *Journal of King Saud University - Computer and Information Sciences*, 28(3), 330–344. <https://doi.org/10.1016/J.JKSUCI.2015.11.003>
- Chrisbiyanto, A. (2018). Perusahaan Ekspedisi Berperan Penting Dorong Pertumbuhan UKM. *SINDONEWS*. Retrieved from

- <https://ekbis.sindonews.com/read/1317290/34/perusahaan-ekspedisi-berperan-penting-dorong-pertumbuhan-ukm-1530201209>
- Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier. *International Journal of Information Engineering and Electronic Business*, 8(4), 54–62. <https://doi.org/10.5815/ijieeb.2016.04.07>
- El-Masri, M., Altrabsheh, N., Mansour, H., & Ramsay, A. (2017). A web-based tool for Arabic sentiment analysis. *Procedia Computer Science*, 117, 38–45. <https://doi.org/10.1016/J.PROCS.2017.10.092>
- Fadlisyah. (2014). *Statistika : Terapannya di Informatika*, 1st ed. Yogyakarta: Graha Ilmu.
- Fanani. (2017). *Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen*. Universitas Gadjah Mada.
- Fauziah, N. (2009). *Aplikasi Fishbone Analysis dalam Meningkatkan Kualitas Produksi Teh Pada PT Rumpun Sari Kemuning, Kabupaten Karanganyar*. Universitas Sebelas Maret.
- Fritz, G. (2016). *Analisa Bad Hike Pada Kran Lavatory Tipe S11234R Menggunakan Metode Nominal Group Technique dan Metode Fishbone di PT Surya Toto Indonesia Tbk*. Universitas Gadjah Mada.
- Han, J., & Kamber, M. (2006). *Data mining: Concepts and Techniques Second Edition*. San Fransisco: Morgan Kauffman.
- Ibrohim, M. O., & Budi, I. (2018). A Dataset and Preliminaries Study for Abusive Language Detection in Indonesian Social Media. *Procedia Computer Science*, 135, 222–229. <https://doi.org/10.1016/j.procs.2018.08.169>
- Jain, V. K., & Kumar, S. (2015). An Effective Approach to Track Levels of Influenza-A (H1N1) Pandemic in India Using Twitter. *Procedia Computer Science*, 70, 801–807. <https://doi.org/10.1016/j.procs.2015.10.120>
- Josi, A., Abdillah, L., & Suryayusra. (2014). Penerapan Teknik Web Scraping Pada Mesin Pencari Artikel Ilmiah. *Jurnal Sistem Informasi (SISFO)*, 05.
- Jumeilah, F. S. (2018). Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE dengan Naïve Bayes. *Jurnal Sistem Informasi Bisnis*, 8(1), 92. <https://doi.org/10.21456/vol8iss1pp92-98>
- Karch. (2019). What is Google Play? Retrieved from [lifewire.com/what-isgoogle-play-](http://lifewire.com/what-isgoogle-play-)

- 1616720.html
- Kotler, & Armstrong. (1996). *Dasar-Dasar Pemasaran Edisi V Jilid 2*. Jakarta: Intermedia.
- Kristiyanti, D. A. (2015). Analisis sentimen review produk kosmetik melalui komparasi feature selection. *Konferensi Nasional Ilmu Pengetahuan Dan Teknologi (KNIT)*, 2(2), 74–81.
- Kunal, S., Saha, A., Varma, A., & Tiwari, V. (2018). Textual Dissection of Live Twitter Reviews using Naive Bayes. *Procedia Computer Science*, 132, 307–313. <https://doi.org/10.1016/J.PROCS.2018.05.182>
- Liu, J., Tian, Z., Liu, P., Jiang, J., & Li, Z. (2016). An approach of semantic web service classification based on naive bayes. *Proceedings - 2016 IEEE International Conference on Services Computing, SCC 2016*, 356–362. <https://doi.org/10.1109/SCC.2016.53>
- Mahadzir, N. H., Omar, M. F., & Nawi, M. N. M. (2018). A Sentiment Analysis Visualization System for the Property Industry. *International Journal of Technology*, 9(8), 1609. <https://doi.org/10.14716/ijtech.v9i8.2753>
- Mohri. (2012). *Foundations of Machine Learning*. Cambridge: MIT Press.
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly: Management Information Systems*, 34(1), 185–200. <https://doi.org/10.2307/20721420>
- Mukherjee, S., & Bala, P. K. (2017). Sarcasm detection in microblogs using Naïve Bayes and fuzzy clustering. *Technology in Society*, 48, 19–27. <https://doi.org/10.1016/j.techsoc.2016.10.003>
- O'Keefe, & Koprinska. (2009). Feature selection and weighting methods in sentiment analysis. *Proceedings of the 14th Australasian Document Computing Symposium*. Sidney.
- Prasetyo. (2006). Training Set, Testing Set, dan Validation Set. Retrieved March 1, 2020, from <https://philips.wordpress.com/2006/08/10/training-set-testing-set-dan-validation-set/>
- Prasetyo, E. (2012). *Data mining Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi.
- Putri, D. (2016). *Implementasi Inferensi Fuzzy Mamdani Untuk Keperluan Sistem Rekomendasi Berita Berbasis Konten*. Universitas Gadjah Mada.

