



**ANALISIS SENTIMEN *REVIEW* PELANGGAN *E-COMMERCE* DI  
INDONESIA MENGGUNAKAN ALGORITMA *NAÏVE BAYES*  
*CLASSIFIER***



UNIVERSITAS  
**Dinamika**

**Oleh:**

**Tito Dwiki Darmawan**

**17410100113**

---

---

**FAKULTAS TEKNOLOGI DAN INFORMATIKA**

**UNIVERSITAS DINAMIKA**

**2022**

**ANALISIS SENTIMEN REVIEW PELANGGAN E-COMMERCE DI  
INDONESIA MENGGUNAKAN ALGORITMA *NAÏVE BAYES*  
*CLASSIFIER***

**TUGAS AKHIR**

**Diajukan sebagai salah satu syarat untuk menyelesaikan  
Program Sarjana**



**UNIVERSITAS  
Dinamika**

**Oleh:**

**Nama : Tito Dwiki Darmawan  
NIM : 17410100113  
Program : S1 Sistem Informasi**

**FAKULTAS TEKNOLOGI DAN INFORMATIKA  
UNIVERSITAS DINAMIKA**

**2022**

## Tugas Akhir

# ANALISIS SENTIMEN REVIEW PELANGGAN E-COMMERCE DI INDONESIA MENGGUNAKAN ALGORITMA *NAÏVE BAYES* *CLASSIFIER*

Dipersiapkan dan disusun oleh

**Tito Dwiki Darmawan**

**NIM: 17410100113**

Telah diperiksa, dibahas dan disetujui oleh Dewan Pembahas

Pada: 9 Agustus 2022

### Susunan Dewan Pembahas

#### Pembimbing

- I. Pantjawati Sudarmaningtyas, S.Kom., M.Eng.  
NIDN. 0712066801
- II. Julianto Lemantara, S.Kom., M.Eng.,  
NIDN. 0722108601

#### Pembahas:

Tutut Wuriyanto, M.Kom.  
NIDN. 0703056702

Universitas  
Dinamika  
2022.08.11  
15:04:38 +07'00'

Digitally signed  
by Julianto  
Date: 2022.08.11  
11:28:55 +07'00'

Date:  
2022.08.11  
15:12:30  
+07'00'

Tugas Akhir ini telah diterima sebagai salah satu persyaratan,

Untuk memperoleh gelar Sarjana

Digitally signed by  
Universitas Dinamika  
Date: 2022.08.11  
17:31:30 +07'00'

**Tri Sagirani, S.Kom., M.MT.**

NIDN. 0731017601

Dekan Fakultas Teknologi dan Informatika

UNIVERSITAS DINAMIKA

## SURAT PERNYATAAN

### PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH

Sebagai mahasiswa Universitas Dinamika, saya:

Nama : Tito Dwiki Darmawan  
Nim : 17410100113  
Program Studi : S1 Sistem Informasi  
Fakultas : Fakultas Teknologi dan Informatika  
Jenis Karya : Tugas Akhir  
Judul Karya : **ANALISIS SENTIMEN REVIEW PELANGGAN E-COMMERCE DI INDONESIA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER**

Menyatakan dengan sesungguhnya bahwa:

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, saya menyetujui memberikan kepada Universitas Dinamika Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalti Free Right*) atas seluruh isi/ sebagian karya ilmiah saya tersebut di atas untuk disimpan, diahlimediakan dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama saya sebagai penulis atau pencipta dan sebagai pemilik hak cipta
2. Karya tersebut di atas adalah karya asli saya, bukan plagiat baik sebagian maupun keseluruhan, kutipan karya atau pendapat orang lain yang ada dalam karya ilmiah ini adalah semata hanya rujukan yang dicantumkan dalam Daftar Pustaka saya
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiat pada karya ilmiah ini, maka saya bersedia untuk menerima pencabutan terhadap gelar kerjasama yang telah diberikan kepada saya.

Demikian surat pernyataan ini saya buat dengan sebenarnya

Surabaya, 6 Juli 2022

Yang menyatakan

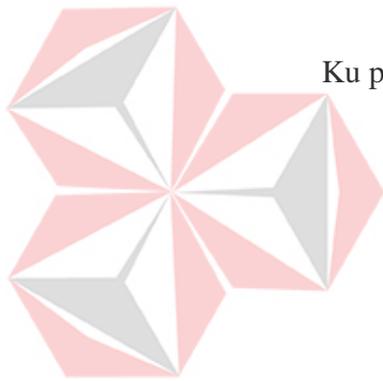


**Tito Dwiki Darmawan**  
NIM: 17410100113



*Lakukanlah Apapun Yang Ingin Kamu Lakukan Dan Jangan Menyesalnya*

UNIVERSITAS  
Dinamika



Ku persembahkan, karya ini untuk orang-orang yang kucintai:

Orangtuaku

dan

Diri Sendiri

UNIVERSITAS  
Dinamika

## ABSTRAK

Berdasarkan lembaga riset asal Inggris yaitu *merchant machine*, didapatkan 10 negara yang memiliki pertumbuhan *e-commerce* tercepat di dunia yaitu pada peringkat pertama adalah Indonesia. berdasarkan survey dari iPrice Group terdapat 11 *e-commerce* di Indonesia dan pada penelitian ini dan dipilih 5 terbesar untuk pengunjung bulanan dari setiap *e-commerce* dengan peringkat pertama yaitu Shopee, Tokopedia, Bukalapak, Lazada, dan Blibli.com. Salah satu komponen penting dalam keberlangsungan bisnis *e-commerce* adalah review dari pelanggan, oleh sebab itu sangat penting bagi bisnis *e-commerce* untuk mengetahui seberapa banyak pelanggan yang memberikan review positif atau negatif. Penelitian ini bertujuan membantu *e-commerce* melakukan analisis sentimen terhadap review pelanggan. Dengan mengetahui sentimen review *e-commerce* di Indonesia, perusahaan dapat melakukan perubahan dan perbaikan untuk meningkatkan layanan kepada pelanggannya. Pengelompokan review pelanggan menjadi sentimen negatif dan positif dilakukan dengan menerapkan metode *naïve bayes classifier* yang menghitung peluang berdasarkan frekuensi pengalaman sebelumnya. Penerapan *naïve bayes classifier* mendapatkan hasil tentang sentimen positif terbesar didapatkan oleh lazada sebesar 97.0%, peringkat kedua bukalapak sebesar 94.6%, peringkat ketiga shopee sebesar 88.5% peringkat keempat blibli.com sebesar 76.1% peringkat kelima tokopedia sebesar 34.4%, lalu hasil akurasi dari metode *naïve bayes* pada lazada sebesar 56.23%, bukalapak sebesar 93.05%, shopee sebesar 87.82%, blibli.com sebesar 55.31%, dan tokopedia sebesar 94.94% Hasil penelitian tersebut telah dievaluasi menggunakan *confusion matrix* dan divalidasi menggunakan *10-fold cross-validation*.

**Kata Kunci:** analisis sentimen, *naïve bayes classifier*, review *e-commerce*.

## KATA PENGANTAR

Segala puji bagi Tuhan yang Maha Esa, Allah SWT atas segala rahmat, karunia dan hidayahNya, sehingga dapat menyelesaikan Laporan Tugas Akhir dengan judul “Analisis Sentimen Review Pelanggan E-Commerce di Indonesia Menggunakan Algoritma Naïve Bayes Classifier” dengan baik dan lancar meskipun penulis sadar bahwa masih banyak kekurangan yang ada didalamnya.

Dalam penyelesaian Tugas Akhir ini, penulis ingin memberikan apresiasi atas bantuannya baik berupa saran, kritik, nasihat maupun dukungan kepada penulis. Oleh karena itu, penulis ingin mengucapkan banyak terima kasih kepada:

1. Tuhan yang Maha Esa, Allah SWT karena selalu mendengarkan doa yang penulis selalu panjatkan supaya diperlancar untuk pengerjaan tugas akhir ini.
2. Orang Tua dan keluarga penulis yang selalu mendoakan dan memberi dukungan kepada penulis
3. Pantjawati Sudarmaningtyas, S.Kom., M.Eng. selaku dosen pembimbing pertama yang selalu sabar dalam memberikan bimbingan dan mendukung dalam menyelesaikan Tugas Akhir.
4. Julianto Lemantara, S.Kom., M.Eng. selaku dosen pembimbing kedua yang selalu sabar dalam memberikan bimbingan dan mendukung dalam menyelesaikan Tugas Akhir.
5. Tutut Wurijanto, M.Kom. selaku dosen pembahas yang telah menyempurnakan laporan tugas akhir ini.
6. Serta kepada Tasya Devi Aryatika yang selalu sabar dan tidak lupa untuk selalu mengingatkan saya jika ada kekurangan saya dalam melakukan pengerjaan Tugas Akhir.
7. dan tidak lupa teman-teman dan kerabat yang tidak bisa disebutkan satu-persatu, saya ucapkan terima kasih banyak.

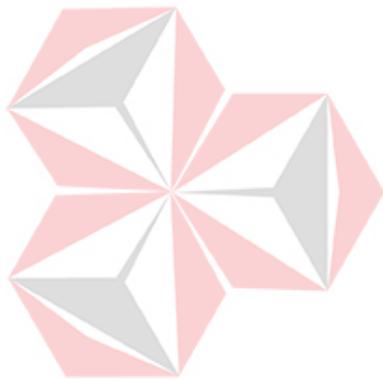
Penulis menyadari bahwa laporan Tugas Akhir ini jauh dari kata sempurna, oleh karena itu penulis ingin meminta maaf sebesar-besarnya jika terdapat banyak kesalahan dalam penulisan laporan. Penulis juga berharap akan masukan, kritik, dan saran dari para pembaca untuk menyempurnakan laporan tugas akhir ini.

Akhir kata saya sampaikan semoga isi dalam laporan ini dapat bermanfaat bagi penulis dan khususnya pada para pembaca.

Surabaya, 18 Juli 2022



Penulis

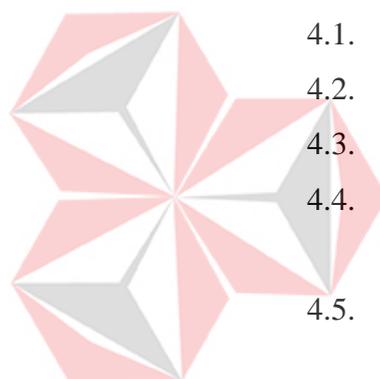


UNIVERSITAS  
**Dinamika**

## DAFTAR ISI

	Halaman
ABSTRAK .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR .....	xiv
DAFTAR LAMPIRAN.....	xv
BAB I : PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan.....	5
1.5 Manfaat.....	5
BAB II : LANDASAN TEORI.....	6
2.1 Penelitian Sebelumnya .....	6
2.2 <i>E-Commerce</i> .....	7
2.3 Analisis Sentimen.....	7
2.4 <i>Web Crawler</i> .....	8
2.4.1. Web Harvy .....	9
2.5 <i>Text Mining</i> .....	9
2.5.1. Text Preprocessing.....	9
2.6 Naïve Bayes.....	10
2.7 <i>Confusion Matrix</i> .....	12
2.7.1. Akurasi.....	13
2.8 <i>RapidMiner</i> .....	13
2.9 Visualisasi.....	14
BAB III : METODOLOGI PENELITIAN .....	16
3.1 Tahap Awal Penelitian .....	17
3.1.1 Studi Literatur .....	17
3.1.2 Teknik Pengumpulan Data.....	17
3.1.3 Jadwal Kerja .....	19

3.2	Tahap Pengembangan.....	19
3.2.1	Analisis .....	19
3.2.2	Penambahan Data .....	20
3.2.3	Pelabelan.....	21
3.2.4	<i>Data Preprocessing</i> .....	21
3.2.5	Data Training dan Data Testing.....	24
3.2.6	Naïve Bayes Classifier.....	24
3.2.7	Evaluasi dan Validasi.....	28
3.2.8	Visualisasi.....	29
3.3	Tahap Akhir Penelitian.....	29
3.3.1.	Kesimpulan .....	29
3.3.2.	Saran .....	30
BAB IV : HASIL DAN PEMBAHASAN .....		31
4.1.	Analisis .....	31
4.2.	Penambahan Data .....	31
4.3.	Pelabelan.....	33
4.4.	Data Preprocessing .....	34
4.4.1.	Frekuensi Kata pada 5 <i>E-Commerce</i> .....	36
4.5.	Implementasi Proses <i>Naïve Bayes</i> .....	37
4.5.1.	Training Data dan Testing Data.....	37
4.5.2	Klasifikasi Naïve bayes .....	38
4.6.	Evaluasi dan Validasi <i>Naïve Bayes</i> .....	39
4.6.1.	Evaluasi Blibli.com.....	39
4.6.2.	Evaluasi Bukalapak.....	40
4.6.3.	Evaluasi Lazada .....	40
4.6.4.	Evaluasi Shopee .....	40
4.6.5.	Evaluasi Tokopedia.....	40
4.6.6.	Hasil Validasi Akurasi <i>Naïve Bayes</i> .....	40
4.7.	Hasil Akhir dari <i>Naïve Bayes</i> .....	41
4.8.	Visualisasi.....	42
4.8.1.	Visualisasi Blibli.com.....	42
4.8.2.	Visualisasi Bukalapak.....	44



4.8.3. Visualisasi Lazada .....	44
4.8.4. Visualisasi Shopee .....	45
4.8.5. Visualisasi Tokopedia.....	45
4.9. Pembahasan .....	45
BAB V : PENUTUP .....	48
5.1. Kesimpulan.....	48
5.2. Saran .....	49
DAFTAR PUSTAKA .....	50
LAMPIRAN.....	54



UNIVERSITAS  
**Dinamika**

## DAFTAR TABEL

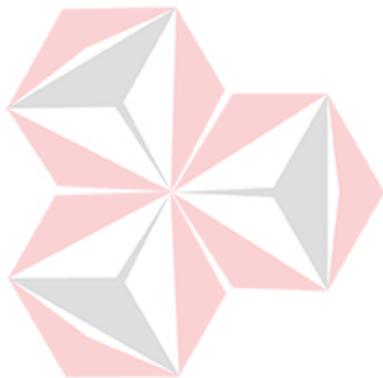
	Halaman
Tabel 1. Penelitian Sebelumnya .....	6
Tabel 2. Confusion Matrix .....	12
Tabel 3. Contoh Data Training Review .....	21
Tabel 4. Contoh Case folding .....	23
Tabel 5. Contoh Tokenizing .....	23
Tabel 6. Contoh Filtering .....	23
Tabel 7. Contoh Stemming .....	24
Tabel 8. Hasil Text Preprocessing .....	26
Tabel 9. Contoh Frekuensi Kata Uji Pada Training Data .....	27
Tabel 10 Contoh Hasil Probabilitas Kata .....	27
Tabel 11. Hasil Klasifikasi .....	28
Tabel 12. Jumlah Review Setiap E-Commerce .....	34
Tabel 13. Hasil Pembagian Training Data dan Testing Data .....	37
Tabel 14. Probabilitas NB Blibli.com .....	39
Tabel 15. Hasil Validasi Akurasi Naïve Bayes .....	40
Tabel 16. Hasil Akhir Naïve Bayes .....	41

## DAFTAR GAMBAR

	Halaman
Gambar 1.1 Pertumbuhan <i>E-Commerce</i> Tercepat .....	1
Gambar 1.2 Pengunjung Situs <i>E-Commerce</i> Tahun 2020 .....	2
Gambar 2.1. <i>Flowchart</i> algoritma <i>Naïve Bayes</i> .....	12
Gambar 3.1 Metode Penelitian.....	16
Gambar 3.2 Tahap Awal .....	17
Gambar 3.3 Rating dan Review <i>E-Commerce</i> dari website Google Play.....	18
Gambar 3.4 Tahap Pengembangan .....	19
Gambar 3.5 <i>Flow Chart Naïve Bayes</i> .....	25
Gambar 3.6 Tahap Akhir.....	29
Gambar 4.1 Contoh Penggunaan <i>RegEx</i> .....	32
Gambar 4.2 Proses Penambangan Data Dengan Web Harvy .....	32
Gambar 4.3 Contoh Hasil Pelabelan .....	33
Gambar 4.4 Tahap Awal <i>Preprocessing</i> .....	34
Gambar 4.5 Tahap <i>Preprocessing</i> .....	35
Gambar 4.6 Frekuensi Kata 5 <i>E-Commerce</i> .....	36
Gambar 4.7 Contoh <i>Training Data</i> dan <i>Testing Data</i> .....	37
Gambar 4.9 <i>Wordcloud</i> Pada Blibli.com .....	42
Gambar 4.10 Data Review <i>E-Commerce</i> .....	43
Gambar 4.11 Persentase Sentimen pada Blibli.com .....	43

## DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Jadwal Kerja .....	54
Lampiran 2. Penambangan Data .....	54
Lampiran 3. Pelabelan.....	55
Lampiran 4. Hasil Data <i>Preprocessing</i> .....	56
Lampiran 5. Hasil <i>Naïve Bayes</i> .....	60
Lampiran 6. Lampiran Visualisasi .....	62
Lampiran 7. Lampiran Daftar Riwayat Hidup .....	70



UNIVERSITAS  
**Dinamika**

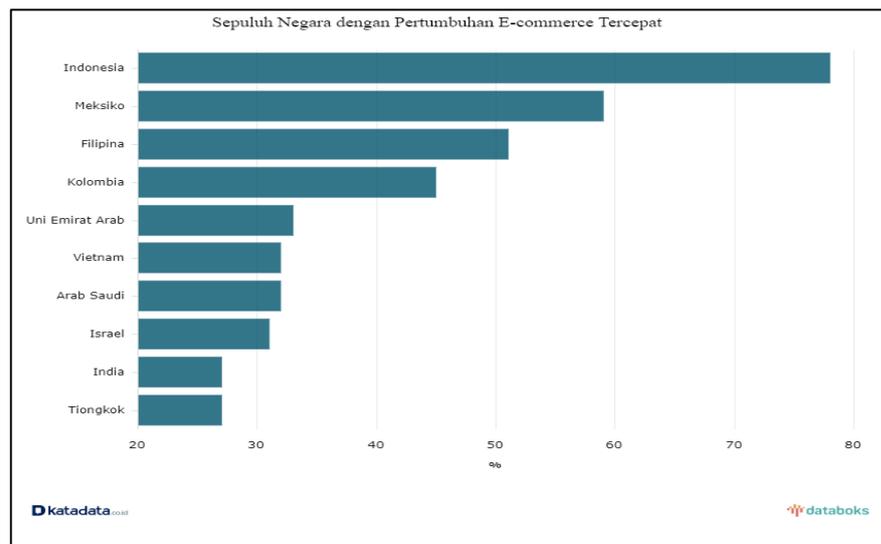
# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Teknologi Informasi adalah suatu hal yang mudah didapatkan pada saat ini, karena teknologi informasi telah masuk ke semua lini dari masyarakat Indonesia. teknologi informasi terus berkembang dan melalui transformasi yang memberikan banyak kemudahan bagi kehidupan bermasyarakat. Kemudahan juga dirasakan dalam dunia ekonomi dan dunia perdagangan. Saat ini Teknologi Informasi bersinergi dengan perdagangan yang tidak mudah lepas, gabungan antara Teknologi Informasi dan Pedagangan maka tercipta istilah *e-commerce* (Hendarsyah, 2019).

