

**ANALISIS SENTIMEN ULASAN APLIKASI ONLINE  
TRAVEL AGENT (OTA) DI GOOGLE PLAY STORE  
MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOUR**

**SKRIPSI**

Diajukan untuk memenuhi persyaratan  
dalam memperoleh gelar Sarjana Komputer  
Program Studi Sistem Informasi



**Disusun Oleh:**

**HASTRI CANTYA DANAHISWARI**

**19082010056**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”  
JAWA TIMUR  
SURABAYA  
2023**

**SKRIPSI**  
**ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT  
(OTA) DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-  
NEAREST NEIGHBOUR**

**Disusun Oleh :**  
**HASTRI CANTYA DANAHISWARI**  
**NPM.19082010056**

**Telah dipertahankan di hadapan dan diterima oleh Tim Pengudi Skripsi  
Program Studi Sistem Informasi Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional “Veteran” Jawa Timur  
Pada Tanggal 27 November 2023**

## Dosen Pembimbing:

1.

## Dosen Penguij:

1.

2.

2.

**Amalia Anjani Arifiyanti, S.Kom, M.Kom**  
**NIP.19841201 202121 2 005**

3.1

**Arista Pratama, S.Kom, M.Kom**  
NPT: 171199 10 320052

3

**Anindo Saka Fitri, S.Kom, M.Kom**  
**NPT. 212199 30 325268**

**Mengetahui,  
Dekan Fakultas Ilmu Komputer  
Universitas Pembangunan “Veteran” Jawa Timur**

**Dr. Ir. Ni Ketut Sari, M.T.**  
**NIP.19650731 199203 2 001**

**LEMBAR PENGESAHAN**  
**ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT**  
**(OTA) DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-**  
**NEAREST NEIGHBOUR**

**Disusun Oleh:**

**Hastri Cantya Danahiswari**

**19082010056**

**Telah disetujui mengikuti Ujian Negara Lisan Gelombang November**

**Periode 2023 pada Tanggal 27 November 2023**

**Menyetujui,**

**Dosen Pembimbing 1**

**Dosen Pembimbing 2**

**Eka Dyar Wahyuni, S.Kom. M.Kom**

**Amalia Anjani Arifiyanti, S.Kom, M.Kom**

**NIP. 19841201 2021212 005**

**NIP. 19920812 2018032 001**

**Mengetahui,**

**Ketua Program Studi Sistem Informasi**

**Fakultas Ilmu Komputer**

**Universitas Pembangunan Nasional “Veteran” Jawa Timur**

**Agung Brastama Putra, S.Kom, M.Kom**

**NIP. 19851124 2021211 003**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR  
FAKULTAS ILMU KOMPUTER  
KETERJAGUNG REVISI**

Kami yang bertanda tangan di bawah ini menyatakan bahwa mahasiswa berikut:

Nama : Hastri Cantya Danahiswari

NPM : 19082010056

## Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 27 November 2023 dengan judul:

# ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT (OTA) DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOUR

Oleh karenanya mahasiswa tersebut diatas dinyatakan bebas revisi Ujian Negara Lisan Skripsi dan diijinkan untuk membukukan laporan SKRIPSI dengan judul tersebut.

Surabaya, 5 Desember 2023

Dosen pengujii yang memeriksa revisi:

1. Rizka Hadiwiyanti, S.Kom, M.Kom, MBA {  
NIP 19860727 2018032 001 }

2. Arista Pratama, S.Kom, M.Kom {  
NPT 171199 10 320052 }

3. Anindo Saka Fitri, S.Kom, M.Kom {  
NPT 212199 30 325268 }

### **Mengetahui.**

Dosen Pembimbing 1

Dosen Pembimbing 2

Eka Dyar Wahyuni, S.Kom. M.Kom    Amalia Anjani Arifiyanti, S.Kom, M.Kom  
NIP. 19841201 2021212 005                  NIP. 19920812 2018032 001

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA  
TIMUR**  
**FAKULTAS ILMU KOMPUTER**  
**KETERANGAN REVISI**

---

**SURAT PERNYATAAN**

Saya yang bertanda tangan di bawah ini:

Nama : Hastri Cantya Danahiswari

NPM : 19082010056

Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir sebagai berikut:

**ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT (OTA)  
DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST  
NEIGHBOUR**

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian orang lain dan juga bukan merupakan Produk / Hasil Karya yang Saya beli dari orang lain.

Saya juga menyatakan bahwa Skripsi / Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustakka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institusi Pendidikan lain. Jika ternyata dikemudian hari pernyataan terbukti benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari.

Surabaya, 12 Desember 2023  
**Hormat Saya,**

**Hastri Cantya Danahiswari**

**NPM. 19082010056**

Judul : ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT (OTA) DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOUR

Pembimbing 1 : Eka Dyar Wahyuni, S.Kom. M.Kom

Pembimbing 2 : Amalia Anjani Arifiyanti, S.Kom, M.Kom

---

## ABSTRAK

Di era digital saat ini, berbagai macam aplikasi Online Travel Agent (OTA) di Google Play Store digunakan oleh Masyarakat Indonesia untuk berbagai macam keperluan seperti berlibur atau travelling. Hingga akhir tahun 2022 aplikasi OTA di Google Play Store seperti aplikasi Traveloka dan Tiket.com telah diunduh oleh 50 juta dan 10 juta pengguna. Penelitian ini bertujuan untuk mempermudah, membantu serta memberikan informasi tentang emosi yang terkandung dalam ulasan tentang aplikasi OTA yang ada pada situs Google Play Store dengan melakukan analisis sentimen dengan algoritma K-Nearest Neighbour sebagai algoritma yang optimal dan melakukan visualisasi berbasis web. Hasil akurasi menunjukkan bahwa nilai  $k=6$  mendapat akurasi yang optimal dimana data tidak overfit maupun underfit. Dari 50 data dilakukan pengujian pada sistem visualisasi terdapat 4 data sentimen yang salah dan 46 data sentimen yang benar sehingga memiliki akurasi yang sama, yaitu 92% dan dilengkapi dengan tampilan wordcloud untuk mengetahui kata yang sering muncul dan grafik batang untuk mengetahui distribusi 50 data sentimen yang diuji.

### Kata kunci:

Analisis Sentimen, *Online Travel Agent*, *K-Nearest Neighbour*

## **KATA PENGANTAR**

Segala puji syukur penulis panjatkan atas kehadirat Allah SWT karena berkat limpahan rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul “ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT (OTA) DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOUR” sebagai sasyarat untuk menyelesaikan program studi Sistem Informasi Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Skripsi ini bisa terselesaikan juga berkat dorongan dan bantuan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada pihak yang telah turut serta membantu selama proses penggerjaan skripsi ini. Untuk itu penulis menyampaikan terima kasih kepada:

1. Bapak, Ibu, dan adik yang mendukung dan memberikan doa agar dimudahkan dalam menimba ilmu terutama saat pembuatan skripsi ini.
2. Ibu Eka Dyar Wahyuni, S.Kom. M.Kom selaku pembimbing I dan Ibu Amalia Anjani Arifiyanti, S.Kom, M.Kom selaku pembimbing II yang telah memberikan bimbingan dan arahan terkait penelitian ini.
3. Bapak Mohammad Irwan Afandi, S.T, M. Sc selaku dosen wali yang selalu membantu saya selama kuliah dalam berbagai hal.
4. Seluruh dosen Sistem Informasi yang telah banyak memberikan ilmu yang sangat bermanfaat selama kuliah dan untuk bekal di masa depan.
5. Eric, Feby, Ale, dan rekan – rekan kantor PT. Pupuk Indonesia (Persero) yang selalu menyemangati, mendukung, dan membantu dalam proses penggerjaan skripsi.

6. Seluruh teman-teman Sistem Informasi angkatan 2019 yang telah memberi semangat untuk menyelesaikan skripsi ini, terutama untuk kelas A.
7. Seluruh pihak yang tidak bisa saya sebutkan satu persatu karena telah memberikan bantuan sekecil apapun.

Semoga Allah SWT membala kebaikan pihak-pihak diatas. Dalam penggeraan skripsi ini penulis menyadari masih banyak sekali kekurangan. Semoga skripsi ini bermanfaat bagi semua pihak.

## **DAFTAR ISI**

ABSTRAK.....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiv
BAB I            PENDAHULUAN .....	2
1.1    Latar Belakang .....	2
1.2    Rumusan Masalah.....	6
1.3    Batasan Masalah .....	7
1.4    Tujuan Penelitian .....	7
1.5    Manfaat Penelitian .....	8
1.6    Sistematika Penulisan .....	8
BAB II            TINJAUAN PUSTAKA.....	10
2.1    Dasar Teori.....	10
2.1.1    Text Mining .....	10
2.1.2    Preprocessing Text.....	11
2.1.3    Analisis Sentimen .....	11
2.1.4    CRISP-DM.....	13
2.1.5    Vader.....	14
2.1.6    Roberta.....	14
2.1.7    Tiket.com .....	15
2.1.8    Traveloka .....	15
2.1.9    Google Play Store .....	16

2.1.10	K-Nearest Neighbour .....	16
2.1.11	Confusion Matrix .....	17
2.2	Penelitian Terdahulu .....	20
<b>BAB III</b>	<b>METODOLOGI PENELITIAN.....</b>	<b>24</b>
3.1	Business Understanding.....	25
3.2	Data Understanding.....	25
3.3	Data Preparation.....	26
3.4	Pemodelan (Modelling).....	29
3.5	Evaluasi (Evaluation).....	29
3.6	Deployment (Visualisasi Web) .....	30
<b>BAB IV</b>	<b>KESIMPULAN DAN SARAN .....</b>	<b>33</b>
4.1	Business Understanding.....	33
4.1.1	Kebutuhan Hardware dan Software .....	33
4.2	Data Understanding.....	34
4.2.1	Pengumpulan Data (Web Scrapping).....	34
4.3	Data Preparation.....	39
4.3.1	Merge CSV Ulasan Aplikasi Traveloka dan Tiket.com.....	39
4.3.2	Filtering Kata Traveloka dan Tiket.com .....	40
4.3.3	Cleansing.....	41
4.3.4	Case Folding .....	43
4.3.5	Tokenization .....	44
4.3.6	Stopword.....	45
4.3.7	Stemming .....	46

4.3.8	Labelling Menggunakan Library Roberta.....	47
4.3.9	Labelling Menggunakan Library Vader.....	48
4.3.10	Labelling Manual .....	51
4.3.11	Perbandingan Labelling .....	54
4.3.12	EDA (Explatory Data Analysis) .....	55
4.4	Pemodelan (Modelling).....	60
4.4.1	Hold Out.....	60
4.4.2	Pemodelan Dengan Algoritma K-Nearest Neighbour.....	60
4.5	Evaluasi.....	62
4.5.1	Confusion Matrix .....	62
4.5.2	Grafik Evaluasi Model KNN .....	63
4.6	Deployment (Web).....	65
4.6.1	Import Model Ke Dalam Bentuk .pkl (berkas) .....	66
4.6.2	Pembuatan Page HTML Menggunakan VsCode .....	66
4.6.3	Import dan build menggunakan flask.....	70
4.6.4	Deployment dan Visualisasi.....	71
BAB V	PENUTUP .....	74
5.1	Kesimpulan .....	74
5.2	Saran .....	75
LAMPIRAN	.....	77
DAFTAR PUSTAKA	.....	82

## DAFTAR GAMBAR

Gambar 1.1 Online Travel Agencies Terpopuler (Katadata, 2022) .....	2
Gambar 1.2 Ulasan Tiket.com di Google Play Store .....	4
Gambar 1.3 Ulasan Traveloka di Google Play Store .....	5
Gambar 2.1 Tahapan dalam analisis sentimen.....	12
Gambar 3.1 Alur Metodologi Penelitian.....	24
Gambar 3.1 Metodologi Penelitian CRISP-DM .....	24
Gambar 3.2 Alur Preprocessing Data .....	26
Gambar 3.3 Ilustrasi Visualisasi Web.....	32
Gambar 4.1 Source Code Pengumpulan Data Ulasan Tiket.com .....	34
Gambar 4.2 Source Code Pengumpulan Data Ulasan Traveloka .....	35
Gambar 4.3 Merge CSV Ulasan Aplikasi.....	39
Gambar 4.4 Filtering Kata Traveloka dan Tiketcom .....	40
Gambar 4.5 Source Code Cleansing .....	42
Gambar 4.6 Source Code Case Folding .....	43
Gambar 4.7 Source Code Tokenization .....	44
Gambar 4.8 Source Code Stopword.....	46
Gambar 4.9 Source Code Stemming.....	46
Gambar 4.10 Source Code Library Roberta.....	47
Gambar 4.11 Source Code Library Translator.....	49
Gambar 4.12 Source Code Library Vader .....	50
Gambar 4.13 Perbandingan data label positive dan negative .....	56
Gambar 4.14 Flow Metode Random Undersampling .....	57
Gambar 4.15 Flow Metode Random Oversampling .....	58
Gambar 4.16 Word Cloud Ulasan OTA.....	58
Gambar 4.17 Word Cloud Positive Ulasan OTA.....	59
Gambar 4.18 Word Cloud Negative Ulasan OTA .....	59
Gambar 4.19 Source Code Word Cloud .....	59
Gambar 4.20 Source Code Hold Out .....	60
Gambar 4.21 Source Code Confusion Matrix.....	62
Gambar 4.22 Confusion Matrix .....	62
Gambar 4.23 Source Code Evaluasi Pengujian Model .....	63
Gambar 4.24 Grafik Evaluasi Model KNN .....	64
Gambar 4.26 Source Code Import Model ke dalam bentuk .pkl.....	66
Gambar 4.27 Source Code Halaman Index.....	67
Gambar 4.28 Source Code Halaman Upload .....	68
Gambar 4.29 Source Code Halaman Script .....	69

Gambar 4.30 Source Code Halaman Head .....	69
Gambar 4.31 Source Code Halaman Navbar .....	69
Gambar 4.32 Source Code app.py .....	70
Gambar 4.33 Tampilan Halaman Index.....	71
Gambar 4.34 Tampilan Halaman Upload Sebelum Submit.....	72
Gambar 4.35 Tampilan Halaman Upload Sesudah Submit .....	72

## **DAFTAR TABEL**

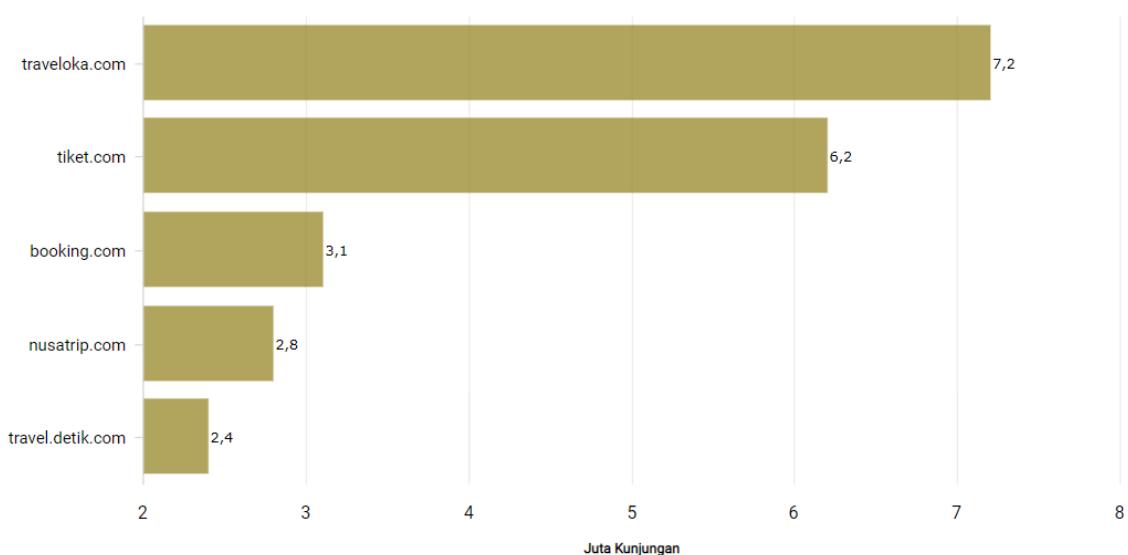
Tabel 2.1 Confusion Matrix .....	18
Tabel 2.2 Rumus Pengukuran Evaluasi .....	18
Tabel 2.3 Perhitungan Confusion Matrix.....	19
Tabel 2.4 Tabel Penelitian Terdahulu .....	20
Tabel 4.1 Tabel Ulasan Aplikasi Tiket.com .....	37
Tabel 4.2 Tabel Ulasan Aplikasi Tiket.com .....	39
Tabel 4.3 Proses Cleansing .....	43
Tabel 4.4 Proses Case Folding.....	44
Tabel 4.5 Proses Tokenization.....	45
Tabel 4.6 Proses Stopword .....	46
Tabel 4.7 Proses Stemming.....	47
Tabel 4.8 Contoh Hasil Labelling Roberta .....	48
Tabel 4.9 Proses Translate Sebelum Labelling Library Vader .....	50
Tabel 4.10 Contoh Hasil Labelling Roberta .....	51
Tabel 4.11 Contoh Hasil Menggunakan Library Roberta.....	52
Tabel 4.12 Contoh Hasil Menggunakan Library Vader.....	53
Tabel 4.13 Perbandingan Hasil Labelling Library Roberta dan Vader.....	55
Tabel 4.14 Hasil Pengujian dan Perbandingan Skenario Ulasan .....	61
Tabel 4.15 Hasil Percobaan Nilai K Skenario Oversampling.....	65
Tabel 4.16 Pengujian Pada Sistem Dengan Hasil Salah .....	73
Tabel 4.17 Pengujian Pada Sistem Dengan Hasil Benar.....	73

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Di era digital saat ini, berbagai macam aplikasi Online Travel Agent (OTA) di Google Play Store digunakan oleh masyarakat Indonesia untuk berbagai macam keperluan seperti berlibur atau *travelling*. Masing – masing aplikasi Online Travel Agent (OTA) memiliki produk dan fitur tersendiri yang membedakan antar aplikasi. Adapun, aplikasi OTA terpopuler di Indonesia, yaitu Traveloka dan Tiket.Com berdasarkan jumlah data pengunjung aplikasi tersebut.



Gambar 1.1 Online Travel Agencies Terpopuler (Katadata, 2022)

Agar konsumen mendapatkan pelayanan terbaik terhadap performa aplikasinya maka, penting untuk meningkatkan pelayanan kepada pengguna saat memesan tiket pada aplikasi tersebut dengan fokus pada kebutuhan dan emosi pengguna ketika menggunakan aplikasi tersebut sebagai salah satu kunci keberhasilan sebuah aplikasi (Maharani, 2022).

Traveloka adalah perusahaan travel yang terdapat aplikasi mobile yang mempermudah dalam pemesanan tiket pesawat secara online, serta menyediakan bermacam kebutuhan dalam perjalanan di satu platform, seperti tiket pesawat, tiket kereta, rekreasi dan lainnya. Traveloka telah bekerjasama lebih 100 maskapai domestic dan juga internasional, melayani lebih 200.000 rute penerbangan di seluruh dunia. Aplikasi mobile Traveloka telah diunduh 30 juta kali, sehingga aplikasi ini menjadi paling popular di Asia Tenggara (Traveloka, 2021).