- Raksanagara, R., Chrisnanto, Y. H., & Hadiana, A. I. (2016). Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naïve Bayes. *Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naïve Bayes*, 19–24.
- Ramadhani, T. G. (2015). *Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier dengan Model Dokumen Bernoulli dan Support Vector Machine*. Universitas Gadjah Mada.
- Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Electrical Power, Electronics, Communications, Controls, and Informatics Seminar (EECCIS)*, 6(1), 37–43.
- Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2019). A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. *Journal of Business Research*, 101, 499–506. <https://doi.org/10.1016/J.JBUSRES.2018.12.051>
- StatCounter. (2020). Mobile Operating System Market Share Indonesia (Des 2018-Des 2019). Retrieved January 5, 2020, from <https://gs.statcounter.com/os-market-share/all/indonesia#monthly-201810-201910>
- Suyanto. (2017). *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung.
- Tanaka, M., & Okutomi, M. (2014). A Novel Inference of a Restricted Boltzmann Machine. *Proceedings - International Conference on Pattern Recognition*, (Cd), 1526–1531. <https://doi.org/10.1109/ICPR.2014.271>
- Turland, M. (2010). *Guide to Web Scraping with PHP*.
- Ulwan, M. N. (2016). *Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine dan Association*. Universitas Islam Indonesia.
- Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes Risa. *Jurnal Evolusi Volume*, 4. <https://doi.org/10.1017/CBO9781107415324.004>
- Widodo, E., Suliztia, M. L., Choirunisa, P., & Werdiningsih, R. Y. (2020). Sentiment Analysis on Ratu Boko Temple Tourist Perceptions Based on Review Content on TripAdvisor Website. *International Journal of Innovative Technology and Exploring Engineering*.

Wongso, R., Luwinda, F. A., Trisnajaya, B. C., Rusli, O., & Rudy. (2017). News Article Text Classification in Indonesian Language. *Procedia Computer Science*, 116, 137–143. <https://doi.org/10.1016/J.PROCS.2017.10.039>