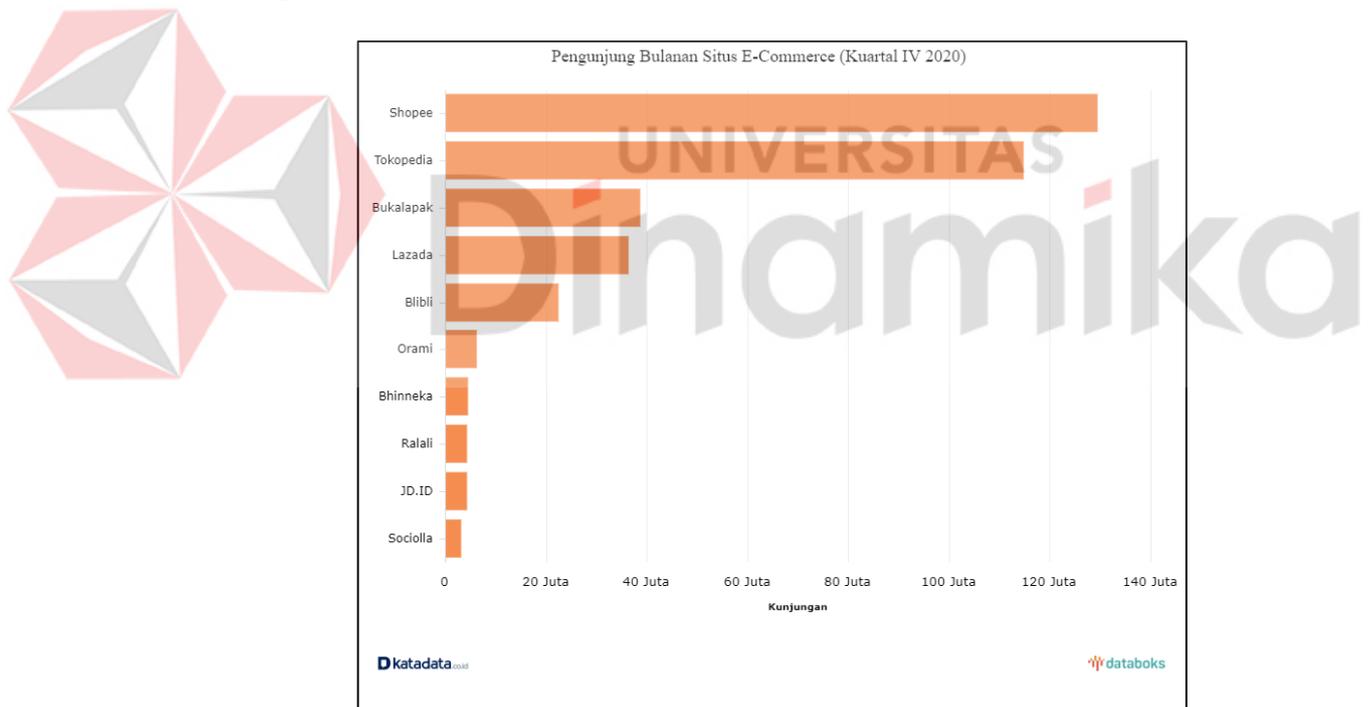
Lembaga riset asal Inggris, *merchant machine*, mengeluarkan sebuah data 10 negara yang memiliki pertumbuhan *e-commerce* tercepat di dunia. Peringkat pertama atas pertumbuhan tercepat di dunia diduduki oleh Indonesia dengan persentase pertumbuhan sebesar 78% pada tahun 2018. Jumlah pengguna internet di Indonesia yang telah mencapai 100 juta pengguna menjadi kekuatan yang mendorong tumbuhnya *e-commerce* di Indonesia (Widowati, 2019). Data tersebut dapat dilihat pada Gambar 1.1.



Gambar 1.1 Pertumbuhan *E-Commerce* Tercepat

Perkembangan *e-commerce* di Indonesia mengalami peningkatan yang sangat pesat, pada tahun 2019 *e-commerce* di Indonesia mengalami peningkatan hingga 500 persen. pernyataan ini juga dibuktikan oleh riset yang dilakukan oleh Google dan termasuk dalam *e-economy* SEA 2018 menunjukkan bahwa ekonomi digital di Indonesia pada tahun 2019 mencapai US\$27 miliar atau setara dengan Rp 391 triliun (Rahayu, 2019).

Menurut Bayu (2020), iPrice Group menobatkan Shopee sebagai *e-commerce* dengan pengunjung situs bulanan terbesar pada kuartal ke 4 tahun 2020. Dengan total pengunjung bulanan Shopee sebanyak 129,3 juta pengunjung. Peringkat selanjutnya sebanyak 114,41 juta pengunjung yaitu Tokopedia selanjutnya Bukalapak sebesar 38,58 juta pengunjung bulanan, Lazada sebesar 36,26 juta dan Blibli sebesar 22,41 juta pengunjung bulanan. Data tersebut dapat dilihat pada Gambar 1.2.



Gambar 1.2 Pengunjung Situs *E-Commerce* Tahun 2020

Kegiatan belanja online melonjak hingga 400% selama pandemi covid-19. Adanya pemberlakuan pembatasan sosial bersekala besar (PSBB) saat pandemi covid-19 menjadi salah satu alasan perilaku konsumen memilih berbelanja secara *online*. Masyarakat menilai belanja secara *online* dapat mengurangi resiko penyebaran virus di tempat ramai. Berbelanja *online* menjadi salah satu solusi untuk

memenuhi kebutuhan sehari-hari (Amelia, Michael, & Mulyadi, 2021). Oleh karena itu seharusnya *e-commerce* sendiri dapat memenuhi ekspektasi dari pelanggannya untuk meningkatkan layanan atau perbaikan dari layanan yang mereka punya agar pelanggan tidak memilih *e-commerce* yang lain dan dapat bersaing dengan yang lain. Karena masih banyak review pelanggan dari beberapa *e-commerce* itu sendiri yang masih belum puas dengan performa atau layanan yang disajikan oleh *e-commerce* yang ada di Indonesia.

Dalam sistem *e-commerce* sendiri, *review* pelanggan merupakan salah satu komponen penting dalam keberlangsungan bisnis perusahaan yang sangat bergantung kepada konsumen atau pelanggan dari *e-commerce* itu sendiri. Maka dari itu menjaga dan mempertahankan konsumen adalah salah satu cara untuk memberikan profit ke perusahaan (Ananda & Yuniawan, 2021). Oleh karena itu, *review* dari pelanggan terutama di media sosial perlu dilakukan analisis sentimen untuk mengetahui tingkat kepuasan pelanggan dari masing – masing *e-commerce*. Agar para *e-commerce* itu sendiri dapat melihat dan mengevaluasi apakah ada yang kurang atau yang perlu di perbaiki dari aplikasi yang mereka punya.

Analisis sentimen atau *opinion mining* yaitu proses memahami, mengekstrak dan mengolah data dari masyarakat tentang suatu produk, jasa ataupun suatu topik secara otomatis yang berdasar pada media sosial atau *website* yang tersedia untuk mendapatkan informasi yang terkandung didalamnya (Buntoro, 2017). Pada penelitian kali ini menggunakan *website* id.priceprice.com dan play.google.com yang diambil mulai dari tahun 2017-2021 dan berdasarkan data diatas dapat diartikan bahwa terdapat 5 (lima) *e-commerce* terpopuler pada saat ini yaitu *Tokopedia*, *Shopee*, *Bukalapak*, *Lazada*, dan *Blibli* yang akan digunakan penulis untuk melakukan penelitian kali ini.

Dalam analisis sentimen terdapat suatu metode *Naïve Bayes* yaitu dibuat untuk mengklasifikasikan data berupa teks. Klasifikasi tersebut digunakan dalam pengelompokan kata berdasarkan kelas pada opini yang sudah diambil. *Naïve Bayes* merupakan teknologi *preprocessing* dalam klasifikasi fitur menambah skalabilitas, akurasi dan efisiensi dari proses klasifikasi teks (Widagdo, Soedijono, & Nasiri, 2020).

Dengan adanya data tersebut, maka perlu adanya solusi untuk mengatasi masalah tersebut. Solusi dari permasalahan tersebut adalah mengetahui *e-commerce* mana yang terbaik dari peringkat 5 (lima) teratas dengan sumber data mengenai *review* pengguna yang bisa didapat melalui *website*. Kemudian data tersebut akan diklasifikasikan menjadi dua jenis klasifikasi yaitu *positive* dan *negative* menggunakan *Naïve Bayes Classifier*. Agar para *e-commerce* dapat memperbaiki kualitas layanan atau performa dari aplikasinya agar pelanggan dapat menikmati aplikasi dari *e-commerce* yang ada di Indonesia.

## 1.2 Rumusan Masalah

Berdasarkan permasalahan diatas maka dapat dirumuskan bahwa masalah yang ada yaitu, bagaimana melakukan analisis sentimen *review* pelanggan *e-commerce* di Indonesia menggunakan algoritma *naïve bayes classifier*.

## 1.3 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

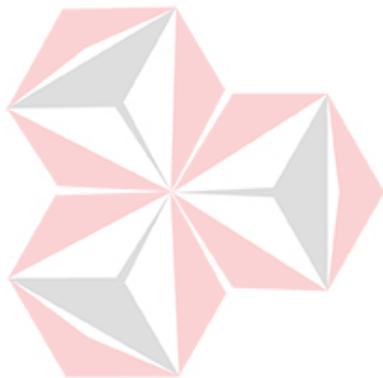
1. Data sekunder yang digunakan pada penelitian ini diambil mulai dari tahun 2017–2021, data tersebut berbasis website [id.priceprice.com](http://id.priceprice.com), dan [play.google.com](http://play.google.com).
2. Analisis ini hanya membahas 5 (lima) *e-commerce* yakni, Shopee, Bukalapak, Tokopedia, Lazada, dan Blibli.
3. Analisis sentimen meliputi pengelompokan negatif dan positif.
4. Tidak membahas analisis sentimen dalam kelompok neutral.
5. Dalam melakukan ekstraksi data, penelitian hanya menggunakan metode text preprocessing pada Text Mining.
6. Proses crawling data menggunakan aplikasi Web Harvy.
7. Proses text preprocessing, penggunaan metode Naïve Bayes, dan Confusion Matrix menggunakan aplikasi RapidMiner.
8. Library yang di masukan pada RapidMiner yaitu Stopwords bahasa Indonesia dan Stemming list bahasa Indonesia.
9. Proses klasifikasi sentimen dilakukan pada review hanya berbahasa Indonesia.

#### **1.4 Tujuan**

Tujuan dari penelitian ini menghasilkan sebuah klasifikasi analisis sentimen pada *review e-commerce* di Indonesia untuk bahan pertimbangan *e-commerce* untuk melihat review dari pelanggan apakah baik atau buruk

#### **1.5 Manfaat**

Berdasarkan analisis sentimen yang dilakukan, diharapkan analisis sentimen dapat digunakan sebagai tolak ukur *e-commerce* untuk melakukan penilaian dari *e-commerce* itu sendiri.



UNIVERSITAS  
**Dinamika**

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Sebelumnya

Tabel 1. Penelitian Sebelumnya

Judul	Penulis	Hasil	Percentage akurasi
Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes	Muljono dan Dian Putri Artanti	Studi kasus: Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes Tools: Testing: Confusion Matrix Sumber data: Twitter Jumlah subjek: E-Commerce: Lazada, Bukalapak, Blibli, dan Tokopedia	Akurasi Naïve Bayes = 93.33% Positive = 95.16% Negative = 91.38%
Analisis Tingkat Kepopuleran E-Commerce Di Indonesia Berdasarkan Sentimen Sosial Media Menggunakan Metode Naïve Bayes	Adika Sri Widagdo	Studi kasus: analisis tingkat kepopuleran e-commerce di indonesia berdasarkan sentimen sosial media menggunakan metode naïve bayes Tools: RapidMiner Testing: Confusion Matrix Sumber data: Twitter E-Commerce: Tidak Disebutkan	Akurasi Naïve Bayes = 97.14% Negative = 99.04% Positive = 96.11%
Implementasi Naïve Bayes Classifier dan Asosiasi Untuk Analisis Sentimen Data Ulasan Aplikasi E-Commerce Shopee Pada Situs Google Play	Zia Ayu Nuansa Gumilang	Studi kasus: implementasi naïve bayes classifier dan asosiasi untuk Analisis sentimen data ulasan aplikasi e-commerce Shopee pada situs google play Tools: RapidMiner Testing: Confusion Matrix Sumber data: Google Play E-Commerce: Shopee	Akurasi Naïve Bayes = 97.4% Negative = 98.2% Positive = 96.4%

Berikut adalah perbedaan penelitian yang peneliti teliti dengan penelitian sebelumnya:

1. Sumber data yang digunakan oleh peneliti dalam penelitian ini adalah data yang di dapat dari dua sumber yaitu id.priceprice.com, dan play.google.com.
2. Data yang digunakan pada penelitian terdahulu adalah *review e-commerce* sebanyak 1 sampai 4 *e-commerce* di Indonesia. Pada penelitian ini data yang digunakan adalah *review 5 (lima) e-commerce* terbesar di Indonesia yakni, Shopee, Tokopedia, Bukalapak, Lazada, dan Blibli dengan menggunakan bahasa Indonesia.
3. Untuk melihat tingkat keakurasian data yang diteliti peneliti menggunakan Rapidminer dengan teknik validasi yang digunakan adalah *10-Fold Cross*

*Validation* dan teknik akurasi algoritma akan diukur menggunakan *Confusion Matrix*.

## 2.2 E-Commerce

*E-Commerce* adalah salah satu alat untuk mengirimkan informasi, produk, layanan, atau pembayaran melalui lini telepon, jaringan komputer atau sarana elektronik lainnya. Dari perspektif layanan, *e-commerce* merupakan suatu alat yang dapat memenuhi keinginan penjual dan pembeli dalam memangkas *service cost* ketika mutu barang dan kecepatan pelayanan ingin ditingkatkan dengan mudah. (Prabowo, 2015).

Menurut Apriadi & Saputra (2017), beberapa kelebihan dari *e-commerce* yang tidak dimiliki oleh transaksi bisnis yang dilakukan secara *offline*, hal tersebut antara lain:

1. Produk: Banyak jenis produk yang bisa dipasarkan dan dijual melalui internet seperti pakaian, kendaraan, barang elektronik, dll.
2. Tempat menjual produk: Tempat menjual yang dimaksud adalah melalui internet yang memerlukan domain dan hosting.
3. Cara menerima pesanan: *Email*, telepon, sms, dll.
4. Cara pembayaran: *Credit card*, *debit card*, *paypal*, dan tunai.
5. Metode pengiriman: Pos Indonesia, JNE, Tiki, EMS.
6. *Customer Service*: *Email*, *contact us*, telepon, *Chatting*.

## 2.3 Analisis Sentimen

Menurut Wibowo & Jumiati (2018), analisis sentimen atau bisa disebut juga dengan *opinion mining* adalah bidang studi yang menganalisis sebuah pendapat, sentimen, sikap, penilaian, evaluasi, emosi orang terhadap suatu produk, organisasi, layanan, masalah, topik, individu, dan atributnya.

Menurut Buntoro (2017), Analisis sentimen atau *opinion mining* yaitu proses memahami, mengekstrak dan mengolah data dari masyarakat tentang suatu produk, jasa ataupun suatu topik secara otomatis yang berdasar pada media sosial atau *website* yang tersedia untuk mendapatkan informasi yang terkandung didalamnya.

Jadi tugas dasar dari analisis sentimen yaitu mengelompokkan teks pada sebuah kalimat atau dokumen untuk menentukan suatu pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif, atau netral. Analisis sentimen juga dapat menyatakan suatu perasaan emosional seperti sedih, gembira, dan marah (Rustiana & Rahayu, 2017).

#### 2.4 Web Crawler

Penambangan data adalah suatu proses penambangan yang menggunakan teknik-teknik tertentu untuk mengekstraksi dan mengidentifikasi informasi yang ada pada database yang besar (Sianturi, Hasugian, Simangunsong, & Nadeak, 2022). *Web Crawler* adalah suatu teknik untuk mempermudah proses penambangan data dari suatu website dengan cara melihat suatu website dan akan muncul indeks yang ada di website tersebut (Napizahni, 2022).

Menurut Amudha (2017), *web crawler* dimulai dengan daftar URL untuk dikunjungi, yang disebut benih. Ketika *crawler* mengunjungi URL ini, ia mengidentifikasi semua *hyperlink* di halaman dan menambahkannya ke daftar URL untuk dikunjungi, yang disebut perbatasan pendataan. URL dari perbatasan adalah dikunjungi secara rekursif sesuai dengan serangkaian kebijakan. Jika *crawler* melakukan pengarsipan situs web, ia menyalin dan menyimpan Informasi. Arsip disimpan secara teratur sedemikian rupa sehingga dapat dilihat, dibaca, dan dinavigasi seperti yang ada di *live web*.

