

**PEMODELAN REGRESI LOGISTIK MULTINOMIAL
HASIL ANALISIS SENTIMEN KONSUMEN TOKO
ONLINE LAZADA INDONESIA**

SKRIPSI

Oleh:
MAILANI PUTRI
145090500111014



**PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2018**

repository.ub.ac.id

HALAMAN JUDUL
PEMODELAN REGRESI LOGISTIK MULTINOMIAL
HASIL ANALISIS SENTIMEN KONSUMEN TOKO
ONLINE LAZADA INDONESIA

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika

Oleh:
MAILANI PUTRI
145090500111014



PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2018

LEMBAR PENGESAHAN SKRIPSI

**PEMODELAN REGRESI LOGISTIK MULTINOMIAL
HASIL ANALISIS SENTIMEN KONSUMEN TOKO *ONLINE*
LAZADA INDONESIA**

Oleh:

**MAILANI PUTRI
145090500111014**

**Setelah dipertahankan di depan Majelis Penguji
pada tanggal 09 April 2018
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Statistika**

Dosen Pembimbing

**Achmad Efendi, S.Si., M.Sc., Ph.D
NIP. 198102192005011001**

**Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA Universitas Brawijaya**

**Ratno Bagus E. W., S.Si., M.Si., Ph.D
NIP. 197509082000031003**

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Mailani Putri
NIM : 145090500111014
Jurusan : Matematika
Penulis Skripsi berjudul :

Pemodelan Regresi Logistik Multinomial Hasil Analisis Sentimen Konsumen Toko *Online* Lazada Indonesia

Dengan ini menyatakan bahwa:

1. Isi dari Skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam Skripsi ini.
2. Apabila dikemudian hari ternyata Skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala risiko yang akan saya terima.

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 09 April 2018
Yang menyatakan,

Mailani Putri
NIM. 145090500111014

repository.ub.ac.id

PEMODELAN REGRESI LOGISTIK MULTINOMIAL HASIL ANALISIS SENTIMEN KONSUMEN TOKO *ONLINE* LAZADA INDONESIA

ABSTRAK

Penerapan analisis sentimen banyak digunakan perusahaan penyedia jasa layanan telekomunikasi. Sejatinya, analisis ini bisa digunakan pada perusahaan lain yang menyediakan media untuk memberikan tanggapan di media sosial, seperti *e-commerce*. Salah satu perusahaan *e-commerce* yang paling populer di Indonesia adalah Lazada. Penelitian ini akan meneliti pengaruh dari komponen tanggapan terhadap pengelompokan kategori sentimen dari tanggapan pelanggan mengenai produk dan jasa pada toko online Lazada. Analisis yang dapat mengakomodir hal tersebut adalah regresi logistik multinomial, karena variabel respon berupa kategori yang memiliki level kategorik lebih dari dua. Tujuan penelitian ini adalah menerapkan analisis sentimen menggunakan *Naive Bayes* dalam mengkategorikan sentimen tanggapan konsumen toko online Lazada dan membentuk model regresi logistik multinomial untuk menjelaskan pengaruh variabel prediktor yang terdiri dari Layanan, Pemesanan dan Promosi terhadap pengelompokan kategori sentimen konsumen. Pembentukan klasifikasi sentimen tanggapan pelanggan Lazada Indonesia pada data latih dengan sentimen paling banyak adalah positif. Hasil dari analisis sentimen yang dibentuk model menggunakan regresi logistik multinomial menunjukkan variabel Pemesanan dan Promosi yang signifikan pada sentimen Negatif dan Promosi yang signifikan pada sentimen Positif.

Kata Kunci: Regresi Logistik Multinomial, *Naive Bayes*, Analisis Sentimen, Lazada Indonesia

repository.ub.ac.id

MULTINOMIAL LOGISTIC REGRESSION MODELING OF CONSUMER SENTIMENT ANALYSIS RESULTS ONLINE SHOP LAZADA INDONESIA

ABSTRACT

Application of sentiment analysis is widely used in the provider of telecommunication services. Indeed, this analysis can be used in any company that provides media to respond in social media such as e-commerce. One of the most popular e-commerce companies in Indonesia is Lazada. This study examines the effect of components response on the grouping of sentiment categories from customer feedback on products and services at Lazada online store. The purpose of this research is to apply sentiment analysis using Naive Bayes algorithm in categorizing consumer response sentiment online store Lazada and forming multinomial logistic regression model to explain influence of component response to grouping of consumer sentiment category. The algorithm used in sentiment classification is Naive Bayes then the result will form Multinomial Logistic Regression model. The formation of the sentence classification of Lazada Indonesia customer response in the training data have neutral sentiment as the most sentiment. The results sentiment analysis is used for form multinomial logistic regression model. It has the predictor variables are Service, Booking and Promotion. The significant variable which Neutral as reference category are Booking and Promotion at Negative category sentiment and Promotion at Positive category sentiment.

Keywords : Multinomial Logistic Regression, Naive Bayes, Sentiment Analysis, Lazada Indonesia

KATA PENGANTAR

Puji syukur dipanjatkan kepada Allah SWT karena atas berkat dan rahmat-Nya penulis dapat menyelesaikan Skripsi yang berjudul **“Pemodelan Regresi Logistik Multinomial Hasil Analisis Sentimen Konsumen Toko *Online* Lazada Indonesia”**. Penulisan Skripsi ini tidak terlepas dari bimbingan, bantuan, dan do’a dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Achmad Efendi, S.Si., M. Sc., Ph.D. selaku Dosen Pembimbing Skripsi atas waktu, saran dan bimbingan yang diberikan.
2. Dr. Dra. Ani Budi Astuti, M.Si. selaku Dosen Penguji I atas waktu, saran dan bimbingan yang diberikan.
3. Dr. Adji Achmad Rinaldo Fernandes., S.Si., M.Sc. selaku Dosen Penguji II atas waktu, saran dan bimbingan yang diberikan.
4. Ratno Bagus Edy Wibowo, S.Si., M.Si., Ph.D selaku Ketua Jurusan Matematika.
5. Semua dosen dan karyawan Jurusan Matematika FMIPA Universitas Brawijaya.
6. Ibunda Irdanel Efni, Ayahanda Jayusman dan kedua kakak, Gusfarina dan Dina Rahmayuni atas doa yang selalu diberikan.
7. Astarina Hartika Murti yang selalu menemani dan memberi semangat dalam melalui tahapan mahasiswa tingkat akhir.
8. Siti Hariati, Tantiana Windy, Shinta Kumala Dewi yang senantiasa membantu, menginspirasi dan mendukung penulis selama penyusunan Skripsi.
9. Semua rekan seperjuangan Statistika 2014 atas dukungan dan bantuannya dan rekan-rekan Direksi, Badan Operasional, dan anggota Studio Statistika 2017.

Dengan segala kerendahan hati, penulis berharap Skripsi ini dapat bermanfaat bagi pembaca yang membutuhkan dan dapat dijadikan referensi bagi penelitian selanjutnya. Penulis juga menyadari bahwa Skripsi ini masih jauh dari kesempurnaan, sehingga penulis mengharapkan saran dan kritik yang membangun terhadap Skripsi ini.

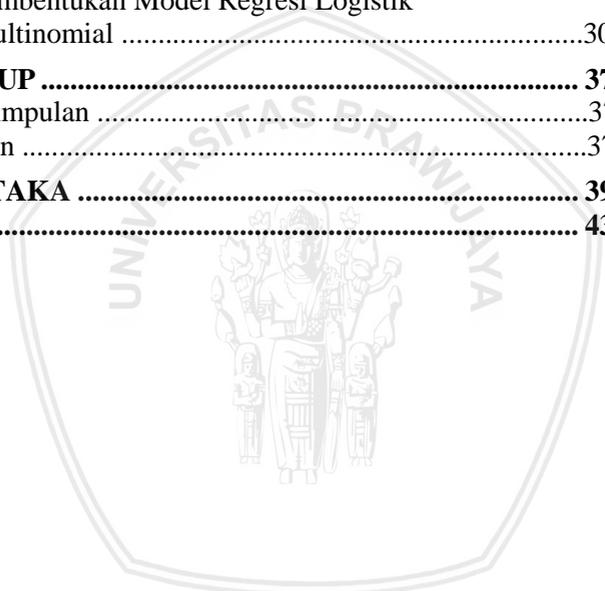
Malang, April 2018

Penulis

DAFTAR ISI

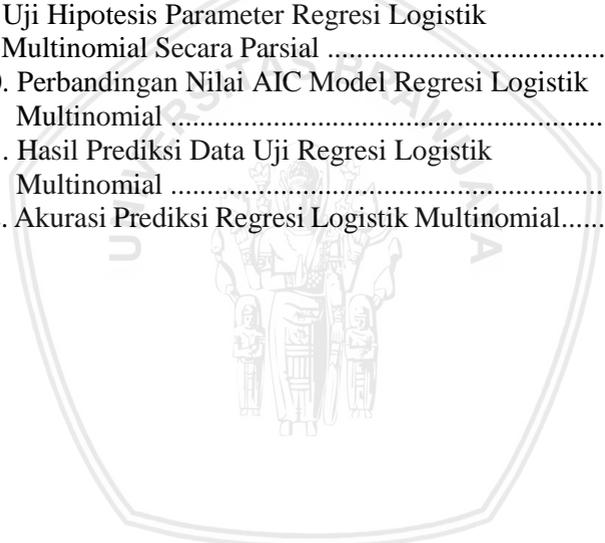
	Hal.
HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI	ii
LEMBAR PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	2
1.5. Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	5
2.1. Analisis Sentimen	5
2.2. Klasifikasi <i>Naive Bayes</i>	6
2.3. <i>Crawling</i> Data	8
2.4. Regresi Logistik Multinomial	8
2.4.1. Defenisi Regresi Logistik Multinomial	8
2.4.2. Pendugaan Parameter Regresi Logistik Multinomial	9
2.4.3. Pengujian Parameter Regresi Logistik Multinomial Secara Simultan	13
2.4.4. Pengujian Parameter Regresi Logistik Multinomial Secara Parsial	13
2.4.5. Keباikan Model	14
2.4.6. Interpretasi Model	14
2.4.7. Non Multikolinieritas	15
2.5. <i>Twitter</i>	15
2.6. <i>Bisnis Online</i> Lazada	16

	Hal
BAB III METODE PENELITIAN.....	19
3.1. Sumber Data	19
3.2. Variabel Penelitian	20
3.3. Metode Analisis	20
3.4. Diagram Alir Penelitian	21
BAB IV HASIL DAN PEMBAHASAN	25
4.1. Deskripsi Data	25
4.2. Pembentukan Klasifikasi Sentimen	25
4.3. Pembentukan Variabel Prediktor Layanan, Pemesanan dan Promosi	30
4.4. Pembentukan Model Regresi Logistik Multinomial	30
BAB V PENUTUP	37
5.1. Kesimpulan	37
5.2. Saran	37
DAFTAR PUSTAKA	39
LAMPIRAN	43



DAFTAR TABEL

	Hal.
Tabel 3.1. Struktur Data Hasil <i>Crawling</i>	19
Tabel 3.2. Struktur Data Analisis Sentimen	20
Tabel 4.1. Data Hasil <i>Crawling</i>	25
Tabel 4.2. Data Hasil <i>Cleaning</i> Teks	26
Tabel 4.3. Kosakata dan Skor	27
Tabel 4.4. Hasil <i>A-Priori Probabilities</i> Kelas Sentimen	28
Tabel 4.5. Hasil Prediksi Klasifikasi Sentimen Data Uji	28
Tabel 4.6. Hasil Prediksi Sentimen <i>Naive Bayes</i>	29
Tabel 4.7. Data Analisis Regresi Logistik Multinomial	30
Tabel 4.8. Hasil Uji Asumsi Non Multikolinieritas Variabel Layanan, Pemesanan dan Promosi	19
Tabel 4.9. Uji Hipotesis Parameter Regresi Logistik Multinomial Secara Parsial	32
Tabel 4.10. Perbandingan Nilai AIC Model Regresi Logistik Multinomial	34
Tabel 4.11. Hasil Prediksi Data Uji Regresi Logistik Multinomial	35
Tabel 4.12. Akurasi Prediksi Regresi Logistik Multinomial.....	36