## LAMPIRAN

### Lampiran 1. Script Preprocessing dengan python Library Sastrawi

```
In [1]: pip install nltk
Requirement already satisfied: nltk in c:\programdata\anaconda3\lib\site-packages (3.4.5)
Requirement already satisfied: six in c:\programdata\anaconda3\lib\site-packages (from nltk) (1.12.0)
Note: you may need to restart the kernel to use updated packages.

In [2]: pip install PySastrawi
Requirement already satisfied: PySastrawi in c:\programdata\anaconda3\lib\site-packages (1.2.0)
Note: you may need to restart the kernel to use updated packages.

In [3]: import pandas as pd
import string
import nltk
import re
import sys
import numpy as np

In [4]: import csv
df = pd.read_csv(r'D:\KULIAH\ta\COBA\COBA.csv', delimiter = ',',encoding = 'unicode_escape')
df.head()

Out[4]:
          Ulasan
0           kecewa
1 bagus tepat waktu kurinya rese menyebalkan n...
2           payah
3           paket belum sampai
4   tidak sesuai estimasi tidak sampai

In [5]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1859 entries, 0 to 1858
Data columns (total 1 columns):
Ulasan    1859 non-null object
dtypes: object(1)
memory usage: 7.3+ KB

In [6]: # from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
# factory = StopWordRemoverFactory()
# stopword = factory.create_stop_word_remover()

# def stop_words_removal(text):
#     return stopword.remove(text)

# print(stop_words_removal("Dengan Menggunakan Python dan Library Sastrawi saya dapat melakukan proses Stopword"))
<REDACTED>

In [7]: # from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
# factory = StemmerFactory()
# stemmer = factory.create_stemmer()

# def stemming(text):
#     return stemmer.stem(text)

# print(stemming("Perekonomian Indonesia sedang dalam pertumbuhan yang membanggakan"))
```

```
In [8]: import nltk
nltk.download('words')
from nltk import word_tokenize
from nltk.corpus import stopwords
nltk.download('stopwords')
set(stopwords.words('indonesian'))

[nltk_data] Downloading package words to
[nltk_data]      C:\Users\user\AppData\Roaming\nltk_data...
[nltk_data] Package words is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\user\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Out[8]: {'ada',
'adalah',
'adanya',
'adapun',
'agak',
'agaknya',
'agar',
'akan',
'akankah',
'akhir',
'akhiri',
'akhirnya',
'aku'.
```

```
In [9]: indo = stopwords.words('indonesian')
```

```
In [10]: Ulasan = [word for word in df['Ulasan'] if not word in indo]
Ulasan
```

```
Out[10]: ['kecawa',
'bagus tepat waktu kurirnya rese menyebalkan ngendok berabad abad ',
'payah',
'paket belum sampai ',
'tidak sesuai estimasi tidak sampai',
'terganggu mengirim kode resi orderan tidak ter cancel',
'buruk pelayanannya taik ',
'buruk kurir tidak menghubungi',
'karyawan pemalas pelit pulsa',
'kecawa paketan kirim',
'kinerjanya jelek tidak becus kiriman mezolitikum sampai tidak nyampe terkendala bersabar gila lemot kurirnya malas paketnya lelet harap perhatikan keluhannya tidak bagus jelek kinerjanya',
'jasa ekspedisi lambat lama dibanding tidak terpaksa males jasa ',
'jatuh miskin pesang paket sampai nuntut suruh ambil pusat daerah non aktiv kurir olshop hitung sampai rat usan kerugian kunjung sampai ',
'jijik kurir paket ditelanlant aramat lengkap alat tidak dipaket malas ogah kecawa pelayan kantor alamat bubar tidak memuaskan ditelepon kantor pusat oke tunggu hubungi nyata dihubungi pusat cabang kelakuananya ',
'merugikan alasan kantor kosong kurir gak gaji mendingan',
'ongkirnya drastis ',
```

```
In [11]: from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()
```

```
In [12]: import re
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
stemmer = StemmerFactory().create_stemmer()
corpus = []
for i in range(0, len(Ulasan)):
    Ulasan = re.sub('[^a-zA-Z]', ' ', df['Ulasan'][i])
    Ulasan = Ulasan.lower()
    Ulasan = Ulasan.split()
    Ulasan = [word for word in Ulasan if not word in indo]
    Ulasan = ' '.join(Ulasan)
    stemming = stemmer.stem(Ulasan)
    corpus.append(stemming)

[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\user\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
In [14]: df = pd.DataFrame(corpus)
csv_data = df.to_csv(index=False)
df.to_csv('baruBANGET.csv', index=False)
```

**Lampiran 2. Script R Preprocessing Data dengan Text Mining**

```
# INSTALL PACKAGES
install.packages("tm") # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordCloud") # word-cloud generator
install.packages("RColorBrewer") # color palettes

# LOAD PACKAGES
library("tm")
library("SnowballC")
library("wordCloud")
library("RColorBrewer")
library(stringr)

setwd("D:\\KULIAH\\ta\\data\\")
docs<-readLines("review.csv")

# LOAD THE DATA AS A CORPUS
docs <- Corpus(VectorSource(docs))

#Inspect the content of the document
inspect(docs)

#Replacing ?/? , ?@? and ?|? with space:
toSpace <- content_transformer(function (x , pattern )
gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\|")

#Cleaning the text
#Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))

#Remove punctuation
```

```

docs <- tm_map(docs, toSpace, "[[:punct:]]")

#Remove numbers
docs <- tm_map(docs, toSpace, "[[:digit:]]")

# add two extra stop words:
myStopwords = readLines("stopwordid.csv")

# remove stopwords from corpus
docs <- tm_map(docs, removeWords, myStopwords)

# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)

# Remove URL
removeURL <- function(x) gsub("http[[:alnum:]]*", " ", x)
docs <- tm_map(docs, removeURL)

#stemming
stem_text <- function(text,mc.cores=1) {
  #stem each word in a block of text
  stem_string <- function(str)
  {
    str <- tokenize(x=str)
    str <- sapply(str, katadasaR)
    str <- paste(str, collapse="")
    return(str)
  }
  #stem each text block in turn
  x <- mclapply(X=text, FUN=stem_string, mc.cores=mc.cores)
  #return stemmed text blocks
  return(unlist(x))
}
docs<- tm_map(docs, stem_text)
inspect(docs[1:10])