*Web crawler* juga dikenal sebagai laba-laba atau web robot adalah program yang secara otomatis melintasi jumlah besar halaman web dengan mengikuti *Hyperlink*. *Web crawling* digunakan oleh mesin pencari untuk melayani tujuan yang sama, banyak pencarian yang dilakukan untuk merancang sebuah laba-laba (*web crawler*) yang sempurna sehingga proses pencarian dan *crawling* dapat ditingkatkan. Terdapat beberapa mesin pencari yang digunakan untuk berbagai jenis *crawler* web misalnya adalah Google menggunakan GoogleBot, *crawler web* Microsoft menggunakan Msnbot/Bingbot, WebHarvey dan sebagainya (Shrivastava, 2018). Pada penelitian kali ini untuk melakukan *web crawling* akan menggunakan aplikasi Web Harvy yang berguna untuk melakukan crawling data pada 2 (dua) *website* yaitu id.priceprice.com dan www.google.play.com.

### 2.4.1. Web Harvy

Web Harvy adalah sebuah aplikasi yang digunakan untuk melakukan pengambilan data dengan mudah dari situs web. Aplikasi ini juga dapat mengidentifikasi pola data secara otomatis yang ada pada halaman web, lalu data tersebut dapat disimpan ke file dalam berbagai format. Web Harvy juga dapat secara otomatis menampilkan dan mengambil data dari beberapa halaman (Sysnucleus, 2021). Pada penelitian kali ini aplikasi Web Harvy hanya digunakan untuk *crawling* data pada *website* yang sudah ditentukan penulis.

*Web Crawler* juga memiliki tahapan yang harus dilalui yaitu yang pertama adalah menentukan *website* yang akan digunakan untuk proses tersebut, lalu mengecek informasi yang ada pada *website* tersebut yang dirasa perlu digunakan, lalu tahapan terakhir adalah *web crawler* mengecek secara otomatis apakah ada yang tidak boleh di tunjukan dan yang boleh di tunjukan pada *website* yang akan di *crawling* (Patria, 2022).

## 2.5 Text Mining

Menurut Kaushik & Naithani (2016), *text mining* atau penambangan teks adalah sub proses *data mining*, yang secara luas digunakan untuk menemukan pola tersembunyi dan informasi signifikan dari sejumlah besar materi tertulis yang tidak terstruktur. Menjamurnya *cloud*, penelitian dan teknologi bertanggung jawab atas penciptaan volume data. Data semacam ini tidak dapat digunakan sampai atau kecuali informasi atau pola spesifik ditemukan. Penambangan teks menggunakan teknik dari berbagai bidang seperti pembelajaran mesin, visualisasi, penalaran berbasis kasus, analisis teks, database statistik teknologi, manajemen pengetahuan, bahasa alami pemrosesan dan pengambilan informasi.

### 2.5.1. Text Preprocessing

Menurut Rosid et al (2019), metode text preprocessing merupakan tahapan yang sangat penting dalam teknik dan penerapan text mining. Langkah pertama dalam proses penambangan teks membahas empat langkah preprocessing utama, yaitu Case Folding, Tokenizing, Filtering, dan stemming.

#### a. Case Folding

Menurut Izzah & Girsang (2021), *Case Folding* adalah proses penyamaan format huruf dalam sebuah dokumen. Tahapan ini dilakukan karena tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu, diperlukan peran *case folding* dalam mengubah keseluruhan teks dalam suatu dokumen dari huruf kapital menjadi bentuk standar yang biasanya berupa huruf kecil.

b. *Tokenizing*

Menurut Izzah & Girsang (2021), *Tokenizing* adalah proses memisahkan setiap kata yang membentuk dokumen atau percakapan. Proses ini juga menghilangkan tanda baca seperti titik (.), Koma (,), tanda tanya (?), dll.

c. *Filtering*

Menurut Izzah & Girsang (2021), *Filtering* adalah proses menghilangkan kata-kata yang tidak dibutuhkan atau kata-kata yang tidak memiliki arti.

d. *Stemming*

Menurut Aribowo et al (2020), *Stemming* adalah proses mengubah kata-kata yang ditempelkan menjadi kata-kata dasar. Proses stemming dapat menghilangkan makna karena suatu kata dapat memiliki arti yang berbeda setelah ditempelkan. Jadi penelitian ini juga akan mencari tahu apakah stemming diperlukan atau tidak.

## 2.6 Naïve Bayes

Menurut Saleh (2015), *naïve bayes* adalah sebuah algoritma klasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan kombinasi nilai dan frekuensi dari data yang ada. Naive Bayes Classifier adalah klasifikasi dengan model statistik untuk menghitung peluang kelas yang memiliki setiap kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal. Dalam metode ini semua atribut akan berkontribusi pada pengambilan keputusan, dengan atribut pembobotan penting yang sama dan setiap atribut independen satu sama lain (Parlina, et al., 2019). Berikut adalah rumus dari *naïve bayes*:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Dimana:

$X$ : Data dengan class yang belum diketahui

$H$ : Hipotesis data merupakan suatu class spesifik

$P(H/X)$ : Probabilitas hipotesis H berdasarkan kondisi X (*posteritori probabilitas*)

$P(H)$ : Probabilitas hipotesis H (*perior probabilitas*)

$P(X/H)$ : Probabilitas X berdasarkan kondisi pada Hipotesis H

$P(X)$ : Probabilitas X

Pada penelitian ini, algoritma *naïve bayes* disesuaikan dengan kelas yang cocok dengan data yang akan diklasifikasi. Berikut adalah rumus *naïve bayes* yang sudah disesuaikan yaitu:

$$P(W_k|C_i) = \frac{n_k + 1}{n + |Vocabulary|} \quad (2)$$

Dimana:

$P(W_k)$ : Probabilitas (peluang) Kemunculan kata

$C_i$ : Kategori Kelas

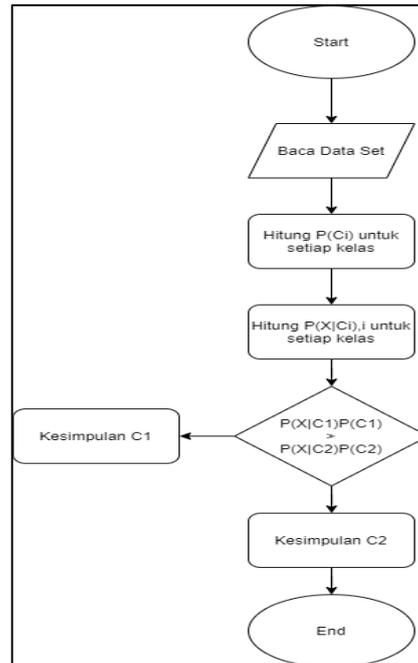
$P(W_k/C_i)$ : Probabilitas kemunculan kata ( $W_k$ ) pada suatu dokumen dengan category kelas ( $C_i$ )

$n_k$ : Nilai Kemunculan kata pada kategori  $C_i$

$n$ : Jumlah keseluruhan kata pada category  $C_i$

*Vocabulary*: Jumlah keseluruhan kata

Dimana alur dari metode *naïve bayes* ini dapat dilihat pada gambar 2.1.



Gambar 2.1. Flowchart algoritma Naïve Bayes

### 2.7 Confusion Matrix

Menurut Holmes, Agius, dan Crompvoets (2020), *confusion matrix* dan indeks yang berasal darinya adalah alat statistik untuk analisis pasangan pengamatan. Ketika tujuannya adalah untuk membandingkan dua data rahasia (dengan proses yang berbeda, berbeda operator, waktu yang berbeda, atau sesuatu yang serupa). *Confusion matrix*, atau bisa disebut juga matriks kesalahan, adalah tabel kemungkinan, yang merupakan alat statistik untuk analisis pengamatan berpasangan. Matriks kebingungan diusulkan dan didefinisikan sebagai ukuran kualitas standar untuk data spasial. Pada Tabel 2 merupakan *confusion matrix* dengan 4 kombinasi nilai prediksi dan nilai aktual yang berbeda.

Tabel 2. Confusion Matrix

TP (True Positive)	FP (False Positive) <i>Type I Error</i>
FN (False Negative) <i>Type II Error</i>	TN (True Negative)

Dimana:

*True Positive* (TP): Merupakan data positif yang diprediksi benar.

*True Negative* (TN): Merupakan data negatif yang diprediksi benar.

*False Positive (FP) – Type I Error*: Merupakan data negatif namun diprediksi sebagai data positif.

*False Negative (FN) – Type II Error*: Merupakan data positif namun diprediksi sebagai data negatif.

### 2.7.1. Akurasi

Akurasi digunakan untuk mengetahui nilai akurasi dari hasil klasifikasi pada penelitian. Berikut adalah rumus dari akurasi *confusion matrix*.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Akurasi = \frac{\text{Jumlah prediksi yang hasilnya benar}}{\text{Total Keseluruhan Jumlah Prediksi}} \quad (4)$$

Menurut Faisal (2016), untuk mengukur baik atau tidaknya algoritma atau teknik klasifikasi dapat dilihat dari akurasi prediksi yang dilakukan oleh algoritma tersebut. Semakin tinggi nilai akurasinya maka semakin baik algoritma yang dijalankan. Jika terdapat dua akurasi dan akurasi yang pertama sebesar 50% dan akurasi yang kedua 70%. Maka akan dipilih akurasi yang kedua karena memiliki akurasi yang paling besar. Sedangkan akurasi pertama sebesar 50% maka itu seperti peluangnya dalam melakukan prediksi benar makin kecil dan peluang prediksi benar hanya sebesar 50%.

## 2.8 RapidMiner

Menurut Silalahi (2018), *RapidMiner* adalah sebuah perangkat lunak yang digunakan untuk mengolah data dengan menggunakan algoritma *data mining*, *RapidMiner* juga mengekstrak pola dari data set yang besar dengan kombinasi metode statistika, kecerdasan buatan, dan *database*.

*RapidMiner* dapat memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan *operator* yang berfungsi untuk memodifikasi data. Data dihubungkan dengan *node* pada *operator*, kemudian pengguna menghubungkan ke *node* hasil untuk dapat melihat hasilnya. Hasil yang diperlihatkan *RapidMiner* dapat ditampilkan secara visual dan grafik,

dengan itu menjadikan *RapidMiner* menjadi salah satu perangkat lunak pilihan untuk melakukan ekstraksi data dengan metode data mining. Pada penelitian kali ini untuk analisis sentimen akan menggunakan aplikasi *RapidMiner* yang digunakan untuk tahap text preprocessing, menganalisis dengan menggunakan menggunakan *Naïve Bayes* serta melakukan pengukuran akurasi dengan menggunakan metode *Confusion Matrix*.

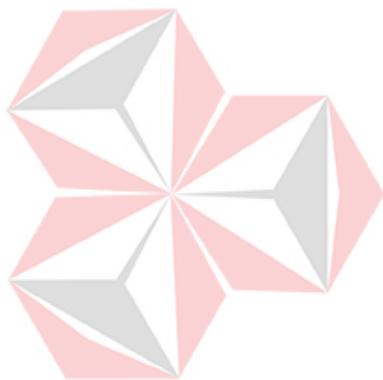
Dalam *RapidMiner* terdapat operator yang berguna untuk melakukan proses yang diperlukan dan terdapat *operator view* yang berguna untuk memudahkan penggunaanya dalam mencari dan menggunakan *operator* yang akan digunakan (Aprilla, Baskoro, Ambarwati, & Wicaksana, 2022). Adapun beberapa kelompok *operator* yang ada pada *RapidMiner* yaitu:

1. *Data Access* : Operator ini berguna untuk mengakses data yang akan digunakan untuk melakukan proses yang ada pada *RapidMiner*.
2. *Process Control / Blending* : Operator ini terdiri dari operator perulangan dan percabangan yang dapat mengatur aliran proses.
3. *Utility* : Operator bantuan, seperti operator *macros*, *loggin*, *subproses*, dan lain-lain.
4. *Extension* : Kelompok ini terdiri dari operator-operator yang dapat digunakan untuk menjalankan sebuah proses yang harus di download pada *repository RapidMiner*.
5. *Modeling* : kelompok ini berisi proses data mining untuk menerapkan model yang dihasilkan menjadi set data yang baru.
6. *Scoring* : kelompok ini berisi operator yang dapat digunakan untuk menghitung kualitas pemodelan dan untuk data baru.
7. *Validation* : Operator ini berisi tentang proses validasi yang diperlukan dalam melakukan proses pada *RapidMiner*.

## 2.9 Visualisasi

Menurut Hayadi (2017), visualisasi adalah bentuk penyampaian informasi yang berguna untuk menjelaskan sesuatu dengan gambar, diagram, ataupun animasi yang bisa dihitung, dieksplor, dan dianalisis datanya. Visualisasi juga merupakan upaya manusia dalam mendeskripsikan maksud tertentu menjadi sebuah bentuk informasi yang mudah dipahami.

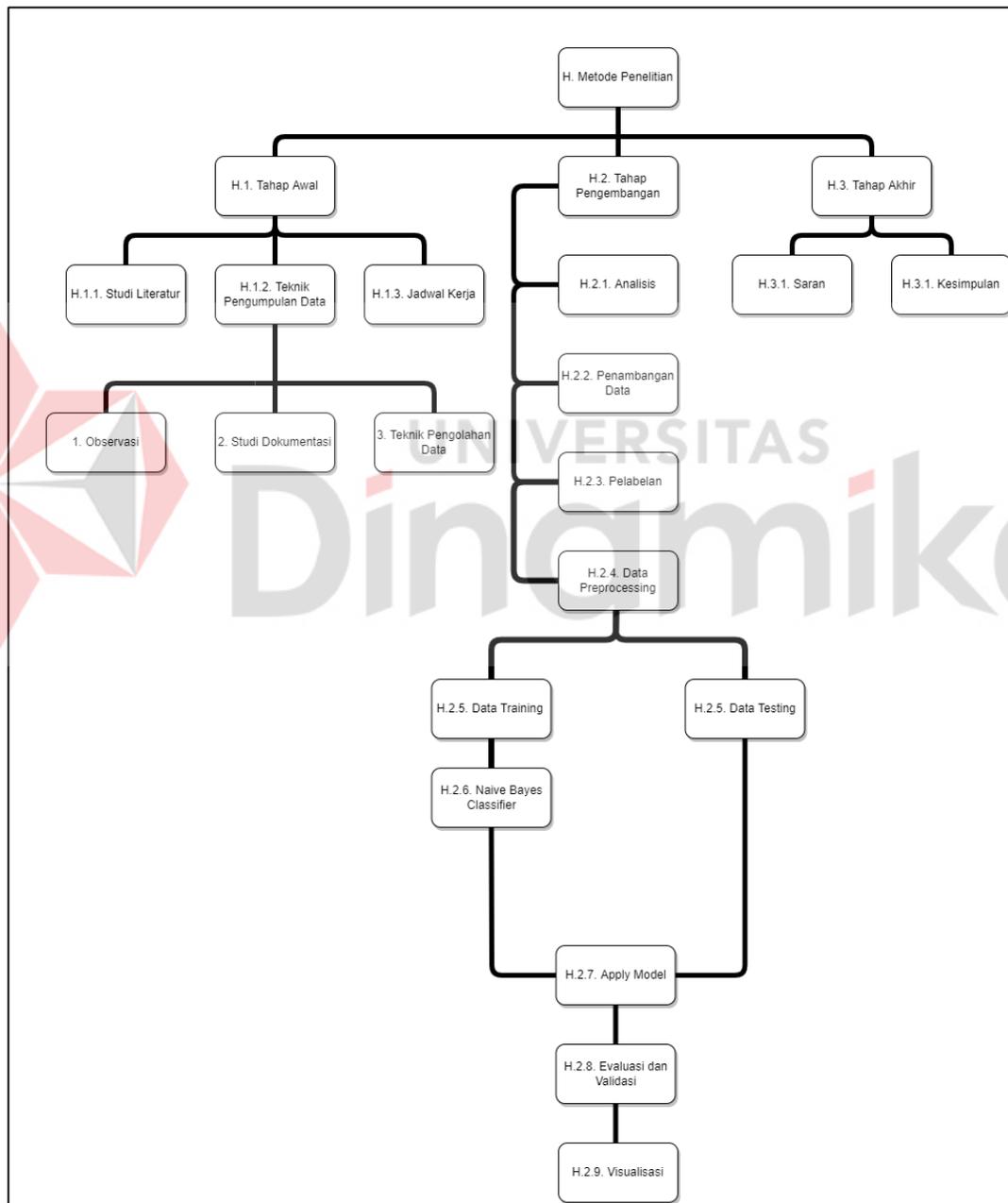
Visualisasi yang dilakukan digunakan penulis untuk mempermudah menjelaskan data yang akan dihasilkan. Agar para *e-commerce* dapat melihat dengan jelas dan tidak perlu banyak penjelasan untuk tau informasi atau data apa yang akan ditunjukkan.



UNIVERSITAS  
Dinamika

### BAB III METODOLOGI PENELITIAN

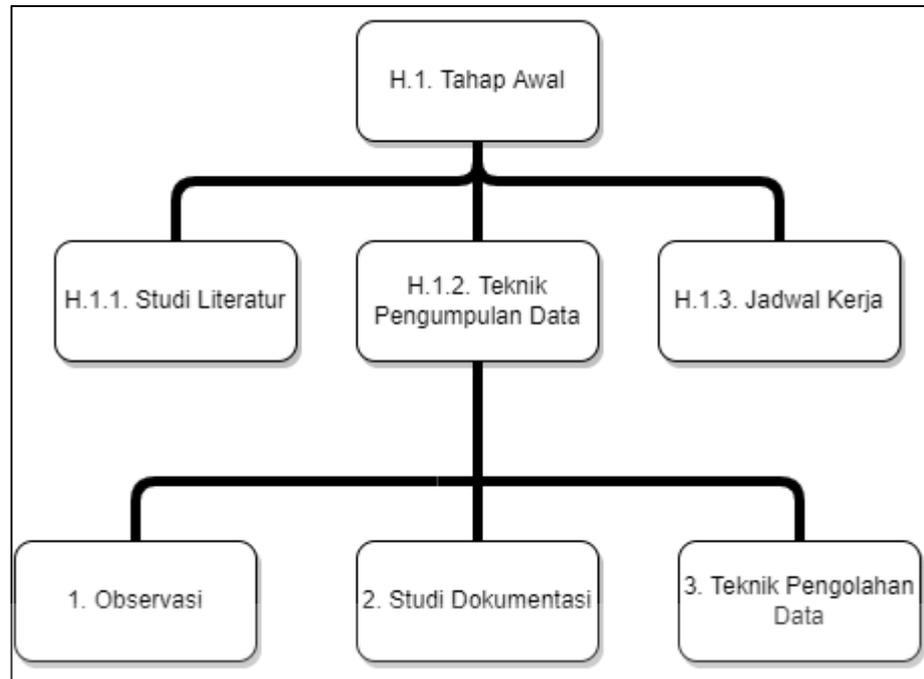
Pada tahap ini menjelaskan tentang tahapan-tahapan yang dilakukan untuk menyelesaikan penelitian. Pada penelitian ini menggunakan tiga tahapan yaitu, Tahap Awal, Tahap Pengembangan, dan Tahap Akhir dapat dilihat pada gambar 3.1.