DAFTAR GAMBAR

	Hal.
Gambar 3.1. Diagram Alir Penelitian	21
Gambar 4.1. Plot Kata dan Frekuensi	26



DAFTAR LAMPIRAN

	Hal.
Lampiran 1. <i>Syntax</i> Crawling Data dari <i>Twitter</i>	43
Lampiran 2. <i>Syntax</i> Fungsi <i>Cleaning</i> Teks.....	45
Lampiran 3. <i>Syntax</i> Fungsi Pembentukan Klasifikasi Sentimen.....	47
Lampiran 4. <i>Syntax</i> Klasifikasi Teks Menggunakan <i>Naive Bayes</i>	49
Lampiran 5. <i>Syntax</i> Pembentukan Variabel Prediktor Layanan, Pemesanan dan Promosi.....	51
Lampiran 6. <i>Syntax</i> Regresi Logistik Multinomial dengan Prediktor Layanan, Pemesanan dan Promosi	53
Lampiran 7. <i>Syntax</i> Pembentukan <i>Terms</i>	55
Lampiran 8. Data Hasil <i>Crawling Twitter</i>	57
Lampiran 9. Data Hasil <i>Cleaning</i>	59
Lampiran 10. Data Latih untuk Pembentukan Sentimen.....	61
Lampiran 11. Data Uji untuk Pembentukan Klasifikasi Sentimen	63
Lampiran 12. Data Uji Hasil Prediksi Sentimen dengan <i>Naive Bayes</i>	65
Lampiran 13. Kosakata Pembentuk Variabel Layanan.....	67
Lampiran 14. Kosakata Pembentuk Variabel Pemesanan.....	69
Lampiran 15. Kosakata Pembentuk Variabel Promosi.....	71
Lampiran 16. Data Latih untuk Regresi Logistik Multinomial Menggunakan Variabel Prediktor Layanan, Pemesanan dan Promosi.....	73
Lampiran 17. Data Hasil Prediksi Regresi Logistik Multinomial.....	75
Lampiran 18. Kosakata Pembentuk Sentimen Positif.....	77
Lampiran 19. Kosakata Pembentuk Sentimen Negatif.....	79
Lampiran 20. Kosakata Pembentuk Sentimen Netral.....	81
Lampiran 21. Output Pembentukan Sentimen Menggunakan <i>Naive Bayes</i>	83
Lampiran 22. Output Pembentukan Model Regresi Logistik Multinomial dari Variabel Layanan, Pemesanan dan Promosi.....	85

BAB I PENDAHULUAN

1.1. Latar Belakang

Analisis sentimen merupakan bagian dari analisis teks. Metode ini berupa analisis yang digunakan pada suatu dokumen yang bisa didapatkan dari media sosial. Teknik analisis ini digunakan untuk melihat kecenderungan tanggapan dari pelanggan terhadap suatu produk atau jasa. Sentimen tanggapan dari konsumen dikategorikan menjadi tiga kelompok, yaitu positif, negatif dan netral. Terdapat beberapa metode yang digunakan pada analisis sentimen, seperti *Naive Bayes Classifier* (NBC), *Support Vector Machine*, *Nearest Neighbour* dan *Decision Tree*. Penelitian ini metode yang digunakan, yaitu *Naive Bayes Classifier* (NBC).

Penerapan analisis sentimen banyak digunakan pada perusahaan *provider* atau penyedia jasa layanan telekomunikasi. Sejatinya, analisis ini bisa digunakan pada setiap perusahaan yang menyediakan media untuk memberikan tanggapan di media sosial seperti *e-commerce*. Saat ini Indonesia sedang marak dengan banyaknya industri *e-commerce* yang berkembang. Berdasarkan data sensus ekonomi dari Badan Pusat Statistik tahun 2016 menyebutkan bahwa *e-commerce* di Indonesia dalam sepuluh tahun terakhir tumbuh sekitar tujuh belas persen dengan total usaha 26,2 juta unit *e-commerce*.

Salah satu perusahaan *e-commerce* yang paling populer di Indonesia adalah Lazada. Menurut survei yang dilakukan oleh W&S *Digital Marketing* Indonesia, Lazada menempati posisi pertama dari *brand e-commerce* yang populer pada tahun 2016, sekitar 43,8% pangsa pasar dikuasai Lazada. Konsumen Lazada tersebar di seluruh Indonesia. Berbagai penelitian yang telah dilakukan terkait dengan kajian *e-commerce*, seperti penelitian Naif (2017) tentang pengaruh kepercayaan dan kualitas pelayanan terhadap keputusan pembelian di Tokopedia yang menggunakan kuesioner dengan analisis regresi berganda. Selanjutnya Pramadhana dkk. (2012) melakukan penelitian mengenai analisis sentimen berdasarkan fitur produk pada toko *online* seperti Ebay.com dan Amazon.com.

Berdasarkan penelitian terdahulu, didapatkan informasi bahwa belum ada penggabungan analisis sentimen yang menggunakan analisis lanjutan dengan memanfaatkan teori statistika seperti Regresi Logistik Multinomial. Oleh karena itu, penelitian ini memberikan ide baru mengenai penggabungan dua jenis analisis tersebut agar dapat

menjelaskan hubungan sentimen dengan tanggapan pelanggan yang bertujuan untuk melihat pengaruh dari variabel prediktor yang terdiri dari variabel Layanan, Pemesanan dan Promosi terhadap sentimen tanggapan pelanggan yang dihasilkan.

Penelitian ini akan mencari pengaruh variabel prediktor terhadap pengelompokan kategori sentimen dari tanggapan pelanggan mengenai produk dan jasa pada toko *online* Lazada. Kedudukan analisis sentimen pada penelitian ini digunakan untuk pengambilan data primer yang telah dikategorikan menjadi tiga jenis sentimen positif, negatif dan netral. Hasil kategorisasi dari analisis sentimen akan digunakan sebagai variabel respon, sehingga akan terdapat tiga kategori pada variabel respon. Kemudian dilanjutkan analisis Regresi Logistik Multinomial yang dapat menghasilkan model untuk menjelaskan pengaruh variabel Layanan, Pemesanan dan Promosi terhadap kategori sentimen.

1.2. Rumusan Masalah

Adapun rumusan masalah berdasarkan latar belakang penelitian antara lain:

1. Bagaimana penerapan analisis sentimen menggunakan algoritma *Naive Bayes* dalam mengkategorikan sentimen tanggapan konsumen toko *online* Lazada?
2. Bagaimana model Regresi Logistik Multinomial dalam menjelaskan pengaruh variabel Layanan, Pemesanan dan Promosi terhadap pengelompokan kategori sentimen konsumen toko *online* Lazada?

1.3. Batasan Masalah

Dalam penelitian ini, terdapat beberapa batasan masalah sebagai berikut:

1. Data tanggapan konsumen diperoleh dari media sosial, *twitter* berupa *tweets* konsumen tiga bulan terakhir.
2. Variabel prediktor dalam penelitian ini dikelompokkan dari topik tanggapan pelanggan yang terdiri variabel Layanan, Pemesanan dan Promosi.

1.4. Tujuan Penelitian

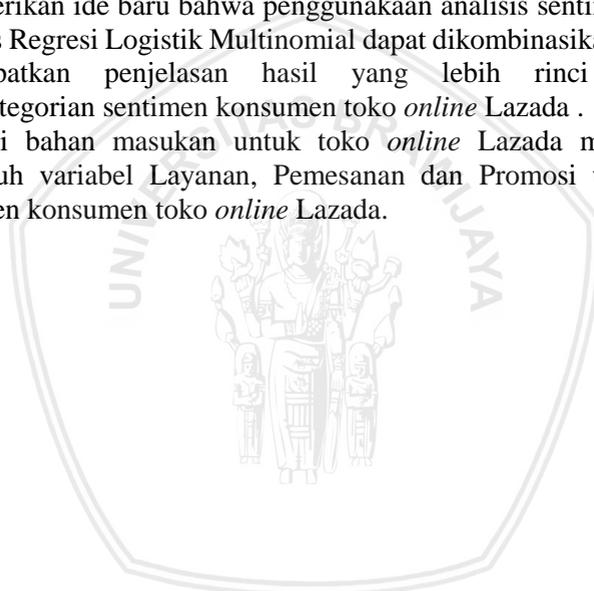
Tujuan yang ingin dicapai dari penelitian ini antara lain:

1. Menerapkan analisis sentimen menggunakan algoritma *Naive Bayes* dalam mengkategorikan sentimen tanggapan konsumen toko *online* Lazada.
2. Membentuk model Regresi Logistik Multinomial untuk menjelaskan pengaruh variabel Layanan, Pemesanan dan Promosi terhadap pengelompokan kategori sentimen konsumen toko *online* Lazada.

1.5. Manfaat Penelitian

Manfaat yang bisa diperoleh dari penelitian ini antara lain:

1. Memberikan ide baru bahwa penggunaan analisis sentimen dan analisis Regresi Logistik Multinomial dapat dikombinasikan untuk mendapatkan penjelasan hasil yang lebih rinci dalam pengkategorian sentimen konsumen toko *online* Lazada .
2. Sebagai bahan masukan untuk toko *online* Lazada mengenai pengaruh variabel Layanan, Pemesanan dan Promosi terhadap sentimen konsumen toko *online* Lazada.



BAB II TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Sentimen bisa diartikan sebagai opini yang disampaikan sesuai dengan perasaan seseorang (Hadna dkk., 2016). Media sosial sebagai salah satu media yang menyimpan dan menyampaikan berbagai macam tulisan yang berbentuk opini seseorang. Cakupan tulisan tersebut sangat luas sehingga tidaklah mudah untuk menentukan jenis opini secara cepat. Indurkha dan Damerau (2010) menyatakan bahwa opini merupakan hal yang sangat penting, karena bisa saja menyangkut pendapat konsumen terhadap layanan dan produk, tentunya hal ini penting bagi pemilik usaha. Analisis sentimen merupakan suatu teknik komputasi untuk mengelompokkan opini atau sentimen dan *emoticon* yang terdapat dalam sebuah teks (Liu, 2010).

Analisis sentimen merupakan bagian dari teks analisis. Teks analisis menurut Mckee (2001) merupakan suatu metodologi untuk mendapatkan dan menganalisis informasi dari media untuk memahami maksud dari sebuah teks. Data teks yang beragam dan memiliki jenis yang banyak termasuk ke dalam data tidak terstruktur. Oleh sebab itu, pengolahan data tidak terstruktur membutuhkan metode yang berbasis komputer atau *machine learning*.

Machine learning berupa kecerdasan buatan yang difokuskan untuk membentuk sistem, minimal dapat menyamai kecerdasan atau kemampuan manusia dalam mempelajari data dan dapat membuat akurasi dari prediksi tersebut (Hall dkk., 2014). *Machine learning* dapat diartikan sebuah komputer yang melalui tahapan proses simulasi berpikir dengan memanfaatkan algoritma. Menurut Berry (2003), saat ini banyak tersedia informasi *online* sehingga dibutuhkan kemampuan untuk mengekstrak dan mengelompokkan kumpulan dokumen secara cepat dengan penerapan algoritma pada dokumen yang sama.