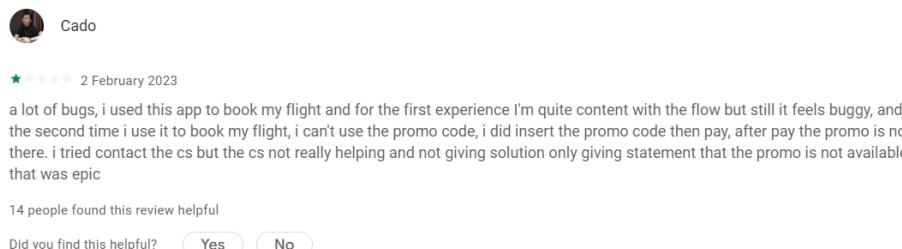
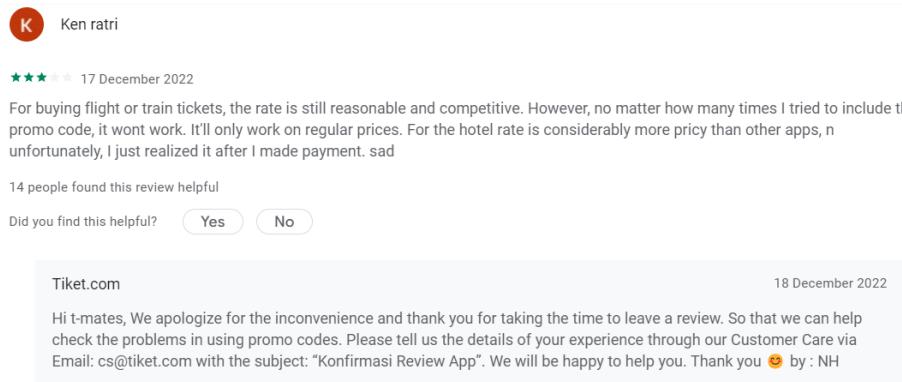
Tiket.com merupakan OTA pertama di Indonesia berdiri tahun 2011, layanan Tiket.com memanfaatkan perkembangan dunia digital untuk memudahkan pemesanan tiket pesawat, cek harga maupun promo pada tiket pesawat dari mana saja dan kapan saja, Tiket.com menjual tiket pesawat dengan berbagai macam maskapai (Tiket.com, 2021). Bantuan teknologi informasi pada pemesanan tiket dapat meringkas yakni biaya dan waktu yang diperlukan saat pemesanan dengan tiket konvensional. Tiket online mengalami perkembangan dari tahun sebelumnya yakni kenaikan mencapai 10,4% menjadi \$694.41 billion (Emarketer, 2018).

Selama ini sebuah predikat untuk aplikasi terbaik diberikan kepada aplikasi dengan jumlah unduhan dan pemberian bintang paling tinggi di Play Store (Erfina et al., 2020). Berdasarkan profil tiket.com dalam Google Play Store bahwa hingga akhir 2022 aplikasi tiket.com telah digunakan oleh lebih dari 10 juta pengguna dan mendapatkan rating 4.3 di Google Play Store. Begitu juga dengan kompetitornya, Traveloka hingga akhir 2023 telah digunakan lebih dari 50 juta pengguna dan mendapatkan rating 4.8. Data tersebut menunjukkan Tiket.com dan Traveloka merupakan aplikasi OTA yang cukup diminati. Akan tetapi, suatu aplikasi selalu memiliki kekurangan dan kelebihan masing-masing yang dapat menimbulkan

berbagai respon dari pengguna aplikasi seperti kepuasan dan kekecewaan terhadap aplikasi tersebut (Giovani et al., 2020).

Kolom ulasan komentar terhadap sebuah aplikasi di Google Play menjadi salah satu tempat untuk melontarkan kepuasaan dan kekecewaan pengguna atau opini terhadap aplikasi tersebut. Hal tersebut dapat dijadikan bahan analisis sentimen terhadap aplikasi tiket.com untuk mengetahui apakah jumlah unduhan dan rating aplikasi tiket.com relevan untuk membuktikan bahwa kolom ulasan komentar merupakan tempat para pengguna aplikasi dalam melontarkan kepuasan dan kekecewaan.

Pada aplikasi pemesanan tiket online seperti Traveloka dan tiket.com terdapat berbagai macam ulasan dari pengguna baik yang bersifat positif, negative, maupun netral yang tertera pada gambar 1.2 dan gambar 1.3.



Gambar 1.2 Ulasan Tiket.com di Google Play Store

The screenshot displays two user reviews for the Traveloka app on the Google Play Store. The first review is from a user named Roxane Varona, posted on 2 March 2023. She complains about price changes and misleading information from a representative named Rose. The second review is from a user named senti senti, posted on 1 March 2023, expressing frustration with frequent push notifications for redeemed Wuling cars. Both reviews include a 'Helpful' count (30 for Roxane's, 18 for senti senti's) and 'Yes' or 'No' buttons for feedback. Below each review is a reply from Traveloka, apologizing for the inconvenience and providing contact information.

Gambar 1.3 Ulasan Traveloka di Google Play Store

Pengungkapan emosi seseorang banyak diekspresikan dalam bentuk tulisan termasuk ungkapan kepuasan maupun kekecewaan yang disampaikan oleh pengguna aplikasi di kolom komentar. Oleh karena itu, proses analisis sentimen yang dilakukan pada penelitian ini menggunakan teknik text mining. Text mining merupakan proses penambangan data (data mining) yang mengambil informasi dari data berupa teks seperti email, dokumen teks, file HTML serta data dari ulasan komentar (Sepri, 2020).

Untuk melakukan analisis sentimen diperlukan teknik klasifikasi dokumen yang memiliki tingkat akurasi yang tinggi. Ada beragam teknik klasifikasi dokumen, beberapa diantaranya yaitu Decision Trees, Naïve Bayes Classifier (NBC), K-Nearest Neighbor, dan Support Vector Machines (SVM). K-Nearest Neighbor (K-NN) merupakan salah satu metode yang dapat digunakan untuk menganalisis sentimen terhadap sebuah produk/aplikasi (Ernawati & Wati, 2018).

Algoritma K-NN merupakan metode lazy learning dimana tidak ada model yang dipelajari dari data testing, jadi hanya belajar dari contoh uji harus diklasifikasikan (Sari, 2020). KNN menghitung kedekatan antara kasus baru (data test) dan kasus lama (data training) yang berdasarkan pencocokan bobot dari sejumlah fitur yang ada. Setelah jarak terhitung, maka jarak yang paling dekat dengan data latih dianggap memiliki kesamaan (Marisa et al., 2021). Algoritma ini mengklasifikasikan data uji yang belum diketahui kategorinya terhadap data latih yang telah diketahui kategorinya dengan menggunakan rumus euclidean dalam dimensi sebesar  $k$  yang mengelilingi data uji (Aisha et al., 2021).

Aplikasi tiket.com sebagai start-up dalam bidang teknologi OTA pernah dijadikan sebagai bahan studi kasus untuk menilai kesuksesan aplikasi (Maharani, 2022). Adapun beberapa contoh penelitian mengenai analisis sentimen terhadap sebuah produk atau aplikasi yang dijadikan referensi dalam penelitian ini belum ada yang membahas mengenai analisis sentimen pada aplikasi android OTA (Online Travel Agent) (Erfina et al., 2020; Ernawati & Wati, 2018; Giovani et al., 2020; Sepri, 2020). Maka dari itu, hasil dari penelitian ini diharapkan dapat mempermudah, membantu serta memberikan informasi tentang emosi yang terkandung dalam ulasan tentang aplikasi OTA yang ada pada situs Google Play Store.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijabarkan sebelumnya maka dapat dirumuskan beberapa permasalahan seperti:

1. Bagaimana melakukan analisis sentimen pada ulasan pengguna aplikasi OTA di google play store dari kolom komentar?

2. Bagaimana menerapkan algoritma K-Nearest Neighbour sebagai algoritma yang optimal?
3. Bagaimana memvisualisasikan hasil dari klasifikasi ke dalam bentuk website?

### **1.3 Batasan Masalah**

Dalam penyusunan penelitian ini terdapat beberapa batasan masalah, diantaranya sebagai berikut:

1. Data yang digunakan dalam penelitian ini merupakan dari aplikasi google play store dengan akun publik.
2. Penelitian ini hanya melakukan proses analisis sentiment terhadap ulasan komentar yang menggunakan Bahasa Indonesia dan pengguna Indonesia.
3. Sentimen yang akan diambil dua, yaitu positif dan negatif mengikuti ketersediaan sentiment oleh library.
4. Data yang diambil yaitu ulasan dari google play store aplikasi Tiket.com dan Traveloka dengan jumlah 7000 untuk tiap studi kasus.
5. Bahasa pemrograman yang digunakan adalah bahasa pemrograman Phyton.

### **1.4 Tujuan Penelitian**

Berdasarkan dari rumusan masalah yang sudah dijelaskan, tujuan dari penelitian ini adalah:

1. Mengetahui tanggapan pengguna aplikasi *Online Travel Agent (OTA)*, yaitu Traveloka dan Tiket.com.
2. Menerapkan skenario algoritma K-Nearest Neighbour sebagai algoritma yang optimal.

3. Menggunakan library Roberta dan Vader dalam melakukan labelling sentimen.
4. Memvisualisasikan hasil klasifikasi ke dalam bentuk website.

## **1.5 Manfaat Penelitian**

Berdasarkan hasil penelitian analisis sentiment yang dilakukan, hasil dari penelitian ini diharapkan dapat mempermudah, membantu serta memberikan informasi tentang emosi yang terkandung dalam ulasan tentang aplikasi OTA yang ada pada situs Google Play Store.

## **1.6 Sistematika Penulisan**

Sistematika penulisan skripsi ini berguna untuk membantu mengarahkan penyusunan laporan skripsi agar tidak menyimpang serta sebagai acuan dalam mencapai tujuan penulisan laporan skripsi sesuai dengan apa yang diharapkan. Tahapan-tahapan dalam proses penyusunan skripsi ini adalah sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini berisi tentang gambaran umum penelitian yang akan dibuat dan berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan dan sistematika penulisan penelitian.

### **BAB II TINJAUAN PUSTAKA**

Bab ini berisi tentang penelitian yang sudah pernah dilakukan sebelumnya yang bertujuan untuk membandingkan penelitian ini dengan penelitian sebelumnya. Penelitian yang dibahas adalah penelitian yang memiliki keterkaitan dengan

penelitian ini seperti *text mining*, *preprocessing text*, *analisis sentiment*, *vader*, *tiket.com*, *google play store*, *naïve bayes*, *k-nearest neighbour*, *TF-IDF*, *confusion matrix*.

### **BAB III METODOLOGI PENELITIAN**

Bab ini berisi tentang metode penelitian yang akan dibuat yaitu meliputi studi literatur, analisis kebutuhan, dan perancangan model dan visualisasi.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini berisi tentang penjelasan penelitian mengenai pengumpulan data, pembangunan model, pengujian sistem, dan membahas hasil pengujian terhadap implementasi model pada sistem.

### **BAB V PENUTUP**

Bab ini berisi kesimpulan dari seluruh hasil penelitian yang telah dilakukan dan saran terhadap penelitian selanjutnya.

### **DAFTAR PUSTAKA**

Bagian ini berisi tentang literatur yang digunakan untuk membantu penyusunan skripsi ini.

### **LAMPIRAN**

Bagian ini berisi tentang data maupun pelengkap yang menunjang penyusunan skripsi ini.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Dasar Teori**

Penelitian ini dibuat dengan menyertakan teori-teori dasar yang berkaitan dengan ruang lingkup permasalahan yang ada. Teori-teori yang disertakan antara lain *text mining, preprocessing text, analisis sentiment, vader, Roberta, Tiket.com, Traveloka, google play store, k-nearest neighbour, TF-IDF, confusion matrix.*

##### **2.1.1 Text Mining**

Text mining merupakan hubungan antar pengguna dengan berbagai macam dokumen dari waktu ke waktu dengan menggunakan metode untuk melakukan analisis (Feldman and Sanger, 2007). Pengertian text mining lainnya yaitu analisis teks yang melingkupi sebuah proses pengolahan informasi yang terpola dan berasal dari sumber data teks, seperti kutipan teks dan sms (tweet) (Hartanto, 2017). Dengan kata lain alur dari text mining yaitu melakukan penambangan data berupa teks dari data yang tidak terstruktur kemudian dicari kata-kata yang sering dibahas pada suatu dokumen untuk dapat menganalisis inti dari dokumen tersebut. Pada proses dilakukannya text mining melibatkan penggunaan Pemrosesan Bahasa Alami (Neuro-Linguistic Programming) atau NLP. Text mining dapat digunakan untuk mengidentifikasi informasi dari media sosial karena text mining memiliki kemampuan untuk mengambil data atau informasi tidak terstruktur dalam jumlah besar yang dapat ditindaklanjuti untuk mencapai suatu tujuan. Saat ini text mining menjadi langkah awal yang sangat penting dalam melakukan prosedur analisis

suatu data atau dokumen. Salah satu langkah awal yang dilakukan pada text mining sebelum menganalisis data adalah melakukan preprocessing text.

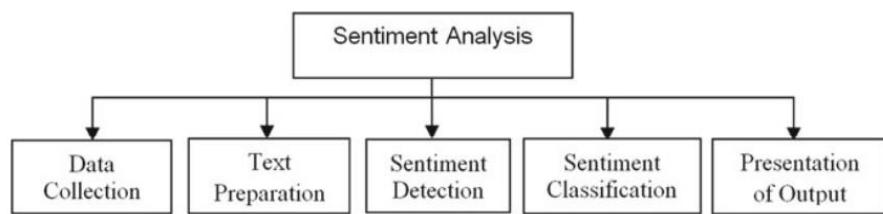
### **2.1.2 Preprocessing Text**

Pada tahap Preprocessing adalah proses perubahan bentuk data yang terstruktur sembarang menjadi data yang terstruktur sesuai kebutuhan untuk proses dalam text mining (Amrullah et al, 2020). Tahapan Preprocessing yang digunakan antara lain: Case Folding yaitu tahap mengubah semua huruf dalam dokumen menjadi huruf kecil dan karakter non-huruf akan dihilangkan, Tokenizing yang merupakan tahap pemotongan string input berdasarkan tiap kata yang menyusunnya (Afuan, 2018) dan secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata, Filtering dengan mengambil kata-kata penting dari hasil tahap tokenizing yang dapat dilakukan dengan menghilangkan stoplist atau stopword (Sugianto & Apandi, 2018), dan Stemming yang mentransformasi suatu kata menjadi kata dasarnya (root word) dengan menggunakan aturan-aturan tertentu.

### **2.1.3 Analisis Sentimen**

Analisis sentimen merupakan sebuah proses dari eksplorasi data yang luas yang umumnya didapat dari *web* untuk mengidentifikasi dan mengkategorikan pandangan pada suatu bagian dari teks. Luaran yang diharapkan dari proses ini yaitu menilai perilaku dari penulis atau narasumber suatu teks terhadap topik tertentu, baik itu film, produk, dll. Hasil dari analisis sentimen berupa kategori positif, negatif, atau netral.

Analisis sentimen umumnya mencakup 5 tahapan untuk memproses data; dimulai dengan pengumpulan data (*data collection*), penyiapan data (*text preparation*), pendekripsi sentimen (*sentiment detection*), klasifikasi sentimen (*sentiment classification*), dan diakhiri dengan presentasi dari hasil akhir (Kotsiantis, 2006).



Gambar 2.1 Tahapan dalam analisis sentimen

### 1. Data Collection

Tahapan pertama dalam analisis sentimen merupakan *data collection*. Pengumpulan data didapat dari bermacam-macam sumber, seperti Twitter, Facebook, Youtube, *blog*, *web* komersial, dll. Dalam laman resmi Twitter Developer, Twitter menyediakan akses khusus untuk mendapatkan data seperti *tweet* dan Mutual Information (MI) secara legal. (Ratnawati, 2018)

### 2. Text Preparation

Dalam tahapan ini, *dataset* diperiksa dan dibersihkan dari *spam* atau kata-kata yang bernilai terlalu buruk agar dalam proses selanjutnya hasil dapat lebih *reliable* (Etaiwi dan Naymat, 2017).

### 3. Sentiment Detection

*Sentiment detection* merupakan proses untuk mencari sentimen yang terekspresikan dari suatu *row* teks dengan menggunakan pendekatan *machine learning*, atau Natural Language Processing (NLP). Deteksi sentimen terdiri dari

pemeriksaan frasa dan kalimat yang diambil dari ulasan atau ide. Semua kalimat yang mengandung ekspresi seperti keyakinan, opini, dan bahkan sesuatu yang merendahkan juga dipertahankan sebagai *key feature* (Alasmari dan Dahab, 2017).

#### **4. *Sentiment Classification***

*Sentiment classification* merupakan tahapan untuk mengekstraksi dan mengklasifikasikan teks. Objektif dari tahapan ini adalah mengklasifikasikan sebuah baris teks memiliki kecenderungan baik dari sisi positif atau negatif, baik atau buruk, maupun intensi suka atau tidaknya terhadap suatu topik (Aqlan dkk, 2019).

#### **5. *Presentation of Output***

Pada tahapan ini, hasil *output* dari keseluruhan proses dirancang sesederhana mungkin agar saat disampaikan pada penerima hasil, informasinya dapat dicerna dengan mudah dan baik.

#### **2.1.4 CRISP-DM**

Cross-Industry Standard Process for Data Mining atau CRISP-DM adalah salah satu model proses datamining (datamining framework) yang awalnya (1996) dibangun oleh 5 perusahaan yaitu Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation dan OHRA. Framework ini kemudian dikembangkan oleh ratusan organisasi dan perusahaan di Eropa untuk dijadikan methodology standard non-proprietary bagi data mining. Versi pertama dari metodologi ini dipresentasikan pada 4th CRISP-DM SIG Workshop di Brussels pada bulan Maret 1999 (Pete Chapman, 1999); dan langkah langkah proses datamining berdasarkan model ini di publikasikan pada tahun berikutnya (Pete Chapman, 2000).

### **2.1.5      Vader**

Valence Aware Dictionary and sEntiment Reasoner (Vader) adalah alat analisis sentimen yang modelnya berbasis simple rule-based dan lexicon, disesuaikan secara khusus untuk ekspresi sentimen yang ada pada media sosial. Dibandingkan dengan alat analisis sentimen berbasis lexicon yang sudah populer seperti Linguistic Inquiry Word Count (LIWC), General Inquirer (GI), Affective Norms for English Words (ANEW), SentiWordNet (SWN), SenticNet (SCN), Word-Sense Disambiguation (WSD) menggunakan WordNet, dan Hu-Liu04 opinion lexicon, Vader memiliki hasil yang cenderung lebih baik karena Vader juga menimbang sentimen dari fitur bawaan natural dari media sosial termasuk emoticon, kata yang disingkat, akronim, slang, dll (Hutto dan Gilbert, 2014).

### **2.1.6      Roberta**

RoBERTa merupakan penyempurnaan model BERT asli bersama dengan manipulasi data dan input. RoBERTa sangat bergantung pada jumlah teks. Jika semakin banyak data maka akan semakin meningkat kinerjanya. Pengklasifikasi Sentimen IndoLEM Basis RoBERTa Indonesia adalah model klasifikasi teks sentimen berdasarkan model RoBERTa. Model ini awalnya merupakan model RoBERTa Base Indonesia yang telah dilatih sebelumnya, yang kemudian disempurnakan berdasarkan kumpulan data Analisis Sentimen Indolem yang terdiri dari tweet berbahasa Indonesia dan ulasan hotel (Fajri K, et al., 2020).