Gambar 3.1 Metode Penelitian

### 3.1 Tahap Awal Penelitian

Tahapan Awal yang dilakukan pada penelitian ini yaitu:



Gambar 3.2 Tahap Awal

Gambar 3.2 merupakan tahapan awal penelitian yang penulis lakukan dalam penelitian ini. Rincian dari tahapan awal adalah sebagai berikut:

#### 3.1.1 Studi Literatur

Studi literatur merupakan tahap pertama dari metode penelitian pada penelitian kali ini. Pada studi literatur yang perlu dilakukan adalah *me-review* penelitian-penelitian sebelumnya dalam bentuk Tugas Akhir (TA), Jurnal, dsb. Setelah selesai dalam melakukan *review* pada penelitian yang sudah ada, penulis dapat menentukan studi kasus dan tahapan-tahapan penyelesaian yang digunakan untuk melakukan penelitian.

#### 3.1.2 Teknik Pengumpulan Data

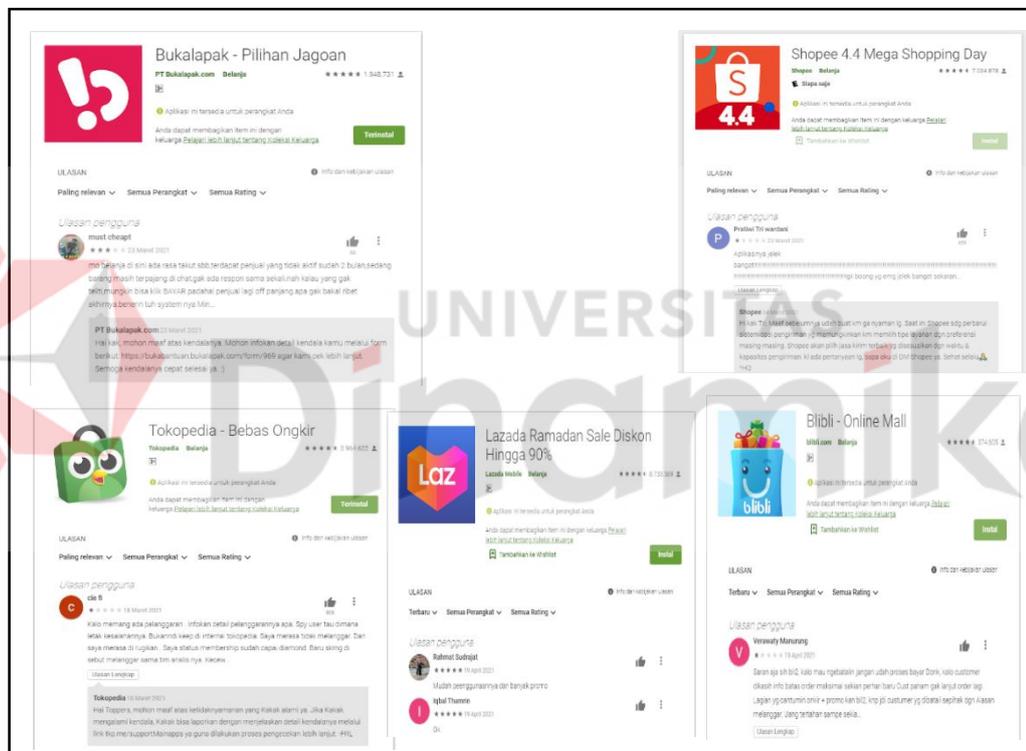
##### 1. Observasi

Observasi merupakan proses kegiatan merencanakan, memperoleh dan menyediakan informasi guna untuk memperoleh data atau informasi yang diperlukan untuk membuat suatu alternatif terhadap penelitian. Dalam penelitian ini dilakukan dengan mengamati *review* masyarakat dari *platform website* [id.priceprice.com](http://id.priceprice.com) dan [play.google.com](http://play.google.com) terdapat banyak macam *review* yang

diberikan oleh pengguna pada aplikasi *e-commerce* yang membuat penulis melakukan penelitian dalam persentase klasifikasi review masyarakat terhadap *e-commerce* yang ada di Indonesia pada *website* [id.priceprice.com](http://id.priceprice.com) dan [play.google.com](http://play.google.com).

## 2. Studi Dokumentasi

Studi Dokumentasi merupakan kegiatan mencari data atau informasi dari berbagai sumber, seperti buku yang memuat berbagai ragam teori yang di butuhkan pada penelitian ini. Dalam studi dokumentasi ini review dari *e-commerce* Bilibli.com, Bukalapak, Lazada, Shopee dan Tokopedia. Data Pada *Google Play* review dapat dilihat pada gambar 3.3.



Gambar 3.3 Rating dan Review *e-commerce* dari *website* *Google Play*

## 3. Teknik Pengolahan Data

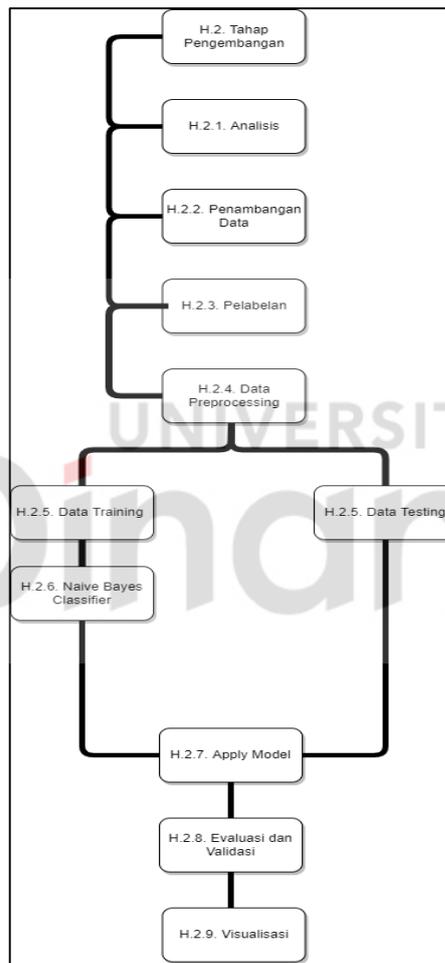
Data awal yang dibutuhkan pada penelitian ini yaitu link dari *website* [id.priceprice.com](http://id.priceprice.com) dan [play.google.com](http://play.google.com) yang akan dijadikan sebagai objek. Selanjutnya untuk mengambil data dari link tersebut menggunakan *tools* WebHarvey, dan data yang diambil hanya review dalam berbahsa Indonesia.

### 3.1.3 Jadwal Kerja

Bagian ini menjelaskan pendjadwalan pengerjaan penelitian dari tahap awal, tahap pengembangan dan tahap akhir. Untuk lebih jelasnya, jadwal kerja terdapat pada tabel L.1 dibawah ini.

### 3.2 Tahap Pengembangan

Tahapan ini menjelaskan tentang tahap pengembangan untuk menyelesaikan penelitian dapat dilihat pada gambar 3.4.



Gambar 3.4 Tahap Pengembangan

#### 3.2.1 Analisis

Tahap analisis bertujuan untuk mengetahui kebutuhan yang di butuhkan dalam penelitian ini. Analisis yang dilakukan dalam penelitian ini terdiri dari 2 (dua) subbab, yaitu identifikasi masalah, dan analisis kebutuhan sistem.

a. Identifikasi Masalah

Identifikasi masalah digunakan sebagai dasar dalam melakukan penelitian. Identifikasi masalah dilakukan dengan mempelajari studi literatur yang serupa dengan penelitian yang dikerjakan, melakukan observasi melalui *website* terkait guna mendapatkan informasi yang digunakan untuk mengetahui *review* atau pendapat dari masyarakat. Dari masalah tersebut, dibutuhkan analisis yang dapat membantu masyarakat untuk mengetahui *review* atau pendapat masyarakat terhadap *e-commerce* di Indonesia.

b. Analisis Kebutuhan Sistem

1) Kebutuhan Perangkat Lunak

Kebutuhan perangkat lunak adalah komponen perangkat lunak yang akan digunakan dalam penelitian ini. Adapun perangkat lunak yang digunakan adalah sebagai berikut:

- i. Tools yang digunakan dalam melakukan *Web Scrapping* untuk website *id.priceprice.com* dan *play.google.com* adalah WebHarvy. Webharvey sangat mudah digunakan dan tidak memerlukan kode untuk penggunanya.

2) Kebutuhan Perangkat Keras

Kebutuhan perangkat keras merupakan komponen dari peralatan fisik yang membentuk suatu sistem komputer terstruktur, yang di dalamnya terdapat beberapa perangkat keras lain yang mendukung jalannya suatu sistem. Adapun kebutuhan perangkat keras yang digunakan adalah sebagai berikut:

- i. Pada penelitian ini menggunakan *Central Processing Unit (CPU) Intel Core I5-5200U*.
- ii. Pada penelitian ini menggunakan *Random Access Memory (RAM)* sebesar 12GB.
- iii. Pada penelitian ini menggunakan *Graphic Processing Unit (GPU) NVIDIA GeForce 820 M*.

### 3.2.2 Penambahan Data

Penambahan data dilakukan untuk mengambil data yang diperlukan dalam penelitian kali ini. Pada penambahan data ini menggunakan *software* WebHarvy

untuk *website* id.priceprice.com dan play.google.com. Adapun cara dari penggunaan *tools* WebHarvy yaitu:

1. Mencari *website* yang sudah ditentukan dengan menggunakan URL dari *website* tersebut *website* tersebut yaitu id.priceprice.com dan play.google.com
2. Lalu dilanjutkan dengan memilih text atau data yang akan di ambil untuk mendukung penelitian ini. yaitu disini saya memilih data *rating*, *review*, penulis, dan tanggal. Pada saat mengambil data rating diperlukan RegEx yang berguna untuk merubah *icon* bintang mengambil kalimat dari aria-label.
3. Menambang data yang akan dipakai dan merubah data tersebut menjadi bentuk data excel.
4. Hasil Penambangan berupa 5 sumber yang berisi review dari Blibli.com, Bukalapak, Lazada, Shopee, dan Tokopedia.

### 3.2.3 Pelabelan

Pelabelan atau penentuan kelas atribut diberikan sesuai dengan subjektifitas peneliti. Dengan cara pengambilan data review yang telah di ambil dari *website* dan pembagian kelas yaitu sentimen positif untuk review *e-commerce* yang bersifat positif dan sentimen negatif untuk review *e-commerce* yang bersifat negatif secara manual tanpa bantuan aplikasi. Contoh dari tahap pelabelan dapat dilihat pada tabel 3 dibawah ini.

Tabel 3. Contoh Data Training Review

Kategori	Review
Negatif	Aplikasi buruk, tidak bisa dipercaya barang lama sampainya...
	Lazada ini gimana kok barang saya baru sampai setelah satu bulan
Positif	Aplikasi lazada ini sangat bagus untuk berbelanja online, dan bisa dipercaya.
	Lazada adalah tempat belanja online terbaik yang pernah ada.!!
	Sangat terpercaya, jika berbelanja di aplikasi ini

### 3.2.4 Data Preprocessing

Tahapan selanjutnya yang dilakukan pada Tahap Pengembangan yaitu *Data Preprocessing*. Tahap *Data Preprocessing* yang dilakukan dalam tahap ini adalah melakukan ekstraksi data. Ekstraksi data bertujuan untuk mengelolah hasil penambangan data yang formatnya tidak terstruktur hingga menjadi format kata dasar. Pada *Data Preprocessing* bertujuan untuk mengolah hasil penambangan data yang tidak terstruktur hingga menjadi format kata dasar. Pada *Data Preprocessing*

pada penelitian kali ini terdiri dari 4 (empat) tahap yaitu, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Tahapan yang dilakukan RapidMiner dalam *Data Preprocessing* yaitu sebagai berikut:

### 1. *Import data penambangan*

Tahapan pertama dalam melakukan penelitian ini yaitu memasukan data hasil penambangan ke *Rapid Miner* yang berguna untuk melakukan *Text Preprocessing*.

### 2. *Proses Text Preprocessing*

Pada tahap *text preprocessing* data yang sudah di masukan ke *Rapid Miner* operator yang di butuhkan adalah sebagai berikut:

1. *Retrieve* data yang berguna untuk memasukan data kedalam model yang akan dijalankan data yang dimasukan adalah data yang telah melakukan tahap pelabelan.
2. Seleksi atribut yang digunakan untuk menyeleksi atribut yang digunakan sebagai data text dalam melakukan *text preprocessing*.
3. Merubah tipe data yang menjadi text dengan menggunakan operator *nominal to text* bila tipe data yang terpilih bukan tipe data text.
4. *Text preprocessing*, dalam proses *text preprocessing* yang berisi
  - *case folding* : adalah tahap yang digunakan untuk mengubah huruf kapital menjadi huruf kecil semua.
  - *tokenizing* : adalah tahap yang digunakan untuk menghapus karakter atau simbol selain huruf A sampai Z.
  - *Filtering* : yaitu menghapus kata-kata yang tidak perlu seperti kata sambung, kata ganti dan kata preposisi.
  - *Stemming* : adalah tahap yang digunakan untuk merubah kata berimbuhan menjadi kata dasarnya.

Juga pada tahap ini terdapat operator *open file* yang berisi rumus sastrawi yang ada di internet digunakan dalam tahap *filtering* dan *stemming* yang dijelaskan prosesnya pada

### 3. *Hasil Text Preprocessing*

Hasil dari *text preprocessing* yaitu data review yang sudah bagus di export menjadi excel dengan 5 sumber yaitu Blibli.com, Bukalapak, Lazada, Shopee,

dan Tokopedia. Adapun frekuensi kata yang muncul pada review tersebut juga terdapat 5 sumber. Contoh hasil dari text preprocessing akan di jelaskan dibawah ini.

### 1. Case Folding

Contoh hasil *Case Folding* dapat dilihat pada tabel 4.

Tabel 4. Contoh *Case folding*

Sebelum	Sesudah
Aplikasi buruk, tidak bisa dipercaya barang lama sampainya...	aplikasi buruk tidak bisa dipercaya barang lama sampainya...
Lazada ini gimana kok barang saya baru sampai setelah satu bulan	lazada ini gimana kok barang saya baru sampai setelah satu bulan
Aplikasi lazada ini sangat bagus untuk berbelanja online, dan bisa dipercaya.	aplikasi lazada ini sangat bagus untuk berbelanja online, dan bisa dipercaya.
Lazada adalah tempat belanja online terbaik yang pernah ada.!!	lazada adalah tempat belanja online terbaik yang pernah ada.!!
Sangat terpercaya, jika berbelanja di aplikasi ini	sangat terpercaya, jika berbelanja di aplikasi ini

### 2. Tokenizing

Contoh Hasil *Tokenizing* dapat dilihat tabel 5.

Tabel 5. Contoh *Tokenizing*

Sebelum	Sesudah
aplikasi buruk tidak bisa dipercaya barang lama sampainya...	aplikasi buruk tidak bisa dipercaya barang lama sampainya
lazada ini gimana kok barang saya baru sampai setelah satu bulan	lazada ini gimana kok barang saya baru sampai setelah satu bulan
aplikasi lazada ini sangat bagus untuk berbelanja online, dan bisa dipercaya.	aplikasi lazada ini sangat bagus untuk berbelanja online dan bisa dipercaya
lazada adalah tempat belanja online terbaik yang pernah ada.!!	lazada adalah tempat belanja online terbaik yang pernah ada
sangat terpercaya, jika berbelanja di aplikasi ini	sangat terpercaya jika berbelanja di aplikasi ini

### 3. Filtering

Contoh hasil *Filtering* dapat dilihat pada tabel 6.

Tabel 6. Contoh *Filtering*

Sebelum	Sesudah
aplikasi buruk tidak bisa dipercaya barang lama sampainya	aplikasi buruk tidak bisa dipercaya barang lama sampainya
lazada ini gimana kok barang saya baru sampai setelah satu bulan	lazada barang saya baru sampai satu bulan
aplikasi lazada ini sangat bagus untuk berbelanja online dan bisa dipercaya	aplikasi lazada sangat bagus berbelanja online bisa dipercaya

lazada adalah tempat belanja online terbaik yang pernah ada	lazada tempat belanja online terbaik pernah ada
sangat terpercaya jika berbelanja di aplikasi ini	sangat terpercaya berbelanja aplikasi
aplikasi buruk tidak bisa dipercaya barang lama sampainya	aplikasi buruk tidak bisa dipercaya barang lama sampainya

#### 4. Stemming

Contoh hasil *Stemming* dapat dilihat pada tabel 7.

Tabel 7. Contoh *Stemming*

Sebelum	Sesudah
aplikasi buruk tidak bisa dipercaya barang lama sampainya	aplikasi buruk tidak bisa percaya barang lama sampai
lazada barang saya baru sampai satu bulan	lazada barang saya baru sampai satu bulan
aplikasi lazada sangat bagus berbelanja online bisa dipercaya	aplikasi lazada sangat bagus belanja online bisa percaya
lazada tempat belanja online terbaik pernah ada	lazada tempat belanja online baik pernah ada
sangat terpercaya berbelanja aplikasi	sangat percaya belanja aplikasi

lalu setelah melakukan *text preprocessing* selanjutnya dilakukan pembagian data secara manual menjadi *data training* dan *data testing* dengan cara manual.