Salah satu ciri penerapan analisis sentimen pada data teks adalah data dibagi menjadi dua bagian, antara lain data latih dan data uji. Data latih digunakan untuk membangun model sedangkan data uji untuk menerapkan model. Teks analisis merupakan bagian dari *data mining*. Menurut Ridwan dkk. (2013), tahapan dalam *data mining* antara lain:

- a. Pembersihan Data (*Data Cleaning*)

Proses menghilangkan kosakata yang tidak diperlukan dan tidak relevan.

b. Integrasi Data

Penggabungan data dari berbagai *database* ke satu *database*.

c. Transformasi Data

Data diubah ke dalam satu format yang sesuai untuk dilakukan tahapan data *mining*.

d. Proses *Mining*

Proses untuk menggali informasi dari data.

e. Evaluasi Pola

Mengidentifikasi pola-pola yang terbentuk pada data.

2.2. Klasifikasi *Naive Bayes*

Algoritma *Naive Bayes* merupakan salah satu metode yang digunakan untuk mengklasifikasikan dokumen. Berdasarkan penelitian Ting dkk. (2011), klasifikasi *Naive Bayes* lebih akurat dibandingkan dengan metode klasifikasi lainnya seperti *Support Vector Machine*, *Nearest Neighbour* dan *Decision Tree*. Asumsi algoritma *Naive Bayes* adalah antar variabel prediktor bersifat independen. Prinsip yang diterapkan algoritma Bayesian adalah peluang bersyarat (Feldman dan Sanger, 2006). Persamaan umum untuk teorema Bayesian seperti pada persamaan (2.1).

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \tag{2.1}$$

di mana,

$P(c|d)$: peluang kelas target saat bersyarat d

$P(c)$: peluang kelas kategori

$P(d)$: peluang d

Setiap dokumen d direpresentasikan oleh atribut $x_1, x_2, x_3, \dots, x_n$, dimana x_i adalah fitur ke- i . Persamaan (2.1) dapat dijabarkan menjadi persamaan (2.2).

$$\begin{aligned} P(c|x_1, \dots, x_n) &= \frac{P(c)P(x_1, \dots, x_n|c)}{P(x_1, \dots, x_n)} = \frac{P(c)P(x_1|c)P(x_2, \dots, x_n|c, x_1)}{P(x_1, \dots, x_n)} \\ &= \frac{P(c)P(x_1|c)P(x_2|c, x_1) \dots P(x_n, \dots, x_n|c, x_1, x_2, \dots, x_{n-1})}{P(x_1, \dots, x_n)} \\ &= \frac{P(c) \prod_{i=1}^n P(x_i|c)}{P(x_i)} \end{aligned} \tag{2.2}$$

Perhitungan $P(x_1, \dots, x_n|c)$ harus memenuhi asumsi, bahwa setiap fitur bersifat independen, artinya antar fitur tidak berkaitan satu sama lain. Berdasarkan asumsi tersebut maka berlaku persamaan (2.3).

$$P(x_i|x_i) = \frac{P(x_i \cap x_i)}{P(x_i)} = \frac{P(x_i)P(x_i)}{P(x_i)} = P(x_i) \quad (2.3)$$

Metode klasifikasi *Naive Bayes* merupakan klasifikasi yang menerapkan teorema Bayes dengan asumsi independen yang kuat. Klasifikasi berdasarkan pada peluang dokumen tertentu masuk ke kategori kelas yang telah ditentukan (Rajeswari dkk., 2017). Jika nilai peluang mencapai nilai yang maksimal maka dokumen d dinyatakan masuk ke kelas c .

Berdasarkan persamaan (2.1), kelas c merupakan kategori dari *tweet* yang terbentuk terdiri dari tiga jenis kategori, yaitu sentimen negatif, netral dan positif yang dapat disimbolkan dengan c_j dimana $j = 1, 2, 3$, yang mewakili jenis sentimen. Sehingga, persamaan (2.1) juga dapat ditulis menjadi persamaan (2.4).

$$P(c_j|x_i) = \frac{P(x_i|c_j)P(c_j)}{P(x_i)} \quad (2.4)$$

Nilai dari $P(c_j)$ dan $P(x_i|c_j)$ didapatkan dari perhitungan data latih sedangkan $P(c_j|x_i)$ merupakan nilai peluang yang dicari pada saat perhitungan pada data uji. Perhitungan nilai $P(c_j)$ dan $P(x_i|c_j)$ menggunakan persamaan (2.5) dan persamaan (2.6).

$$P(c_j) = \frac{docs_j}{ndocs} \quad (2.5)$$

$$P(x_i|c_j) = \frac{n_k + 1}{n + kata} \quad (2.6)$$

di mana,

$docs_j$: banyaknya *tweet* setiap kategori ke- j

$ndocs$: banyaknya *tweet* dari semua kategori

n_k : frekuensi kemunculan setiap kata k

n : frekuensi kemunculan kata dari semua kategori

$kata$: kata dari semua kategori

Hasil perhitungan data uji merupakan nilai prediksi dari model *Naive Bayes*. Ketepatan prediksi dinilai dari akurasi. Akurasi merupakan jumlah dokumen yang diklasifikasikan dengan benar (Hadna dkk., 2016). Menurut Perhitungan akurasi seperti pada persamaan (2.7).

$$\text{Akurasi} = \frac{tp + tn}{tp + fp + tn + fn} \quad (2.7)$$

di mana,

tp : *true positive*, data Positif yang diklasifikasikan Positif.

tn : *true negative*, data Negatif yang diklasifikasikan Negatif.

fp : *false positive*, data Negatif yang diklasifikasikan Positif.

fn : *false negative*, data Positif yang diklasifikasikan Negatif.

2.3. *Crawling* Data

Crawling data merupakan tahapan dalam penelitian yang bertujuan untuk mengumpulkan data dari suatu basis data atau dari media sosial seperti *twitter*. Data yang dikumpulkan bisa sebagian maupun keseluruhan. *Crawling* sangat bermanfaat di era *big data*. Data dari media sosial yang memiliki volume besar, kecepatan tinggi dan variasi yang beragam memudahkan pengguna untuk mengakses data tersebut menggunakan metode *crawling*.

Berbagai perusahaan berusaha mengembangkan web maupun *software* yang mendukung untuk kebutuhan *crawling* data, seperti *googlebot* dan *teleport pro* (Zuliarso dan Mustofa, 2009). *Software* yang digunakan untuk *crawling* data adalah R ataupun Phyton. Penggunaan *software* R untuk proses *crawling* memanfaatkan *Package TwitteR*.

2.4. Regresi Logistik Multinomial

2.4.1. Definisi Regresi Logistik Multinomial

Analisis regresi logistik digunakan untuk melihat pengaruh hubungan variabel respon dan variabel prediktor. Variabel respon pada regresi logistik berupa data kategorik. Regresi logistik yang dengan pembeda banyaknya level kategori pada respon dikelompokkan menjadi dua yaitu regresi logistik biner dan Regresi Logistik Multinomial. Regresi logistik biner memiliki dua level kategori pada variabel respon sedangkan Regresi Logistik Multinomial memiliki lebih dari dua level kategori untuk variabel respon (Hosmer dan Lemeshow, 2000).

Salah satu dari tiga level pada variabel respon dijadikan sebagai referensi sedangkan variabel lain sebagai pembanding (Kleinbaum dan Klein, 2002). Bentuk dasar dari persamaan Regresi Logistik Multinomial ditunjukkan pada persamaan (2.8) (Hosmer dan Lemeshow, 2013).

$$\pi(\mathbf{x}) = P(Y = j|\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}} \quad (2.8)$$

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

di mana,

- \mathbf{x} : vektor dari observasi variabel prediktor
- $\boldsymbol{\pi}(\mathbf{x})$: vektor peluang untuk kategori respon ke- j pada prediktor ke- i
- $g(\mathbf{x})$: fungsi logit variabel respon untuk kategori ke- j prediktor ke- i
- j : indeks untuk kategori variabel respon ($j = 0, 1, 2, \dots, j-1$)
- k : indeks untuk variabel prediktor ($k = 0, 1, 2, \dots, p$)
- i : indeks untuk pengamatan ($i = 0, 1, 2, \dots, n$)

2.4.2. Pendugaan Parameter Regresi Logistik Multinomial

Pendugaan parameter Regresi Logistik Multinomial memanfaatkan metode maksimum *likelihood*. Prinsip pendugaan parameter Regresi Logistik Multinomial hampir sama dengan logistik biner. Regresi logistik biner memiliki dua level kategori respon yang terdiri dari kejadian sukses dan gagal. Bentuk persamaan umum regresi logistik biner sebagai berikut.

$$\pi(\mathbf{x}) = P(Y = 1|\mathbf{x}) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Transformasi dari $\pi(\mathbf{x})$ akan membentuk fungsi transformasi logit seperti pada persamaan berikut.

$$g(\mathbf{x}) = \ln \left[\frac{P(Y = 1|\mathbf{x})}{P(Y = 0|\mathbf{x})} \right] = \beta_0 + \beta_1 x$$

Pendugaan parameter Regresi Logistik Multinomial dilakukan dengan cara membandingkan dua level. Satu level kategori digunakan sebagai referensi sedangkan sisanya sebagai variabel pembanding. Jika level Regresi Logistik Multinomial tersebut terdiri dari j kategori, maka persamaan yang akan terbentuk sebanyak $j-1$. Apabila $j = 3$, akan terbentuk dua persamaan logit seperti pada persamaan (2.9) dan persamaan (2.10).

$$\begin{aligned}
 g_1(\mathbf{x}) &= \ln \left[\frac{P(Y = 1|\mathbf{x})}{P(Y = 0|\mathbf{x})} \right] & (2.9) \\
 &= \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 \\
 &= \mathbf{x}' \boldsymbol{\beta}_1
 \end{aligned}$$

dan

$$\begin{aligned}
 g_2(\mathbf{x}) &= \ln \left[\frac{P(Y = 2|\mathbf{x})}{P(Y = 0|\mathbf{x})} \right] & (2.10) \\
 &= \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 \\
 &= \mathbf{x}' \boldsymbol{\beta}_2
 \end{aligned}$$

di mana,

$\boldsymbol{\beta}_j$: vektor koefisien kategori ke- j

\mathbf{x} : vektor yang berisi k prediktor dan sebuah konstan, dengan panjang vektor $p + 1$, di mana $x_0 = 1$

Persamaan untuk setiap kategori terdapat pada persamaan (2.11), persamaan (2.12) dan persamaan (2.13).

$$\pi_0(\mathbf{x}) = \frac{1}{1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})}} \quad (2.11)$$

$$\pi_1(\mathbf{x}) = \frac{e^{g_1(\mathbf{x})}}{1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})}} \quad (2.12)$$

$$\pi_2(\mathbf{x}) = \frac{e^{g_2(\mathbf{x})}}{1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})}} \quad (2.13)$$

Bentuk umum dari peluang tiga kategori variabel respon seperti berikut.

$$\pi_j(\mathbf{x}) = \frac{e^{g_j(\mathbf{x})}}{\sum_{j=0}^2 e^{g_j(\mathbf{x})}}$$

Fungsi pendugaan parameter metode *Maximum Likelihood* pada persamaan (2.14).

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi_0(\mathbf{x})^{y_{0i}} \pi_1(\mathbf{x})^{y_{1i}} \pi_2(\mathbf{x})^{y_{2i}} \dots \pi_{j-1}(\mathbf{x})^{y_{(j-1)i}} \quad (2.14)$$

di mana $\boldsymbol{\beta} = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jp})$.