```

```
#Build a term-document matrix
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)

dataframe<-data.frame(text=unlist(sapply(docs, `[`)),
stringsAsFactors=F)

write.csv(dataframe, "D:\\KULIAH\\ta\\data\\cleaning.csv")
save.image()
```

**Lampiran 3.** Script R Pelabelan dan Pembobotan

```

library(tm)
setwd("D:\\KULIAH\\ta\\data")
kalimat2<-read.csv("cleaning.csv", header=TRUE)

#Scoring data berdasarkan kamus kata
positif <- scan("positif.txt",what="character",comment.char=";")
negatif <- scan("negatif.txt",what="character",comment.char=";")
kata.positif = c(positif)
kata.negatif = c(negatif)
score.sentiment = function(kalimat2, kata.positif, kata.negatif,
.progress='none')
{
  require(plyr)
  require(stringr)
  scores = laply(kalimat2, function(kalimat, kata.positif, kata.negatif)
  {
    kalimat = gsub('[[:punct:]]', '', kalimat)
    kalimat = gsub('[[:cntrl:]]', '', kalimat)
    kalimat = gsub('\\d+', '', kalimat)
    kalimat = tolower(kalimat)

    list.kata = str_split(kalimat, '\\s+')
    kata2 = unlist(list.kata)
    positif.matches = match(kata2, kata.positif)
    negatif.matches = match(kata2, kata.negatif)
    positif.matches = !is.na(positif.matches)
    negatif.matches = !is.na(negatif.matches)
    score = sum(positif.matches) - (sum(negatif.matches))
    return(score)
  }, kata.positif, kata.negatif, .progress=.progress )
  scores.df = data.frame(score=scores, text=kalimat2)
  return(scores.df)
}

```

```
hasil = score.sentiment(kalimat2$text, kata.positif, kata.negatif)
head(hasil)
View(hasil)

#CONVERT SCORE TO SENTIMENT
hasil$klasifikasi<- ifelse(hasil$score<0, "Negatif","Positif")
hasil$klasifikasi
View(hasil)

#Tukar Row
data <- hasil[c(3,1,2)]
View(data)
write.csv(data, file = "D:\\KULIAH\\ta\\data\\label.csv")

#Menyimpan Data Positif dan Negatif
data.pos <- hasil[hasil$score>0,]
View(data.pos)
write.csv(data.pos, file = "D:\\KULIAH\\ta\\data\\positive.csv")
data.neg <- hasil[hasil$score<0,]
View(data.neg)
write.csv(data.neg, file = "D:\\KULIAH\\ta\\data\\negative.csv")
```