#### 3.2.5 Data Training dan Data Testing

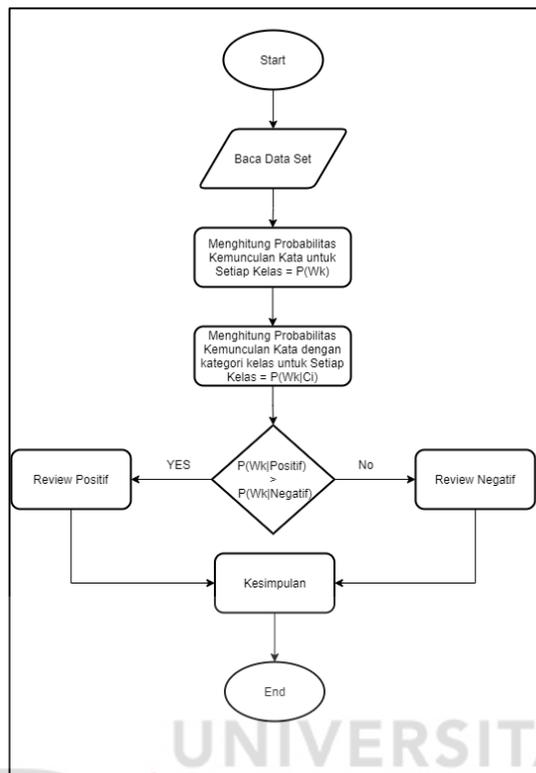
Bagian ini untuk proses training atau pelatihan data dan untuk proses testing atau pengujian. Pada data training, data yang digunakan harus sudah terlebih dahulu dilabeli dan juga harus melalui proses *preprocessing* untuk dijadikan input dari penggunaan metode *Naïve Bayes Classifier*.

Pada proses ini menggunakan data yang telah dilakukan *Data Preprocessing* lalu dilakukan pembagian data secara manual tanpa bantuan aplikasi dengan cara membagi data dalam bentuk *excel* untuk *data training* dan *data testing* secara manual

#### 3.2.6 Naïve Bayes Classifier

Tahap pengembangan selanjutnya adalah melakukan klasifikasi data menggunakan *Naïve Bayes Classifier* yang berguna untuk melakukan klasifikasi data dengan cara menghitung setiap kata dari kalimat apakah kata kalimat tersebut termasuk kalimat negatif atau positif. Klasifikasi data pada penelitian ini memiliki fungsi untuk menentukan class pada data yaitu *Positive* dan *Negative*. Dibawah ini

adalah flowchart *naïve bayes* untuk klasifikasi kalimat negatif dan positif dapat dilihat pada gambar 3.5.



Gambar 3.5 Flow Chart Naïve Bayes

Tahapan atau model yang akan dilakukan *RapidMiner* dalam melakukan analisis sentimen adalah sebagai berikut:

1. *Import Training Data dan Testing Data*

Tahap ini yaitu melakukan *import data training* dan *data testing* dari 5 sumber yang telah melalui tahap *Preprocessing* dan tahap pembagian data yang dilakukan secara manual pada excel dan dimasukkan kedalam *RapidMiner*.

2. *Proses Naive Bayes*

Pada proses *naïve bayes* hal yang dilakukan yaitu memasukan operator atau tools yang diperlukan dalam menjalankan analisis sentimen, adapun operator tersebut seperti:

- a. *Retrieve training data dan testing data* yang berguna untuk menampilkan data yang akan di uji untuk melakukan analisis sentimen.
- b. *Set role* yang berguna untuk melakukan menentukan role yang akan dirubah dari kelas menjadi label.

- c. *Validation* yang berguna untuk menghitung sebanyak 10 kali untuk memastikan akurasi yang dijalankan dan didalamnya terdapat model yang akan dilakukan.
  - d. *Naïve Bayes* yaitu berguna untuk melakukan analisis sentimen yang akan menghitung review dari 5 sumber yang telah disediakan.
3. Evaluasi Naive Bayes
- Pada tahap evaluasi *naïve bayes* hal yang dilakukan yaitu melakukan evaluasi yang memerlukan operator sbagai berikut
- a. *Apply Model* yang berguna untuk menjalankan menerapkan suatu model yang telah di lakukan.
  - b. *Performance* berguna untuk menghitung akurasi dari model yang dilakukan.

Di bawah ini adalah contoh proses dari proses *naïve bayes* yang akan dilakukan.

Setelah melakukan tahap *Preprocessing* maka berikut adalah hasil dari *Text Preprocessing* review pelanggan yang telah dilakukan diatas seperti tabel 8.

Tabel 8. Hasil *Text Preprocessing*

Kategori	Review	Jumlah Kata
Negatif	Aplikasi buruk tidak bisa percaya barang lama sampai	8
	Lazada barang saya baru sampai satu bulan	7
Positif	Aplikasi lazada sangat bagus belanja online bisa percaya	8
	Lazada tempat belanja online baik pernah ada	7
	Sangat percaya belanja aplikasi	4

Setelah melakukan *text preprocessing* maka dilakukan perhitungan probabilitas *review* positif dan negatif. Dengan total *review* sebanyak 5 maka probabilitas *review* positif sebesar  $P(\text{positif}) = \frac{3}{5} = 0.6$  dan probabilitas *review* negatif sebesar  $P(\text{negatif}) = \frac{2}{5} = 0.4$ . Setelah itu dilakukan pembagian setiap kata yang ada pada *review*. Dari data diatas diketahui jumlah kata yang terdapat pada data tersebut sebanyak 34 kata dan masing – masing kata dari *review* positif sebanyak 19 sedangkan *review* negatif sebanyak 15 kata. Selanjutnya dihitung setiap kata dengan rumus *Naïve Bayes* sesuai dengan frekuensi kata pada training data.

Tabel 9. Contoh Frekuensi Kata Uji Pada *Training Data*

No	Kata	<i>Training Negative</i>	<i>Training Positive</i>
1	Aplikasi	1	2
2	Buruk	1	0
3	Tidak	1	0
4	Bisa	1	1
5	Percaya	1	2
6	Barang	2	0
7	Lama	1	0
8	Sampai	2	0
9	Lazada	1	1
10	Saya	1	0
11	Baru	1	0
12	Satu	1	0
13	Bulan	1	0
14	Sangat	0	2
15	Bagus	0	1
16	Belanja	0	2
17	Online	0	2
18	Tempat	0	1
19	Baik	0	1
20	Pernah	0	1
21	Ada	0	1
22	Sangat	0	1
23	Percaya	0	2

Dari data tabel 9 tersebut mempunyai nilai yang dapat dicari nilai probabilitas dengan menggunakan rumus  $P(W_k|C_i)$  yaitu sebagai berikut:

$$P(\text{Aplikasi}|\text{negatif}) = (1+1) / (15+34) = 0.04082$$

$$P(\text{Aplikasi}|\text{positif}) = (2+1) / (19+34) = 0.05661$$

$$P(\text{Barang}|\text{negatif}) = (2+1) / (15+34) = 0.06122$$

$$P(\text{Barang}|\text{positif}) = (0+1) / (19+34) = 0.01887$$

$$P(\text{Bagus}|\text{negatif}) = (0+1) / (15+34) = 0.02041$$

$$P(\text{Bagus}|\text{positif}) = (1+1) / (19+34) = 0.03773$$

Hasil dari Perhitungan  $P(W_k|C_i)$  dapat dilihat sebagai berikut:

Tabel 10 Contoh Hasil Probabilitas Kata

No	Kata	$N_k \text{Negatif}$	$N_k \text{Positif}$	$P(W_k C_i)$ <i>Negatif</i>	$P(W_k C_i)$ <i>Positif</i>
1	Aplikasi	1	2	0.04082	0.05661
2	Buruk	1	0	0.04082	0.01887
3	Tidak	1	0	0.04082	0.01887
4	Bisa	1	1	0.04082	0.03773
5	Percaya	1	2	0.04082	0.05661
6	Barang	2	0	0.06122	0.01887
7	Lama	1	0	0.04082	0.01887
8	Sampai	2	0	0.06122	0.01887
9	Lazada	1	1	0.04082	0.03773
10	Saya	1	0	0.04082	0.01887

No	Kata	$N_k Negatif$	$N_k Positif$	$P(W_k C_i)$ Negatif	$P(W_k C_i)$ Positif
11	Baru	1	0	0.04082	0.01887
12	Satu	1	0	0.04082	0.01887
13	Bulan	1	0	0.04082	0.01887
14	Sangat	0	2	0.02041	0.05661
15	Bagus	0	1	0.02041	0.03773
16	Belanja	0	2	0.02041	0.05661
17	Online	0	2	0.02041	0.05661
18	Tempat	0	1	0.02041	0.03773
19	Baik	0	1	0.02041	0.03773
20	Pernah	0	1	0.02041	0.03773
21	Ada	0	1	0.02041	0.03773
22	Sangat	0	1	0.02041	0.03773
23	Percaya	0	2	0.02041	0.05661

Berdasarkan nilai probabilitas pada tabel 10 maka dapat diketahui probabilitas perkata. Jika ingin melihat hasilnya maka akan dilakukan dengan data testing yang akan digunakan untuk test model yang sudah di training sebelumnya

Tabel 11. Hasil Klasifikasi

Review Data Testing	Lazada Aplikasi Belanja Online Bisa Percaya
Training Negatif	0.00000000023 / $2.3 \times 10^{-10}$
Training Positif	0.00000000877 / $8.77 \times 10^{-9}$
Kategori	Positif

Hasil tabel 11 diketahui sebagai berikut:

Negatif:  $p(\text{negatif}) \times p(\text{lazada}|\text{negatif}) \times p(\text{aplikasi}|\text{negatif}) \times p(\text{belanja}|\text{negatif}) \times p(\text{online}|\text{negatif}) \times p(\text{bisa}|\text{negatif}) \times p(\text{percaya}|\text{negatif})$ .

Negatif:  $0.4 \times 0.04082 \times 0.04082 \times 0.02041 \times 0.02041 \times 0.04082 \times 0.02041 = 0.00000000023$

Positif:  $p(\text{positif}) \times p(\text{lazada}|\text{positif}) \times p(\text{aplikasi}|\text{positif}) \times p(\text{belanja}|\text{positif}) \times p(\text{online}|\text{positif}) \times p(\text{bisa}|\text{positif}) \times p(\text{percaya}|\text{positif})$ .

Positif:  $0.6 \times 0.03773 \times 0.05661 \times 0.05661 \times 0.05661 \times 0.03773 \times 0.05661 = 0.00000000877$

Perhitungan diatas merupakan proses klasifikasi sentimen yang didasari oleh hasil perhitungan probabilitas yang mempunyai nilai tertinggi yaitu  $8.77 \times 10^{-9}$  sehingga tweet diatas dikategorikan sebagai tweet "Positif".

### 3.2.7 Evaluasi dan Validasi

Evaluasi dapat dilakukan untuk mengetahui akurasi, akurasi akan diukur dengan menggunakan *confusion matrix* ..... Teknik validasi digunakan untuk

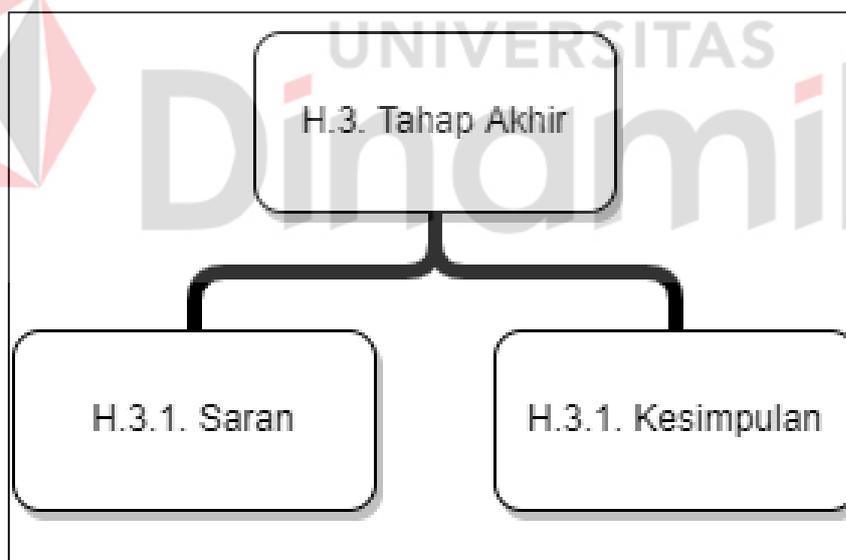
menguji data, dalam penelitian kali ini validasi menggunakan *10-folds cross-validation* dengan membagi *training* data menjadi beberapa bagian.

### 3.2.8 Visualisasi

Tahap visualisasi data dilakukan dengan memvisualkan data dengan *Worldcloud*, *bar chart*, *pie chart* dan *table*. Visualisasi data dengan *worldcloud* menghasilkan kata yang sering muncul dalam teks yang telah dianalisis sebelumnya, penggunaan *font-size* huruf yang berbeda menggambarkan frekuensi kemunculan kata pada data yang dianalisis. Visualisasi data dengan *bar chart* digunakan untuk mengetahui jumlah data dari 5 *e-commerce* yang bisa menentukan *e-commerce* paling banyak diulas. Visualisasi data dengan *pie chart* digunakan untuk hasil presentase *naïve bayes*.

### 3.3 Tahap Akhir Penelitian

Bagian ini menjelaskan tentang tahap akhir untuk menyelesaikan penelitian terdapat pada gambar 3.6.



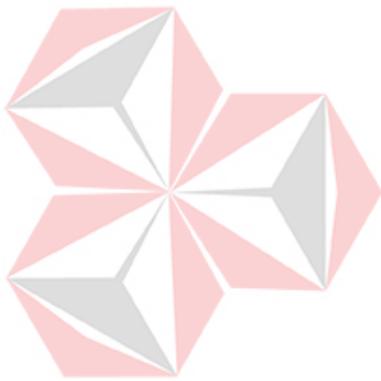
Gambar 3. 6 Tahap Akhir

#### 3.3.1. Kesimpulan

Tahapan selanjutnya yang dilakukan pada tahap akhir yaitu kesimpulan, tahap kesimpulan bertujuan untuk mengetahui perbandingan hasil analisis sentimen dan hasil evaluasi pada 5 (lima) *e-commerce* yaitu *Shopee*, *Bukalapak*, *Tokopedia*, *Lazada*, dan *Blibli*.

### 3.3.2. Saran

Tahap selanjutnya yang dilakukan pada tahap akhir yaitu saran. Tahap saran bertujuan untuk memberi masukan untuk kepada pihak *e-commerce* bagaimana untuk meningkatkan kepercayaan para pelanggan untuk membeli produk yang terjual pada *e-commerce* seperti *Shopee*, *Bukalapak*, *Tokopedia*, *Lazada*, dan *Blibli*.



UNIVERSITAS  
**Dinamika**

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

Pada bab ini akan dibahas hasil analisis berdasarkan pengolahan data yang telah dilakukan pada website *id.priceprice.com* dan *play.goolge.com* yang memuat komentar mengenai lima *E-Commerce* yang ada di Indonesia yaitu, *Shopee*, *Tokopedia*, *Bukalapak*, *Blibli.com*, dan *Lazada*.

#### **4.1. Analisis**

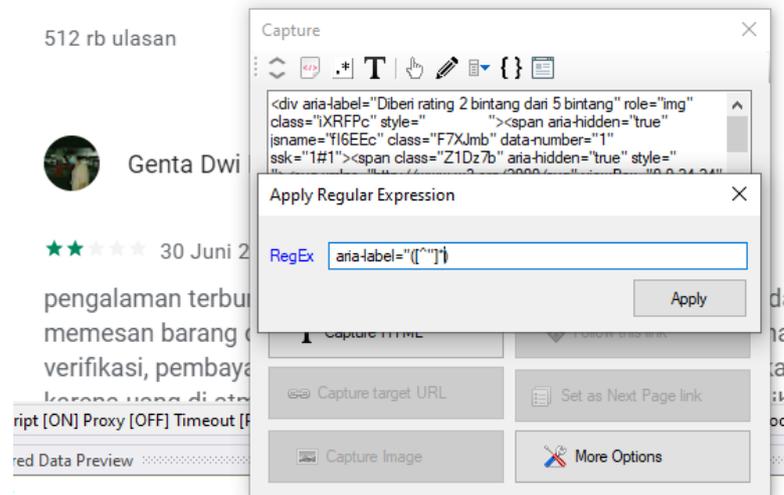
Analisis dimulai dengan identifikasi masalah yang digunakan sebagai dasar untuk pengerjaan proses yang akan dimulai dalam identifikasi masalah penulis menemukan bahwa perkembangan *e-commerce* di Indonesia mengalami peningkatan dan terdapat 5 *e-commerce* di Indonesia yang mempunyai pengunjung situs terbanyak yaitu *Shopee*, *Tokopedia*, *Bukalapak*, *Lazada*, dan *Blibli.com*. Review pelanggan juga merupakan salah satu kompoen yang penting dalam keberlangsungan bisnis untuk mempertahankan kualitas suatu bisnis atau produk. Oleh karena itu penelitian ini ditujukan untuk para *e-commerce* untuk mengetahui apakah layanan yang diberikan sudah memenuhi kebutuhan dari pengguna atau belum dengan mengetahui review positif dan review negatif dari setiap *e-commerce*.

Analisis tugas akhir ini dikerjakan dengan menggunakan 2 perangkat lunak, yaitu ada *WebHarvy* dan *Rapid Miner*. *WebHarvy* digunakan untuk melakukan *Web Scrapping* yang berguna untuk penambangan data. Sedangkan *Rapid Miner* digunakan untuk melakukan analisis dari data yang sudah diambil.

#### **4.2. Penambangan Data**

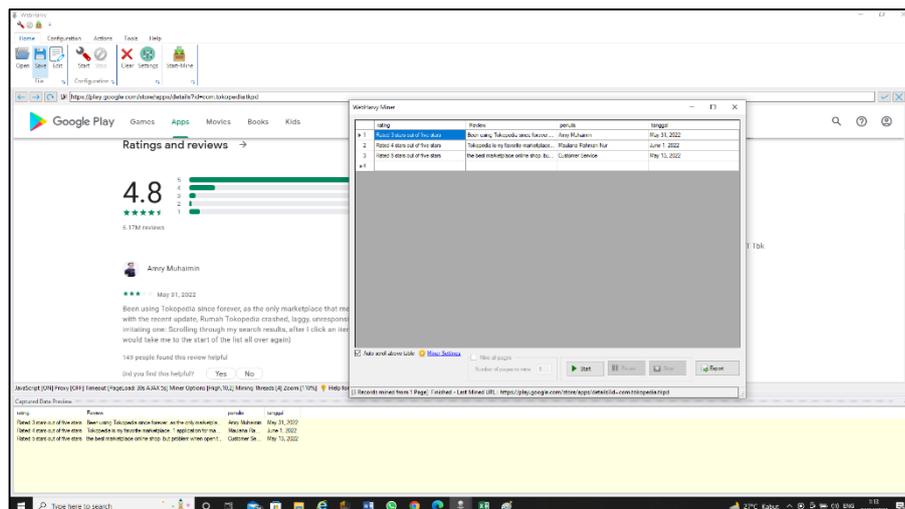
Setelah melakukan tahap analisis, proses dilanjutkan dengan melakukan penambangan data. Penambangan data sendiri digunakan untuk *scraping* data dari *source* yang digunakan pada penelitian ini. Alat untuk melakukan penambangan data yaitu menggunakan *software* *WebHarvy*, dimana cara melakukan *WebHarvy* adalah mengambil data yang ada pada website yaitu berupa *rating*, *review*, nama,

dan tanggal yang ada pada website. Pada saat melakukan penambangan *rating* menggunakan RegEx berupa rumus yang berguna untuk merubah icon bintang dan mengambil kalimat dari aria-label. Contohnya seperti gambar 4.1.