Agar dapat membentuk fungsi likelihood tiga kategori variabel respon maka dibuat kondisi berikut: jika $Y = 0$ maka $Y_0 = 1, Y_1 = 0, Y_2 = 0$; jika $Y = 1$ maka $Y_0 = 0, Y_1 = 1, Y_2 = 0$; jika $Y = 2$ maka $Y_0 = 0, Y_1 = 0, Y_2 = 1$. Sehingga $\sum_{j=0}^2 Y_j = 1$ untuk setiap i , fungsi likelihood dari $\boldsymbol{\beta}$ untuk n observasi ditunjukkan pada persamaan (2.15).

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi_0(\mathbf{x})^{y_{0i}} \pi_1(\mathbf{x})^{y_{1i}} \pi_2(\mathbf{x})^{y_{2i}} \quad (2.15)$$

Logaritma natural fungsi *likelihood* dari persamaan (2.15) ditunjukkan pada persamaan (2.16).

$$L(\boldsymbol{\beta}) = \ln \prod_{i=1}^n \pi_0(\mathbf{x})^{y_{0i}} \pi_1(\mathbf{x})^{y_{1i}} \pi_2(\mathbf{x})^{y_{2i}} \quad (2.16)$$

Susbsitusikan persamaan (2.11), persamaan (2.12) dan persamaan (2.13) ke persamaan (2.16). Logaritma dari persamaan (2.16) seperti pada persamaan (2.17).

$$\begin{aligned} L(\boldsymbol{\beta}) &= \ln \prod_{i=1}^n \pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}} \\ &= \ln \prod_{i=1}^n \left[\left(\frac{1}{(1 + e^{g_1(x_i)} + e^{g_2(x_2)})} \right)^{y_{0i}} \left(\frac{e^{g_1(x_i)}}{(1 + e^{g_1(x_i)} + e^{g_2(x_2)})} \right)^{y_{1i}} \left(\frac{1}{(1 + e^{g_1(x_i)} + e^{g_2(x_2)})} \right)^{y_{0i}} \right] \\ &= \ln \prod_{i=1}^n \left[(e^{g_1(x_i)})^{y_{1i}} (e^{g_2(x_i)})^{y_{2i}} \left(\frac{1}{(1 + e^{g_1(x_i)} + e^{g_2(x_2)})} \right)^{y_{0i} + y_{1i} + y_{2i}} \right] \\ &= \ln \left[(e^{g_1(x_i)})^{\sum_{i=1}^n y_{1i}} (e^{g_2(x_i)})^{\sum_{i=1}^n y_{2i}} (1 + e^{g_1(x_i)} + e^{g_2(x_2)})^{-\sum_{i=1}^n (y_{0i} + y_{1i} + y_{2i})} \right] \\ &= \ln \left((e^{g_1(x_i)})^{\sum_{i=1}^n y_{1i}} (e^{g_2(x_i)})^{\sum_{i=1}^n y_{2i}} \right) + \ln \left(1 + e^{g_1(x_i)} + e^{g_2(x_2)} \right)^{-\sum_{i=1}^n (y_{0i} + y_{1i} + y_{2i})} \quad (2.17) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n y_{1i} \ln(e^{g_1(x_i)}) + \sum_{i=1}^n y_{2i} \ln(e^{g_2(x_i)}) - \sum_{i=1}^n (y_{0i} + y_{1i} + y_{2i}) \ln(1 + e^{g_1(x_i)} + e^{g_2(x_2)}) \\
 &= \sum_{i=1}^n [y_{1i} g_1(x_i) + y_{2i} g_2(x_i)] - \sum_{i=1}^n \ln(1 + e^{g_1(x_i)} + e^{g_2(x_2)})
 \end{aligned}$$

Persamaan *likelihood* didapatkan dari turunan pertama parsial $L(\beta)$.

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^n x_{ki} (y_{ji} - \pi_{ji}) \tag{2.18}$$

di mana,

x_{ki} : nilai variabel prediktor ke- k pada pengamatan ke- i

y_{ji} : nilai variabel respon kategori ke- j pada pengamatan ke- i

π_{ji} : nilai peluang kategori respon ke- j pada pengamatan ke- i

Hasil dari turunan persamaan *log likelihood* dilanjutkan dengan iterasi Newton-Rapshon untuk mendapatkan nilai pendugaan parameter dengan langkah antara lain menentukan penduga awal dari $\hat{\beta}^t, \hat{\beta}^1 = \mathbf{0}$, selanjutnya iterasi untuk mendapatkan penduga $\hat{\beta}^t$ yang baru, $\hat{\beta}^{t+1} = \hat{\beta}^t - \mathbf{H}^{-1} L'(\hat{\beta}^t)$, di mana \mathbf{H}^{-1} adalah matriks Hessian yang merupakan turunan kedua dari fungsi dari fungsi log likelihood atau turunan kedua dari fungsi dari fungsi *log likelihood* dan $L'(\hat{\beta}^t)$, merupakan turunan pertama dari fungsi *log likelihood* penduga parameter. Iterasi berhenti saat iterasi mencapai iterasi maksimum dan nilai panjang \mathbf{d} atau $\|\mathbf{d}\| < 10^{-6}$. Rumus $\|\mathbf{d}\|$ sebagai berikut.

$$\|\mathbf{d}\| = \sqrt{\sum_{k=1}^p (\hat{\beta}_{ik}^{t+1} - \hat{\beta}_{ik}^t)^2}$$

di mana, $\mathbf{d} = \hat{\beta}_{ik}^{t+1} - \hat{\beta}_{ik}^t = \begin{bmatrix} \beta_{10}^{t+1} \\ \cdot \\ \cdot \\ \beta_{1k}^{t+1} \end{bmatrix} - \begin{bmatrix} \beta_{10}^t \\ \cdot \\ \cdot \\ \beta_{1k}^t \end{bmatrix}$

2.4.3. Pengujian Parameter Regresi Logistik Multinomial Secara Simultan

Pengujian parameter Regresi Logistik Multinomial secara simultan menggunakan statistik uji G. Hipotesis yang digunakan yaitu,

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 ; \beta_{jk} = 0 \text{ vs}$$

$$H_1 : \text{paling tidak terdapat satu } \beta_{jk} \neq 0$$

Jika H_0 benar maka statistik uji G sebagaimana ditunjukkan persamaan (2.19) (Hosmer dan Lemeshow, 2013).

$$G = -2 \log \left(\frac{L_0}{L_1} \right) \sim \chi_p^2 \quad (2.19)$$

di mana,

p : banyaknya variabel prediktor dalam model

L_0 : nilai *likelihood* tanpa variabel prediktor

L_1 : nilai *likelihood* dengan variabel prediktor

Keputusan menolak H_0 jika $G > \chi_{(p,\alpha)}^2$ atau saat nilai peluang statistik uji G kurang dari nilai α $P(\chi_p^2 > G) < \alpha$.

2.4.4. Pengujian Parameter Regresi Logistik Multinomial Secara Parsial

Uji *Wald* dapat digunakan untuk menguji parameter Regresi Logistik Multinomial untuk masing-masing variabel prediktor. Hipotesis yang digunakan uji *Wald* sebagai berikut,

$H_0 : \beta_k = 0$ atau variabel prediktor ke- k tidak berpengaruh terhadap variabel respon vs

$H_1 : \beta_k \neq 0$ atau variabel prediktor ke- k berpengaruh terhadap variabel respon

Jika H_0 benar, maka statistik uji Wald ditunjukkan pada persamaan (2.20).

$$W = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \sim N(0,1) \quad (2.20)$$

di mana,

k : 1, 2, ..., p

p : banyaknya variabel prediktor dalam model

$\hat{\beta}_k$: penduga parameter β_k

$SE(\hat{\beta}_k)$: salah baku penduga parameter β_k yang bernilai sama dengan $\sqrt{var(\beta)}$.

dimana,

$$var(\beta) = (X'VX)^{-1}$$

$$V = \begin{bmatrix} \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & \dots & 0 \\ 0 & \hat{\pi}_2(1 - \hat{\pi}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \hat{\pi}_n(1 - \hat{\pi}_n) \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}$$

Keputusan menolak H_0 jika statistik uji $W > Z_{\alpha/2}$ atau saat nilai peluang statistik uji W kurang dari nilai α atau $P(Z_{\alpha/2} > W) < \alpha$.

2.4.5. Kebaikan Model

Kebaikan model Regresi Logistik Multinomial bisa ditentukan dari *Akaike Information Criteria* (AIC) dengan menggunakan rumus (Chatterjee dan Hadi, 2006), sebagaimana persamaan (2.21)

$$AIC = -2(\text{Log likelihood model}) + 2p \quad (2.21)$$

Semakin kecil nilai AIC yang didapatkan semakin bagus model yang terbentuk. Perbandingan nilai AIC bisa digunakan pada kombinasi variabel prediktor seperti membandingkan AIC saat variabel prediktor berjumlah dua buah dengan tiga variabel prediktor.

2.4.6. Interpretasi Model

Menurut Agresti (2007), regresi logistik diinterpretasikan menggunakan *odds ratio* (OR) yang nilainya sama dengan $\exp(\beta_{jk})$. *Odds ratio* adalah untuk mengetahui tingkat kecenderungan atau risiko hubungan peubah prediktor terhadap peubah respon. Ukuran ini digunakan untuk mengetahui risiko kecenderungan mengalami suatu kejadian tertentu. *Odd ratio* dilambangkan dengan ψ (Ψ).

$$\psi = \exp(\hat{\beta}_{jk}) \quad (2.22)$$

dimana $\hat{\beta}_{jk}$ merupakan koefisien model kategori ke- j peubah prediktor ke- k . Menurut Kleinbaum dan Klein (2002), pada Regresi Logistik Multinomial dengan peubah respon yang memiliki j kategori akan didapat kan sebanyak $j-1$ *odds ratio*. *Odds ratio* pertama adalah perbandingan peluang antara peubah respon kategori satu dengan peubah respon kategori pembanding dan *odds ratio* kedua adalah perbandingan peluang antara peubah respon kategori dua dengan peubah respon kategori pembanding.

2.4.7. Non Multikolinieritas

Asumsi analisis Regresi Logistik Multinomial adalah *non multikolinieritas*. Berbeda dengan analisis regresi linier, Regresi Logistik Multinomial tidak membutuhkan asumsi linieritas antar variabel, homoskedastisitas dan normalitas galat (Garson, 2008). Multikolinieritas menunjukkan adanya hubungan yang kuat antar variabel prediktor sehingga akan terdapat bias dalam menentukan pengaruh variabel prediktor terhadap variabel respon (Chatterjee dan Hadi, 2006). Oleh sebab itu, agar model regresi logistik dapat digunakan untuk prediksi maka antar variabel prediktor diharuskan tidak memiliki hubungan yang erat atau *non multikolinieritas*.

Adapun cara untuk mendeteksi multikolinieritas bisa dilihat dari nilai VIF atau *Variance Inflation Factor*. Jika terdapat variabel prediktor X_j , nilai VIF-nya dapat dicari menggunakan rumus pada persamaan (2.23)

$$VIF_j = \frac{1}{1 - R_j^2}, \quad j = 1, 2, \dots, p \quad (2.23)$$

Sesuai dengan persamaan (2.23) nilai j bergerak dari 1 sampai p , di mana p merupakan banyaknya variabel prediktor dan R_j^2 adalah nilai koefisien korelasi berganda yang didapatkan dari X_j yang diregresikan dengan prediktor lainnya. Istilah ini juga dikenal dengan koefisien determinasi. Nilai VIF semakin besar saat R_j^2 mendekati satu. Apabila nilai VIF lebih dari 10 dapat disimpulkan bahwa terdapat multikolinieritas antar variabel (Chatterjee dan Hadi, 2006).