### **2.1.7      Tiket.com**

Tiket.com merupakan perusahaan yang pada awalnya hanya menyediakan website dalam transaksi pembelian tiket pesawat dan hotel saja, Hingga pertumbuhannya pada bulan Juli 2014 Tiket.com meluncurkan mobile apps dan sudah diunduh sebanyak 1,7 juta kali. Sehingga akhir 2016 Blibli mengakuisisi Tiket.com dengan tujuan bisa menumbuhkan Tiket.com. Meskipun begitu, website Tiket.com masih bisa diakses dan sudah diselaraskan dengan versi aplikasinya sehingga lebih user friendly dibandingkan sebelumnya. (Tiket.com, 2022). Pada tahun 2018 merupakan tahun dimana Tiket.com mengalami pertumbuhan yang sangat signifikan karena telah mencatat rekor dengan berhasil melakukan transaksi sebanyak 8 juta kali mengalahkan pesaingnya (Haryanto, 2017). Namun, dengan adanya pandemi Covid-19 ini membuat bisnis Travel Agent, Maskapai penerbangan, dan sektor wisata mengalami penurunan secara drastis karena setiap harinya penjualan terkait kebutuhan perjalanan ini semakin dihindari karena masyarakat takut dengan adanya pandemi, peraturan pemerintah yang mengharuskan adanya pemeriksaan masyarakat sebelum melakukan perjalanan, serta adanya pembatasan dalam kapasitas penumpang untuk transportasi maupun sektor wisata sehingga hal ini sangat berdampak pada pembelian travel agent yang akan semakin sepi. Hal ini disetujui oleh World Travel & Tourism Council (WTTC) dimana industri ini akan pulih pada tahun 2022 (Prasetio dan Wita, 2022).

### **2.1.8      Traveloka**

Traveloka adalah perusahaan yang bergerak pada bidang pelayanan pemesanan tiket pesawat, kereta api, bus dan travel serta reservasi hotel secara online. Traveloka pertama kali didirikan oleh Ferry Unardi, Derianto Kusuma, dan

Albert Zhang yang pada awalnya hanya sebuah konsep untuk melihat dan membandingkan harga. Peluncuran aplikasi traveloka pada 31 Juli 2014 bertujuan untuk menghemat waktu dalam melakukan pemesanan tiket (Traveloka, 2021).

### **2.1.9 Google Play Store**

Google Play menyajikan koleksi aplikasi dan game yang dipersonalisasi kepada pengguna berdasarkan kriteria seperti aktivitas pengguna yang lalu, tindakan yang ingin mereka selesaikan, lokasi, dan peristiwa utama. Koleksi ini dipilih secara otomatis dan oleh tim editorial Google Play untuk memastikan koleksi menyertakan aplikasi dan game terbaik yang ditawarkan. Hal yang tidak kalah penting 14 pada google play adalah adanya fitur ulasan yang berisi review dari para pengguna.

### **2.1.10 K-Nearest Neighbour**

Algoritma K-Nearest Neighbor merupakan sebuah algoritma yang sering digunakan untuk klasifikasi teks dan data. Tujuan dari algoritma ini adalah mengklasifikasikan objek berdasarkan atribut dan training sample. Classifier tidak menggunakan apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik query, akan ditemukan sejumlah  $k$  obyek atau (titik training) yang paling dekat dengan titik query. Klasifikasi menggunakan voting terbanyak antara klasifikasi dari  $k$  obyek. Algoritma K-Nearest Neighbor (K-NN) menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari query instance yang baru (Lidya, Opim, & Syahril, 2015). KNN digunakan untuk mengklasifikasi dokumen dengan cara menghitung jarak antara dokumen dan semua dokumen dalam pelatihan yang ditentukan menggunakan ukuran variasi atau kesamaan. Kemudian menemukan

tetangga k terdekat dan menentukan dokumen ke dalam kategori dengan jumlah tetangga terdekat yang terbanyak. Nilai k dari algoritma KNN adalah faktor menunjukkan jumlah dokumen training terdekat dengan dokumen testing. Klasifikasi menentukan jarak vektor antar dokumen menggunakan persamaan.

(Alfani, A et al, 2021)

$$d(x, y) = \sqrt{\sum_{r=1}^n (a_{rx} - a_{ry})^2}$$

Dimana  $d(x,y)$  adalah jarak antara dua dokumen yang dinamakan jarak euclidean. Pengukuran jarak euclidean adalah salah satu metode pengukuran yang umum digunakan.  $a_{rx}$  adalah bobot term r dalam dokumen x,  $a_{ry}$  adalah bobot term r dalam dokumen y. (Cahyo, J 2022).

Metode klasifikasi algoritma k-nearest neighbor merupakan salah satu metode pengklasifikasian data yang memiliki konsistensi yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot (Kusrini dan Lutfhi, 2009). Algoritma ini lebih efektif dalam melakukan training data yang besar dan dapat menghasilkan data yang lebih akurat (Sumarlin, 2015).

### 2.1.11 Confusion Matrix

Confusion Matrix atau juga bisa disebut matriks klasifikasi merupakan suatu metode perhitungan yang biasa digunakan dalam supervised learning (Manning, Raghavan and Schütze, 2009). Setelah submit ulasan komentar aplikasi tiket.com di google play store, sistem membutuhkan ukuran untuk menentukan seberapa tepat klasifikasi telah dibuat oleh sistem. Akurasi yang baik dalam pembelajaran mesin bersifat subjektif. Namun menurut obviously.ai, angka yang

lebih besar dari 70% adalah performa model yang hebat. Faktanya, ukuran akurasi antara 70%-90% tidak hanya ideal, tetapi juga realistik. Hal ini juga konsisten dengan standar industri. (Obviously.ai, 2022)

Tabel 2.1 Confusion Matrix

Classification	Predicted Positives	Predicted Negatives
Actual Positive Cases	<b>True Positive (TP)</b>	<b>False Negative (FN)</b>
Actual Negative Cases	<b>False Positive (FP)</b>	<b>True Negative (TN)</b>

Dimana:

**True Positive (TP):** Ketika prediksi positif dan hasilnya benar positif.

**True Negative (TN):** Ketika prediksi positif dan hasilnya benar negative.

**False Positive (FP):** Ketika prediksi negative dan hasilnya benar positif.

**False Negative (FN):** Ketika prediksi negative dan hasilnya benar negative.

Tabel 2.2 Rumus Pengukuran Evaluasi

Pengukuran	Formula
<i>Accuracy, recognition rate</i>	$\frac{TP + TN}{P + N}$
<i>Error rate, misclassification rate</i>	$\frac{FP + FN}{P + N}$
<i>Sensitivity, true positive rate, recall</i>	$\frac{TP}{P}$
<i>Specificity, true negative rate</i>	$\frac{TN}{N}$
<i>Precision</i>	$\frac{TP}{TP + FP}$

Dimana:

**Accuracy:** Merupakan rasio dari subjek yang diberi label dengan benar ke seluruh kumpulan subjek.

**Recall/Sensitivity:** Merupakan rasio dari label positif yang benar oleh program untuk semua penderita pada kenyataannya

**Specificity:** Merupakan rasio dari label negatif yang benar oleh program dengan semua orang yang sehat dalam kenyataan.

**Precision:** Merupakan rasio positif yang diberi label dengan benar oleh program dengan semua label positif.

### 1. Studi Kasus Perhitungan Confusion Matrix Ulasan Komentar

Dari hasil membuat sebuah model yang telah dilatih untuk memprediksi apakah komentar tersebut positif atau negative. Dengan asumsi jumlah komentar sebanyak 1000 komentar. Dari model classifier memprediksi komentar positif sebanyak 715 dan komentar negative sebanyak 285. Namun, pada kenyataanya 700 komentar positif dan 300 komentar negatif.

Classification N=1000	Predicted Negatives	Predicted Positives	Total
Actual Negative Cases	270 (TN)	15 (FP)	285
Actual Positive Cases	15 (FN)	700 (TP)	715
<b>Total</b>	<b>285</b>	<b>715</b>	<b>1000</b>

Tabel 2.3 Perhitungan Confusion Matrix

- Accuracy

$$\begin{aligned}
 &= (TP + TN) / (P + N) \\
 &= (700 + 270) / (1000) \\
 &= 0.97
 \end{aligned}$$

- Error Rate

$$\begin{aligned}
 &= (FP + FN) / (P + N) \\
 &= (15 + 15) / 1000 \\
 &= 0.03
 \end{aligned}$$

- Sensitivity

$$\begin{aligned}
 &= TP / P \\
 &= 700 / 715
 \end{aligned}$$

$$= 0.98$$

- Specificity

$$= TN / N$$

$$= 270 / 285$$

$$= 0.95$$

- Precision

$$= TP / (TP + FP)$$

$$= 700 / (700 + 15)$$

$$= 0.98$$

## 2.2 Penelitian Terdahulu

Tabel 2.4 Tabel Penelitian Terdahulu

Peneliti	Tahun	Judul	Tujuan	Metode	Hasil
Sri Rahayu, Yumarlin MZ, Jemmy Edwin Bororing, dan Rahmat Hadiyat	2022	Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP	Menganalisis sentimen pengguna terhadap aplikasi Flip sehingga dapat diketahui apakah ulasan pengguna aplikasi Flip juga menunjukkan nilai positif seperti rating yang diterima	Penelitian ini menggunakan rangkaian proses text mining terhadap data ulasan pengguna aplikasi Flip di Google Play Store dan algoritma klasifikasi yang digunakan adalah K-Nearest Neighbor dengan pembobotan TF-IDF	Hasil penelitian menunjukkan 77,67% dari data uji benar terklasifikasi ke dalam kelas ulasan positif dengan nilai precision dan recall tinggi yaitu 82,67% dan 86,92%. Selain itu, dari hasil penerapan metode dalam mengklasifikasi data ulasan pengguna Flip dengan perbandingan data latih dan data uji sebesar 80%:20% diperoleh

					tingkat akurasi klasifikasi dengan algoritma K-Nearest Neighbor sebesar 76,68%.
Muhammad Farid El Firdaus, Nurfaizah, dan Sarmini	2022	Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor	Melakukan sentimen analisis atas performa aplikasi yang didasarkan pada komentar pengguna aplikasi, dimana dataset yang digunakan dalam penelitian ini sebanyak 992 komentar.	Algoritma yang digunakan dalam penelitian ini yaitu algoritma naïve bayes dan k-nearest neighbor	Hasil penelitian menunjukkan tingkat accuracy dari penggunaan algoritma naïve bayes sebesar 75,30% dan tingkat akurasi algoritma k-nearest neighbor didapatkan hasil sebesar 86,09%.
Rizki Wahyudi, Gilang Kusumawardhana	2021	Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine	Menganalisis review pengguna aplikasi Grab pada Google Play Store, menggunakan analisis sentimen	Analisis review pengguna ini menggunakan metode Support Vector Machine (SVM). Evaluasi yang diusulkan dilakukan pada lebih dari 1.000 review pengguna yang dikumpulkan dari aplikasi Grab Indonesia di	Hasil dari analisis menggunakan Support Vector Machine menghasilkan akurasi 85,54% dan Hasil Review positif yang paling sering diulas adalah "ovo", sedangkan review negatif yang paling sering diulas adalah "driver".

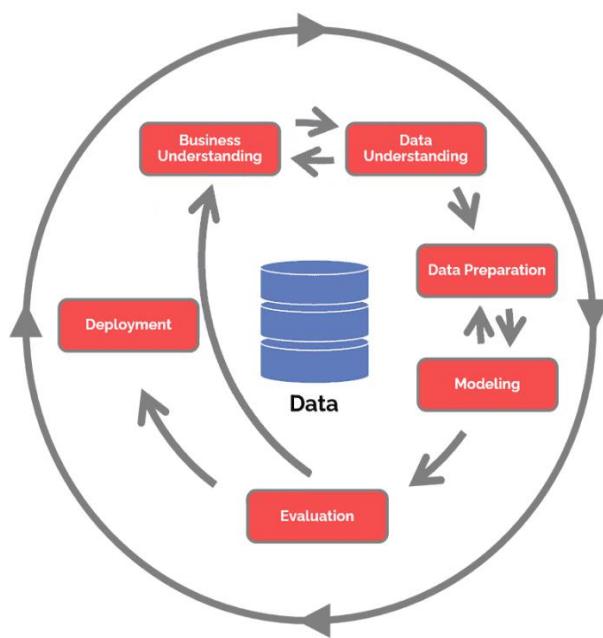
				Google play store	
Siti Ernawati dan Risa Wati	2018	Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel	Peninjauan agen perjalanan pengolahan data menggunakan algoritma K-Nearest Neighbor (K-NN) yang menggunakan 100 review positif dan 100 review negatif dengan enam kata yang berhubungan dengan sentimen yaitu: Fast, Good, Great, Buruk, Cencel, dan Tunggu.	Menggunakan metode K-Nearest Neighbor (KNN) dan pembobotan untuk setiap kata menggunakan TF-IDF	Menggunakan algoritma K-Nearest Neighbor (K-NN), ia mencapai hasil akurasi terbaik dan berdasarkan perhitungan yang dinyatakan dalam aplikasi. Titik akurasi peninjauan agen perjalanan menggunakan K-Nearest Neighbor (K-NN) algoritma telah mencapai 87,00% dan titik AUC adalah 0,916, titik AUC milik kelompok Klasifikasi Excellent sehingga dinyatakan bahwa K-Nearest Neighbor (K-NN) memiliki hasil yang akurat dalam menganalisis sentimen ulasan agen perjalanan
Herda Andriana, Shofa Shofia Hilabi, dan Agustia Hananto	2023	Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter Terhadap KTT G20 di Indonesia	Tujuan dilakukan ya penelitian ini adalah untuk analisis opini publik di Twitter pada KTT G20.	Pengolahananya menggunakan perangkat lunak Orange, library vader, dan metode K-Nearest Neighbor	Hasil berupa 1107 data tweet masuk pada kelompok original dengan tingkat accuracy, F1 score, Recall, Precision rata-rata 99%. Dan hasil sentimen analisis 89 tweet masuk kedalam kategori negatif,

					614 netral, dan 404 positif dengan mayoritas tingkat emosional joy, surprise, dan fear.
--	--	--	--	--	--

## BAB III

### METODOLOGI PENELITIAN

Pada Bab Metodologi Penelitian ini berisi tentang tahapan-tahapan yang digunakan dalam penelitian ini agar terstruktur dengan baik. Tahapan-tahapan metodologi yang dilakukan adalah seperti gambar berikut.



Gambar 3.1 Metodologi Penelitian CRISP-DM

Agar tahapan analisis bisa lebih terstruktur peneliti menerapkan Cross-Industry Standard Process for Data Mining (CRISP-DM) yang merupakan metodologi standar industri yang netral dalam proses data mining yang dikembangkan pada tahun 1996 oleh Daimler Chrysler di Eropa. (M. Brumer, 2007). Pada metode ini terdiri dari 6 tahapan sebagai berikut

### **3.1 Business Understanding**

Pemahaman bisnis merupakan tahapan awal dimana peneliti melakukan observasi terkait topik penelitian yaitu analisis sentimen pada ulasan komentar di Google Play Store dan menyusun rencana atau langkah untuk mencapai tujuan penelitian. Melakukan studi literatur terkait aplikasi Online Travel Agent dan menentukan ulasan aplikasi yang akan digunakan. Adapun, ulasan aplikasi Tiket.com dan Traveloka yang digunakan sebagai representasi aplikasi Online Travel Agent. Hal ini dipilih sesuai penelitian yang dilakukan oleh Katadata pada tahun 2022 sektor aplikasi Online Travel Agent (OTA) 2 aplikasi teratas, yaitu Tiket.com dan Traveloka.

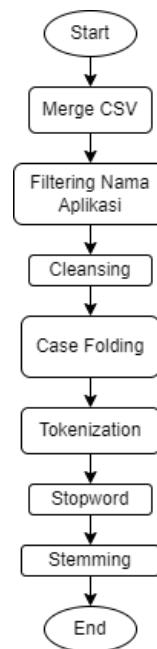
### **3.2 Data Understanding**

Setelah melakukan pemahaman bisnis Langkah selanjutnya, yaitu mengumpulkan ulasan komentar aplikasi di Google Play Store. Pada tahap ini dimulai tahap pengumpulan data ulasan komentar aplikasi tiket.com dimulai dari dengan total data sebanyak 7000 ulasan komentar Bahasa Indonesia dan pengguna dari Indonesia. Selanjutnya, dimulai tahap pengumpulan data ulasan komentar aplikasi Traveloka dengan total data sebanyak 7000 ulasan komentar Bahasa Indonesia dan pengguna dari Indonesia. Adapun, cara web scraping akan menggunakan tools Google Colab dengan Bahasa pemrograman Python dengan memanfaatkan link gits dari Google Play Store. Data yang telah di scrape akan di generate menjadi CSV (Comma Separated Value) yang dapat dibuka pada Microsoft Excel. Pada tahapan ini dilakukan pemahaman data dengan melakukan pengecekan pada kolom apa saja yang didapatkan dari hasil pengumpulan data (web

scrapping) pada google play store. Selanjutnya, dilakukan pengecekan kolom yang telah didapatkan dari hasil web scrapping, yaitu UserName, Content, Score, reviewCreatedVersion, at (timestamp Review user), repliedAt (timestamp review developer), replyContent, appVersion.

### 3.3 Data Preparation

Setelah data sudah dikoleksi dan dilakukan pemahaman data, maka langkah selanjutnya tahap preprocessing. Tahap preprocessing adalah proses perubahan bentuk data yang terstruktur sembarang menjadi data yang terstruktur sesuai kebutuhan untuk proses dalam text mining (Takdirillah, 2020). Tahapan Preprocessing yang digunakan antara lain:



Gambar 3.2 Alur Preprocessing Data

a. Merge CSV

Setelah Scrapping didapatkan data berupa file csv untuk aplikasi Tiket.com dengan jumlah 7000 data dan ulasan aplikasi Traveloka sejumlah 7000, sehingga total jumlah data menjadi 14.000.

b. Filtering Nama Aplikasi

Filtering ini dilakukan untuk memfilter kata “Traveloka”, “Traveloka”, “TRAVELOKA”, “Tiket.com”, “tiket.Com”, “tiket.com”, “TIKET.COM”, “Tiketcom”, “tiketcom” agar pada ulasan fokus pada kalimat sentimen dan tidak menimbulkan ambigu terutama pada kata “Tiket” karena nama aplikasinya “Tiket.com”.

c. Cleansing

Pada tahapan cleansing ini dilakukan dengan:

- a. Menghilangkan URL situs lain atau URL gambar dan video  
Untuk menghilangkan URL situs lain, gambar, dan video dari teks.
- b. Menghilangkan paragraf baru.
- c. Menghilangkan tanda baca dan angka.
- d. Menghilangkan spasi di depan dan di belakang untuk kalimat yang repetisi.
- e. Menghilangkan tab pada kalimat.
- f. Mengubah huruf pada kalimat menjadi huruf kecil.

d. Case Folding

Case folding dilakukan dengan mengubah semua karakter teks menjadi huruf kecil. Hal ini dilakukan untuk mengurangi kompleksitas dan perbedaan kapitalisasi teks.

e. Tokenization

Tokenizing dilakukan dengan pemisahan teks menjadi kata, frasa, simbol menjadi elemen yang berbeda biasanya proses pemisahan setiap kata dipisahkan dengan tanda koma (,).

f. Stopword

Stop Removal digunakan untuk menghilangkan kata umum yang tidak memiliki arti penting dan tidak digunakan dengan tujuan mengurangi jumlah kata yang disimpan oleh sistem (Sugianto & Apandi, 2018).

g. Stemming

Stemming dilakukan dengan menyatukan atau mengubah varian bentuk kata dari umum menjadi kata dasar (Kannan and Gurusamy, 2015).