Gambar 4.1 Contoh Penggunaan RegEx

pada gambar 4.1 dapat dilihat bahwa penggunaan RegEx hanya dalam mengcapture atau menambang bagian *rating* saja. Dengan rumus “aria-label=”([^\"]\*)” maka *rating* yang sebelumnya icon bintang lalu mengambil suatu kalimat dari aria-label. Setelah melakukan hal tersebut lalu dilanjutkan untuk mengcapture atau menambang bagian review, nama dan tanggal. Contoh dari penambangan tersebut terdapat pada gambar 4.2.



Gambar 4.2 Proses Penambangan Data Dengan Web Harvy

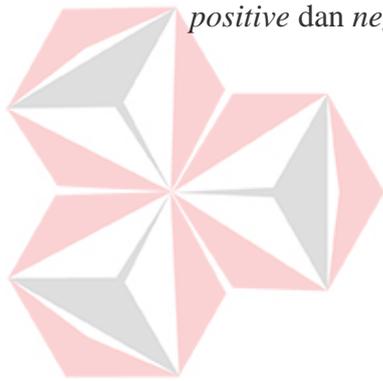
Pada gambar 4.2 dapat dilihat proses penambangan data dengan WebHarvy yang telah dilakukan. Setelah proses diatas dijalankan maka data di ambil dan hasil

data yang didapatkan dari penambangan diunduh dalam bentuk file excel (.xls) seperti gambar L.1.

Gambar L.1 adalah contoh hasil penambangan data pada *software* WebHarvey. Terdapat beberapa tabel yaitu ada rating, review, penulis, dan tanggal. Hasil data tersebut diunduh dalam bentuk file excel (.xls).

### 4.3. Pelabelan

Setelah melakukan proses penambangan data, maka proses selanjutnya yaitu proses pelabelan. Pada proses pelabelan ini dilakukan secara manual dan sesuai dengan subjektifitas penulis. Sebelum melakukan proses pelabelan, data yang telah ditambang di *excel* di ambil reviewnya dan ditaruh ke *excel* yang berbeda agar dapat melakukan proses pelabelan dengan mudah. Lalu dilakukan pemberian *name* label pada data review yang sudah ada, lalu label yang akan diberikan adalah *positive* dan *negative* contoh hasil pelabelan dapat dilihat pada gambar 4.3.



	Review	Label
1	Review	
2	bagus banget donlowt	positif
3	customer retur alasan bukti kuat shopee mediasi video blibli customer retur terima bukti video bukti customer kuat akun band saldo kembalikan janji balikin gak balikin aja seller hati buka toko tunggu etikat ubah ulasan bintangnv	negatif
4	pengalaman top up emoney buruk gak kayak marketplace gk mood langsung tunggu jam transaksi batal giliran refund x jam aplikasi blm payahh nihh mending pindah market place	negatif
5	parah aplikasi bli bli pembatalan pesanan refund skrg gak refund gak tuker gak customer paksa pembelian pengirimannya slogannya pengiriman no pembelian pengiriman parah aplikasi bli gak peduli customer	negatif
6	aplikasi mohon diperbaiki alamat pengiriman catatan pengiriman membingungkan kurir alamat pengiriman no tlp alamat pengiriman terhubung no yg terdaftar digopay membeli	negatif
7	kecewa banget blibli order an pakai someday pickup jam malem ordernya kemarin udah komplain opsi same day pakai rara ekspedisi nggak estimasi ngapain pakai opsi someday pengirimannya kaya reguler e commerce barangnya nggak order blibli	negatif
8	aplikasi payah not recommended ngelag berhenti pas ngecek deskripsi terpaksa pake aplikasi ga tersedia yg	negatif
9	warung gersang semenjak blibli memberlakukan pembatalan pesanan level infinite dipertahankan blibli belanja kontan ngutang mohon blibli dikembalikan pembatalan kasih saran blibli voucher discount voucher cashback dimanfaatkan kabarnya puluhan akun	negatif
10	pake blibli enak bingungin krn blm terbiasa tp skrng udah mayan gunain btw ekonomis bngtt lbh murah si oren hehe trus tuh dah otomatis gratis ongkir gitu klo memenuhi t c belanjaan jd ga capek nyariin voucher hehe jd langganan kedepannya hehe	positif

Gambar 4.3 Contoh Hasil Pelabelan

Terdapat jumlah keseluruhan review yang ada pada 5 *E-Commerce* dapat dilihat pada tabel 12.

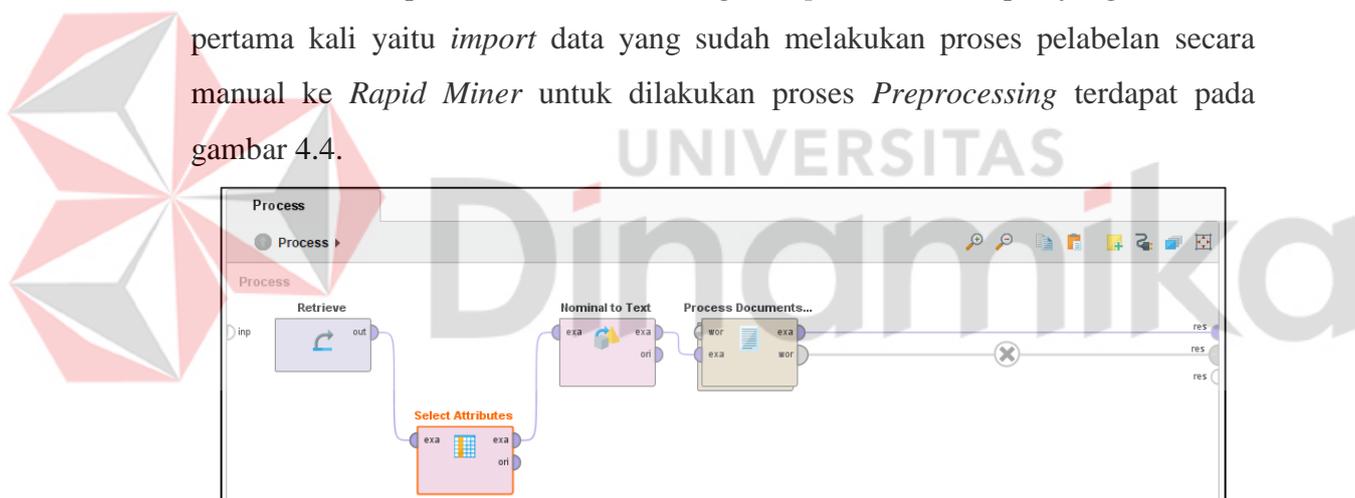
Tabel 12. Jumlah Review Setiap *E-Commerce*

	Blibli.com	Bukalapak	Lazada	Shopee	Tokopedia
Jumlah Review	1800	3217	3277	2601	2613

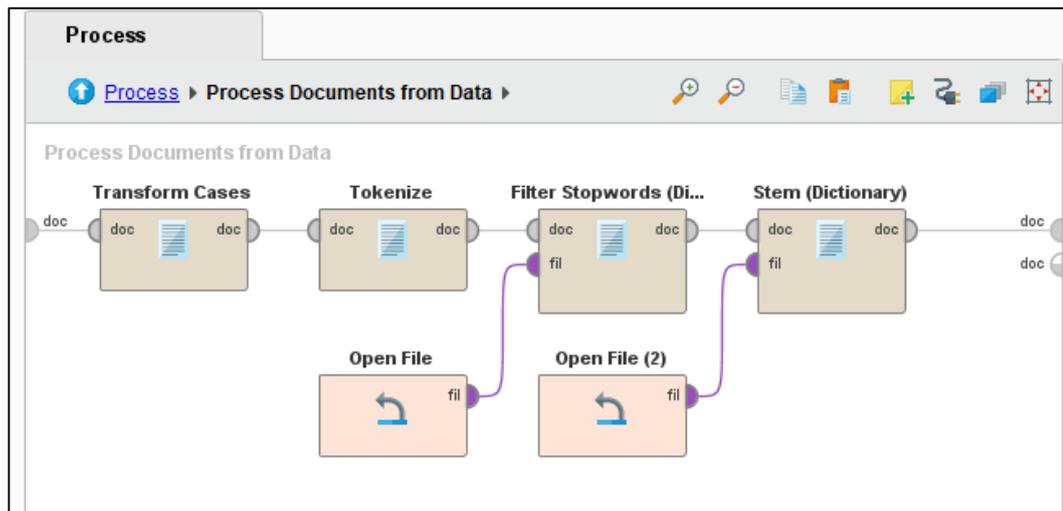
Pada tabel 12 dapat dilihat bahwa Blibli.com mempunyai data review sebesar 1800, Bukalapak mempunyai data review sebesar 3217, Lazada mempunyai data review sebesar 3277, Shopee mempunyai data review sebesar 2601, dan Tokopedia mempunyai data review sebesar 2613.

#### 4.4. Data Preprocessing

Proses selanjutnya adalah *data preprocessing*, pada tahap ini dilakukan pengolahan terhadap hasil penambangan yang formatnya tidak terstruktur menjadi kata dasar. Pada proses ini dilakukan dengan *Rapid Miner*, tahapan yang dilakukan pertama kali yaitu *import* data yang sudah melakukan proses pelabelan secara manual ke *Rapid Miner* untuk dilakukan proses *Preprocessing* terdapat pada gambar 4.4.

Gambar 4.4 Tahap Awal *Preprocessing*

Gambar 4.4 merupakan tahapan dari proses *preprocessing*, yang pertama yaitu ada *retrieve* data yang berguna untuk mengambil data review dari penambangan data untuk menjalankan proses, yang kedua yaitu *select attributes* yang berguna untuk memilih atribut review yang dilakukan tahap *preprocessing*, lalu ada proses *nominal to text* yang berguna merubah tipe data yang sebelumnya nominal menjadi text, Selanjutnya tahapan *preprocessing* dengan menggunakan operator Proses Dokumen yang didalamnya terdapat sub proses yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming* seperti gambar 4.5.



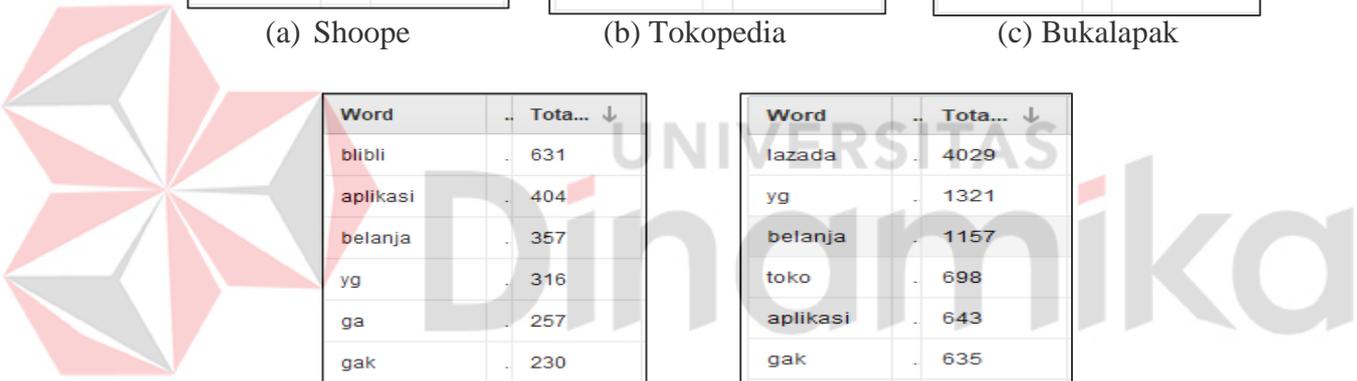
Gambar 4.5 Tahap *Preprocessing*

Pada gambar 4.5 merupakan tahapan proses *Preprocessing* dimana ada operator yaitu dapat dijelaskan dibawah ini yaitu *Case folding* untuk mengubah semua review menjadi huruf kecil (*lowercase*), tahap *tokenizing* yang dilakukan yaitu pemotongan kalimat menjadi beberapa token atau bagian dan pengalihan *symbol*, tanda baca, dan karakter bukan huruf, tahap *filtering* untuk penghilangan kata yang tidak relevan, penghilangan kata yang tidak memiliki arti seperti kata ganti dan kata sambung. Pada tahap ini menggunakan operator *open file* yang berisi tentang *stopword* dari rumus sastra yang ada. Setelah itu ada tahapan *stemming* yang dilakukan untuk proses pencarian kata dasar dengan menghilangkan imbuhan. Pada tahap ini juga menggunakan operator *open file* yang berisi tentang rumus dari sastra. Setelah melakukan tahapan tersebut maka dihasilkan 5 sumber review yang telah melalui tahap *preprocessing* dan hasil tersebut dimasukkan kembali ke excel untuk melakukan proses pembagian data ke *data training* dan *data testing*.

Selain menghasilkan data review yang bagus pada tahap *data preprocessing*, dapat juga diketahui *word frequency* tiap kata. *Word frequency* menunjukkan kata yang paling populer dari kata atau data yang telah melakukan *data preprocessing*. Frekuensi data yang sering muncul sangat penting untuk memprediksi kelas sentimen. Frekuensi data dijelaskan pada subab yang ada dibawah ini.

#### 4.4.1. Frekuensi Kata pada 5 E-Commerce

Sepuluh kata yang sering muncul pada hasil frekuensi kata 5 E-Commerce dengan hasil keseluruhan review yang terdapat pada tabel 12, dapat dilihat pada gambar 4.6



Word	...	Tota... ↓
shopee	?	3152
yg	?	1228
ulasan	?	856
lengkap	?	848
gak	?	746
aplikasi	?	653
belanja	?	632
akun	?	576
ga	?	548
aja	?	538

(a) Shoope

Word	..	Tota... ↓
yg	.	1519
tokopedia	.	1476
aplikasi	.	872
lengkap	.	822
ulasan	.	804
belanja	.	690
gak	.	581
ongkir	.	541
ga	.	515
udah	.	502

(b) Tokopedia

Word	..	Tota... ↓
bukalapak	.	2553
yg	.	1172
gak	.	886
aplikasi	.	791
buka	.	776
aja	.	616
ga	.	572
belanja	.	556
pelapak	.	536
udah	.	519

(c) Bukalapak

Word	..	Tota... ↓
bibli	.	631
aplikasi	.	404
belanja	.	357
yg	.	316
ga	.	257
gak	.	230
aja	.	195
bagus	.	165
pengiriman	.	154
bli	.	148

(d) Blibli.com

Word	..	Tota... ↓
lazada	.	4029
yg	.	1321
belanja	.	1157
toko	.	698
aplikasi	.	643
gak	.	635
ga	.	609
bagus	.	525
online	.	524
banget	.	516

(c) Lazada

Gambar 4.6 Frekuensi Kata 5 E-Commerce

Pada gambar 4.6 dapat dilihat frekuensi kata 5 E-commerce yang diperoleh dari tahapan *text preprocessing* setelah selesai melakukan tahapan tersebut maka akan dilanjutkan dalam proses implementasi dari *naïve bayes*

## 4.5. Implementasi Proses *Naive Bayes*

Pada tahap ini dilakukan implementasi proses *naive bayes* yang dimulai dari membagi data menjadi data *training* dan data *testing* dari keseluruhan data yang telah melalui tahap *preprocessing*

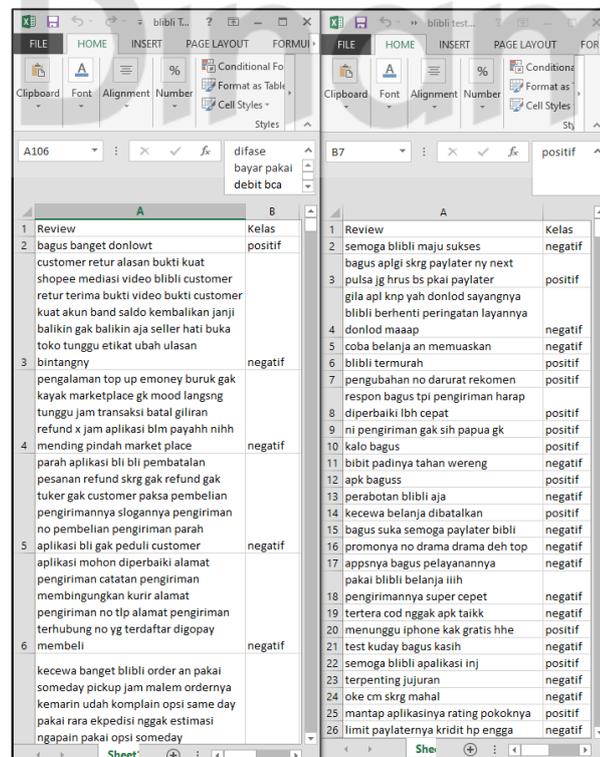
### 4.5.1. Training Data dan Testing Data

Pada tabel 12 dapat dilihat keseluruhan data review yang didapat pada saat penambahan lalu data review dibagi menjadi *training data* dan *testing data* dengan cara manual dan menghasilkan data dapat dilihat pada tabel 13.

Tabel 13. Hasil Pembagian *Training Data* dan *Testing Data*

	Blibli.com	Bukalapak	Lazada	Shopee	Tokopedia
Training	1620	2804	2680	2340	2353
Testing	180	314	298	261	260

Pada tahapan pembagian data dilakukan secara manual dengan membagi setiap data menjadi 2 secara manual yaitu data *training* dan data *testing*, pada tahap pembagian data ini juga memiliki 5 sumber review. Contoh dari *training data* dan *testing data* terdapat pada gambar 4.7.