2.5. Twitter

Twitter sudah ada sejak tahun 2006 diciptakan oleh Jack Dorsey, Evan Williams, Biz Stone dan Noah Glass. *Twitter* dapat digunakan untuk menyampaikan pesan atau dikenal dengan *tweets* atau kicauan.

repository.ub.ac.id

Twitter merupakan salah satu contoh media sosial. Media sosial dapat diartikan sebagai media *online* di mana pengguna dapat memanfaatkan aplikasi *online* untuk menyampaikan pesan berupa tulisan, gambar, maupun video (Humas Kementerian Perdagangan, 2014). Meskipun sekarang telah banyak media sosial baru yang digunakan masyarakat Indonesia, tapi *twitter* masih termasuk sosial media yang banyak digunakan.

Penelitian yang dilakukan oleh Daniel dkk. (2017) menyatakan bahwa *twitter* merupakan salah satu media sosial sangat berpengaruh dalam dunia pasar. Pernyataan ini juga didukung dari hasil penelitian yang dilakukan oleh Yang dkk. (2014), adanya hubungan timbal balik yang kuat antara sosial media dan pergerakan harga yang dibentuk suatu komunitas finansial. Hal ini dapat disimpulkan berlakunya sentimen secara umum pada pesan yang disampaikan konsumen terhadap bisnis tertentu.

Twitter yang hanya mampu menampung maksimal 140 karakter dimanfaatkan untuk mengungkapkan percakapan melalui tulisan atau *emoticon*. Tulisan tersebut bisa meliputi rasa senang, sedih, sindiran dan kekecewaan atau istilah ini biasa kita kenal dengan sentimen. Tulisan dari penggunaan *twitter* tersebut tersebut dapat ditentukan jenis sentimennya (Go dkk., 2009).

2.6. Bisnis Online Lazada

E-commerce merupakan salah satu perdagangan barang atau jasa yang berbasis internet (Kasim, 2001). Laudon dan Traver (2013) menyebutkan *e-commerce* memanfaatkan internet atau web untuk melakukan transaksi bisnis. Berdasarkan pengertian tersebut, *e-commerce* dapat didefinisikan sebagai kegiatan jual beli produk, jasa atau informasi melalui internet atau sarana *online* lain. *E-commerce* seperti Lazada Indonesia memiliki karakteristik informasi dan pelayanan yang diberikan bersifat umum dengan mekanisme yang dapat digunakan oleh masyarakat berdasarkan permohonan (*on demand*) dengan artian konsumen melakukan inisiatif dan produsen harus siap memberikan respon sesuai dengan permohonan.

Lazada Indonesia diluncurkan pada bulan Maret 2012 dan merupakan salah satu bagian dari jaringan retail *online* Lazada Group yang beroperasi di enam negara di Asia Tenggara yaitu Indonesia, Malaysia, Thailand, Vietnam, Singapura dan Filipina dengan total pengguna 550 juta pengguna dari total keseluruhan enam negara tersebut. Lazada merupakan perusahaan yang bergerak di bidang

layanan jual beli *online* dan retail *e-commerce*. Berbagai produk yang dijual diantaranya adalah produk elektronik, dekorasi rumah, produk kesehatan hingga produk kecantikan. Lazada Indonesia memiliki prioritas tertinggi yaitu menciptakan pengalaman belanja *online* terbaik untuk setiap pelanggan di Indonesia.



BAB III METODE PENELITIAN

3.1. Sumber Data

Penelitian ini menggunakan data yang diperoleh dengan cara *crawling* data dari *twitter* dengan mengambil *tweets* terkait Lazada Indonesia. *Crawling* data menggunakan *software* R dengan memanfaatkan *Package TwitteR* untuk mengambil data secara otomatis dari *twitter*. Data *tweets* konsumen hasil *crawling* berisi teks *tweets* seperti yang terdapat Tabel 3.1.

Tabel 3.1. Struktur Data Hasil *Crawling*

No	Teks	<i>favorited</i>	Banyak favorit	<i>replyToSN</i>
1	@LazadaID min ada kode voucher promo kah hari ini? Gak ada yg bisa dipakai/valid mau checkout neh	FALSE	0	LazadaID
2	Keren @LazadaID dan @xiaomiindonesiabarukmren order mi A1, sekarang udah datang aja barangnya, posisi di Bandung padahal, mantap!	FALSE	0	LazadaID
⋮	⋮	⋮	⋮	⋮
1.000	@LazadaID Kok pesanan saya belum sampai - sampai ya?		0	LazadaID

Data pada Tabel 3.1 akan diklasifikasikan menggunakan analisis sentimen. Struktur data untuk membentuk analisis sentimen terdapat pada Tabel 3.2.

Tabel 3.2. Struktur Data Analisis Sentimen

No Observasi	Kosakata	Kategori Sentimen
1	Lama	Negatif
2	Bagus	Positif
⋮	⋮	⋮
1.000	Apa	Netral

3.2. Variabel Penelitian

Pemodelan Regresi Logistik Multinomial hasil klasifikasi sentimen dari *tweet* pelanggan toko *online* Lazada diperlukan adanya penentuan variabel prediktor. Hasil analisis sentimen dibentuk dari kosakata yang dijadikan sebagai penentu dalam mengkategorikan sentimen *tweet*. Kosakata tersebut dikelompokkan sesuai dengan jenis sentimennya dan berdasarkan topik yang sama. Kosakata yang dikelompokkan berdasarkan jenis topik yang sama akan dijadikan untuk pembentukan variabel prediktor dalam pemodelan Regresi Logistik Multinomial. Variabel prediktor terdiri dari variabel Layanan, Pemesanan dan Promosi. Variabel respon terdiri dari tiga level kategori antara lain Positif, Negatif dan Netral.

3.3. Metode Analisis

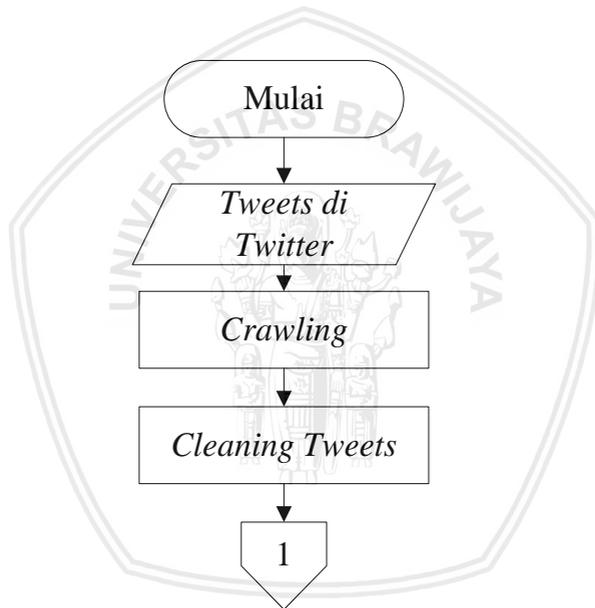
Prosedur dari penelitian ini terdiri dari beberapa langkah sebagai berikut:

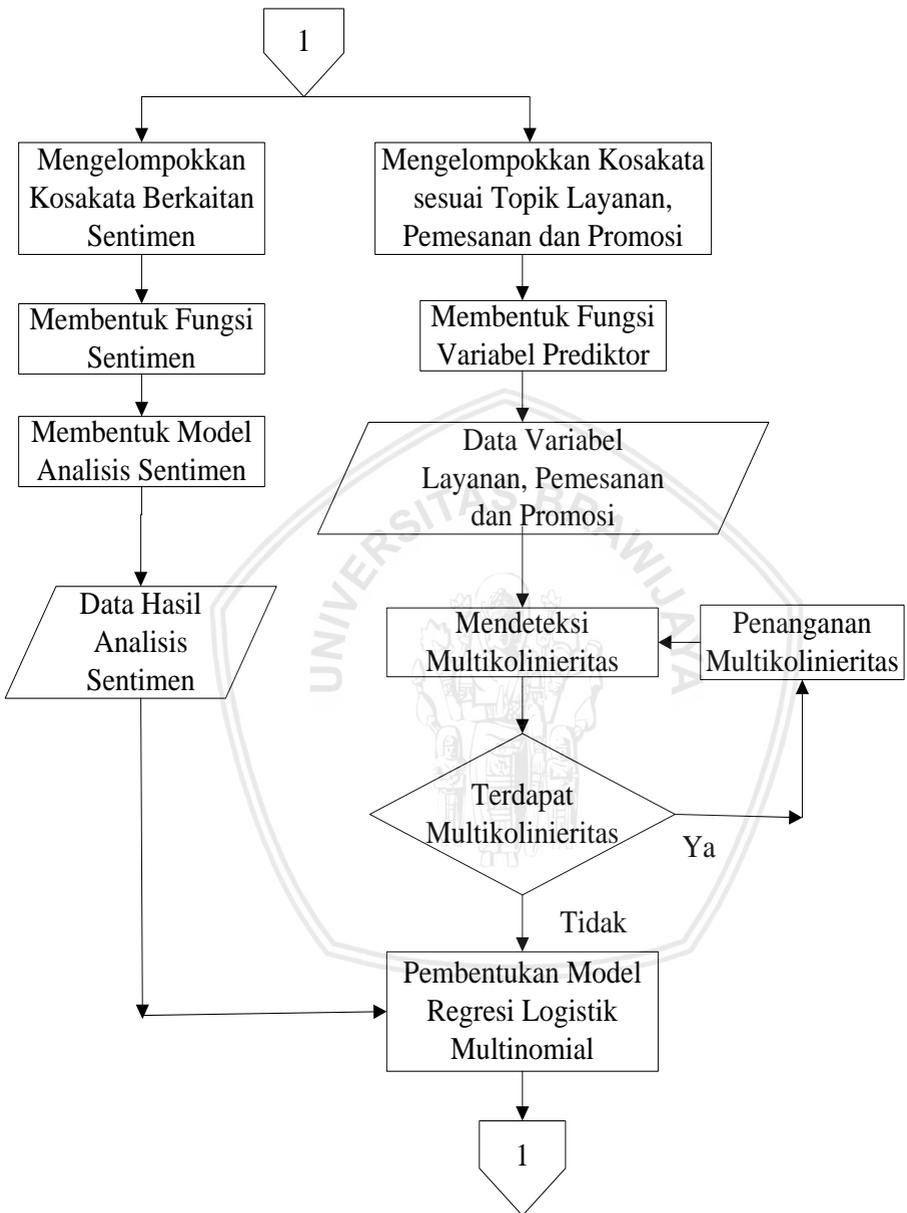
1. *Crawling* data dari *twitter* menggunakan Rstudio, penjelasan tentang *crawling* data pada sub bab 2.3.
2. *Cleaning tweets* pada data *twitter* sesuai penjelasan pada sub bab 2.1.
3. Membuat koding untuk analisis sentimen berdasarkan *keyword* dari *tweets* konsumen Lazada.
4. Mengelompokkan kosakata pada *tweets* sehingga dapat membentuk beberapa variabel prediktor.
5. Menjalankan koding analisis sentimen sehingga masing-masing observasi dikelompokkan berdasarkan tiga jenis sentimen, struktur data terdapat pada Tabel 3.2.