**Lampiran 4.** *Script R Klasifikasi dengan menggunakan Metode Naïve Bayes Classifier (NBC)*

```
# INSTALL PACKAGES
install.packages("devtools")
library(dev)
install_github("cran/maxent")
install_github("cran/RTextTools")
library(RTextTools)
library(SparseM)
library(RTextTools)
install.packages("tm")

# Load required libraries
library(tm)
library(SparseM)
library(RTextTools)
library(e1071)
library(stats)
library(dplyr)
library(NLP)
library(caret)
library(maxent)
library(readr)

# Library for parallel processing
library(doMC)
registerDoMC(cores=detectCores())

input data
library(readr)
df <- read.csv("datanaive.csv", stringsAsFactors=FALSE, encoding="UTF-
8")
View(df)
glimpse(df)
```

```
set.seed(1)
df <- df[sample(nrow(df)), ]
df <- df[sample(nrow(df)), ]
glimpse(df)
df$class <- as.factor(df$class)
corpus <- Corpus(VectorSource(df$text))
corpus
inspect(corpus[1:3])

corpus.clean <- corpus %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removeWords, stopwords(kind="en")) %>%
  tm_map(stripWhitespace)

dtm <- DocumentTermMatrix(corpus)
inspect(dtm[40:50, 10:15])

#PEMBAGIAN DATA LATIH 60% DAN DATA UJI 40%
df.train <- df[1:1106,]
df.test <- df[1107:1843,]
dtm.train <- dtm[1:1106,]
dtm.test <- dtm[1107:1843,]
corpus.train <- corpus[1:1106]
corpus.test <- corpus[1107:1843]

dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
fivefreq
dtm.train.nb <- DocumentTermMatrix(corpus.train, control=list(dictionary
= fivefreq))
dim(dtm.train.nb)
```

```

dtm.test.nb <- DocumentTermMatrix(corpus.test, control=list(dictionary =
fivefreq))
dim(dtm.train.nb)

#Fungsi Pelabelan
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))
  y
}
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)

library(naivebayes)
naive=system.time( classifier <- naiveBayes(trainNB, df.train$class,
laplace = 0) )
print(naive)
system.time( pred <- predict(classifier, newdata=testNB) )

# Confusion Matrix
conf.mat <- confusionMatrix(table("Predictions"= pred, "Actual" =
df.test$class ))
conf.mat
conf.mat$byClass
conf.mat$overall
conf.mat$overall['Accuracy']

#PEMBAGIAN DATA LATIH 70% DAN DATA UJI 30%
df.train <- df[1:1290,]
df.test <- df[1291:1843,]
dtm.train <- dtm[1:1290,]
dtm.test <- dtm[1291:1843,]
corpus.train <- corpus[1:1290]
corpus.test <- corpus[1291:1843]

```

```

dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
fivefreq
dtm.train.nb <- DocumentTermMatrix(corpus.train, control=list(dictionary =
= fivefreq))
dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.test, control=list(dictionary =
fivefreq))
dim(dtm.train.nb)

#Fungsi Pelabelan
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))
  y
}
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)

library(naivebayes)
naive=system.time( classifier <- naiveBayes(trainNB, df.train$class,
laplace = 0) )
print(naive)
system.time( pred <- predict(classifier, newdata=testNB) )

# Confusion Matrix
conf.mat <- confusionMatrix(table("Predictions"= pred, "Actual" =
df.test$class ))
conf.mat
conf.mat$byClass
conf.mat$overall
conf.mat$overall['Accuracy']

```

```
#PEMBAGIAN DATA LATIH 80% DAN DATA UJI 20%
df.train <- df[1:1474,]
df.test <- df[1475:1843,]
dtm.train <- dtm[1:1474,]
dtm.test <- dtm[1475:1843,]
corpus.train <- corpus[1:1474]
corpus.test <- corpus[1475:1843]

dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
fivefreq
dtm.train.nb <- DocumentTermMatrix(corpus.train, control=list(dictionary = fivefreq))
dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.test, control=list(dictionary = fivefreq))
dim(dtm.train.nb)

#Fungsi Pelabelan
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))
  y
}
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)

library(naivebayes)
naive=system.time( classifier <- naiveBayes(trainNB, df.train$class,
laplace = 0) )
print(naive)
system.time( pred <- predict(classifier, newdata=testNB) )
```

```

# Confusion Matrix
conf.mat <- confusionMatrix(table("Predictions"= pred, "Actual" =
df.test$class ))
conf.mat
conf.mat$byClass
conf.mat$overall
conf.mat$overall['Accuracy']

#PEMBAGIAN DATA LATIH 85% DAN DATA UJI 15%
df.train <- df[1:1567,]
df.test <- df[1568:1843,]
dtm.train <- dtm[1:1567,]
dtm.test <- dtm[1568:1843,]
corpus.train <- corpus[1:1567]
corpus.test <- corpus[1568:1843]

dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
fivefreq
dtm.train.nb <- DocumentTermMatrix(corpus.train, control=list(dictionary =
fivefreq))
dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.test, control=list(dictionary =
fivefreq))
dim(dtm.train.nb)

#Fungsi Pelabelan
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))
  y
}
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)