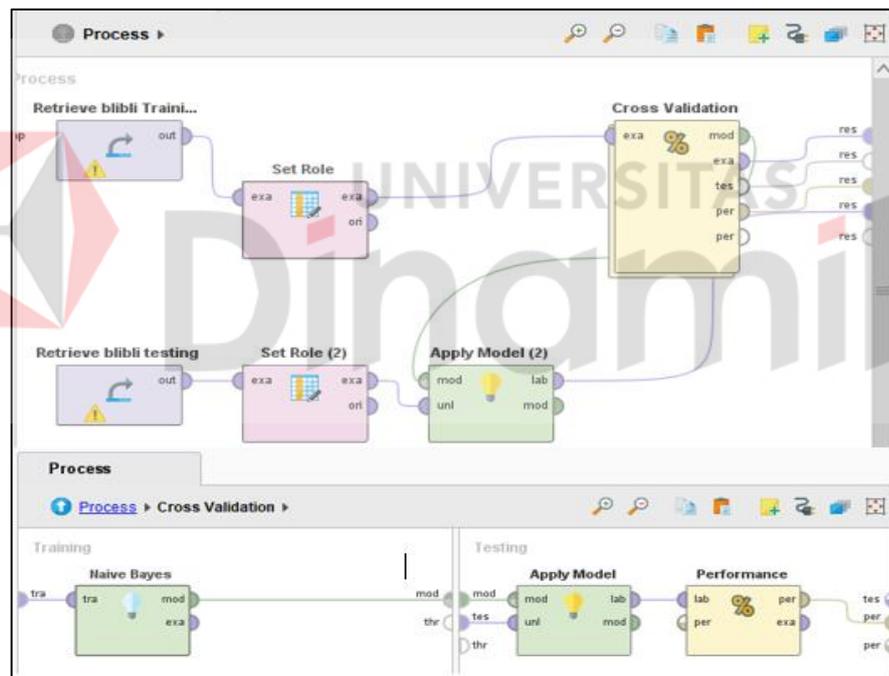
Review	Kelas
Review	Kelas
bagus banget donlwt	positif
customer retur alasan bukti kuat shopee mediasi video blibli customer retur terima bukti video bukti customer kuat akun band saldo kembalikan janji balikin gak balikin aja seller hati buka toko tunggu etiket ubah ulasan	negatif
bintangny	negatif
pengalaman top up emoney buruk kayak marketplace gk mood langsung tunggu jam transaksi batal gliran refund x jam aplikasi blm payahh nihh	negatif
mending pindah market place	negatif
parah aplikasi bli bli pembatalan pesanan refund skrg gak refund gak tuker gak customer paksa pembelian pengirimannya slogannya pengiriman no pembelian pengiriman parah	negatif
aplikasi bli gak peduli customer aplikasi mohon diperbaiki alamat pengiriman catatan pengiriman membingungkan kurir alamat pengiriman no tlp alamat pengiriman terhubung no yg terdaftar digopay	negatif
membeli	negatif
kecewa banget blibli order an pakai someday pickup jam malem ordernya kemarin udah komplain opsi same day pakai rara ekspedisi nggak estimasi ngapain pakai opsi someday	negatif
semoga blibli maju sukses	negatif
bagus aplgi skrg paylater ny next pulsa jg hrus bs pkai paylater	positif
gila apl knp yah donlod sayangnya blibli berhenti peringatan layannya	negatif
donlod maaap	negatif
coba belanja an memuaskan	negatif
blibli termurah	positif
pengubahan no darurat rekomen	positif
respon bagus tpi pengiriman harap	positif
diperbaiki lbh cepat	positif
ni pengiriman gak sih papua gk	positif
kalo bagus	positif
bi bit padinya tahan wereng	negatif
apk bagus	positif
perabotan blibli aja	negatif
kecewa belanja dibatalkan	positif
bagus suka semoga paylater blibli	negatif
promonya no drama drama deh top	negatif
appsnya bagus pelayanannya	negatif
pakai blibli belanja liih	negatif
pengirimannya super cepet	negatif
tertera cod nggak apk taikk	negatif
menunggu iphone kak gratis hhe	positif
test kuday bagus kasih	negatif
semoga blibli aplikasi inj	positif
terpenting jujur	negatif
oke cm skrg mahal	negatif
mantap aplikasinya rating pokoknya	positif
limit paylaternya kridit hp engga	negatif

Gambar 4.7 Contoh Training Data dan Testing Data

Dalam gambar 4.7 dapat dijelaskan bahwa data tersebut adalah data dari *e-commerce* Blibli.com yang meliputi data excel yang sebelah kiri adalah *training data* yang mempunyai data sebesar 1620 data review yang ada. Sedangkan data excel yang sebelah kanan adalah *testing data* yang mempunyai data sebesar 180 data review yang ada. Jika digabungkan maka data review dari Blibli.com sebesar 1800 data.

#### 4.5.2 Klasifikasi Naïve bayes

*Naïve Bayes* adalah metode yang digunakan untuk memprediksi peluang dari kelas yang akan muncul. Hal yang dilakukan pertama kali untuk mengklasifikasi model adalah *import training data* dan *testing data* dari 5 sumber yang telah disediakan. Desain model *naïve bayes* terdapat pada gambar 4.8.



Gambar 4.8 Desain Model Validation Naïve Bayes Classifier

Pada gambar 4.8 digunakan review dari sumber Blibli.com, proses diawali dengan operator *retrieve* yang berfungsi untuk menginput atau memasukan data review ke RapidMiner pada input data ini yang di input adalah data *training*. Setelah menginput data review maka dilanjutkan operator *set role* yang berguna untuk membedakan penamaan atribut dan prediksi posisi yang akan masuk kedalam label atau kolom sentimen, gunanya supaya data label tidak ikut terhitung dan jika tidak melakukannya maka bisa menyebabkan hasilnya akan berubah. Setelah itu

dilakukan operator *validation* menggunakan *10-fold cross validation* yang berguna untuk mengevaluasi kinerja dari model *Naïve Bayes*. Didalam *validation* terdapat dua proses, yaitu proses *training* dan proses *testing*. Dalam proses *training* terdapat algoritma klasifikasi yang akan diterapkan yaitu *Naïve Bayes*, sedangkan di dalam proses *testing* terdapat operator *Apply Model* yang berguna untuk menjalankan model *Naïve Bayes* dan operator *Performance* yang berguna untuk mengukur performa dari model *Naïve Bayes* itu sendiri. Lalu ada terdapat operator *retrieve data testing* yang berguna untuk mencek data yang sudah ada dengan model *naive bayes* yang sudah melakukan proses *training* dengan operator *apply model* agar mendapat hasil dari data yang sudah di *testing*.

#### 4.6. Evaluasi dan Validasi *Naïve Bayes*

Tahap ini merupakan hasil pemodelan yang diproses dengan *Tools RapidMiner*. Setelah dilakukan pemodelan *Naïve Bayes*, maka akan dapat diketahui probabilitas klasifikasi dengan metode *Naïve Bayes* pada 5 *e-commerce* yang ada di Indonesia. Jika probabilitas sentimen negatif lebih kecil dibandingkan dengan probabilitas sentimen positif, maka review tersebut akan menghasilkan keputusan positif begitu juga dengan sebaliknya.

##### 4.6.1. Evaluasi Blibli.com

Tabel 14. Probabilitas NB Blibli.com

Review	Label	Prediction
1	Negatif	Positif
2	Positif	Positif
3	Negatif	Negatif
4	Negatif	Positif
5	Positif	Positif
6	Positif	Positif
7	Positif	Positif
8	Positif	Positif
...	...	...
180	Positif	Positif

Pada tabel 14 dapat dilihat bahwa data review pelanggan pertama yang sudah dilabeli manual diisi dengan review negatif dan hasil prediksi dari analisis sentimen juga dapat dilihat pada tabel tersebut.

#### 4.6.2. Evaluasi Bukalapak

Pada tabel L.3 dapat dilihat bahwa data review pelanggan keempat yang telah diberi label manual diisi dengan review negatif dan hasil dari prediksi analisis sentimen yang dapat dilihat pada tabel

#### 4.6.3. Evaluasi Lazada

Pada tabel L.4 dapat dilihat bahwa data review pelanggan ketujuh yang telah diberi label manual diisi dengan review positif dan hasil prediksi dari analisis sentimen yang dapat dilihat pada tabel

#### 4.6.4. Evaluasi Shopee

Pada tabel L.5 dapat dilihat bahwa data review pelanggan kedelapan yang telah diberi label manual diisi dengan review positif dan hasil prediksi dari analisis sentimen yang dapat dilihat pada tabel

#### 4.6.5. Evaluasi Tokopedia

Pada tabel L.6 dapat dilihat bahwa data review pelanggan kedelapan yang telah diberi label manual diisi dengan review positif dan hasil prediksi dari analisis sentimen yang dapat dilihat pada tabel

#### 4.6.6. Hasil Validasi Akurasi *Naïve Bayes*

Langkah selanjutnya adalah validitas data untuk mengetahui tingkat keakuratan data yang akan dihasilkan dengan menggunakan RapidMiner. Untuk melakukan pengujian validasi data yaitu menggunakan *10-Fold Cross Validation*, data dibagi menjadi dua yaitu terdapat data *training* dan data *testing* dengan perbandingan pembagiannya dapat dilihat pada tabel 13. Hasil uji validitas data menggunakan metode *Naïve Bayes* pada 5 *e-commerce* di Indonesia dapat dilihat pada tabel 15.

Tabel 15. Hasil Validasi Akurasi *Naïve Bayes*

Fold ke	Blibli	Bukalapak	Lazada	Shopee	Tokopedia
2	55.31%	91.58%	56.08%	97.01%	93.46%
3	53.89%	92.51%	56.23%	97.22%	94.31%
4	52.65%	92.76%	56.23%	97.52%	94.69%
5	52.84%	92.87%	56.16%	97.52%	94.64%
6	51.05%	92.97%	56.23%	97.74%	94.73%
7	49.94%	92.69%	56.23%	97.74%	94.81%
8	50.68%	93.05%	56.23%	97.82%	94.94%
9	50.68%	92.87%	56.23%	97.69%	94.69%

Fold ke	Blibli	Bukalapak	Lazada	Shopee	Tokopedia
10	51.73%	92.94%	56.18%	97.69%	94.56%
AVG	52.08%	92.69%	56.2%	97.55%	94.54%

Dari tabel 15 dapat diketahui bahwa rata-rata akurasi dari setiap *e-commerce* memiliki tingkat akurasi yang berbeda. Peringkat pertama terdapat *Shopee* yang memiliki tingkat akurasi sebesar 97.55%. Peringkat kedua terdapat *Tokopedia* yang memiliki tingkat akurasi sebesar 94.56%. Lalu peringkat ketiga terdapat *Bukalapak* yang memiliki tingkat akurasi sebesar 92.69%. Setelah itu diperingkat keempat terdapat *Lazada* yang memiliki tingkat akurasi sebesar 56.2%. Pada peringkat kelima terdapat *Blibli.com* yang memiliki tingkat akurasi sebesar 52.73%. Untuk rata-rata yang sudah mencapai 90% ke atas dapat lebih dipercaya dikarenakan akurasi yang dihasilkan sudah hampir mencapai nilai tertinggi. Sedangkan akurasi yang rata-ratanya masih 50% masih perlu di pertanyakan untuk tingkat kebenaran dari hasil klasifikasi *naïve bayes*.

#### 4.7. Hasil Akhir dari *Naïve Bayes*

Setelah dilakukan validasi menggunakan *10-Fold Cross Validation* maka diambil tingkat akurasi paling tinggi dari hasil tersebut, maka hasil dari metode yang telah dikerjakan adalah sebagai berikut:

Tabel 16. Hasil Akhir *Naïve Bayes*

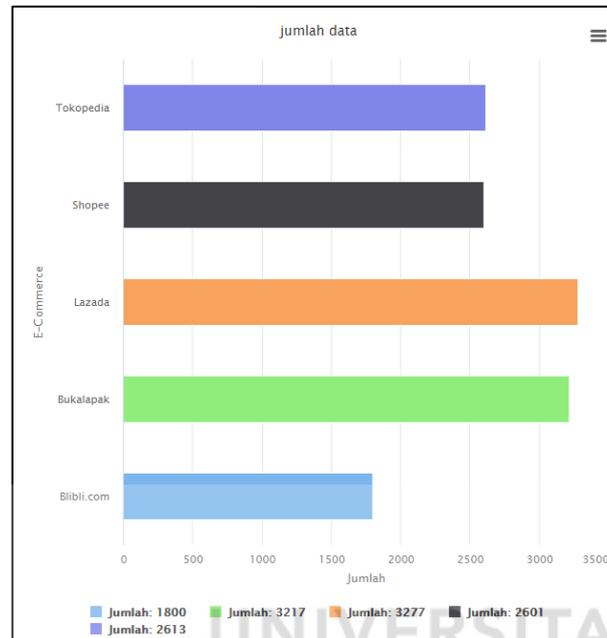
	Blibli.com <i>Naïve Bayes</i>	Bukalapak <i>Naïve Bayes</i>	Lazada <i>Naïve Bayes</i>	Shopee <i>Naïve Bayes</i>	Tokopedia <i>Naïve Bayes</i>
<i>Accuracy</i>	55.31%	93.05%	56.23%	97.82%	94.94%
<i>Prec</i>	35.45%	94.95%	100%	97.52%	92.95%
<i>Sens</i>	30.35%	52.54%	0.34%	99.85%	87.72%
<i>Positive</i>	76.1%	94.6%	97.0%	88.5%	34.4%
<i>Negative</i>	23.9%	5.4%	3.0%	11.5%	65.6%

Hasil dari tabel 16 dapat diketahui bahwa:

1. Hasil dari sentimen positif dan negatif dapat dilihat bahwa sentimen positif terbesar didapatkan oleh *E-Commerce* *Lazada* sebesar 97.0%. Sedangkan sentimen negatif terbesar didapatkan oleh *E-Commerce* *Tokopedia* sebesar 65.6%.
2. Hasil akurasi dari model *naive bayes* mempunyai rata-rata sebesar 78.62%. Diantara kelima *e-commerce* yang diteliti, akurasi paling besar didapatkan pada *e-commerce* *Shopee* sebesar 97.82%. Sedangkan akurasi yang paling kecil didapatkan pada *e-commerce* *Blibli.com* sebesar 55.31%.

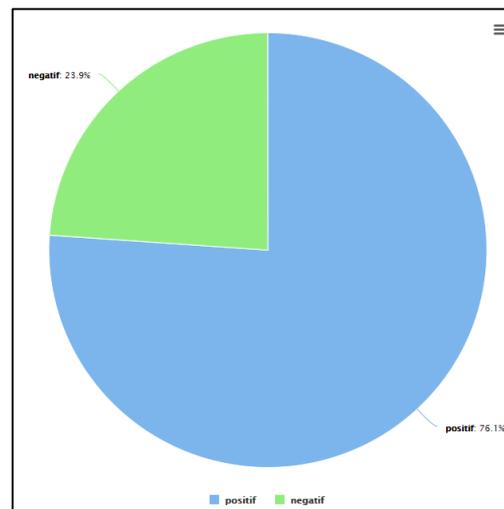


Pada gambar 4.9, menunjukkan bahwa kata “blibli” adalah kata yang paling sering diucapkan dari jumlah data yang ada pada penelitian ini atau yang sudah dicrawl oleh penulis, kata penting pada penelitian ini yaitu seperti kata positif yaitu “Bagus” dan “sukses” sedangkan kata negatif seperti “Kecewa” dan “nunggu”.



Gambar 4.9 Data Review *E-Commerce*

Pada gambar 4.10, dapat dilihat jumlah data review dari masing-masing *E-Commerce* yang telah dicrawl menggunakan *tools* WebHarvey. Dapat dilihat juga bahwa aplikasi Blibli.com mendapat jumlah data review sebesar 1800 data review.



Gambar 4.10 Persentase Sentimen pada Blibli.com

Pada gambar 4.11, menunjukkan persentase review yang sudah dilakukan klasifikasi oleh *naive bayes*. Dari gambar diatas dapat dilihat menunjukkan bahwa review positif lebih besar dari pada review negatif. Secara keseluruhan presentase dari aplikasi Blibli.com yaitu review positif sebesar 76.1% sedangkan review negatif sebesar 23.9%.

#### 4.8.2. Visualisasi Bukalapak

Pada gambar L.11, menunjukkan bahwa kata yang paling banyak disebutkan dari review aplikasi bukalapak yang sudah dicrawl adalah kata “bukalapak”. Lalu kata penting yaitu seperti kata positif yaitu “lengkap” dan “cepat” sedangkan kata negatif yaitu “kecewa” dan “susah”.

Pada gambar L.9, dapat dilihat jumlah data review dari masing-masing E-Commerce yang telah dicrawl menggunakan *tools* WebHarvey. Dapat dilihat juga bahwa aplikasi Bukalapak mendapat jumlah data review sebesar 3217 data review.

Pada gambar L.12, menunjukkan persentase review yang sudah dilakukan klasifikasi oleh *naive bayes*. Dari gambar diatas dapat dilihat menunjukkan bahwa review positif lebih besar dari pada review negatif. Secara keseluruhan presentase dari aplikasi Bukalapak yaitu review positif sebesar 94.6% sedangkan review negatif sebesar 5.4%.

#### 4.8.3. Visualisasi Lazada

Pada gambar L.13, menunjukkan bahwa kata yang paling banyak disebutkan dari review aplikasi Lazada yang sudah dicrawl adalah kata “Lazada”. Lalu kata penting yaitu seperti kata positif “lengkap” dan “bagus” sedangkan kata negatif seperti “kecewa” dan “jelek”.

Pada gambar L.9, dapat dilihat jumlah data review dari masing-masing E-Commerce yang telah dicrawl menggunakan *tools* WebHarvey. Dapat dilihat juga bahwa aplikasi Lazada mendapat jumlah data review sebesar 3277 data review.

Pada gambar L.14, menunjukkan persentase review yang sudah dilakukan klasifikasi oleh *naive bayes*. Dari gambar diatas dapat dilihat menunjukkan bahwa review positif lebih besar dari pada review negatif. Secara keseluruhan presentase dari aplikasi Lazada yaitu review positif sebesar 97.0% sedangkan review negatif sebesar 3.0%.