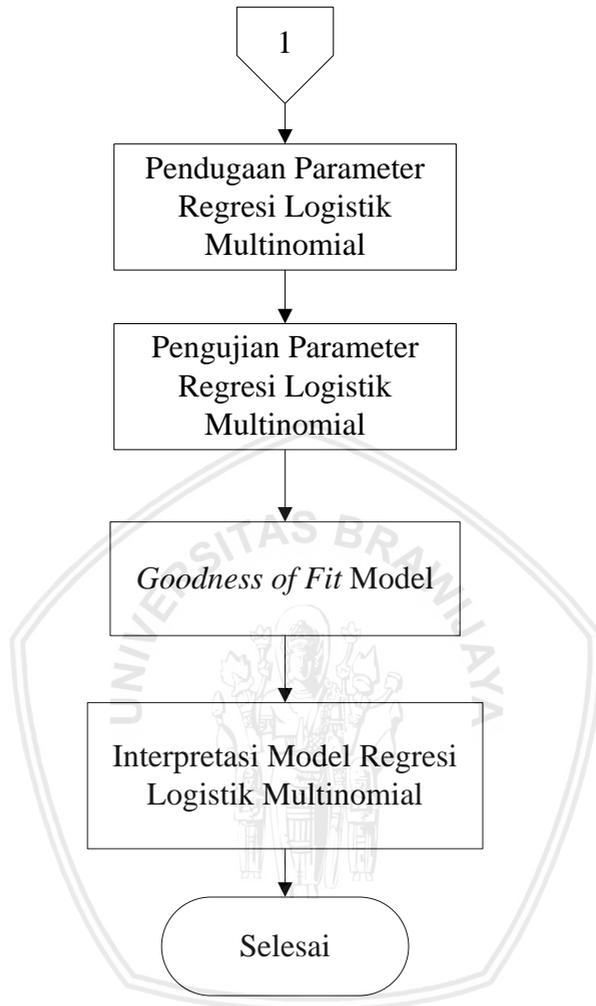
6. Mendeteksi multikolinieritas pada data hasil analisis sentimen sesuai persamaan (2.23).
7. Pendugaan parameter Regresi Logistik Multinomial sesuai persamaan (2.17).
8. Menentukan *goodness of fit* atau kebaikan model sesuai persamaan (2.21).
9. Interpretasi model Regresi Logistik Multinomial yang terbentuk.

3.4. Diagram Alir Penelitian

Berdasarkan metode analisis dapat dibentuk diagram alir penelitian seperti pada Gambar 3.1.







Gambar 3.1. Diagram Alir Penelitian

BAB IV HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Data yang digunakan pada penelitian ini berasal dari *twitter* yang memuat kata LazadaID dengan cara mengumpulkan atau *crawling tweets* menggunakan program R yang diwakili dengan fungsi : `dtweet = searchTwitter("LazadaID", n=1.000)`. Fungsi “dtweet” merupakan nama variabel yang menyimpan data yang telah dikumpulkan sebanyak 1.000 *tweets*. *Tweets* diambil bulan Januari tahun 2018. Hasil *crawling* merupakan *tweets* yang berasal dari pelanggan Lazada Indonesia dan akun LazadaID sendiri. Dalam penelitian ini *tweets* yang diperlukan hanya *tweets* yang berasal dari pelanggan Lazada Indonesia. Oleh sebab itu, dari 1.000 *tweets* didapatkan 812 *tweets* yang berasal dari pelanggan Lazada Indonesia. Data *tweets* pelanggan Lazada Indonesia seperti pada Tabel 4.1.

Tabel 4.1. Data Hasil *Crawling*

No	Teks
1	@LazadaID sudah saya DM data2nya, terimakasih
2	@LazadaID sudah 3 hari pesanan saya kok masih dalam proses pengemasan?kpn dikirimnya?!
3	@LazadaIDCare @LazadaID ga jadi min. slow respond banget di tanya nya . lebih dr 24 jam respon nya.
⋮	⋮
812	@LazadaID Kok pesanan saya belum sampai - sampai ya?

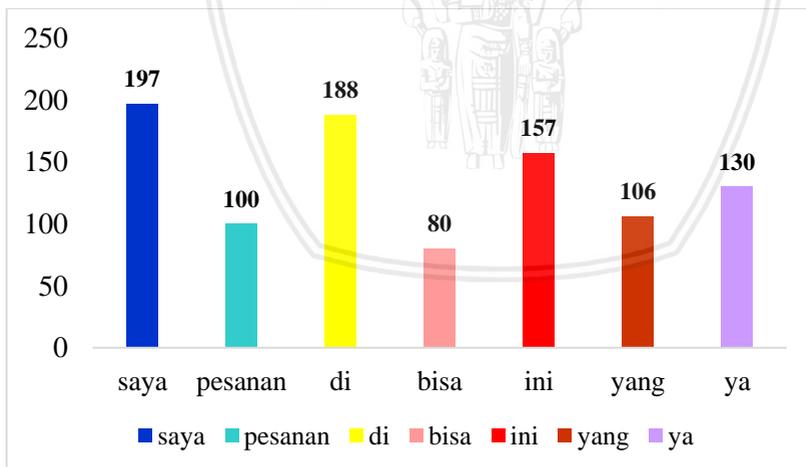
Data *tweets* hasil *crawling* pada Tabel 4.1 masih memiliki tanda baca yang tidak diperlukan untuk proses analisis selanjutnya. Tanda baca yang dimaksud seperti “@”, “<” dan “/”. Kata lain yang perlu dihilangkan seperti tautan dari sebuah halaman *website* dan merubah huruf kapital menjadi huruf kecil. Oleh sebab itu, tanda baca dan kata-kata yang tidak diperlukan tersebut harus dihilangkan melalui proses pembersihan teks. Pembersihan teks memanfaatkan fungsi “gsub” yang terdapat di dalam fungsi buatan “Teksbersih”. Analogi berpikir dari *syntax* “gsub” seperti contoh *syntax* berikut ini, `oke = gsub("@\\w+", "", oke)`, artinya data “oke” yang memiliki tanda baca “@”

yang diikuti oleh huruf lainnya akan diganti dengan kata yang diinisialisasikan setelahnya atau dalam kasus ini dihilangkan. Fungsi buatan untuk pembersihan teks terdapat pada Lampiran 2. Hasil pembersihan teks dicantumkan pada Tabel 4.2.

Tabel 4.2. Data Hasil *Cleaning* Teks

No	Teks
1	sudah saya dm datanya terimakasih
2	sudah hari pesanan saya kok masih dalam proses pengemasan kpn dikirimnya
3	ga jadi min slow respond banget di tanya nyalebih drjam respon nya
⋮	⋮
812	kok pesanan saya belum sampaisampai ya

Data berupa teks pada Tabel 4.2, disusun oleh kata-kata yang sering muncul dalam setiap *tweets*. Fungsi untuk membentuk kata dan frekuensinya terdapat pada Lampiran 7 dengan hasil seperti pada Gambar 4.1 berikut.



Gambar 4.1. Plot Kata dan Frekuensi Minimal 100

Masing-masing kata penyusun setiap *tweets* akan ditentukan bobotnya sesuai dengan jenis klasifikasi sentimen yang akan dibentuk. Gambar 4.1 menampilkan kata-kata dan frekuensinya dengan frekuensi minimal adalah 100. Kata yang paling banyak muncul adalah kata “saya” dan “di”. Namun, dua kata tersebut tidak bisa dijadikan skor untuk pembobot sentimen karena kata tersebut hanya berupa kata penghubung dalam *tweets*. Kata “ya” dengan frekuensi 130 dan “bisa” frekuensinya 80, dapat digunakan untuk kosakata pembentuk sentimen Positif.

4.2. Pembentukan Klasifikasi Sentimen

Klasifikasi sentimen dibentuk dengan memanfaatkan fungsi buatan “cb”, seperti yang terdapat pada Lampiran 3. Langkah awal dari fungsi ini adalah menginisialisasi kosakata berdasarkan skor. Pemberian skor -1 untuk kata yang memiliki sentimen Negatif, skor 0 untuk kata yang memiliki sentimen Netral dan skor 1 untuk kata yang memiliki sentimen Positif, seperti yang ditunjukkan pada Tabel 4.3 berikut.

Tabel 4.3. Kosakata dan Skor

No	Kosakata	Skor
1	hi	0
2	tidak	-1
3	promo	1
⋮	⋮	⋮
279	cepat	1

Kosakata yang dijadikan untuk pembobot dalam menentukan klasifikasi sentimen terdapat 279 kata. Dari jumlah total kosakata tersebut, terdapat 130 kata Negatif, 60 kata Netral dan 88 kata Positif, seperti dicantumkan pada Lampiran 18 hingga Lampiran 20. Tahap membentuk klasifikasi sentimen dibagi menjadi tahap pembentukan model dan tahap prediksi. Oleh sebab itu, total data yang terdiri dari 812 *tweets* dibagi menjadi 70% untuk data latih dan 30% untuk data uji. Terdapat 243 *tweets* yang digunakan pada data uji dan pembentukan model *Naive Bayes* menggunakan data latih yang terdiri dari 569 *tweets*, seperti pada Lampiran 10. Fungsi “cb” diterapkan untuk semua data. Analogi dari fungsi “cb” seperti pada data ketiga di Tabel 4.2 yang memiliki kalimat “ga jadi min slow respond banget di

tanya nyalebih drjam respon nya”. Pada kalimat tersebut terdapat dua kata yang mengandung kosakata sentimen Negatif, “ga” dan “slow” sehingga *tweets* tersebut diklasifikasikan sebagai sentimen Negatif. Hasil pembentukan sentimen menggunakan fungsi buatan “cb” dimanfaatkan untuk membentuk model klasifikasi *Naive Bayes*.

Pembentukan model menggunakan *Naive Bayes* memanfaatkan *package* e1071 pada R. *Syntax* yang digunakan seperti Lampiran 4. Berdasarkan persamaan (2.5), nilai $P(c_j)$ merupakan rasio dari banyak sentimen tertentu dengan total semua *tweets* atau nilai *A-priori probabilities*, sesuai dengan Tabel 4.4 berikut.

Tabel 4.4. Hasil *A-Priori Probabilities* Kelas Sentimen

Sentimen	<i>A-priori probabilities</i>	Jumlah <i>Tweets</i>
Negatif	0,337	192
Netral	0,288	164
Positif	0,374	213

Berdasarkan Tabel 4.4, peluang awal sentimen Negatif bernilai 0,337, dengan jumlah *tweets* 192. Peluang sentimen Netral 0,288 dengan 164 *tweets* dan peluang awal sentimen Positif adalah 0,374 dengan jumlah *tweets* 213.

Hasil perhitungan dari data latih digunakan untuk memprediksi klasifikasi sentimen pada data uji, seperti yang terdapat pada Lampiran 12. Hasil prediksi berupa nilai peluang untuk setiap sentimen dengan perhitungan berdasarkan persamaan (2.4). Hasil prediksi untuk tujuh data pertama akan ditampilkan pada Tabel 4.5.

Tabel 4.5. Hasil Prediksi Klasifikasi Sentimen Data Uji

No	Negatif	Netral	Positif	Sentimen
1	0,133	0,687	0,179	Netral
2	0,962	0,036	0,002	Negatif
3	0,962	0,036	0,002	Negatif
4	0,133	0,687	0,179	Netral
5	0,560	0,349	0,089	Negatif
	⋮	⋮	⋮	⋮
243	0,133	0,687	0,179	Netral

Berdasarkan Tabel 4.5, kelas sentimen ditentukan dari nilai peluang terbesar dari setiap sentimen. Diambil contoh untuk data pertama, peluang sentimen Negatif yang diperoleh 0,133, Netral 0,687 dan Positif 0,179. Dari ketiga nilai peluang tersebut, nilai peluang maksimal adalah sentimen Netral, sehingga data pertama diklasifikasikan menjadi sentimen Netral. Hasil prediksi sentimen paling banyak diklasifikasikan sebagai sentimen Netral sebanyak 98 *tweets*, sentimen Negatif 79 *tweets* dan sentimen Positif 66 *tweets*. Kesesuaian hasil prediksi sentimen dengan klasifikasi sentimen aktual atau sebenarnya ditunjukkan pada Tabel 4.6.