```

```
library(naivebayes)
naive=system.time( classifier <- naiveBayes(trainNB, df.train$class,
laplace = 0) )
print(naive)
system.time( pred <- predict(classifier, newdata=testNB) )

# Confusion Matrix
conf.mat <- confusionMatrix(table("Predictions"= pred, "Actual" =
df.test$class ))
conf.mat
conf.mat$byClass
conf.mat$overall
conf.mat$overall['Accuracy']

#PEMBAGIAN DATA LATIH 90% DAN DATA UJI 10%
df.train <- df[1:1659,]
df.test <- df[1660:1843,]
dtm.train <- dtm[1:1659,]
dtm.test <- dtm[1660:1843,]
corpus.train <- corpus[1:1659]
corpus.test <- corpus[1660:1843]

dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
fivefreq
dtm.train.nb <- DocumentTermMatrix(corpus.train, control=list(dictionary =
fivefreq))
dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.test, control=list(dictionary =
fivefreq))
dim(dtm.train.nb)
```

```
#Fungsi Pelabelan
convert_count <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))
  y
}
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)

library(naivebayes)
naive=system.time( classifier <- naiveBayes(trainNB, df.train$class,
laplace = 0) )
print(naive)
system.time( pred <- predict(classifier, newdata=testNB) )

# Confusion Matrix
conf.mat <- confusionMatrix(table("Predictions"= pred, "Actual" =
df.test$class ))
conf.mat
conf.mat$byClass
conf.mat$overall
conf.mat$overall['Accuracy']
```

**Lampiran 5.** Script R Visualisasi dan Asosiasi Teks (Ulasan Positif)

```
# LOAD PACKAGES
library(NLP)
library(tm)
library(SnowballC)
library(RColorBrewer)
library(wordCloud)
library(stringr)
library(wordCloud2)

docs <- read.csv("positive.csv", encoding = "UTF-8")
docs<- Corpus(VectorSource(docs$text))
docs<- tm_map(docs, removeWords,
c("tidak", "gak", "cod", "tdk", "yes", "biar",
"tau", "nama", "tlp", "alamat", "coba", "lama", "barangnya", "mohon", "semoga", "jawa",
"trus", "klw", "dimana"))
docs <- tm_map(docs, stripWhitespace)
docs[[1]][1]

dtm<- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 70)

set.seed(1234)
wordCloud(words = d$word, freq = d$freq, min.freq = 1,
max.words=50, random.order=FALSE, rot.per=0.35,
colors=brewer.pal(8, "Dark2"))
findFreqTerms(dtm, lowfreq = 4)

v<-as.list(findAssocs(dtm, terms
=c("kirim", "paket", "kurir", "bagus", "aplikasi", "cepat", "layan",
"terima", "bantu", "bintang"),
```

```
corlimit = c(0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15))

v
#Find related words (one by one)
v<-as.list(findAssocs(dtm, terms =c("kirim"),
                      corlimit = c(0.10)))
View(v$kirim)
v<-as.list(findAssocs(dtm, terms =c("paket"),
                      corlimit = c(0.15)))
View(v$paket)
v<-as.list(findAssocs(dtm, terms =c("kurir"),
                      corlimit = c(0.15)))
View(v$kurir)

v<-as.list(findAssocs(dtm, terms =c("bagus"),
                      corlimit = c(0.10)))
View(v$bagus)

v<-as.list(findAssocs(dtm, terms =c("aplikasi"),
                      corlimit = c(0.10)))
View(v$aplikasi)

v<-as.list(findAssocs(dtm, terms =c("cepat"),
                      corlimit = c(0.15)))
View(v$cepat)

v<-as.list(findAssocs(dtm, terms =c("layan"),
                      corlimit = c(0.15)))
View(v$layan)

v<-as.list(findAssocs(dtm, terms =c("terima"),
                      corlimit = c(0.15)))
View(v$terima)

v<-as.list(findAssocs(dtm, terms =c("bantu"),
                      corlimit = c(0.15)))
```

```
View(v$bantu)

v<-as.list(findAssocs(dtm, terms =c("bintang"),
                      corlimit = c(0.15)))

View(v$bintang)

#barplot
k<-barplot(d[1:10,]$freq, las = 2, names.arg =
             d[1:10,]$word,cex.axis=1.2,cex.names=1.2,
             main ="Most frequent words",
             ylab = "Word frequencies",col = terrain.colors(20))
```