#### 4.8.4. Visualisasi Shopee

Pada Gambar L.15, menunjukkan bahwa kata yang paling banyak disebutkan dari review aplikasi Shopee yang sudah dicrawl adalah kata “Shopee”. Lalu kata penting yaitu seperti kata positif “lengkap” dan “bagus” sedangkan kata negatif seperti “kecewa” dan “jelek”.

Pada gambar L.9, dapat dilihat jumlah data review dari masing-masing E-Commerce yang telah dicrawl menggunakan *tools* WebHarvey. Dapat dilihat juga bahwa aplikasi Shopee mendapat jumlah data review sebesar 2601 data review.

Pada gambar L.16, menunjukkan persentase review yang sudah dilakukan klasifikasi oleh *naive bayes*. Dari gambar diatas dapat dilihat menunjukkan bahwa review positif lebih besar dari pada review negatif. Secara keseluruhan presentase dari aplikasi Shopee yaitu review positif sebesar 88.5% sedangkan review negatif sebesar 11.5%.

#### 4.8.5. Visualisasi Tokopedia

Pada gambar L.17, menunjukkan bahwa kata yang paling banyak disebutkan dari review aplikasi Tokopedia yang sudah dicrawl adalah kata “yg”. Lalu kata penting yaitu seperti kata positif yaitu “lengkap” dan “bagus” sedangkan kata negatif seperti “susah” dan “kecewa”.

Pada gambar L.9, dapat dilihat jumlah data review dari masing-masing E-Commerce yang telah dicrawl menggunakan *tools* WebHarvey. Dapat dilihat juga bahwa aplikasi Tokopedia mendapat jumlah data review sebesar 2601 data review.

Pada gambar L.18, menunjukkan persentase review yang sudah dilakukan klasifikasi oleh *naive bayes*. Dari gambar diatas dapat dilihat menunjukkan bahwa review negatif lebih besar dari pada review positif. Secara keseluruhan presentase dari aplikasi Tokopedia yaitu review positif sebesar 34.4% sedangkan review negatif sebesar 65.6%.

### 4.9. Pembahasan

Dari penjelasan semua subbab dapat disimpulkan bahwa:

1. *E-Commerce* Lazada mendapatkan hasil sentimen positif terbanyak dari review yang telah dilakukan. Lalu dilanjutkan dengan Bukalapak pada posisi kedua

dalam mendapatkan hasil sentimen positif terbanyak. Lalu ada Shopee yang menempati peringkat ketiga dalam mendapatkan hasil sentimen positif. Pada peringkat keempat ada Blibli.com dalam mendapatkan hasil sentimen positif. Peringkat kelima dalam mendapatkan hasil sentimen positif yaitu Tokopedia.

2. Dalam peringkat pertama terdapat Lazada yang mendapatkan tingkat hasil sentimen positif sebesar 97.0% dikarenakan banyak ulasan bagus yang ada. Seperti “bagus”, “lengkap”, “murah”, dan “suka”. Pada *E-Commerce* tersebut juga terdapat ulasan negatif yang didapat seperti “kecewa”, “jelek”, “tipu”, dan “gagal”. Oleh karena itu yang perlu di tingkatkan dari aplikasi ini adalah tentang para penjual yang masih ada penipuan ataupun gagal dalam melakukan transaksi. Yang perlu dipertahankan pada aplikasi ini adalah para pelanggan telah banyak yang suka terhadap aplikasi ini karena lengkap atau banyak barang yang dapat dicari.
3. Dalam peringkat kedua terdapat Bukalapak yang mendapatkan tingkat hasil sentimen positif sebesar 94.6% dikarenakan banyak ulasan bagus yang ada. Seperti “cepat”, “lengkap”, dan “promo”. Pada *E-Commerce* tersebut juga terdapat ulasan negatif yang didapat seperti “ribet”, “kecewa”, “lemot”, dan “mahal”. Oleh karena itu yang perlu di tingkatkan dari aplikasi ini adalah tentang aplikasinya yang mungkin masih sulit untuk dimengerti dan juga lamban untuk penggunaan aplikasinya. Lalu yang perlu dipertahankan pada aplikasi ini adalah para pelanggan telah banyak yang mendapat promo dan cepat dalam pengiriman dan juga mendapat keuntungan uang kembali.
4. Dalam peringkat ketiga terdapat Shopee yang mendapatkan tingkat hasil sentiment positif sebesar 88.5% dikarenakan banyak ulasan bagus yang ada. Seperti “gratis”, “lengkap”, “puas”, dan “promo”. Pada *E-Commerce* tersebut juga terdapat ulasan negatif yang didapat seperti “susah”, “gagal”, “habis”, dan “jelek”. Oleh karena itu yang perlu di tingkatkan dari aplikasi ini adalah tentang aplikasinya yang mungkin sama seperti yang sebelumnya yaitu masih sulit untuk dimengerti untuk pelanggan dan juga dalam melakukan transaksi masih ada yang gagal dan tidak bisa melanjutkannya. Lalu yang perlu dipertahankan pada aplikasi ini adalah para pelanggan telah banyak yang mendapat promo

dan dalam aplikasi ini terdapat banyak variasi barang yang diinginkan oleh pelanggan.

5. Dalam peringkat keempat terdapat Blibli.com yang mendapatkan tingkat hasil sentiment positif sebesar 76.1% dikarenakan banyak ulasan bagus yang ada. Seperti “gratis”, “refund”, “voucher”, dan “cepat”. Pada *E-Commerce* tersebut juga terdapat ulasan negatif yang didapat seperti “nunggu”, “pembatalan”, “rugi”, dan “mahal”. Oleh karena itu yang perlu di tingkatkan dari aplikasi ini adalah tentang aplikasinya yang mungkin masih terdapat terjadi pembatalan secara sepihak sehingga pelanggan tidak nyaman dan terdapat pelanggan yang rugi dalam melakukan transaksi di aplikasi ini. Lalu yang perlu dipertahankan pada aplikasi ini adalah para pelanggan telah banyak yang mendapat voucher dan mendapat gratis pengiriman yang ada pada aplikasi ini.
6. Dalam peringkat kelima terdapat Tokopedia yang mendapatkan tingkat hasil sentimen positif sebesar 34.4% dikarenakan banyak ulasan bagus yang ada. Seperti “lengkap”, “gratis”, “murah”, dan “cashback”. Pada *E-Commerce* tersebut juga terdapat ulasan negatif yang didapat seperti “susah”, “ribet”, “dibatalkan”, dan “kecewa”. Oleh karena itu yang perlu ditingkatkan dari aplikasi ini adalah tentang aplikasinya yang mungkin masih sulit untuk dimengerti dan juga ada fitur yang terlalu berbelit dalam menjalankannya. Lalu yang perlu dipertahankan pada aplikasi ini adalah para pelanggan telah menemukan beberapa variasi yang diperlukan dan pelanggan juga mendapatkan harga murah dari aplikasi ini

## BAB V PENUTUP

Bab ini berisi tentang hasil dari penelitian terkait pengerjaan tugas akhir pada penelitian kali ini. Subbab kesimpulan, memilisi isi tentang kesimpulan hasil pengerjaan tugas akhir. Lalu pada subbab saran berisi tentang saran untuk mengembangkan penelitian tugas akhir kali ini.

### 5.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilaksanakan, dapat disimpulkan bahwa:

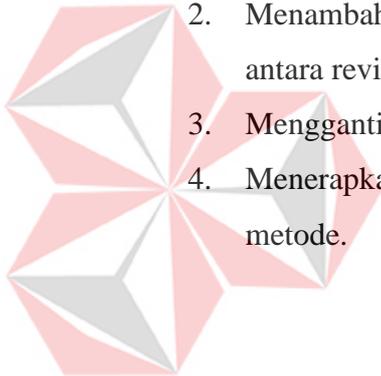
1. Sesuai dengan tabel 16, data yang telah diolah menggunakan Algoritma *Naïve Bayes* mempunyai *accuracy* dengan rata-rata sebesar 78.62%. lalu hasil akurasi dari setiap *e-commerce* dalam metode *naïve bayes* yaitu lazada sebesar 56.23%, bukalapak sebesar 93.05%, shopee sebesar 87.82%, blibli.com sebesar 55.31%, dan tokopedia sebesar 94.94%. *Naïve Bayes Classifier* pada penelitian kali ini jika *accuracy* memiliki angka 50% maka model *Naïve Bayes Classifier* pada penelitian kali ini bisa dikatakan kurang memenuhi dalam melakukan analisis sentimen pada penelitian ini. dan jika tingkat *accuracy* 80% - 90% maka model yang dilakukan pada saat itu dapat dikatakan cukup bagus.
2. Hasil visualisasi dari persentase sentimen positif didapatkan kepada *e-commerce Lazada* yang mendapatkan hasil sentimen positif terbesar didapatkan oleh lazada sebesar 97.0%, peringkat kedua bukalapak sebesar 94.6%, peringkat ketiga shopee sebesar 88.5% peringkat keempat blibli.com sebesar 76.1% peringkat kelima tokopedia sebesar 34.4%,
3. Hasil visualisasi dari persentase sentimen negatif didapatkan kepada *e-commerce Tokopedia* yang mendapatkan hasil sentimen negatif sebesar 65.6%. Peringkat kedua terdapat *Blibli.com* sebesar 23.9%. Peringkat ketiga terdapat Shopee sebesar 11.5%. Peringkat keempat terdapat *Bukalapak* sebesar 5.4% dan yang terakhir adalah *Lazada* sebesar 3.0%.
4. Terdapat beberapa kata-kata negatif yang berguna untuk *e-commerce* dalam mengembangkan aplikasinya, seperti pada Lazada terdapat kata negatif

“kecewa”, “jelek”, “tipu”. Pada Bukalapak terdapat kata negatif seperti “ribet”, “kecewa”, “lemot”. Shopee juga terdapat kata negatif seperti “susah”, “gagal”, “habis”. Pada Bilibli.com terdapat kata negatif yaitu “nunggu”, “rugi”, “mahal” dan pada Tokopedia terdapat kata negatif yaitu “susah”, “ribet”, “kecewa”.

## 5.2. Saran

Dalam pengerjaan tugas akhir tentunya masih memiliki beberapa kekurangan. Oleh karena itu, berikut adalah beberapa saran yang dapat digunakan untuk pengembangan dari penelitian ini agar dapat lebih baik kedepannya, maka dapat diberikan sebagai berikut:

1. Menambahkan kamus tentang bahasa gaul atau kata-kata singkatan, dikarenakan pada review masih banyak bahasa yang kurang baku.
2. Menambahkan review berbahasa Inggris agar dapat melakukan perbandingan antara review berbahasa Indonesia dan review berbahasa Inggris.
3. Mengganti pembagian proporsi penggunaan *training data* dan *testing data*.
4. Menerapkan suatu metode yang berguna untuk meningkatkan akurasi suatu metode.



## DAFTAR PUSTAKA

- Amelia, R. D., Michael, & Mulyadi, M. R. (2021). ANALISIS ONLINE CONSUMER REVIEW TERHADAP KEPUTUSAN PEMBELIAN PADA E-COMMERCE KECANTIKAN. *Jurnal Indonesia Sosial Teknologi*, 275-278.
- Amudha, S. (2017). Web Crawler For Mining Web Data. *International Research Journal Of Engineering And Technology*, 128.
- Ananda, R. P., & Yuniawan, A. (2021). Studi Empiris Kepuasan Pelanggan E-Commerce Secara Global. *Jurnal Ilmiah Indonesia*, 3500.
- Apriadi, D., & Saputra, A. Y. (2017). E-COMMERCE BERBASIS MARKETPLACE DALAM UPAYA MEMPERSINGKAT DISTRIBUSI PENJUALAN HASIL PERTANIAN. *RESTI*, 132.
- Aprilla, D., Baskoro, D. A., Ambarwati, L., & Wicaksana, I. S. (2022). Belajar Data Mining dengan Rapid Miner. *Repository dinus*, 22.
- Ardianti, A. N., & Widiartanto. (2019). Pengaruh Online Customer Review dan Online Customer Rating terhadap Kepuasan Pembelian melalui Marketplace Shopee. *ejurnal3 undip*, 2.
- Aribowo, A. S., Basiron, H., Herman, N. S., & Khomsah, S. (2020). An Evaluation of Preprocessing Steps and Tree-based Ensemble Machine Learning for Analysing Sentiment on Indonesian Youtube Comments. *International Journal of Advanced Trends in Computer Science and Engineering*, 7080.
- Bayu, D. J. (2020, 2 11). *DataBoks KataData*. Retrieved from Databoks: <https://databoks.katadata.co.id/datapublish/2021/02/11/10-e-commerce-dengan-pengunjung-terbesar-pada-kuartal-iv-2020>
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter. *Integer Journal*, 33.

- Faisal, R. (2016, 10 10). Retrieved from <https://staf.ulm.ac.id/rezafaisal/2016/10/10/imbalanced-class-accuracy/>
- Hayadi, B. H. (2017). Visualisasi Konsep Umum Sistem Pakar Berbasis Multimedia. *Riau Journal Of Computer Science*, 17-18.
- Hendarsyah, D. (2019). E-Commerce di Era Industri 4.0 dan Society 5.0. *Jurnal Ilmiah Ekonomi Kita*, 172.
- Holmes, J., Agius, C., & Cromptvoets, J. (2020). 3rd International Workshop On Spatial Data Quality. *Europian Spatial Dara Research*, 51-52.
- Izzah , K. I., & Girsang, A. S. (2021). Modified TF-Assoc Term Weighting Method For Text Classification on News Dataset from Twitter. *International Journal of Computer Science*, 48.
- Kaushik, A., & Naithani, S. (2016). A Comprehensive Study of Text Mining Approach. *International Journal of Computer Science and Network Security*, 69.
- Muljono, Artanti, D. P., Syukur, A., Prihandono, A., & Setiadi, D. M. (2018). Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja online menggunakan Algoritma Naive Bayes. *STMIK Atma Luhur Pangkalpinang*, 1.
- Napizahni, M. (2022, June 21). *Web Crawler: Pengertian, Fungsi, dan Cara Kerjanya*. Retrieved from DEWAWEB: <https://www.dewaweb.com/blog/apa-itu-web-crawler/>
- Pariadi, D. (2018). Pengawasan E Commerce Dalam Undang-Undang Perdagangan dan Undang-Undang Perlindungan Konsumen. *Jurnal Hukum dan Pembangunan* , 651.
- Parlina, I., Arnol, M. Y., Febriati, N. A., Dewi , R., Wanto, A., Lubis, M. R., & Susiani. (2019). Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category. *Journal of Physics: Conference Series*, 3.

- Patria, R. (2022, January 22). *Web Crawler: Definisi, Cara Kerja, dan Kaitannya Dengan SEO*. Retrieved from Domainsia:  
<https://www.domainsia.com/berita/web-crawler/>
- Prabowo, D. (2015). WEBSITE E-COMMERCE MENGGUNAKAN MODEL VIEW CONTROLLER (MVC) DENGAN FRAMEWORK CODEIGNETER. *Jurnal ilmiah DASI*, 23.
- Prasetyo, H. D., & Purbawati, D. (2017). Pengaruh e-service quality dan e-security seals terhadap e-satisfaction melalui keputusan pembelian konsumen e-commerce. *Jurnal ilmu Administrasi Bisnis*, 1-9.
- Purfini, A. P. (2018). Pengaruh cuaca dan review pelanggan terhadap penjualan e-commerce di indonesia menggunakan association rules berdasarkan objek evaluasi. *Majalah Ilmiah UNIKOM*, 189.
- Putri, C. (2018). Perbandingan Loyalitas Toko Antara Pemasaran Online dan Offline Pada Produk High Involvement. *eprints ums*, 1-12.
- Rahayu, N. (2019, Februari). *Pertumbuhan E-Commerce Pesat di Indonesia*. Retrieved from Warta Ekonomi .co.id:  
<https://www.wartaekonomi.co.id/read216302/pertumbuhan-e-commerce-pesat-di-indonesia.html>
- Rosid, M. A., Fitriani, A. S., Astutik, I. R., Mulloh, N. I., & Gozali, H. A. (2019). Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi. *ICETsAS* , 3.
- Rustiana, D., & Rahayu, N. (2017). Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naive Bayes. *Jurnal Simetris*, 113.
- Saleh, A. (2015). IMPLEMENTASI METODE KLASIFIKASI NAIVE BAYES DALAM MEMPEREDIKSI BESARNYA PENGGUNAAN LISTRIK RUMAH TANGGA. *CItec Journal*, 209.
- Shrivastava, V. (2018). A Methodical Study of Web Crawler. *ijera*, 1.

Sianturi, F. A., Hasugian, P. M., Simangunsong, A., & Nadeak, B. (2022). DATA MINING: Teori dan Aplikasi Weka. In *DATA MINING: Teori dan Aplikasi Weka* (p. 21).

Silalahi, M. (2018). ANALISIS CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS TERHADAP PENJUALAN PRODUK PADA PT. BATAMAS NIAGA JAYA. *CBIS Journal*, 25-26.

Suleman, D. (2018). FAKTOR PENENTU KEPUTUSAN KONSUMEN INDONESIA MEMILIH TEMPAT BELANJA DISEBUAH E-COMMERCE. *Journal JDM*, 1-9.

Sysnucleus. (2021). *Web Harvey*. Retrieved from Web Harvey:  
<https://www.webharvy.com/index.html>

Wibowo, A. P., & Jumiati, E. (2018). Sentiment Analysis Masyarakat Pekalongan Terhadap Pembangunan Jalan Tol Pemalang-Batang Di Media Sosial. *IC-Tech*, 42-43.

Widagdo, A. S., Soedijono, B., & Nasiri, A. (2020). Analisis Tingkat Kepopuleran E-Commerce di Indonesia Berdasarkan Sentimen Sosial Media Menggunakan Metode Naive Bayes. *Jurnal INFORMA Politeknik Indonusa Surakarta*, 1.

Widowati, H. (2019, 4 25). *Indonesia Jadi Negara Dengan Pertumbuhan E-Commerce Tercepat di Dunia1*. Retrieved from DataBooks:  
<https://databoks.katadata.co.id/datapublish/2019/04/25/indonesia-jadi-negara-dengan-pertumbuhan-e-commerce-tercepat-di-dunia>

Zulfa, L. (2018). Analisis Pengaruh Presepsi Risiko, Kualitas Situs Web, dan Kepercayaan Konsumen terhadap keputusan pembelian Konsumen E-Commerce Shopee di Kota Semarang. *UEJS*, 4.