Tahapan – tahapan diatas merupakan tahapan penting sebelum labelling. Dimana langkah selanjutnya, akan dilakukan labelling menggunakan library vader dan Roberta. Labelling menggunakan library Roberta dilakukan langsung dengan memanggil library. Hal ini berbeda dengan library vader karena, library vader tidak menyediakan library untuk Bahasa Indonesia, sehingga perlu diterjemahkan dengan menggunakan library translator. Setelah dilakukan penerjemahan maka, library vader akan dipanggil. Hasil dari library Roberta dan Vader tidak jauh berbeda dimana keduanya menghasilkan labelling positif dan negatif dari ulasan.

### **3.4 Pemodelan (Modelling)**

Sebelum dilakukannya pemodelan, dari hasil data yang telah melewati preprocessing pada tahapan data preparation akan dilakukan pengecekan EDA. Hal ini diperlukan untuk mengetahui apakah terjadi imbalanced data. Apabila terjadinya imbalanced data, maka dilakukan skenario undersampling dan oversampling agar dataset menjadi seimbang (Cici H et al, 2023). Oversampling bertujuan untuk meningkatkan sampel kelas minoritas sampai sama dengan kelas mayoritas lain dengan menduplikasi secara acak sampel kelas minoritas(H. He et al, 2018). Sedangkan random undersampling yaitu menghasilkan sub sampel acak dari instance kelas mayoritas. Undersampling merupakan metode sampling secara acak memilih sampel di kelas mayoritas dan menambahkannya ke kelas minoritas, membentuk sebuah dataset pelatihan baru (H. He et al, 2018). Kemudian, Langkah selanjutnya melakukan perhitungan nilai perbandingan antara nilai data yang sudah dilakukan preprocessing dan sebelum dilakukan preprocessing, yang terdiri dari beberapa tahapan yaitu pemecahan dataset menggunakan split data yang terdiri dari data training dan data testing dengan nilai perbandingan 75:25 dengan menerapkan algoritma k-nearest neighbors sebagai nilai perbandingan mana hasil nilai akurasi tertinggi dengan pengecekan nilai k yang beragam.

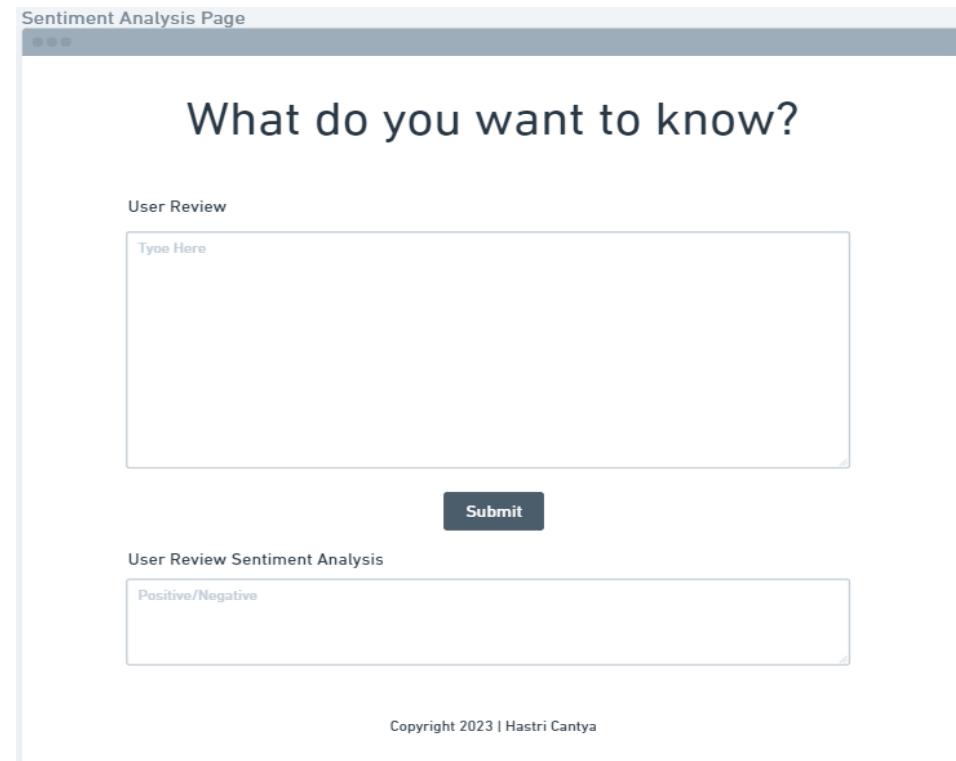
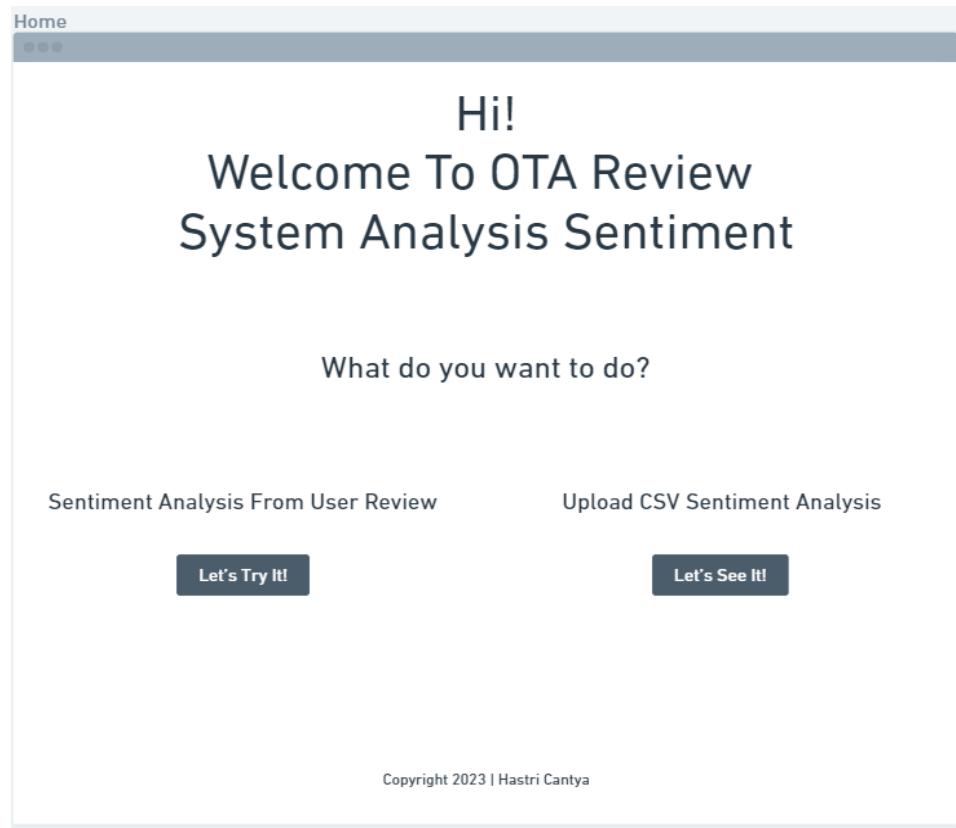
### **3.5 Evaluasi (Evaluation)**

Pada tahapan ini peneliti melakukan evaluasi terhadap hasil perhitungan algoritma k-nearest neighbors yang dilakukan terhadap ulasan komentar yang sudah melewati proses data preparation dan data preprocessing dengan menggunakan confusion matrix yang digunakan untuk menghitung nilai akurasi dari proses

klasifikasi. Evaluasi menggunakan confusion matrix untuk mengetahui perkiraan yang benar dan salah dari hasil klasifikasi. Evaluasi digunakan untuk mengetahui seberapa akurat model yang sudah dibuat. Hasil akurasi tersebut didapatkan dengan membandingkan data yang sudah di beri label benar atau salah pada model dengan hasil prediksi label. Tahap pengujian dilakukan dengan menggunakan data uji. Evaluasi model klasifikasi dianalisis dengan membandingkan nilai akurasi secara keseluruhan dan waktu yang diperlukan untuk melakukan training model. Lalu, dilakukannya pengecekan apakah dengan nilai  $k$  yang telah dimasukkan pada pemodelan, dataset mengalami overfit atau underfit dengan melihat grafik evaluasi K-Nearest Neighbour.

### **3.6 Deployment (Visualisasi Web)**

Setelah hasil dari tingkat akurasi kedua metode klasifikasi didapat maka langkah selanjutnya yaitu visualisasi (Ilmawan, et al, 2015). Proses visualisasi disini berarti proses pengkodean dan pembuatan tampilan implementasi antar muka agar para pengguna dapat memahami hasil dari klasifikasi ini. Hasil dari klasifikasi data dengan kedua metode tersebut akan divisualisasikan ke dalam bentuk grafik yang mudah dipahami dan dari klasifikasi juga ditampilkan kata yang menjadi pemicu sentimen positif, negative, dan netral dalam beropini yang akan diimplementasikan dalam suatu aplikasi berbasis web menggunakan Bahasa PHP. Fitur dari visualisasi yaitu website bisa memprediksi label dari ulasan yang dimasukkan ke dalam inputan dan menampilkan line graphic berdasarkan tren tiap bulan dengan hasil label ulasan.





Gambar 3.3 Ilustrasi Visualisasi Web

## **BAB IV**

### **KESIMPULAN DAN SARAN**

Pada Bab Hasil dan Pembahasan dari penelitian yang telah dilakukan sebelumnya, yaitu terkait analisis sentimen ulasan aplikasi OTA di google play store menggunakan Algoritma K-Nearest Neighbour.

#### **4.1 Business Understanding**

Untuk melaksanakan penelitian analisis sentimen ulasan aplikasi OTA di google play store menggunakan Algoritma K-Nearest Neighbour diperlukan data yang akan diolah serta spesifikasi perangkat lunak dan perangkat keras.

Data yang diperlukan dalam penelitian ini adalah data dari ulasan aplikasi OTA (Online Travel Agent) di google play store, yaitu aplikasi Tiket.com dan Traveloka. Dengan ulasan dari aplikasi Tiket.com sejumlah 7000 dan ulasan aplikasi Traveloka sejumlah 7000, sehingga jumlah data 14.000. Adapun kebutuhan hardware dan software pada penelitian ini sebagai berikut.

##### **4.1.1 Kebutuhan Hardware dan Software**

- Kebutuhan Hardware

Dalam menjalankan penelitian ini dibutuhkan perangkat keras berupa laptop Lenovo Ideapad Slim 3 dengan spesifikasi RAM 4GB + 4GB serta penyimpanan internal SSD sebesar 512GB.

- Kebutuhan Software

Software yang digunakan yaitu web browser Chrome, Microsoft Excel untuk melihat data, dan Google Colab untuk menjalankan bahasa pemrograman Phyton.

## 4.2 Data Understanding

### 4.2.1 Pengumpulan Data (Web Scrapping)

Pada tahap pengumpulan data, digunakan Source Code berikut sebagai alat untuk mengumpulkan ulasan komentar aplikasi Tiket.com dan Traveloka di Google Play Store.

```
● ● ●

!pip install google-play-scraper
from google_play_scraper import app

import pandas as pd

import numpy as np
from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.tiket.gits',
    lang='id', # defaults to 'en'
    country='id', # defaults to 'us'
    sort=Sort.NEWEST, # defaults to Sort.MOST_RELEVANT you can use Sort.NEWEST to get newest reviews
    count=7000, # defaults to 100
    filter_score_with=None # defaults to None(means all score) Use 1 or 2 or 3 or 4 or 5 to select
certain score
)
df_busu = pd.DataFrame(np.array(result),columns=['review'])

df_busu = df_busu.join(pd.DataFrame(df_busu.pop('review').tolist()))

df_busu.head()
len(df_busu.index) #count the number of data we got
df_busu[['userName', 'content','score', 'reviewCreatedVersion', 'at', 'replyContent', 'repliedAt',
'appVersion']].head() #preview userName, rating, date-time, and reviews only

new_df = df_busu[['userName', 'content','score', 'reviewCreatedVersion', 'at', 'replyContent',
'repliedAt', 'appVersion']]
sorted_df = new_df.sort_values(by='at', ascending=False) #Sort by Newest, change to True if you want to
sort by Oldest.
sorted_df.head()

my_df = sorted_df[['userName', 'content','score', 'reviewCreatedVersion', 'at', 'replyContent',
'repliedAt', 'appVersion']] #get userName, rating, date-time, and reviews only

my_df.to_csv("scrapped_data.csv", index = False) #Save the file as CSV , to download: click the folder
icon on the left. the csv file should be there.
```

Gambar 4.1 Source Code Pengumpulan Data Ulasan Tiket.com

```

● ● ●

!pip install google-play-scraper
from google_play_scraper import app

import pandas as pd

import numpy as np
from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.traveloka.android',
    lang='id', # defaults to 'en'
    country='id', # defaults to 'us'
    sort=Sort.NEWEST, # defaults to Sort.MOST_RELEVANT you can use Sort.NEWEST to get newest reviews
    count=7000, # defaults to 100
    filter_score_with=None # defaults to None(means all score) Use 1 or 2 or 3 or 4 or 5 to select
certain score
)
df_busu = pd.DataFrame(np.array(result),columns=['review'])

df_busu = df_busu.join(pd.DataFrame(df_busu.pop('review').tolist()))

df_busu.head()
len(df_busu.index) #count the number of data we got
df_busu[['userName', 'content','score', 'reviewCreatedVersion', 'at', 'replyContent', 'repliedAt',
'appVersion']].head() #preview userName, rating, date-time, and reviews only

new_df = df_busu[['userName', 'content','score', 'reviewCreatedVersion', 'at', 'replyContent',
'repliedAt', 'appVersion']]
sorted_df = new_df.sort_values(by='at', ascending=False) #Sort by Newest, change to True if you want to
sort by Oldest.
sorted_df.head()

my_df = sorted_df[['userName', 'content','score', 'reviewCreatedVersion', 'at', 'replyContent',
'repliedAt', 'appVersion']] #get userName, rating, date-time, and reviews only

my_df.to_csv("scrapped_data.csv", index = False) #Save the file as CSV , to download: click the folder
icon on the left. the csv file should be there.

```

Gambar 4.2 Source Code Pengumpulan Data Ulasan Traveloka

Gambar 4.1 dan gambar 4.2 menunjukkan langkah – langkah untuk mengambil data ulasan aplikasi Tiket.com dan Traveloka hingga ke dalam bentuk csv. Data yang diambil merupakan ulasan terbaru dengan masing – masing aplikasi diambil 7000 data dengan kriteria ulasan oleh akun dengan negara Indonesia dan bahasa Indonesia. Kolom yang diambil terdiri dari username, content, score, reviewCreatedVersion, at, replyContent, repliedAt, dan appVersion.

Berikut ini merupakan contoh ulasan aplikasi Tiket.com yang diperoleh:

Username	Content	Score	Review Created Version	at	Reply Content	Replied At	appVersion
frenky Chandra	Aplikasi yg pusing, susah di pakai nya	1	4.60.0	2023-07-03 02:09:20	Hai Frenky Chandra, Terima kasih ya sudah luangkan waktumu berikan ulasan untuk kami. Jika kamu ada alami kendala saat pesan atau bingung dengan apps kami jangan sungkan hubungi customer care kami ya sobat. Kami siap bantu kamu 24/7 atau kamu bisa kirimkan email ke cs@ dengan subject : "Konfirmasi Review App". Terima kasih & by : FA	2023-07-03 08:01:33	4.60.0
Gishela Dintya	Suka banget sama promonya, apalagi pas masa-masa war tiket konser, pertama kali war konser di tikom langsung dapet paling depan	5	4.60.0	2023-07-03 00:52:57	-	-	4.60.0
Syukur Syukur	Beberapa kali booking hotel, selalu tidak	1	4.52.0	2023-07-02 14:57:24	Hai Syukur, Mohon maaf atas ketidaknyamanan yang dan terima kasih sudah meluangkan	2023-07-02 18:44:55	4.52.0

	sesuai dg harga di hotelnya, akhirnya diminta tambah sama resepsonis nya... Mungkin lebih ditekankan lagi dengan mitra hotelnya,, agar setelah bayar di aplikasi tidak dimintai tambah biayanya oleh pihak hotelnya.... Terimakasih				waktu untuk memberikan ulasan. Boleh kamu ceritakan detail kendalanya kepada kami via Email: cs@ dengan subject: "Konfirmasi Review App". Kami tunggu konfirmasinya dan dengan senang hati kami akan membantu kamu. Terima kasih 😊 by : AR		
FZ	Lumayan efisien buat pesan tiket pesawat dan liat harga tiket ditanggal tertentu, tapi "Gila sih dana refund potongan nya gede banget hampir 50%"	2	4.60.0	2023-07-01 11:02:28	Hai sobat tiket, Mohon maaf atas ketidaknyamanan yang dan terima kasih sudah meluangkan waktu untuk memberikan ulasan. Bisa kamu informasikan Order ID dan alamat email pemesanan kamu kepada kami via Email: cs@ dengan subject: "Konfirmasi Review App". Kami tunggu konfirmasinya dan dengan senang hati kami akan membantu kamu. Terima kasih 😊 by : PY	2023-07-01 12:58:52	4.60.0
Good People (Bestfriend)	Suka aneh2 mancing2 hrg murah,tp skali masuk hrg lgsg berubah	2	4.60.0	2023-06-30 11:02:28	Hai sobat tiket, Kami memahami kendala yang kamu alami. Mohon maaf atas ketidaknyamanan yang dan terima kasih sudah meluangkan waktu untuk memberikan ulasan. Boleh kamu ceritakan detail kendalanya kepada kami via Email: cs@ dengan subject: "Konfirmasi Review App". Dengan senang hati kami akan membantu kamu. Terima kasih 😊 by : RD	2023-06-30 23:49	4.60.0

Tabel 4.1 Tabel Ulasan Aplikasi Tiket.com

Berikut ini merupakan contoh ulasan aplikasi Traveloka yang diperoleh:

Usernam e	Content	Sco re	Revie w Creat ed Versi on	at	Reply Content	Replied At	appVers ion
Asep Awaludin	Sangat membantu dan diskon harganya pantastis	5	3.79.0	2023-07-03 07:31:52	Halo Kak, terima kasih banyak atas ratingnya, semoga Kakak tetap nyaman dan selau setia menjadi pelanggan . Yuk Kak pantengin terus Lifestyle Super App karna kami selalu berikan promo-promo yg menarik looh, jangan sampai ketinggalan ya :)	2023-07-03 07:42:09	3.79.0
Franky Sidjaya	mantap	5	3.79.0	2023-07-03 06:10:13	Halo Kak, terima kasih banyak atas ratingnya, semoga Kakak tetap nyaman dan selau setia menjadi pelanggan . Yuk Kak pantengin terus Lifestyle Super App karna kami selalu berikan promo-promo yg menarik looh, jangan sampai ketinggalan ya :)	2023-07-03 06:12:14	3.79.0
Sisca Nandita	trimakaasih h smg mjd kebahagiaan semua umat.....Ami n.	5	3.79.0	2023-07-03 03:51:34	Halo Kak, terima kasih banyak atas ratingnya, semoga Kakak tetap nyaman dan selau setia menjadi pelanggan . Yuk Kak pantengin terus Lifestyle Super App karna kami selalu berikan promo-promo yg menarik looh, jangan sampai ketinggalan ya :)	2023-07-03 04:12:13	3.79.0
Rofians yah Perman a	Kenapa tidak bisa pake paylater	1	3.79.0	2023-07-03 03:29:11	Hai Rofiansyah. Mohon maaf atas kendala yang kamu alami ya Kak. Supaya bisa mimin bantu cek lebih lanjut, mohon infokan detail kendala kamu ke akun IG PayLater <a href="https://www.instagram.com/paylater/">https://www.instagram.com/paylater/</a> ya. Terima kasih. - KK	2023-07-03 03:48:06	3.79.0
Zahra san & Kayla san di Japan	Sangat kecewa saya beli tiket dgn paylater sa ya reschedule tglnya dr 26 Juli ke 8 Juli karena ibu saya sakit keras di indonesia.saya membelinya dengan paylater lg tetapi saya di kenakan tagihan dobel tiket 1 & yg baru ga masalah bayar tapi yg 1 sdh	2	3.79.0	2023-07-03 04:05:13	Hai, Zahra. Sebelumnya mohon maaf atas pengalaman yang kamu alami. Pada dasarnya untuk pegajuan reschedule akan dikenakan biaya sesuai kebijakan maskapai. Serta untuk pembelian menggunakan PayLater tetap harus dibayarkan ya Kak, karena tidak ada pengembalian dana ke limit PayLater. Terima kasih.-TK	2023-07-03 04:20:08	3.79.0

	tidak ada tiketnya Khan sdh dirubah.sdn gkan biaya reschedule 60 rb perorang tp say tetap diharuskan membayar paylater yg sdh di rubah jadwalnya Khan aneh.jadi berasa beli tiket 40 jt kalu sya byar dobel.sejak semalam sampe skrg no tanggapan					
--	---	--	--	--	--	--

Tabel 4.2 Tabel Ulasan Aplikasi Tiket.com

#### 4.3 Data Preparation

Untuk melaksanakan penelitian analisis sentimen ulasan aplikasi OTA di google play store menggunakan Algoritma K-Nearest Neighbour diperlukan data sebelum diberi label telah di preprocess agar ketika pelabelan dapat menghasilkan akurasi yang baik.