Tabel 4.6. Hasil Prediksi Sentimen *Naive Bayes*

Data	Aktual				Total
	Sentimen	Negatif	Netral	Positif	
Uji	Negatif	74	1	4	79
	Netral	3	90	5	98
	Positif	0	11	55	66
Total		77	102	64	243

Berdasarkan Tabel 4.6, dapat disimpulkan dari 79 total sentimen Negatif pada data uji terdapat 74 *tweets* yang diklasifikasikan dengan benar. Kesalahan klasifikasi sentimen Negatif pada data uji yaitu satu *tweet* yang seharusnya diklasifikasikan sentimen Negatif pada data uji menjadi sentimen Netral dan empat *tweets* seharusnya masuk sentimen Negatif diklasifikasikan sentimen Positif. Sentimen Netral dengan tepat diklasifikasikan sebanyak 90 *tweets*, sedangkan tiga *tweets* yang seharusnya menjadi sentimen Netral diklasifikasikan sentimen Negatif dan lima *tweets* salah diklasifikasikan menjadi sentimen Positif. Sentimen Positif yang diklasifikasikan dengan benar adalah 55 *tweets* dan terdapat 11 *tweets* yang seharusnya masuk sentimen Netral. Secara keseluruhan akurasi dari hasil prediksi 90,1%. Nilai ini didapatkan sesuai dengan persamaan (4.1).

$$\text{Akurasi} = \frac{74 + 90 + 55}{243} = 0,901 \quad (4.1)$$

4.3. Pembentukan Variabel Prediktor Layanan, Pemesanan dan Promosi

Hasil klasifikasi sentimen menggunakan metode *Naive Bayes* dijadikan sebagai variabel respon untuk pembentukan model Regresi Logistik Multinomial dengan variabel prediktor Layanan (X_1), Pemesanan (X_2) dan Promosi (X_3). *Syntax* untuk pembentukan nilai variabel ini terdapat pada Lampiran 6. Kosakata setiap variabel diinisialisasikan seperti pada Lampiran 13, Lampiran 14 dan Lampiran 15. Data untuk variabel prediktor dicantumkan seperti pada Tabel 4.7. Tabel 4.7 menjelaskan bahwa saat terdapat tiga kata yang berkaitan dengan variabel Layanan, satu kata berkaitan variabel Pemesanan dan nol untuk variabel Promosi, maka sentimen *tweets* tersebut adalah Positif. Pada *tweet* berikutnya, terdapat tiga kata berkaitan Promosi dan variabel lainnya bernilai nol, sentimennya Positif. Pada data ke-569, satu kata berkaitan variabel Layanan, satu kata variabel Pemesanan sentimennya Negatif. Oleh sebab, itu perlu dilakukan suatu uji untuk melihat apakah terdapat pengaruh variabel kategori Layanan, Pemesanan dan Promosi terhadap klasifikasi sentimen dari *tweets* tersebut.

Tabel 4.7. Data Analisis Regresi Logistik Multinomial

No	Layanan atau X_1	Pemesanan atau X_2	Promosi atau X_3	Sentimen atau Y
1	3	1	0	Positif
2	0	0	3	Positif
3	3	1	0	Positif
4	3	4	0	Positif
5	2	1	0	Negatif
6	0	0	0	Positif
7	0	0	3	Positif
⋮	⋮	⋮	⋮	⋮
569	1	1	0	Negatif

4.4. Pembentukan Model Regresi Logistik Multinomial

Variabel yang dianalisis pada Regresi Logistik Multinomial memiliki variabel prediktor yang terdiri dari variabel Layanan (X_1),

Pemesanan (X_2) dan Promosi (X_3). Nilai variabel prediktor dalam pembentukan model Regresi Logistik Multinomial ini berisi banyaknya kata dalam satu *tweets* yang berkaitan dengan setiap variabel. Variabel respon berupa kelas sentimen yang terdiri dari tiga kelas sentimen yaitu Negatif, Netral dan Positif. Asumsi yang harus terpenuhi untuk analisis Regresi Logistik Multinomial adalah asumsi non multikolinieritas. Hasil uji asumsi non multikolinieritas terdapat pada Tabel 4.8 .

Tabel 4.8. Hasil Uji Asumsi Non Multikolinieritas Variabel Layanan, Pemesanan dan Promosi

Variabel	VIF
Layanan	3,9
Pemesanan	2,8
Promosi	1,2

Merujuk pada persamaan (2.23), multikolinieritas antar variabel prediktor terjadi apabila keeratan hubungan antar variabel sangat besar. Hal tersebut ditandai dengan nilai R_j^2 mendekati satu atau dapat dimisalkan dengan 0,9. Saat $R_j^2 = 0,9$, maka nilai *Variance Inflation Factor* (VIF) akan sama dengan 10. Oleh sebab itu, apabila nilai VIF dari variabel prediktor lebih dari 10, dapat dikatakan bahwa antar variabel memiliki keeratan hubungan yang kuat atau terdapat multikolinieritas. Nilai VIF variabel Layanan pada Tabel 4.8, didapatkan dari meregresikan variabel Layanan sebagai variabel respon terhadap variabel Pemesanan dan Promosi. Koefisien determinasi dari regresi tersebut dimasukkan pada persamaan (2.23). Begitu juga dengan perhitungan VIF untuk variabel Pemesanan, yang menjadi variabel prediktornya adalah variabel Layanan dan Promosi.. Semua variabel prediktor memiliki nilai VIF kurang dari 10 sehingga dapat disimpulkan bahwa tidak terdapat multikolinieritas antar variabel.

Berdasarkan persamaan (2.9) dan persamaan (2.10), terdapat nilai kelas sentimen yang dijadikan sebagai referensi untuk pembuatan model. Pada penelitian ini menggunakan sentimen Netral sebagai nilai referensi. Sehingga akan terbentuk dua model persamaan Regresi Logistik Multinomial seperti pada persamaan (4.2) dan persamaan (4.3).

$$\ln \left[\frac{Y=\text{Negatif}}{Y=\text{Netral}} \right] = 0,09 - 0,06 \text{ Layanan} + 0,30 \text{ Pemesanan} - 0,68 \text{ Promosi} \quad (4.2)$$

$$\ln \left[\frac{Y=\text{Positif}}{Y=\text{Netral}} \right] = -0,14 + 0,15 \text{ Layanan} + 0,14 \text{ Pemesanan} + 0,35 \text{ Promosi} \quad (4.3)$$

Parameter model yang terdapat pada persamaan (4.2) dan persamaan (4.3) di atas, diuji secara simultan menggunakan persamaan (2.17) yang mengikuti sebaran *Chi Square*. Nilai statistik uji G, 38,058 dan nilai *p-value* dari statistik uji tersebut adalah $1,09 \times 10^{-6}$, artinya jika dibandingkan dengan tingkat kesalahan 5%, nilai *p-value* sangat kecil sehingga sudah cukup bukti untuk menyatakan bahwa secara bersama-sama, ketiga variabel prediktor yang terdapat dalam model sangat berpengaruh terhadap klasifikasi sentimen. Agar mendapatkan keputusan untuk masing-masing variabel prediktor yang berpengaruh, maka dilakukan uji hipotesis parameter Regresi Logistik Multinomial secara parsial sesuai dengan persamaan (2.20) menggunakan statistik uji *Wald*. Hasil uji hipotesis tersebut ditunjukkan pada Tabel 4.9.

Tabel 4.9. Uji Hipotesis Parameter Regresi Logistik Multinomial Secara Parsial

Sentimen	Prediktor	Koefisien	P-Value	Keputusan	Odd Ratio
Negatif	Intersep	0,09	0,59	Terima H ₀	1,09
	Layanan	-0,06	0,52	Terima H ₀	0,94
	Pemesanan	0,30	0,01	Tolak H ₀	1,34
	Promosi	- 0,68	0,01	Tolak H ₀	0,51
Positif	Intersep	-0,14	0,43	Terima H ₀	0,87
	Layanan	0,15	0,98	Terima H ₀	1,16
	Pemesanan	0,14	0,24	Terima H ₀	1,15
	Promosi	0,35	0,01	Tolak H ₀	1,42

Keterangan : Sentimen Netral sebagai referensi

Hasil uji hipotesis pada Tabel 4.9 mengikuti distribusi *t*. Rumus statistik uji *Wald* pada persamaan (2.20) mengikuti distribusi normal

baku. Distribusi t setara dengan normal baku apabila memiliki derajat bebas yang sangat besar. Oleh sebab itu, perhitungan uji hipotesis parameter pada penelitian ini bisa didekati dengan distribusi normal baku. Agar dapat mengetahui makna dari hasil pengujian, maka perlu diperhatikan nilai p -value. Variabel Pemesanan berpengaruh signifikan pada kategori sentimen Negatif, jika dibandingkan dengan sentimen Netral. Sehingga, rata-rata banyaknya kata yang berkaitan dengan Pemesanan pada sebuah *tweet* di Lazada Indonesia akan memberikan klasifikasi sentimen yang berbeda antara sentimen Negatif dan Netral. Variabel Promosi juga berpengaruh signifikan pada kategori sentimen Negatif. Oleh sebab itu, rata-rata banyaknya kata yang berkaitan dengan Promosi akan memberikan klasifikasi sentimen yang berbeda antara sentimen Negatif dan Netral. Klasifikasikan *tweets* sentimen Positif dipengaruhi oleh variabel Promosi, artinya variabel Promosi dengan nilai tertentu dapat dijadikan acuan dalam mengklasifikasikan *tweets* menjadi sentimen Positif maupun Netral.

Besaran dari pengaruh variabel, dapat dihitung menggunakan persamaan (2.20). Hasil dari persamaan tersebut berupa nilai *odd ratio* yang merupakan tingkat kecenderungan antara variabel Layanan, Pemesanan dan Promosi terhadap klasifikasi sentimen Negatif dan Positif. Nilai dari *odd ratio* masing-masing variabel terdapat pada Tabel 4.9. Nilai *odd ratio* kategori sentimen Negatif untuk variabel Pemesanan sebesar 1,34, artinya penambahan jumlah kata-kata berkaitan dengan Pemesanan meningkatkan kecenderungan sentimen Negatif 1,34 kali dari sentimen Netral. *Odd ratio* untuk variabel Promosi pada kategori sentimen Negatif adalah 0,51. Nilai tersebut dapat diartikan dengan penambahan kata berkaitan dengan variabel Promosi yang terdapat di sebuah *tweets* akan meningkatkan kecenderungan sentimen Negatif 0,51 kali dari sentimen Netral atau meningkatkan kecenderungan sentimen Positif 1,96 kali.

Kategori sentimen Positif untuk variabel Promosi memiliki *odd ratio* sebesar 1,42. Penambahan kata-kata berkaitan dengan variabel Promosi di sebuah *tweets* akan meningkatkan kecenderungan sentimen Positif sebesar 1,42 kali dibandingkan sentimen Netral. Dapat kita artikan bahwa tanggapan pelanggan Lazada Indonesia mengenai variabel Promosi cenderung Positif daripada Netral. Oleh sebab itu, dapat dikatakan pelanggan Lazada Indonesia mengapresiasi Positif hal-hal yang berkaitan dengan Promosi yang diadakan Lazada Indonesia di *twitter*.

Kebaikan model Regresi Logistik Multinomial yang terbentuk dilihat dari nilai AIC sesuai dengan persamaan (2.21). Nilai AIC model saat dipengaruhi oleh variabel Layanan, Pemesanan dan Promosi dibandingkan dengan AIC model yang hanya memiliki dua variabel prediktor dari ketiga variabel prediktor tersebut. Secara rinci ditampilkan pada Tabel 4.10.