**Lampiran 6.** *Script R Visualisasi dan Asosiasi Teks (Ulasan Negatif)*

```
# LOAD PACKAGES
library(NLP)
library(tm)
library(SnowballC)
library(RColorBrewer)
library(wordCloud)
library(stringr)
library(wordCloud2)

docs <- read.csv("negative.csv", encoding = "UTF-8")
docs<- Corpus(VectorSource(docs$text))
docs<- tm_map(docs, removeWords, c("tidak","gak","tdk","yes","jnt"))
docs <- tm_map(docs, stripWhitespace)
docs[[1]][1]

dtm<- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 70)

set.seed(1234)
wordCloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=50, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

findFreqTerms(dtm, lowfreq = 4)

v<-as.list(findAssocs(dtm, terms
=c("kirim", "paket", "kurir", "kecewa", "layan", "jasa", "buruk",
"aplikasi", "parah", "lambat"),
corlimit = c(0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15)))
v
#Find related words (one by one)
v<-as.list(findAssocs(dtm, terms =c("kirim"),
```

```
corlimit = c(0.15)))  
View(v$kirim)  
  
v<-as.list(findAssocs(dtm, terms =c("paket"),  
                      corlimit = c(0.15)))  
View(v$paket)  
  
v<-as.list(findAssocs(dtm, terms =c("kurir"),  
                      corlimit = c(0.15)))  
View(v$kurir)  
  
v<-as.list(findAssocs(dtm, terms =c("kecewa"),  
                      corlimit = c(0.10)))  
View(v$kecewa)  
  
v<-as.list(findAssocs(dtm, terms =c("layan"),  
                      corlimit = c(0.15)))  
View(v$layan)  
  
v<-as.list(findAssocs(dtm, terms =c("lambat"),  
                      corlimit = c(0.1)))  
View(v$lambat)  
  
v<-as.list(findAssocs(dtm, terms =c("jasa"),  
                      corlimit = c(0.15)))  
View(v$jasa)  
  
v<-as.list(findAssocs(dtm, terms =c("buruk"),  
                      corlimit = c(0.15)))  
View(v$buruk)  
  
v<-as.list(findAssocs(dtm, terms =c("aplikasi"),  
                      corlimit = c(0.15)))  
View(v$aplikasi)
```

```
v<-as.list(findAssocs(dtm, terms =c("parah"),
                      corlimit = c(0.15)))
View(v$parah)

#barplot
k<-barplot(d[1:10,]$freq, las = 2, names.arg =
             d[1:10,]$word,cex.axis=1.2,cex.names=1.2,
             main ="Most frequent words",
             ylab = "Word frequencies",col = terrain.colors(20))
```



**Lampiran 7.** Script R Visualisasi dan Asosiasi Teks Ulasan Keseluruhan

```

library(NLP)
library(tm)
library(SnowballC)
library(RColorBrewer)
library(wordcloud)
library(stringr)
library(wordcloud2)

docs <- read.csv("label.csv", encoding = "UTF-8")
docs<- Corpus(VectorSource(docs$text))
docs<- tm_map(docs, removeWords,
c("bom","pke","org","mulu","masak","ngga","malang","reg","karna","bandung",
"sya","tpi","gitu","gin","nih","lot","tidak","gak","cod","tdk","yes",
"biar",
"tau","nama","tlp","alamat","coba","lama","barangnya","mohon","semoga",
"jawa", "trus", "klw", "dimana"))
docs <- tm_map(docs, stripWhitespace)
docs[[1]][1]

dtm<- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 100)

set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=100, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
findFreqTerms(dtm, lowfreq = 4)

v<-as.list(findAssocs(dtm, terms
=c("kirim","paket","kurir","layan", "aplikasi", "jasa"),
corlimit = c(0.05,0.05,0.05,0.05,0.05,0.05)))

```

```

v
#Find related words (one by one)
v<-as.list(findAssocs(dtm, terms =c("kirim"),
                      corlimit = c(0.05)))
View(v$kirim)

v<-as.list(findAssocs(dtm, terms =c("paket"),
                      corlimit = c(0.05)))
View(v$paket)

v<-as.list(findAssocs(dtm, terms =c("kurir"),
                      corlimit = c(0.05)))
View(v$kurir)

v<-as.list(findAssocs(dtm, terms =c("layan"),
                      corlimit = c(0.05)))
View(v$layan)

v<-as.list(findAssocs(dtm, terms =c("aplikasi"),
                      corlimit = c(0.05)))
View(v$aplikasi)

v<-as.list(findAssocs(dtm, terms =c("jasa"),
                      corlimit = c(0.05)))
View(v$jasa)

#barplot
k<-barplot(d[1:10,]$freq, las = 2, names.arg =
            d[1:10,]$word,cex.axis=1.2,cex.names=1.2,
            main ="Most frequent words",
            ylab = "Word frequencies",col = terrain.colors(20))

```