##### 4.3.1 Merge CSV Ulasan Aplikasi Traveloka dan Tiket.com

Setelah Scrapping didapatkan data berupa file csv untuk aplikasi Traveloka dengan jumlah 7000 data dengan kurun waktu dan ulasan aplikasi Traveloka sejumlah 7000 dengan kurun waktu.



```
# Define the URLs of the CSV files
url1 = 'https://raw.githubusercontent.com/hastricantya/OTA/main/Final_LabelledTravelokaRemoved.csv'
url2 = 'https://raw.githubusercontent.com/hastricantya/OTA/main/Final_LabelledTiketcomRemoved.csv'

# Read the CSV files from the URLs into DataFrames
df1 = pd.read_csv(url1)
df2 = pd.read_csv(url2)

# Merge the two DataFrames using all columns as they have the same column names
merged_df = pd.concat([df1, df2], ignore_index=True)

# Save the merged DataFrame to a new CSV file
merged_df.to_csv('Final_Data_OTA.csv', index=False)
```

Gambar 4.3 Merge CSV Ulasan Aplikasi

### 4.3.2 Filtering Kata Traveloka dan Tiket.com

Filtering ini dilakukan untuk memfilter kata “Traveloka”, “Traveloka”, “TRAVELOKA”, “Tiket.com”, “tiket.Com”, “tiket.com”, “TIKET.COM”, “Tiketcom”, “tiketcom” agar pada ulasan fokus pada kalimat sentimen dan tidak menimbulkan ambigu terutama pada kata “Tiket” karena nama aplikasinya “Tiket.com”.



```
● ● ●

import csv

def remove_traveloka_from_csv(input_file, output_file):
    with open(input_file, 'r', newline='') as infile, open(output_file, 'w', newline='') as outfile:
        reader = csv.reader(infile)
        writer = csv.writer(outfile)
        for row in reader:
            # Assuming the column to check is the first one (@ index)
            modified_row = [cell.replace('traveloka', '').replace('tiketcom', '').replace('Tiketcom',
            '').replace('Tiket.com', '').replace('tiket.com', '').replace('TIKETCOM', '').replace('TIKET.COM', '')]
            for cell in row:
                writer.writerow(modified_row)

if __name__ == "__main__":
    input_file = "tiket_scrapped_data.csv" # Replace with your input CSV file path
    output_file = "output.csv" # Replace with your output CSV file path
    remove_traveloka_from_csv(input_file, output_file)
/
```

Gambar 4.4 Filtering Kata Traveloka dan Tiketcom

Ulasan aplikasi OTA yang telah dilakukan filtering:

No.	Sebelum Filtering	Setelah Filtering
1.	Sangat kecewa saya beli tiket dgn paylater Traveloka saya reschedule tglnya dr 26 Juli ke 8 Juli karena ibu saya sakit keras di indonesia.saya membelinya dengan paylater lg tetapi saya di kenakan tagihan dobel tiket 1 & yg baru ga masalah bayar tapi yg 1 sdh tidak ada tiketnya Khan sdh dirubah.sdngkan biaya reschedule 60 rb perorang tp say tetap diharuskan membayar paylater yg sdh di rubah jadwalnya Khan aneh.jadi berasa beli tiket 40 jt kalu sya byar dobel.sejak semalam sampe skrg no tanggapan	Sangat kecewa saya beli tiket dgn paylater saya reschedule tglnya dr 26 Juli ke 8 Juli karena ibu saya sakit keras di indonesia.saya membelinya dengan paylater lg tetapi saya di kenakan tagihan dobel tiket 1 & yg baru ga masalah bayar tapi yg 1 sdh tidak ada tiketnya Khan sdh dirubah.sdngkan biaya reschedule 60 rb perorang tp say tetap diharuskan membayar paylater yg sdh di rubah

		jadwalnya Khan aneh.jadi berasa beli tiket 40 jt kalu sya byar dobel.sejak semalam sampe skrg no tanggapan
2.	Alhamdulillah beberapa bln yg lalu pke Traveloka sangat memudahkan saya pesan hotel, semoga kali ini jd memuaskan	Alhamdulillah beberapa bln yg lalu pke sangat memudahkan saya pesan hotel, semoga kali ini jd memuaskan
3.	Tepat hari ini, Sabtu 1 Juli 2023, saya tidak bisa mengakses aplikasi traveloka, baik utk melihat tiket pesawat, hotel dll.....mohon petunjuk admin, terimakasih	Tepat hari ini, Sabtu 1 Juli 2023, saya tidak bisa mengakses aplikasi, baik utk melihat tiket pesawat, hotel dll.....mohon petunjuk admin, terimakasih
4.	Biaya platform sama adm total 70k, tapi pas waktunya war crash mulu. Gimana sih tiketcom mahal mahal bayar biaya gituan tapi GAJELASS!!!!!!	Biaya platform sama adm total 70k, tapi pas waktunya war crash mulu. Gimana sih mahal mahal bayar biaya gituan tapi GAJELASS!!!!!!
5.	Terima Kasih tiketcom selalu memberikan diskon	Terima Kasih selalu memberikan diskon

#### 4.3.3 Cleansing

Pada proses cleansing, terdapat beberapa langkah-langkah yang akan dijelaskan berdasarkan source code dan hasilnya. Langkah-langkah tersebut meliputi:

- Menghilangkan URL situs lain atau URL gambar dan video  
Untuk menghilangkan URL situs lain, gambar, dan video dari teks.
- Menghilangkan paragraf baru.
- Menghilangkan tanda baca dan angka.

- d. Menghilangkan spasi di depan dan di belakang untuk kalimat yang repetisi.
- e. Menghilangkan tab pada kalimat.
- f. Mengubah huruf pada kalimat menjadi huruf kecil.

Berikut merupakan contoh source code cleansing:

```

● ● ●
!pip install nltk
!pip install sastrawi
import pandas as pd
import re

data =
pd.read_csv("https://raw.githubusercontent.com/hastricantya/OTA/main/traveloka_scraped_data.csv")
data.head()

df = data
df = df.drop('userName', axis=1)
df = df.drop('reviewCreatedVersion', axis=1)
df = df.drop('replyContent', axis=1)
df = df.drop('repliedAt', axis=1)
df = df.drop('appVersion', axis=1)
df = df.drop('score', axis=1)
df = df.drop('at', axis=1)

df.head()

def text_cleaning(data):
    temporary_text = []

    for text in data:
        # removal of links[https://blabala.com]
        text = re.sub(r"HTTP\S+", "", text)

        # removal of new line
        text = re.sub('\n', '', text)

        # removal of punctuations and numbers
        text = re.sub("[^a-zA-Z]", " ", text)
        text = re.sub("{2,}", " ", text)

        # remove leading and trailing whitespace
        text = text.strip()

        # remove whitespace with a single space
        text = re.sub(r'\s+', ' ', text)

        # convert text to Lowercase
        text = text.lower();
        temporary_text.append(text)
    return temporary_text

```

Gambar 4.5 Source Code Cleansing

No.	Sebelum Cleansing	Setelah Cleansing
1.	praktis & banyak manfaatnya... 	praktis banyak manfaatnya
2.	aplikasi mantap 100% untuk ini semuanya	aplikasi mantap untuk ini semuanya
3.	sulit sekali pesan disini not recommend:(	sulit sekali pesan disini not recommend

4.	jelek sekali sering error banyak bug	jelek sekali sering error banyak bug
5.	mantap .. 🤙 pemesanan cepat sekali sayang harganya kurang murah	mantap pemesanan cepat sekali sayang harganya kurang murah

Tabel 4.3 Proses Cleansing

#### 4.3.4 Case Folding

```
● ● ●
df['case_folded'] = text_cleaning(df['content'])
df.head(25)
```

Gambar 4.6 Source Code Case Folding

Gambar 4.6 di atas menjelaskan proses case folding pada teks dalam kolom ‘Content’ dari file CSV. Case folding dilakukan dengan mengubah semua karakter teks menjadi huruf kecil. Hal ini dilakukan untuk mengurangi kompleksitas dan perbedaan kapitalisasi teks. Hasil dari case folding disimpan dalam kolom baru ‘case\_folded’ dan disimpan dalam file CSV baru. Berikut contoh hasil dari proses case folding

No.	Sebelum Case Folding	Setelah Case Folding
1.	Sangat kecewa saya beli tiket dgn paylater saya reschedule tglnya dr 26 Juli ke 8 Juli karena ibu saya sakit keras di 43ndonesia.saya membelinya dengan paylater lg tetapi saya di kenakan tagihan dobel tiket 1 & yg baru ga masalah bayar tapi yg 1 sdh tidak ada tiketnya Khan sdh dirubah.sdngkan biaya reschedule 60 rb perorang tp say tetap diharuskan membayar paylater yg sdh di rubah jadwalnya Khan aneh.jadi berasa beli tiket 40 jt kalu sya byar dobel.sejak semalam sampe skrg no tanggapan	Sangat kecewa saya beli tiket dgn paylater saya reschedule tglnya dr 26 Juli ke 8 Juli karena ibu saya sakit keras di 43ndonesia.saya membelinya dengan paylater lg tetapi saya di kenakan tagihan dobel tiket 1 & yg baru ga masalah bayar tapi yg 1 sdh tidak ada tiketnya khan sdh dirubah.sdngkan biaya reschedule 60 rb perorang tp say tetap diharuskan membayar paylater yg sdh di rubah jadwalnya khan aneh.jadi berasa beli tiket 40 jt kalu sya byar dobel.sejak semalam sampe skrg no tanggapan

2.	Sangat membantu dan diskon harganya pantastis	sangat membantu dan diskon harganya pantastis
3.	Beberapa kali booking hotel, selalu tidak sesuai dg harga di hotelnya, akhirnya diminta tambah sama resepsionis nya... Mungkin lebih ditekankan lagi dengan mitra hotelnya,, agar setelah bayar di aplikasi tidak dimintai tambah biayanya oleh pihak hotelnya... Terimakasih	beberapa kali booking hotel, selalu tidak sesuai dg harga di hotelnya, akhirnya diminta tambah sama resepsionis nya... mungkin lebih ditekankan lagi dengan mitra hotelnya,, agar setelah bayar di aplikasi tidak dimintai tambah biayanya oleh pihak hotelnya... terimakasih
4.	Ku kira jaringan internet ku lelet, ternyata aplikasi nya yang lelet. Hadeh ga jadi deh beli tiket disini	ku kira jaringan internet ku lelet, ternyata aplikasi nya yang lelet. Hadeh ga jadi deh beli tiket disini
5.	Lebih murah dibandingkan dengan yg lain	lebih murah dibandingkan dengan yg lain

Tabel 4.4 Proses Case Folding

#### 4.3.5 Tokenization

Setelah melakukan case folding, selanjutnya yaitu tahapan tokenisasi.

Berikut adalah source code tokenization:

```
● ● ●

import nltk
nltk.download('punkt')

def tokenized(data):
    return data.apply(nltk.word_tokenize)

df['tokenized'] = tokenized(df['case_folded'])
df.head(25)
```

Gambar 4.7 Source Code Tokenization

Gambar 4.7 di atas merupakan implementasi tokenisasi menggunakan library NLTK dalam pemrograman Python. Pertama, library yang diperlukan seperti pandas, nltk, dan TweetTokenizer dari NLTK diimport. Selanjutnya, Google Drive di-mount untuk mengakses file CSV yang akan diproses. File CSV dibaca menggunakan pandas dan disimpan dalam DataFrame yang disebut df. Fungsi

tokenize\_text didefinisikan untuk melakukan tokenisasi pada teks. Hasil tokenisasi diterapkan pada kolom ‘tokenized’. Kolom baru ‘tokenized’ digunakan untuk menyimpan hasil tokenisasi. Contoh hasil tokenisasi adalah sebagai berikut:

No.	Sebelum Tokenization	Setelah Tokenization
1.	praktis banyak manfaatnya	praktis,banyak,manfaatnya
2.	semoga berhasil tidak mengecewakan untuk pemesanan orang tua	semoga,berhasil,tidak,mengecewakan,untuk,pemesanan,orang,tua
3.	sangat kecewa saya beli tiket dgn paylater saya reschedule tglnya dr juli ke juli karena ibu saya sakit keras di 45ndonesia saya membelinya dengan paylater lg tetapi saya di kenakan tagihan dobel tiket yg baru ga masalah bayar tapi yg sdh tidak ada tiketnya khan sdh dirubah sdngkan biaya reschedule rb perorang tp say tetap diharuskan membayar paylater yg sdh di rubah jadwalnya khan aneh jadi berasa beli tiket jt kalu sya byar dobel sejak semalam sampe skrg no tanggapan	sangat,kecewa,saya,beli,tiket,dgn,paylater,45ndonesia,saya,reschedule,tglnya,dr,juli,ke,juli,karena,ibu,saya,sakit,keras,di,45ndonesia,saya,membelinya,dengan,paylater,lg,tetapi,saya,di,kenakan,tagihan,dobel,tiket,yg,baru,ga,masalah,bayar,tapi,yg,sdh,tidak,ada,tiketnya,khan,sdh,dirubah,sdngkan,biaya,reschedule,rb,perorang,tp,say,tetap,diharuskan,membayar,paylater,yg,sdh,di,rubah,jadwalnya,khan,aneh,jadi,berasa,beli,tiket,jt,kalu,sya,byar,dobel,sejak,semalam,sampe,skrg,no,tanggapan
4.	bagus banyak promo	bagus,banyak,promo
5.	cepat proses pemasana hotel	cepat,proses,pemasana,hotel

Tabel 4.5 Proses Tokenization

#### 4.3.6 Stopword

Setelah melewati proses tokenisasi selanjutnya yaitu proses stopword removal. Untuk melakukan stopword removal, maka source code yang digunakan adalah sebagai berikut:

```

● ● ●

from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
factory = StopWordRemoverFactory()
stopwords = factory.get_stop_words()
print(stopwords)

def stopword_text(tokens):
    cleaned_tokens = []
    for token in tokens:
        if token not in stopwords:
            cleaned_tokens.append(token)
    return cleaned_tokens

df['stop'] = df['tokenized'].apply(stopword_text)
df.head()
df.tail()

```

Gambar 4.8 Source Code Stopword

Adapun hasil dari stopword sebagai berikut:

No.	Sebelum Stopword	Setelah Stopword
1.	[sangat, membantu, dan, diskon, harganya, pantastis]	[sangat, membantu, diskon, harganya, pantastis]
2.	[lengkap, dan, cepat]	[lengkap, cepat]
3.	[kenapa, tidak, bisa, pake, paylater]	[pake, paylater]
4.	[bagus, banyak, promo]	[bagus, banyak, promo]
5.	[cepat, proses, pemasana, hotel]	[cepat, proses, pemasana, hotel]

Tabel 4.6 Proses Stopword

#### 4.3.7 Stemming

Setelah selesai dengan proses stopwords removal, angkah berikutnya adalah stemming. Untuk melakukan stemming, berikut adalah contoh source code yang dapat digunakan:

```

● ● ●

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
stem_factory = StemmerFactory()
stemmer = stem_factory.create_stemmer()

def stemming_text(tokens):
    hasil = [stemmer.stem(token) for token in tokens]
    return hasil

df['stemmed'] = df['stop'].apply(stemming_text)
df.head(100)

df['FINAL_READY'] = df['stemmed'].apply(lambda x: ' '.join(x))
df.head(25)

```

Gambar 4.9 Source Code Stemming

Adapun contoh dari hasil stemming sebagai berikut:

No.	Sebelum Stemming	Setelah Stemming
1.	[cepat, proses, pemasana, hotel]	[cepat, proses, pemasana, hotel]
2.	[semoga, berhasil, mengecewakan, pemesanan, orang, tua]	[moga, hasil, kecewa, mesan, orang, tua]
3.	[trimakaasihh, smg, mjd, kebahagiaan, semua, umat]	[trimakaasihh, smg, mjd, 47ahagia, semua, umat]
4.	[bagus, banyak, promo]	[bagus, banyak, promo]
5.	[oke, baik, kak]	[oke, baik, kak]

Tabel 4.7 Proses Stemming

#### 4.3.8 Labelling Menggunakan Library Roberta

Data yang telah melewati preprocessing data, selanjutnya akan diberi label menggunakan library Roberta. Library Roberta merupakan library yang dibuat untuk memberi sentimen pada data berbahasa Indonesia. Berikut merupakan contoh source code dalam menggunakan library Roberta.

```
● ● ●

from transformers import pipeline

pretrained_name = "wllwo[indonesian-roberta-base-indolem-sentiment-classifier-fold-0"]

nlp = pipeline(
    "sentiment-analysis",
    model=pretrained_name,
    tokenizer=pretrained_name
)

review = df_filtered_new['FINAL_FIX']
data = []

for i in review:
    data.append(nlp(i))

data

predicted_label = []
confidence = []

for my_list in data:
    for item in my_list:
        predicted_label.append(item['label'])
        confidence.append(item['score'])

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
df_filtered_new['sentiment'] = le.fit_transform(predicted_label)
df_filtered_new['confidence'] = confidence
df_filtered_new

def get_analysis(score):
    if score == 0:
        return 'Negative'
    else:
        return 'Positive'

df_filtered_new['label'] = df_filtered_new['sentiment'].apply(get_analysis)
df_filtered_new
```

Gambar 4.10 Source Code Library Roberta

Pada source code diatas library Roberta digunakan dengan memanggil library Roberta, dengan data yang digunakan kolom komentarnya saja. Lalu, diberi label, score, dan nilai confidence nya. Apabila score = 0 maka, ulasan tersebut memiliki sentimen negatif, sedangkan score = 1 maka, ulasan tersebut memiliki sentimen positif.