Tabel 4.10. Perbandingan Nilai AIC Model Regresi Logistik Multinomial

Variabel Prediktor	<i>Log likelihood</i>	AIC
Layanan, Pemesanan dan Promosi	-602,87	1211,74
Pemesanan dan Promosi	-606,08	1216,16
Layanan dan Promosi	-606,34	1216,68
Layanan dan Pemesanan	-616,49	1236,98

Nilai AIC yang paling kecil dikatakan sebagai model terbaik. Pada Tabel 4.10, nilai AIC terkecil 1211,74 dengan variabel prediktor Layanan, Pemesanan dan Promosi. Oleh sebab itu, ketiga variabel tersebut dapat membentuk model paling bagus dibandingkan dengan dua kombinasi variabel prediktor lainnya.

Hasil prediksi klasifikasi sentimen pada data uji menggunakan model Regresi Logistik Multinomial, seperti Tabel 4.11 menjelaskan saat variabel Layanan memiliki nilai satu, variabel Pemesanan juga bernilai satu dan Promosi nilainya nol, maka *tweets* diklasifikasikan menjadi sentimen Negatif. Pada contoh data terakhir saat variabel Layanan memiliki nilai tiga, variabel Pemesanan nilainya satu dan Promosi nol, maka *tweets* diklasifikasikan menjadi sentimen Negatif. Hasil prediksi dari model Regresi Logistik Multinomial tidak terdapat *tweets* yang diklasifikasikan ke dalam sentimen Netral. Dari 243 data

uji, 111 *tweets* diklasifikasikan menjadi sentimen Negatif dan *tweets* yang diklasifikasikan menjadi sentimen Positif sebanyak 132 *tweets*.

Tabel 4.11. Hasil Prediksi Data Uji Regresi Logistik Multinomial

No	Tweets	Layanan	Pemesanan	Promosi	Sentimen
1	sudah saya dm datanya terimakasih	1	1	0	Negatif
2	sudah hari pesanan saya kok masih dalam proses pengemasan kpn dikirimnya	2	1	0	Positif
3	ga jadi min slow respond banget di tanya nyalebih drjam responnya	1	0	0	Negatif
:	:	:	:	:	:
243	duluuu pernah kayak gini waktu beli di ada promo ultah sekarang terulang lagi saat harbolnas	3	1	0	Negatif

Ketepatan prediksi model Regresi Logistik Multinomial ditunjukkan dari nilai akurasi seperti pada Tabel 4.12.

Tabel 4.12. Akurasi Prediksi Regresi Logistik Multinomial

Model Regresi Logistik Multinomial	Data Aktual				Total
	Sentimen	Negatif	Netral	Positif	
Data Uji	Negatif	44	39	28	111
	Positif	33	63	36	132
Total		77	102	64	243

Berdasarkan Tabel 4.12 dapat disimpulkan secara umum bahwa model Regresi Logistik Multinomial yang terbentuk dari variabel prediktor Layanan, Pemesanan dan Promosi untuk data uji tidak dapat mengklasifikasikan sentimen Netral. Model tidak dapat mengklasifikasikan tweets untuk sentimen Netral dengan tepat. Pada sentimen Negatif hanya terdapat 44 tweets yang diklasifikasikan dengan benar. Terdapat 39 tweets sentimen Negatif di data uji seharusnya masuk ke sentimen Netral dan 28 tweets seharusnya masuk ke sentimen Positif. Tweets sentimen Positif yang diklasifikasikan secara benar hanya 36 tweets, 33 tweets seharusnya diklasifikasikan menjadi sentimen Negatif dan sisanya 63 tweets sentimen Netral. Sehingga, dapat dihitung akurasi hasil prediksi secara keseluruhan sebesar 32,9%, berdasarkan persamaan (4.4).

$$\text{Akurasi} = \frac{44 + 36}{243} = 0,329 \quad (4.4)$$

BAB V PENUTUP

5.1. Kesimpulan

Berdasarkan hasil dari penelitian pemodelan Regresi Logistik Multinomial dari hasil analisis sentimen *tweets* dari pelanggan toko *online* Lazada Indonesia dapat disimpulkan antara lain,

1. Pembentukan klasifikasi sentimen tanggapan pelanggan Lazada Indonesia pada data latih dengan sentimen Negatif sebesar 192, sentimen Netral 164 dan sentimen Positif 213. Sedangkan hasil prediksi data uji menggunakan model *Naive Bayes* yang terbentuk memiliki ketepatan prediksi sentimen Negatif 74 *tweets* dari 79 *tweets*, sentimen Netral 90 *tweets* dari 98 *tweets* dan sentimen Positif 55 *tweets* dari 66. Sehingga akurasi dari hasil prediksi pada data uji 90,1%.
2. Hasil dari analisis sentimen yang dibentuk model menggunakan Regresi Logistik Multinomial dengan variabel prediktor Layanan, Pemesanan dan Promosi. Model Regresi Logistik Multinomial yang terbentuk sebagai berikut.

$$\ln \left[\frac{Y=\text{Negatif}}{Y=\text{Netral}} \right] = 0,09 - 0,06 \text{ Layanan} + 0,30 \text{ Pemesanan} - 0,68 \text{ Promosi}$$

$$\ln \left[\frac{Y=\text{Positif}}{Y=\text{Netral}} \right] = -0,14 + 0,15 \text{ Layanan} + 0,14 \text{ Pemesanan} + 0,35 \text{ Promosi}$$

3. Variabel prediktor yang berpengaruh signifikan pada kategori sentimen Negatif jika dibandingkan sentimen Netral yaitu variabel Pemesanan dan Promosi. Kategori sentimen Positif jika dibandingkan dengan sentimen Netral hanya variabel Promosi yang berpengaruh signifikan. Ketepatan prediksi model Regresi Logistik Multinomial sebesar 32,9%, berdasarkan hasil prediksi pada data uji, tidak terdapat *tweets* yang diklasifikasikan ke dalam sentimen Netral. Hal inilah yang menyebabkan nilai akurasi prediksi jauh dari 100%.

5.2. Saran

Saran untuk penelitian selanjutnya agar membentuk sebuah fungsi di R untuk memisahkan antara *tweets* pelanggan dan Lazada Indonesia. Serta perlu kajian lebih mendalam mengenai penentuan variabel prediktor dan model regresi logistik ordinal disarankan untuk

penelitian selanjutnya. Hasil penelitian ini juga dapat digunakan sebagai bahan pertimbangan untuk meningkatkan respon Lazada Indonesia terhadap proses Pemesanan, karena kecenderungan sentimen pelanggan mengenai Pemesanan lebih besar diklasifikasikan menjadi sentimen Negatif daripada sentimen Netral.



DAFTAR PUSTAKA

- Agresti, A. 2007. *An Introduction to Categorical Data Analysis, Second Edition*. John Wiley and Sons, Inc, New York.
- Berry, M. W. 2003. *Survey of Text Mining*. Springer, New York.
- Chatterjee, S dan Hadi, A. S. 2006. *Regression by Example Fourth Edition*. John Willey & Son, New York.
- Daniel, M., Neves, R. F. dan Horta, N., 2017. “ Company Event Popularity for Financial Markets using Twitter and Sentiment Analysis ”. *Expert System with Application*. Vol 71 hal 111-124.
- Feldman, R. dan Sanger, J. 2006. *The Text Mining Handbook*. Cambridge University Press, United Kingdom.
- Garson, G. D. 2008. *Logistics Regression*. Diakses pada tanggal 20 September 2017 < <https://www.chass.ncsu.edu> >
- Go, A., Bhayani, R. dan Huang, L. 2009. “ Twitter Sentiment Classification Using Distant Supervision “. *Twitter Sentiment Classification Supervision*. Hal 12.
- Hadna, N. M. S., Santosa, P.I dan Winarno, W.W. “Studi Literatur Tentang Perbandingan Metode untuk Proses Analisis Sentimen di Twitter”. *Seminar Nasional Teknologi Informasi dan Komunikasi 2016*. Diakses pada tanggal 24 Oktober 2017 <<https://fti.uajy.ac.id>>.
- Hall, P., Dean, J., Kabul, I. K. dan Silva, J. 2014. *An Overview of Machine Learning with SAS Enterprise Miner*. Diakses pada tanggal 4 September 2017 <<https://support.sas.com>>.
- Hosmer, D. W dan Lemeshow, S. 2000. *Applied Logistic Regression Second Edition*. John Willey & Son, New York.
- Hosmer, D.W dan Lemeshow, S. 2013. *Applied Logistic Regression Third Edition*. John Willey & Son, New York.
- Humas Kementerian Perdagangan Republik Indonesia. 2014. *Panduan Optimalisasi Media Sosial*. Jakarta, Kementerian Perdagangan Republik Indonesia. Diakses pada tanggal 20 September 2017 < <https://www.kemendag.go.id> >

- Indurkhya, N dan Damerau, F. J. 2010. *Natural Language Processing Second Edition*. Taylor dan Francis Group LLC, USA.
- Kasim, H. 2001. *Kiat Memulai dan Mengelola E-Commerce Sendiri*. Elex Media Komputindo, Jakarta,
- Kleinbaum, D.G dan Klein, M. 2002. *Logistic Regression A Self Learning Text Second Edition*. Springer, Jakarta.
- Laudon, K. C. dan Traver, C. G. 2013. *E-Commerce Business Technology Society*. Pearson, USA.
- Liu, B. 2010. "Sentiment Analysis and Subjectivity ". *Handbook of Natural Language Processing*. Diakses pada tanggal 24 Oktober 2017 <<https://www.cs.uic.edu>>.
- McKee, A. 2001. "A Beginner's guide to Textual Analysis ". *Metro Magazine*. Hal 138-149. Diakses pada tanggal 8 September 2017 <<https://www.researchgate.net>>.
- Naif, A. 2017. *Analisis Pengaruh Kepercayaan dan Kualitas Pelayanan terhadap Keputusan Pembelian di Tokopedia.com*. Dalam skripsi Jurusan Manajemen Fakultas Ekonomi Universitas Islam Negeri Maulana Malik Ibrahim.
- Pramadhana, B I., Maharani, W. Dan Kurniati, A. P. 2012. *Analisis Sentimen Berdasarkan Fitur Produk Menggunakan Opinion Lexicon dan Wordnet*. Dalam tugas akhir Jurusan Teknik Informatika Fakultas Teknik Informatika Universitas Telkom.
- Rajeswari, R. P., Kavitha, J. dan Ardhana. 2017. "Classification for Student Data Set Using Naive Bayes Classifier and KNN Classifier ". *International Journal of Computer Trends and Technology (IJCTT)*. Vol. 43 No. 1.
- Ridwan, M., Suyono, H. dan Sarosa, M. 2013. Penerapan *Data Mining* untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naive Bayes Classifier*. *Jurnal EECCIS*. Vol 1. No.7.
- Ting, S. L., Ip, W. H. dan Tsang, A. H. C. 2011. "Is Naive Bayes a Good Classifier for Document Classification". *International Journal of Software Engineering and Its Applications*. Vol. 5. No. 3.

- Yang, S.Y., Mo, S. Y. K. dan Zhu, X. 2014. “An Empirical Study of The Financial Community Network on twitter”. *Computational Intelligent for Financial Engineering & Economics*. Hal 55-62.
- Zuliarso, E dan Mustofa, K. 2009. “Crawling Web Berdasarkan Ontology”. *Jurnal Teknologi Informasi Dinamik*. Vol. 14. No. 2. Hal. 105-112.