Berikut merupakan contoh hasil dari labelling menggunakan library Roberta:

No.	FINAL_FIX	Sentiment	Confidence	Label
1.	sangat bantu diskon harga pantastis	1	0.999755	Positive
2.	sangat kecewa beli tiket dgn paylater reschedule...	0	0.999918	Negative
3.	travel yg tanggung jawab walaupun yg lbh murah...	1	0.987701	Positive
4.	trimakaasihh smg mjd bahagia semua umat amin	1	0.999656	Positive
5.	limit paylater kok turun kecewaaa	0	0.999933	Negative

Tabel 4.8 Contoh Hasil Labelling Roberta

#### 4.3.9 Labelling Menggunakan Library Vader

Data yang telah melewati preprocessing, selanjutnya akan diberi label menggunakan library vader. Library Vader merupakan library yang dibuat untuk memberi label pada data berbahasa Inggris. Pada data penelitian ini menggunakan data berbahasa Indonesia. Maka dari itu, data perlu di translate terlebih dahulu sebelum di labelling.

Berikut merupakan contoh source code dalam menggunakan library translator:

```

● ● ●

!pip install -U deep-translator

df =
pd.read_csv('https://raw.githubusercontent.com/hastricantya/OTA/main/Labelled_DataTravelokaRemoved_Skripsi.csv')
df

from deep_translator import GoogleTranslator

data_indo = df['FINAL_FIX']

process_data_indo = []

for i in data_indo:
    process_data_indo.append(GoogleTranslator(source='id', target='en').translate(i))

process_data_indo

# Create a new DataFrame with the processed data
df_processed = pd.DataFrame({'Translated_TEXT': process_data_indo})

# Save the DataFrame as a CSV file
output_file_path = 'processed_data.csv'
df_processed.to_csv(output_file_path, index=False)

print(f"Processed data saved to {output_file_path}.")

```

Gambar 4.11 Source Code Library Translator

Berikut merupakan contoh ulasan yang melewati terjemahan:

No.	Sebelum Translate	Setelah Translate
1.	sangat kecewa beli tiket dgn paylater reschedule tglnya dr juli juli ibu sakit keras 49recommend bel paylater lg kena tagih dobel tiket yg baru ga masalah bayar yg sdh tiket khan sdh rubah sdngkan biaya reschedule rb orang tp say tetap harus bayar paylater yg sdh rubah jadwal khan aneh jadi asa beli tiket jt kalu sya byar dobel sejak malam sampe skrg no tanggap	very disappointed to buy a ticket with a rescheduled paylater date from July to July, the mother is seriously ill in Indonesia, the paylater bell is subject to being billed for double the new ticket, no problem paying for those who have already changed the ticket, while the rescheduling fee is thousands of people but I still have to pay the paylater, who has changed the schedule, it's strange so I hope to buy a ticket if I pay double since the evening until now no response
2.	sangat bantu diskon harga pantastis	very helpful discount reasonable prices
3.	lama lama jelek bgt aplikasi kmrn beli tiket pesawat pke aplikasi tulis blm pesan aktif skrg 49ecomme tiket dapat masalah koneksi ga 49recommended lama lama aplikasi	it took a long time, the application was really bad, I bought plane tickets using the application, I haven't ordered yet, now I want cents for tickets, I got connection problems, it doesn't take long to recommend the application
4.	kecewa pesan hotel bayar tiba pihak e voucher gagal terbit proses refundnya sd hari kerja konsumen harus proses refund wajib langsung	Disappointed, the hotel ordered the payment to arrive, the e-voucher failed to issue the refund process on working days, the consumer had to

	masuk rekening hari karna salah pihak nya customer servicenya solusi	process the refund, it was mandatory to go directly to the account on the day because the customer service was the wrong solution
5.	makasih yg setia teman baik hati dgn bonus	Thank you, loyal, kind-hearted friend with a bonus

Tabel 4.9 Proses Translate Sebelum Labelling Library Vader

Setelah di translate, selanjutnya akan diberi label menggunakan library vader. Berikut merupakan contoh source code library vader:

```
● ● ●

!pip install -U deep-translator

df =
pd.read_csv('https://raw.githubusercontent.com/hastricantya/OTA/main/Labelled_DataTravelokaRemoved_Skripsi.csv')
df

from deep_translator import GoogleTranslator

data_indo = df['FINAL_FIX']

process_data_indo = []

for i in data_indo:
    process_data_indo.append(GoogleTranslator(source='id', target='en').translate(i))

process_data_indo

# Create a new DataFrame with the processed data
df_processed = pd.DataFrame({'Translated_TEXT': process_data_indo})

# Save the DataFrame as a CSV file
output_file_path = 'processed_data.csv'
df_processed.to_csv(output_file_path, index=False)

print(f"Processed data saved to {output_file_path}.")
```

Gambar 4.12 Source Code Library Vader

Berikut merupakan contoh hasil setelah labelling library Vader:

No.	Translated_Text	Sentimen
1.	very disappointed to buy a ticket with a rescheduled paylater date from July to July, the mother is seriously ill in Indonesia, the paylater bell is subject to being billed for double the new ticket, no problem paying for those who have already changed the ticket, while the rescheduling fee is thousands of people but I still have to pay the paylater, who has changed the schedule, it's strange so I hope to buy a ticket if I pay double since the evening until now no response	Negative
2.	very helpful discount reasonable prices	Positive
3.	it took a long time, the application was really bad, I bought plane tickets using the application, I haven't ordered yet, now I want cents for tickets, I got	Negative

	connection problems, it doesn't take long to recommend the application	
4.	Disappointed, the hotel ordered the payment to arrive, the e-voucher failed to issue the refund process on working days, the consumer had to process the refund, it was mandatory to go directly to the account on the day because the customer service was the wrong solution	Negative
5.	Thank you, loyal, kind-hearted friend with a bonus	Positive

Tabel 4.10 Contoh Hasil Labelling Roberta

#### 4.3.10 Labelling Manual

Data yang telah melewati preprocessing, selanjutnya dilakukan labelling manual. Data yang diambil untuk labelling manual merupakan 15 data teratas dan 15 data terbawah library Roberta dan library Vader. Berikut ulasan menggunakan labelling library Roberta:

No.	Ulasan	Sentimen
1.	sangat bantu diskon harga pantastis	Positive
2.	moga hasil kecewa mesan orang tua	Negative
3.	sangat kecewa beli tiket dgn paylater reschedule tglnya dr juli juli ibu sakit keras indonesia bel paylater lg kena tagih dobel tiket yg baru ga masalah bayar yg sdh tiket khan sdh rubah sdngkan biaya reschedule rb orang tp say tetap harus bayar paylater yg sdh rubah jadwal khan aneh jadi asa beli tiket jt kalu sya byar dobel sejak malam sampe skrg no tanggap	Negative
4.	trimakaasihh smg mjd bahagia semua umat amin	Positive
5.	bagus jadi makin mudah aja	Positive
6.	sangat mudah guna sangat bantu	Positive
7.	lama lama jelek bgt aplikasi kmrn beli tiket pesawat pke aplikasi tulis blm pesan aktif skrg mau sen tiket dapat masalah koneksi ga recomended lama lama aplikasi	Negative
8.	lmyn gampang ringkes utk aplikasi nya	Positive
9.	udah tawar tukar point produk produk kalau memang gabisa pakai kode vouchernya tukar point ichitan gabisa tukar kode cs hubung komplain masalah gak selesai	Negative
10.	bagus cepat layan nya tiket cari	
11.	saldo uang cair transfer rekening bank guna paylater kemudian langsung bayar laku refund dana dana refund sebut paylater tagih ditindaklanjuti lebih cepat jangan langgan pergi satu per satu	Negative
12.	layan cepat lebih efisien bagus	Positive

13.	keren cepat bgt cuma sayang cari bangku blkg sopir paling depan susah bgt mudah an yg sendiri mau tukeran sy sama ank dua biar gak mabok anak klo depan	Negative
14.	ubah harga nya dong harga nya naik turun terus mahal banget coi	Negative
15.	ijin ss save desain nya kak	Positive
16.	jangan lupa update beli po bus malam rosalia indah juragan	Positive
17.	aplikasi tipu kedok inestasi emas	Negative
18.	selalu beri mudah utk cari kamar hotel pergi makas	Positive
19.	limit gak pakai terus dpt notif terus biar pake limit kan aneh refundnya lamaaaaaaaaaaaaaaaa banget	Negative
20.	bagus pokok lancar jaya sukses trus	Positive
21.	mudah sesuai expetasi kamar layan	Positive
22.	kalau tambah tiket kapal laut pak	Positive
23.	harga nya lebih murah dr aplikasi yg metode beli gak ribet	Positive
24.	no hp blokir kok gk ganti verifikasi lewat email aja sungguh aneh	Negative
25.	puas dgn aplikasi mudah mesan hotel jalan keluar kota selalu dapat hotel baik	Positive
26.	cepet proses nya terima kasih	Positive
27.	konsisten bantu konsumen kmarin reques kursi pesawat tp sesuai yg reques seblum padahal bayar kok sangat kecewa	Negative
28.	travel yg tanggung jawab walaupun yg lbh murah tetap	Positive
29.	limit paylater kok turun kecewaaa	Negative
30.	dulu check in online mudah sekarang makin lama makin sulit wbsite la bandara la lebih apa mesti uang ubah jam aja mesti bayar asuransi aja bukan tanggung maskapai mesti yg bayar apa beda nya	Negative

Tabel 4.11 Contoh Hasil Menggunakan Library Roberta

Berikut ulasan menggunakan labelling library Vader:

No.	Ulasan	Sentimen
1.	very helpful discount reasonable prices	Positive
2.	I hope the results will be disappointed by the parents' orders	Negative
3.	thank you, may all the people be happy, amen	Positive
4.	it's good that it's easier	Positive
5.	very easy to use very helpful	Positive
6.	it took a long time, the application was really bad, I bought plane tickets using the application, I haven't ordered yet, now I want cents for tickets, I got connection problems, it doesn't take long to recommend the application	Negative
7.	lmyn is easy to use for its application	Positive
8.	I've bargained to exchange product points if you really can't use the voucher code to exchange point ichitan can't exchange cs codes contact complaints the problem isn't resolved	Negative
9.	it's good, fast service, looking for tickets	Positive

10.	liquid money balance transfer bank account using paylater then immediately pay refund behavior fund refund call paylater bill followed up faster don't subscribe go one by one	Negative
11.	it's really cool, fast, it's just a shame to find a seat behind the driver at the front, it's really difficult, it's easy for the one who wants to exchange me with two children so that the child doesn't get drunk in the front	Positive
12.	permission to save the design sis	Positive
13.	the pylater list never keeps rejecting results	Negative
14.	The application is very helpful, the application is better for the verification menu, the steps for hacker activity are very vicious, please consider it, thank you	Positive
15.	Thank God, a few months ago, it was very easy to order a hotel, I hope you are satisfied	Positive
16.	a thousand read discount is clear, the minimum buy is million, I want to be surprised, hopefully the application is better, bored with the apk	Positive
17.	very accurate to help with inpatient problems	Positive
18.	Ah, it sucks that I want to pay for the latter, it doesn't matter if it fails, it won't repeat, if you don't have enough capital, it's better not to pretend to pay for the latter	Negative
19.	enter ready to pay via bsi	Positive
20.	refund, don't take too long, please give the funds during the day you pay for the transfer, just give it a few hours, transfer, pay cash, not credit, the other day the flight is canceled, the airline has to refund it faster, go back, use it, buy a new ticket, subscribe faithfully since if you want to find an agent	Negative
21.	don't forget to update, buy the beautiful Rosalia night bus po, skipper	Positive
22.	always make it easy to find a hotel room, go to Makas	Positive
23.	If you don't use the limit, you get notifications, so if you use the limit, it's weird, the refund takes so long	Negative
24.	good, smooth, successful, successful then	Positive
25.	easy according to the service room expectations	Positive
26.	the cellphone number is blocked, how come it doesn't change verification via email, it's really strange	Negative
27.	Satisfied with the easy application to order hotels, way out of town, always get good hotels	Positive
28.	fast process thank you	Positive
29.	consistently helping consumers yesterday requesting airplane seats but according to the request before even though they paid how come they were very disappointed	Negative
30.	responsible travel even though the cheaper ones are still	Positive

Tabel 4.12 Contoh Hasil Menggunakan Library Vader

#### 4.3.11 Perbandingan Labelling

Setelah melakukan labelling menggunakan library Roberta, Vader, dan manual, Langkah selanjutnya, yaitu melakukan perbandingan hasil dari ketiga label tersebut. Berikut merupakan tabel perbandingan labelling:

No.	Ulasan	Labelling Vader	Labelling Roberta	Labelling Manual
1.	sangat bantu diskon harga pantastis	Positive	Positive	Positive
2.	moga hasil kecewa mesan orang tua	Negative	Negative	Negative
3.	sangat kecewa beli tiket dgn paylater reschedule tglnya dr juli juli ibu sakit keras indonesia bel paylater lg kena tagih dobel tiket yg baru ga masalah bayar yg sdh tiket khan sdh rubah sdngkan biaya reschedule rb orang tp say tetap harus bayar paylater yg sdh rubah jadwal khan aneh jadi asa beli tiket jt kalu sya byar dobel sejak malam sampe skrg no tanggap	Positive	Negative	Negative
4.	trimakaasihh smg mjd bahagia smua umat amin	Positive	Positive	Positive
5.	bagus jadi makin mudah aja	Negative	Positive	Positive
6.	sangat mudah guna sangat bantu	Negative	Positive	Positive
7.	lama lama jelek bgt aplikasi kmrn beli tiket pesawat pke aplikasi tulis blm pesan aktif skrg mau sen tiket dapat masalah koneksi ga recomended lama lama aplikasi	Negative	Negative	Negative
8.	Imyn gampang ringkes utk aplikasi nya	Negative	Positive	Positive
9.	udah tawar tukar point produk produk kalau memang gabisa pakai kode vouchernya tukar point ichitan gabisa tukar kode cs hubung komplain masalah gak selesai	Negative	Negative	Negative
10.	bagus cepat layan nya tiket cari	Negative	Positive	Positive
11.	saldo uang cair transfer rekening bank guna paylater kemudian langsung bayar laku refund dana dana refund sebut paylater tagih ditindaklanjuti lebih cepat jangan langgan pergi satu per satu	Negative	Negative	Negative
12.	layan cepat lebih efisien bagus	Negative	Positive	Positive
13.	keren cepat bgt cuma sayang cari bangku blkg sopir paling depan susah bgt mudah an yg sendiri mau tukeren sy sama ank dua biar gak mabok anak klo depan	Negative	Positive	Negative
14.	ubah harga nya dong harga nya naik turun terus mahal banget coi	Negative	Positive	Negative

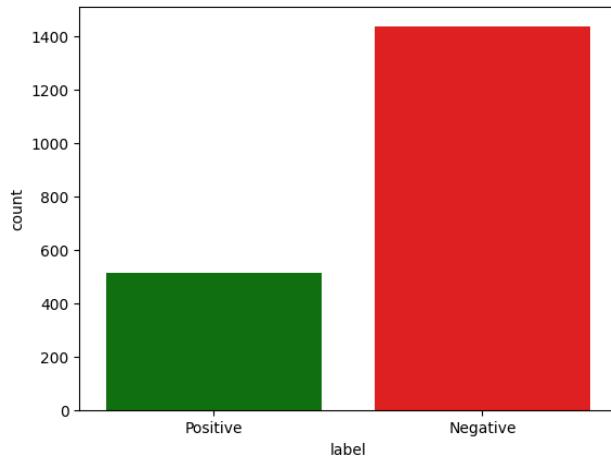
15.	ijin ss save desain nya kak	Negative	Positive	Positive
16.	jangan lupa update beli po bus malam rosalia indah juragan	Negative	Positive	Positive
17.	aplikasi tipu kedok inevstasi emas	Negative	Positive	Negative
18.	selalu beri mudah utk cari kamar hotel pergi makas	Positive	Positive	Positive
19.	limit gak pakai terus dpt notif terus biar pake limit kan aneh refundnya lamaaaaaaaaaaaaaaaaaaaaaa banget	Negative	Negative	Negative
20.	bagus pokok lancar jaya sukses trus	Positive	Positive	Positive
21.	mudah sesuai expetasi kamar layan	Negative	Positive	Positive
22.	kalau tambah tiket kapal laut pak	Positive	Negative	Positive
23.	harga nya lebih murah dr aplikasi yg metode beli gak ribet	Positive	Positive	Positive
24.	no hp blokir kok gk ganti verifikasi lewat email aja sungguh aneh	Negative	Negative	Negative
25.	puas dgn aplikasi mudah mesan hotel jalan keluar kota selalu dapat hotel baik	Positive	Positive	Positive
26.	cepet proses nya terima kasih	Negative	Positive	Positive
27.	konsisten bantu konsumen kmarin reques kursi pesawat tp sesuai yg reques seblum padahal bayar kok sangat kecewa	Negative	Negative	Negative
28.	travel yg tanggung jawab walaupun yg lbh murah tetap	Positive	Positive	Positive
29.	limit paylater kok turun kecewaaa	Negative	Negative	Negative
30.	dulu check in online mudah sekarang makin lama makin sulit wibsite la bandara la lebih apa mesti uang ubah jam aja mesti bayar asuransi aja bukan tanggung maskapai mesti yg bayar apa beda nya	Negative	Negative	Negative

Tabel 4.13 Perbandingan Hasil Labelling Library Roberta dan Vader

Dari tabel diatas dapat dilihat bahwa hasil dari library roberta dan vader tidak jauh berbeda dan dilakukan untuk validasi.

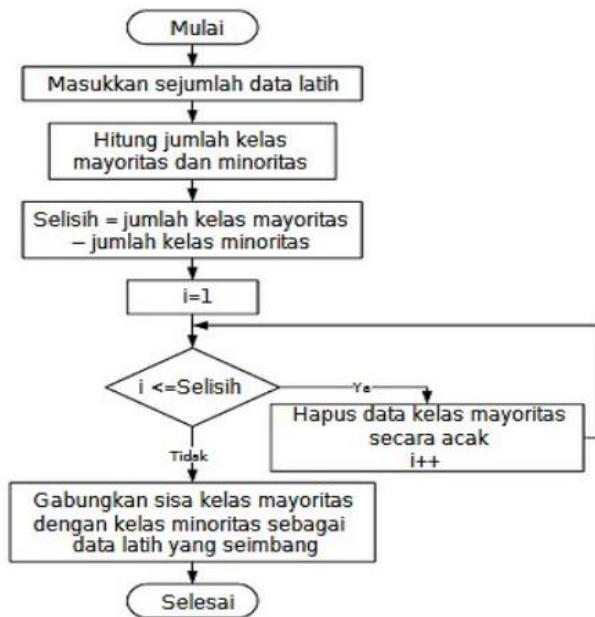
#### 4.3.12 EDA (Explatory Data Analysis)

Pada penelitian ini akan dilakukan beberapa skenario data ulasan karena setelah dilakukan preprocessing dan dilakukan pengecekan jumlah perbandingan data positive dan negative didapat hasil berupa:



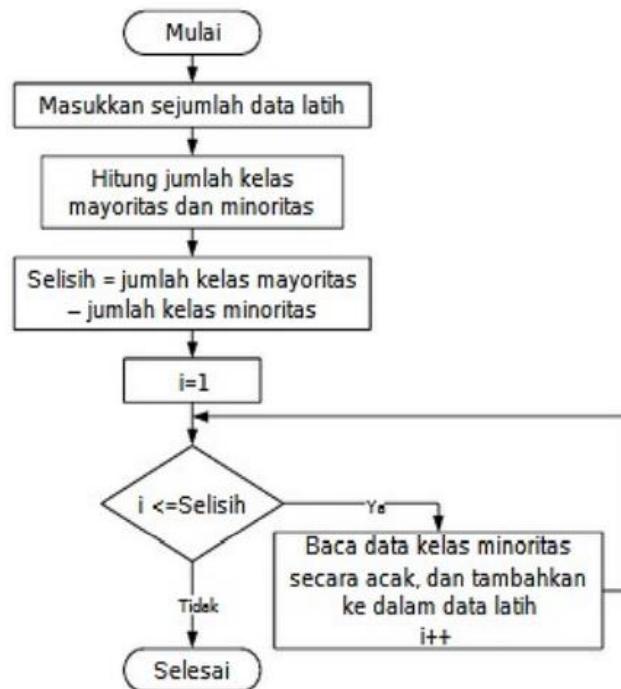
Gambar 4.13 Perbandingan data label positive dan negative

Pada grafik tersebut dapat terlihat bahwa data ulasan yang telah melewati preprocessing data mengalami imbalanced dataset (tidak seimbang). Maka dari itu, perlu dilakukan penyeimbangan data seperti, undersampling dan oversampling. Oversampling bertujuan untuk meningkatkan sampel kelas minoritas sampai sama dengan kelas mayoritas lain dengan menduplikasi secara acak sampel kelas minoritas(H. He et al, 2018). Sedangkan random undersampling yaitu menghasilkan sub sampel acak dari instance kelas mayoritas. Undersampling merupakan metode sampling secara acak memilih sampel di kelas mayoritas dan menambahkannya ke kelas minoritas, membentuk sebuah dataset pelatihan baru (H. He et al, 2018) seperti yang ditunjukkan pada Gambar 4.14.



Gambar 4.14 Flow Metode Random Undersampling

**Oversampling:** Dalam metode ini, contoh dari kelas minoritas dipilih secara acak dan diduplikasi. Instance yang dihasilkan hanya meningkatkan besarnya jumlah kelas minoritas dengan hanya mereplikasi informasi yang sama. Salah satu teknik oversampling yang sering digunakan dalam mengatasi ketidakseimbangan kelas yaitu SMOTE (Synthetic Minority Over-sampling Technique).

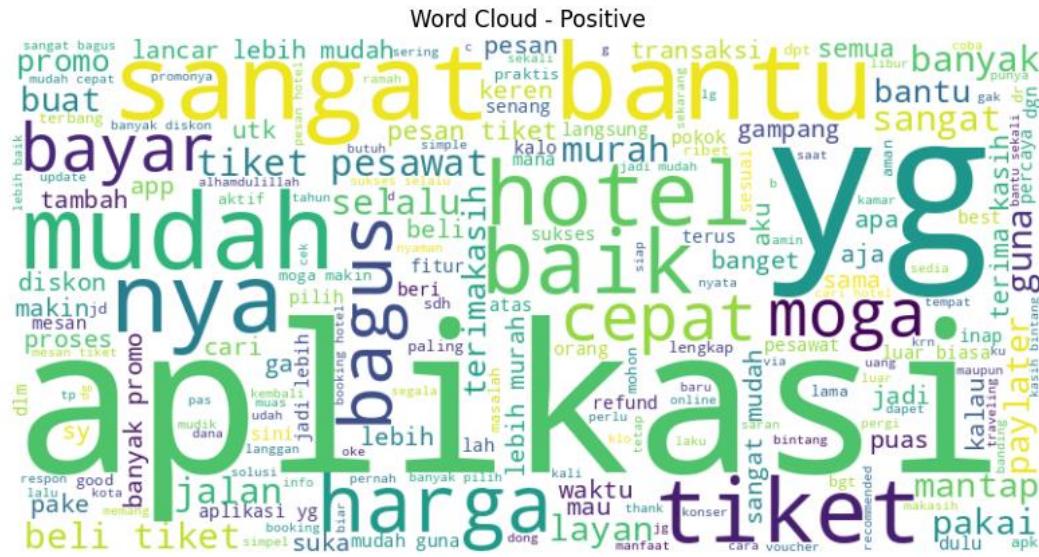


Gambar 4.15 Flow Metode Random Oversampling

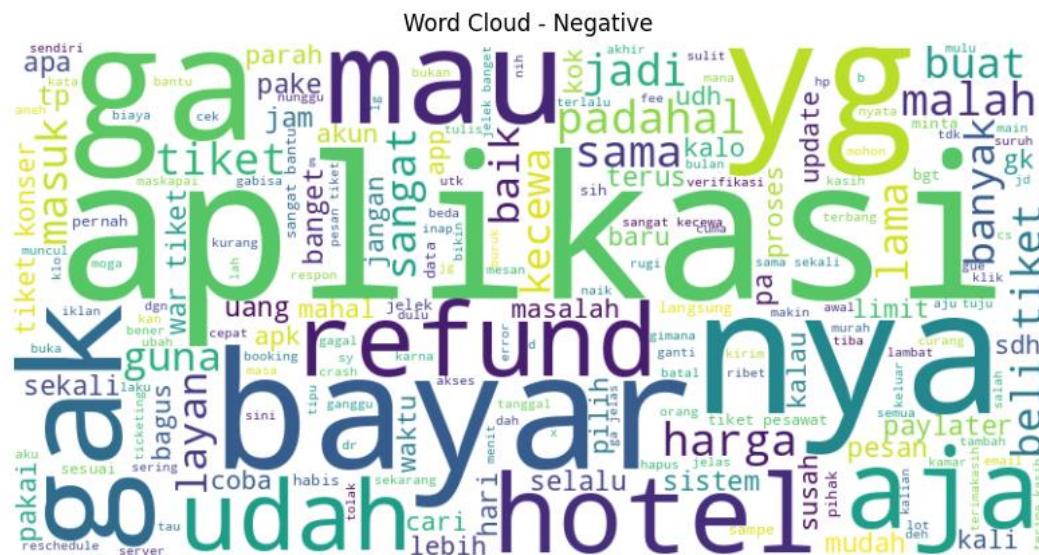
Adapun dari data tersebut dilakukan pengecekan menggunakan word cloud untuk mengetahui kata yang sering muncul pada ulasan OTA.



Gambar 4.16 Word Cloud Ulasan OTA



Gambar 4.17 Word Cloud Positive Ulasan OTA



Gambar 4.18 Word Cloud Negative Ulasan OTA

Berikut merupakan contoh source code untuk word cloud:

```
positive_values_df = df_filtered_new[df_filtered_new['label'] == 'Positive']
positive_values_df

negative_values_df = df_filtered_new[df_filtered_new['label'] == 'Negative']
negative_values_df = negative_values_df[:600]
negative_values_df

final_df = pd.concat([positive_values_df, negative_values_df], ignore_index=True)
final_df_shuffled = final_df.sample(frac=1).reset_index(drop=True)
final_df_shuffled.to_csv('Final_LabelledTravelokaRemoved.csv')
final_df_shuffled
```

Gambar 4.19 Source Code Word Cloud

## 4.4 Pemodelan (Modelling)

### 4.4.1 Hold Out

Langkah selanjutnya setelah melakukan skenario untuk balancing data, yaitu hold out. Hold Out dilakukan untuk menguji model yang akan digunakan. Pada metode ini, dataset yang kita miliki akan dibagi menjadi dua bagian yaitu data latih (training) dan data uji (testing). Umumnya, dataset dibagi dengan perbandingan 70:30, 75:25, dan 80:20. Pada penelitian ini digunakan perbandingan 75:25. Berikut merupakan contoh source code hold out:

```
# Assuming your CSV has a 'text' column for input text and a 'label' column for labels
X = data['Content']
y = data['label']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

# TF-IDF vectorization
tfidf_vectorizer = TfidfVectorizer(max_features=1000) # You can adjust max_features as needed
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)
```

Gambar 4.20 Source Code Hold Out

### 4.4.2 Pemodelan Dengan Algoritma K-Nearest Neighbour

Langkah selanjutnya setelah melakukan hold out untuk menguji data latih, yaitu pemodelan dengan algoritma K-Nearest Neighbour. Pada metode ini, dilakukan pengujian pemodelan KNN nilai k=1 hingga k=25 dengan menggunakan skenario data mentah, data labelling Vader, dan data labelling Roberta. Berikut merupakan hasil pengujian nilai akurasi dengan algoritma KNN.

No.	Nilai K	Roberta Under Sampling	Vader Under Sampling	Data Mentah (Imbalanced)	Roberta Over Sampling	Vader Over Sampling
1.	1	0.64	0.78	0.77	0.97	0.77
2.	2	0.60	0.65	0.74	0.87	0.74
3.	3	0.50	0.63	0.75	0.94	0.75
4.	4	0.49	0.74	0.75	0.91	0.73

5.	5	0.49	0.71	0.75	0.82	0.72
6.	6	0.49	0.75	0.74	0.92	0.71
7.	7	0.48	0.70	0.74	0.72	0.71
8.	8	0.69	0.75	0.74	0.72	0.71
9.	9	0.65	0.72	0.74	0.49	0.71
10.	10	0.74	0.76	0.74	0.80	0.71
11.	11	0.72	0.74	0.74	0.79	0.71
12.	12	0.75	0.78	0.74	0.80	0.71
13.	13	0.74	0.76	0.74	0.76	0.71
14.	14	0.77	0.76	0.75	0.79	0.71
15.	15	0.75	0.76	0.74	0.76	0.73
16.	16	0.77	0.77	0.75	0.81	0.72
17.	17	0.77	0.76	0.74	0.79	0.74
18.	18	0.79	0.75	0.65	0.80	0.73
19.	19	0.77	0.75	0.77	0.78	0.74
20.	20	0.78	0.75	0.73	0.80	0.74
21.	21	0.78	0.75	0.74	0.78	0.76
22.	22	0.79	0.77	0.75	0.80	0.75
23.	23	0.79	0.75	0.74	0.80	0.77
24.	24	0.79	0.76	0.74	0.82	0.76
25.	25	0.80	0.75	0.78	0.80	0.78

Tabel 4.14 Hasil Pengujian dan Perbandingan Skenario Ulasan

Dari hasil pengujian dan perbandingan diatas, dapat dilihat bahwa nilai akurasi terbaik diantara skenario balancing data, yaitu Roberta oversampling dengan nilai  $k=1$  dan akurasi tertinggi, yaitu 97%. Namun, setelah dilakukan evaluasi pada grafik menunjukkan nilai  $k=6$  dengan akurasi 92% yang optimal pada studi kasus. Terlihat bahwa nilai  $k=6$  tidak mengalami underfit maupun overfit dan menemukan titik temu dari grafik testing dan training. Selain itu, pada process time, nilai  $k=6$  mendapat nilai tercepat dibandingkan dengan nilai  $k$  lainnya. Sedangkan, pada tabel lainnya, nilai akurasi yang berwarna kuning merupakan akurasi tertinggi dari masing - masing skenario.

## 4.5 Evaluasi

### 4.5.1 Confusion Matrix

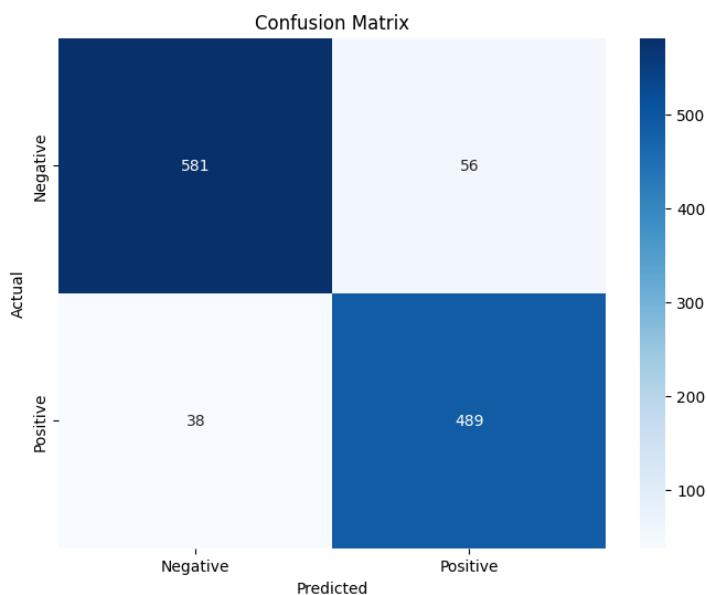
Langkah selanjutnya, yaitu confusion matrix. Confusion Matrix atau juga bisa disebut matriks klasifikasi merupakan suatu metode perhitungan yang biasa digunakan dalam supervised learning (Manning, Raghavan and Schütze, 2009).

Berikut merupakan contoh source code untuk confusion matrix:

```
● ● ●  
from sklearn.metrics import confusion_matrix  
import seaborn as sns  
import matplotlib.pyplot as plt  
  
confusion = confusion_matrix(y_test, y_pred)  
  
plt.figure(figsize=(8, 6))  
sns.heatmap(confusion, annot=True, fmt='d', cmap='Blues', xticklabels=knn_classifier.classes_,  
            yticklabels=knn_classifier.classes_)  
plt.xlabel('Predicted')  
plt.ylabel('Actual')  
plt.title('Confusion Matrix')  
plt.show()
```

Gambar 4.21 Source Code Confusion Matrix

Pada penelitian ini dilakukan perhitungan confusion matrix dengan hasil sebagai berikut:



Gambar 4.22 Confusion Matrix

Dengan jumlah data 4653 dengan data testing yang digunakan 25%, sehingga hasil perhitungan confusion matrix didapatkan nilai:

- a. TP (True Positive) dimana prediksi benar positif dengan jumlah 581
- b. FP (False Positive) dimana prediksi salah positif dengan jumlah 56
- c. FN (False Negative) dimana prediksi salah negative dengan jumlah 38
- d. TN (True Negative) dimana prediksi benar negative dengan jumlah 489

#### 4.5.2 Grafik Evaluasi Model KNN

Langkah selanjutnya, yaitu melakukan evaluasi pengujian dengan mengecek akurasi data training dan data testing. Berikut merupakan contoh source code untuk melakukan evaluasi pengujian model:

```
● ● ●

import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.neighbors import KNeighborsClassifier

X = data['Content']
y = data['Label']
# Generate some example data (replace this with your own data)
# X should be your feature matrix, and y should be your target variable
# X, y = np.random.rand(100, 2), np.random.randint(0, 2, 100)

# Split the data into a training set and a test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

tfidf_vectorizer = TfidfVectorizer(max_features=1000) # You can adjust max_features as needed
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)
# Define a range of K values to try
k_values = list(range(1, 11))

# Initialize lists to store cross-validation scores and training/testing scores
cv_scores = []
train_scores = []
test_scores = []

for k in k_values:
    # Create a KNN classifier with k neighbors
    knn = KNeighborsClassifier(n_neighbors=k)

    # Train the KNN model on the training data
    knn.fit(X_train_tfidf, y_train)

    # Calculate training and testing accuracy
    train_accuracy = knn.score(X_train_tfidf, y_train)
    test_accuracy = knn.score(X_test_tfidf, y_test)

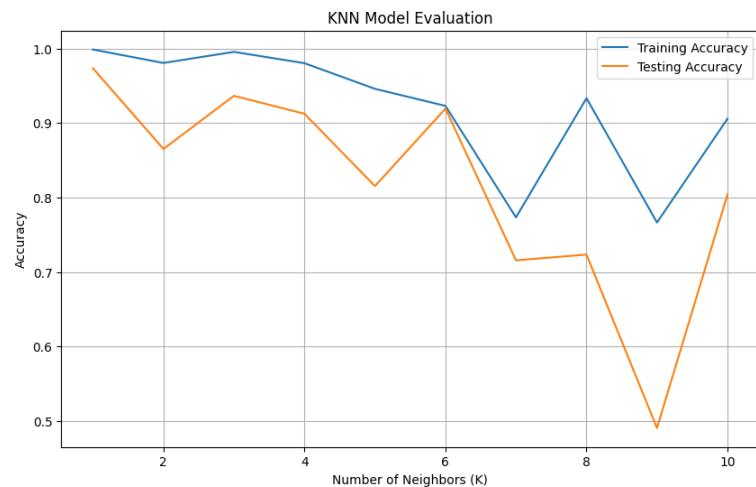
    train_scores.append(train_accuracy)
    test_scores.append(test_accuracy)

# Plot the results
plt.figure(figsize=(10, 6))
plt.plot(k_values, train_scores, label='Training Accuracy')
plt.plot(k_values, test_scores, label='Testing Accuracy')
plt.xlabel('Number of Neighbors (K)')
plt.ylabel('Accuracy')
plt.legend()
plt.title('KNN Model Evaluation')
plt.grid(True)
plt.show()
```

Gambar 4.23 Source Code Evaluasi Pengujian Model

Pada source diatas digunakan untuk mengevaluasi model. Dimana data yang diambil merupakan kolom content dan label. Dengan hold out 75:25 untuk perbandingan data train dan data testing. Dengan memasukkan range nilai k untuk menampilkan nilai k neighbours untuk menghitung train accuracy dengan test accuracy. Lalu, pembentukan grafik training accuracy dan testing accuracy, sehingga dapat mengetahui nilai k=6 adalah yang menemukan titik temu antara training accuracy dan testing accuracy.

Berikut merupakan hasil berbentuk grafik evaluasi nilai k=6 pada model KNN yang telah dibuat:



Gambar 4.24 Grafik Evaluasi Model KNN

Pada grafik tersebut dapat terlihat bahwa nilai k=6 akurasi training dan akurasi testingnya menemukan titik Tengah dimana nilai akurasinya = 92%. Adapun, pada grafik tersebut dapat terlihat bahwa nilai k=1 data trainnya memiliki akurasi = 100% namun, testing akurasinya memiliki akurasi = 97%. Maka dari itu, nilai k yang digunakan pada penelitian ini, yaitu k=6.

Untuk informasi lebih rinci mengenai report hasil klasifikasi dari setiap skenario dapat dilihat dari tabel berikut.

Percobaan nilai K		Akurasi	Waktu Proses	Presisi	Recall	F1-Score	Support
k=1	Negative	97%	1.97 s	96%	100%	98%	637
	Positive			99%	95%	97%	527
k=5	Negative	82%	1.42 s	93%	72%	81%	637
	Positive			73%	93%	82%	527
k=6	Negative	92%	1.19 s	94%	91%	93%	637
	Positive			90%	93%	91%	527
k=7	Negative	72%	1.94 s	93%	52%	67%	637
	Positive			62%	95%	75%	527

Tabel 4.15 Hasil Percobaan Nilai K Skenario Oversampling

Dari tabel berikut, telah dilakukan percobaan nilai k = 1, 5, 6, dan 7.

Adapun, hasil yang didapatkan, yaitu k = 1 dengan nilai akurasi tertinggi 97% dan waktu proses 1.97 detik. Lalu, k=5 dengan nilai akurasi 82% dan waktu proses 1.42 detik. Selanjutnya k=6 dengan nilai akurasi 92% dan waktu proses tercepat 1.19 detik. Dan k=7 dengan nilai akurasi 72% dan waktu proses 1.94 detik. Dari hasil tersebut, nilai k memiliki akurasi tertinggi namun menunjukkan similarity dengan 1 instance lain dan memiliki waktu proses yang paling lama, yaitu 1.97 detik. Lalu, akurasi terbaik lainnya yaitu k=6 dengan nilai akurasi 92% dengan waktu proses tercepat, yaitu 1.19 detik. Nilai k=5 dan 7 menunjukkan akurasi tidak lebih baik dari k=6. Maka dari itu, pada penelitian ini digunakan nilai k=6 sebagai nilai k yang memiliki akurasi yang tinggi dan waktu proses yang tercepat.

#### 4.6 Deployment (Web)

Langkah selanjutnya, yaitu deployment web. Tahapan ini merupakan pengaplikasian dari analisis sentimen dalam bentuk web. Pada web ini terdapat beberapa fitur, seperti input ulasan per kalimat, input ulasan menggunakan csv sesuai template, menampilkan hasil sentimen, menampilkan hasil sentimen berupa word cloud, grafik, dan tabel sentimen. Dalam tahapan deployment web ini terdapat beberapa tahapan, yaitu:

#### **4.6.1 Import Model Ke Dalam Bentuk .pkl (berkas)**

Pada tahapan ini diperlukan berkas berbentuk pkl agar model yang telah dibuat dapat dilakukan deployment menggunakan flask. Berikut merupakan contoh source code import model ke dalam bentuk .pkl:



```
import pickle
filename = 'knn_classifier.pkl'
with open(filename, 'wb') as file:
    pickle.dump(knn, file)
print(f"KNN classifier with {k} neighbors saved to {filename}")
```

Gambar 4.26 Source Code Import Model ke dalam bentuk .pkl

Pada source code tersebut digunakan untuk mengimport ke dalam bentuk berkas. Dimana yang diperlukan adalah model knn dengan nilai k-nya. Dengan nama berkas knn\_classifier.pkl.

#### **4.6.2 Pembuatan Page HTML Menggunakan VsCode**

Pada tahapan ini diperlukan pembuatan landing page agar user dapat berinteraksi menggunakan sistem analisis sentiment. Berikut merupakan contoh source code pembuatan html:

```

● ● ●

<!DOCTYPE html>
<html lang="en">

{&gt; include "includes/head.html" %}

<body>
{&gt; include "components/navbar.html" %}
<div class="container">
<div class="row justify-content-center text-center">
<div class="col-lg-8 p-5">

<h3>OTA Sentiment</h3>
<p>Masukkan review komentar tentang OTA</p>

<div class="mb-3">
<form action="/" method="post">
<div class="form-group">
<label for="review" class="form-label"></label>
<textarea class="form-control" name="review" id="review" cols="5" rows="5">
</textarea>
</div>
<button class="btn btn-primary mt-3" type="submit" id="submitForm">Cek Sentimen!
</button>
</form>
{&gt; if sentiment %}
<p class="mt-3">{{ review }}</p>
<h3 class="mt-3">{{ sentiment }}</h3>
{&gt; endif %}
</div>
</div>
</div>
{&gt; include "includes/script.html" %}
</body>
</html>

```

Gambar 4.27 Source Code Halaman Index

Pada source code diatas digunakan untuk membuat halaman index. Dengan memanggil head.html dan navbar.html. Lalu, menggunakan class div dengan memasukkan gambar generator static, heading 3 judul aplikasi OTA Sentiment, dan paragraph masukkan review komentar tentang OTA. Menggunakan method post untuk user dapat submit text form yang dapat diisi dengan kalimat ulasan OTA. Lalu, memanggil env py dengan % % dan halaman script.html yang berisi link bootstrap.

```

● ● ●

<!DOCTYPE html>
<html lang="en">

{& include "includes/head.html" &}

<style>
    tr {
        text-align: center !important;
    }
</style>

<body>
    {& include "components/navbar.html" &}
    <div class="container">
        <div class="row justify-content-center text-center">
            <div class="col-lg-12 p-5">
                
                <h3>OTA Sentiment</h3>
                <p>Masukkan review komentar tentang OTA</p>
                <p>Template csv yang sesuai dengan web untuk upload klik download <a href="https://intip.in/TemplateSentimentAppOTA">disini</a></p>

                <div class="mb-3">
                    <form action="/upload" method="post" enctype="multipart/form-data">
                        <div class="form-group">
                            <input type="file" name="file" class="form-control">
                        </div>
                        <button class="btn btn-primary mt-3" type="submit" id="submitForm">Cek Sentimen!</button>
                    </form>
                </div>
                <div class="row">
                    {& if wordcloud &}
                    <div class="col-lg-6 py-3">
                        
                    </div>
                    {& endif &}
                    {& if countplot &}
                    <div class="col-lg-6 py-3">
                        
                    </div>
                    {& endif &}
                </div>
                <div class="pt-5 text-center">
                    {& if table &}
                    {{ table | safe }}
                    {& endif &}
                </div>
            </div>
        </div>
    </div>
    {& include "includes/script.html" &}
</body>
</html>

```

Gambar 4.28 Source Code Halaman Upload

Pada source diatas digunakan untuk membuat halaman upload. Dengan memanggil head.html, navbar.html. Lalu, menggunakan class div dengan memasukkan gambar generator static, heading 3 judul aplikasi OTA Sentiment, paragraph masukkan review komentar tentang OTA, dan paragraph template csv yang sesuai dengan web untuk upload klik download disini. Menggunakan method

post untuk user dapat submit text form yang dapat diisi dengan kalimat ulasan OTA.

Lalu dalam hasil submit form akan menampilkan word cloud, plot grafik. Lalu, memanggil env py dengan % % dan halaman script.html yang berisi link bootstrap.



```
<script src="https://cdn.jsdelivr.net/npm/bootstrap@5.3.2/dist/js/bootstrap.bundle.min.js"
integrity="sha384-C6RzsynM9kWDrNNeT87bh950GNYZPhcTNXj1NW7RuBCsyN/oJlpcV8Qyq46cDfL"
crossorigin="anonymous"></script>
```

Gambar 4.29 Source Code Halaman Script

Pada source code diatas digunakan untuk membuat halaman script. Yang berisi link bootstrap.



```
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>OTA Sentiment</title>
    <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.3.2/dist/css/bootstrap.min.css" rel="stylesheet"
integrity="sha384-T3c6CoI6uLrA9TneNEoa7RxnatzcDSmG1MXxSR1GAsXEVDwykc2MPKBM2HN"
crossorigin="anonymous">
</head>
```

Gambar 4.30 Source Code Halaman Head

Pada source code diatas digunakan untuk membuat halaman head yang berisi title dan link bootstrap.



```
<nav class="navbar navbar-expand-lg bg-dark">
    <div class="container-fluid">
        <a class="navbar-brand text-white" href="/">OTA Sentiment</a>
        <button class="navbar-toggler" type="button" data-bs-toggle="collapse" data-bs-target="#navbarNav"
aria-controls="navbarNav" aria-expanded="false" aria-label="Toggle navigation">
            <span class="navbar-toggler-icon"></span>
        </button>

        <div class="collapse navbar-collapse" id="navmenu">
            <ul class="navbar-nav ms-auto">
                <li class="nav-item">
                    <a href="/" class="nav-link text-white">Home</a>
                </li>
                <li class="nav-item">
                    <a href="/upload" class="nav-link text-white">Upload</a>
                </li>
            </ul>
        </div>
    </div>
</nav>
```

Gambar 4.31 Source Code Halaman Navbar

Pada source code diatas digunakan untuk membuat halaman navbar yang berisi panggilan navbar untuk membuat navbar berisi direct ke halaman lainnya.

### 4.6.3 Import dan build menggunakan flask

```
● ● ●

import base64
import io
import pickle

import matplotlib
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from flask import Flask, render_template, request
from wordcloud import WordCloud

matplotlib.use("Agg")

app = Flask(__name__)

with open('knn_classifier.pkl', 'rb') as file:
    knn = pickle.load(file)

with open('tfidf_vectorizer.pkl', 'rb') as tfidf_file:
    tfidf_vectorizer = pickle.load(tfidf_file)

@app.route("/", methods=["GET", "POST"])
def index():
    if request.method == "POST":
        review = request.form["review"]
        review_tfidf = tfidf_vectorizer.transform([review])
        prediction = knn.predict(review_tfidf)[0]
        return render_template("index.html", review=review, sentiment=prediction)
    else:
        return render_template("index.html")

@app.route("/upload", methods=["GET", "POST"])
def upload():
    if request.method == "POST":
        df = pd.read_csv(request.files["file"])
        text_data = df['content']
        text_tfidf = tfidf_vectorizer.transform(text_data)
        predictions = knn.predict(text_tfidf)
        df['sentiment'] = predictions

        wordcloud = WordCloud(
            width=800, height=400, background_color='white').generate(' '.join(text_data))
        img_data = io.BytesIO()
        plt.figure(figsize=(8, 4))
        plt.imshow(wordcloud, interpolation='bilinear')
        plt.axis('off')
        plt.savefig(img_data, format='png')
        img_data.seek(0)
        img_base64 = base64.b64encode(img_data.read()).decode()

        plt.figure(figsize=(8, 4))
        sns.countplot(data=df, x='sentiment')
        plt.title('Sentiment Distribution')
        plt.xlabel('Sentiment')
        plt.ylabel('Count')
        countplot_img_data = io.BytesIO()
        plt.savefig(countplot_img_data, format='png')
        countplot_img_data.seek(0)
        countplot_img_base64 = base64.b64encode(
            countplot_img_data.read()).decode()

        df = df.to_html(classes="table table-striped", index=False)
        return render_template("upload.html", table=df, wordcloud=img_base64,
countplot=countplot_img_base64)
    else:
        return render_template("upload.html")

if __name__ == "__main__":
    app.run()
```

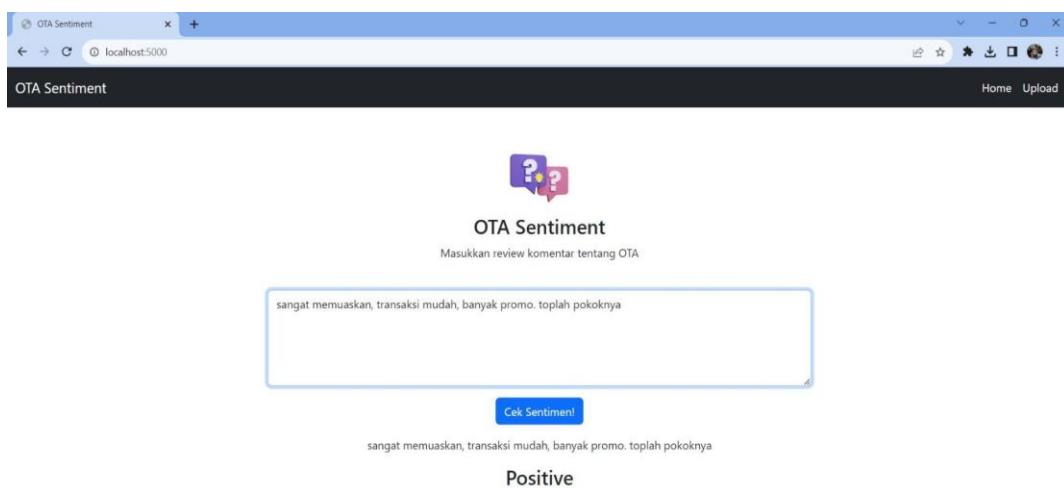
Gambar 4.32 Source Code app.py

Pada source code diatas digunakan untuk app python. Dengan import base64 untuk encode dan decode 64. Import io untuk manage file input dan output.

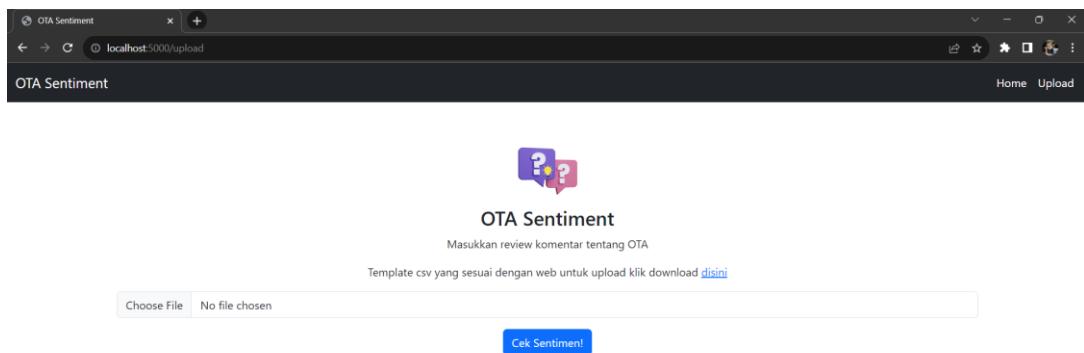
Import pickle untuk import berkas .pkl. import matplotlib untuk membuat grafik. From flask dengan menggunakan flask template dan requestnya. Lalu, membuka berkas knn\_classifier, tf\_idf\_vectorizer, dengan menggunakan method post untuk submit yang menghasilkan hasil sentimen dari ulasan yang disubmit. Lalu, terdapat generate word cloud dan grafik dari ulasan yang diupload.

#### 4.6.4 Deployment dan Visualisasi

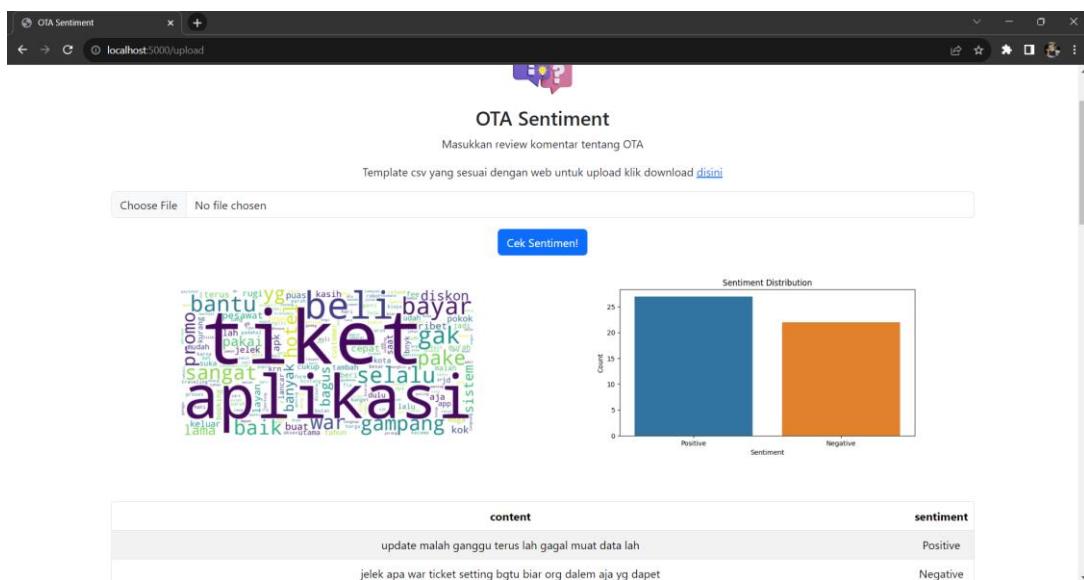
Pada tahapan deployment dan visualisasi dalam bentuk web yang dapat digunakan oleh user. Terdapat dua halaman, yaitu halaman index untuk user mengecek sentimen per kalimat dan halaman upload untuk user mengecek sentimen per csv.



Gambar 4.33 Tampilan Halaman Index



Gambar 4.34 Tampilan Halaman Upload Sebelum Submit



Gambar 4.35 Tampilan Halaman Upload Sesudah Submit

Pada sistem OTA sentiment yang telah di deploy dilakukan pengecekan dari data testing yang telah melewati preprocessing sejumlah 50 data. Dari 50 data tersebut, terdapat 4 data yang sentimennya salah, sehingga akurasinya 92%.

Berikut merupakan data yang sentimennya salah:

No.	Ulasan	Sentimen Salah	Sentimen Benar	Analisis Mengapa Terjadinya Kesalahan
1.	update malah ganggu terus lah gagal muat data lah	Positive	Negative	Karena kata “update” terdapat pada kata pertama dan oleh sistem update=terbaru=baik
2.	biaya platform besar asli besar	Positive	Negative	Karena kalimat tersebut tidak menunjukkan kalimat yang memiliki sentimen negatif untuk sistem seperti, mahal.
3.	gara elo gue gak nonton konser suga	Positive	Negative	Karena kata “gak” tidak dibaca oleh sistem dan hanya membaca “nonton”.
4.	apknya gak cocok buat war tiket war aja dah crash pas jam padahal biaya platfrom nya mahal gak nampung banyak customer	Positive	Negative	Karena kata “gak” tidak dibaca oleh sistem dan hanya membaca “nonton”.

Tabel 4.16 Pengujian Pada Sistem Dengan Hasil Salah

Berikut merupakan data yang sentimennya benar:

No.	Ulasan	Sentimen Salah
1.	sangat manfaat cara pakai gampang fiktur yg sedia lengkap pokok good	Positive
2.	fee doang yg gedeeee sistem ancoorrr	Negative
3.	alhamdulillah bagus terus smoga bnyk diskon utama customer yg mungkin sdh tahan pakai aplikasi brp tahun	Positive
4.	beli tiket bus nyata habis saat selesai bayar online lalu uang balik hubung costumer service nya robot chat otomatis bantu hati hati kawan masalah sistem ganti rugi pwnumpang	Negative

Tabel 4.17 Pengujian Pada Sistem Dengan Hasil Benar

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berikut adalah ringkasan tahapan membangun model analisis sentimen ulasan aplikasi online travel agent di google play store:

1. Tahapan membangun model analisis sentimen adalah studi literatur, analisis kebutuhan, perancangan model yang meliputi pengumpulan data, penyaringan data, text preprocessing, EDA, pelabelan data, klasifikasi model, pembagian data hold out, evaluasi model klasifikasi, dan visualisasi. Pada tahapan pengumpulan data dilakukan dengan login pada web Google Play Store menggunakan akun pribadi dan melakukan code pengambilan data menggunakan library Google Play Scraper. Lalu, melakukan filtering nama aplikasi sebelum ke tahapan preprocessing. Pada tahapan preprocessing dilakukan cleansing, case folding, tokenization, stopword, dan stemming. Selanjutnya, dilakukan labelling menggunakan library Roberta dan vader. Setelah itu, dilakukan EDA dalam bentuk grafik batang hasil labelling dan wordcloud. Lalu, dilakukan pembagian data hold out 75:25, dan dilakukan evaluasi model dengan confusion matrix. Lalu, visualisasi menggunakan web (html).
2. Dalam pengembangan model klasifikasi menggunakan algoritma KNN (K-Nearest Neighbour) menggunakan berbagai macam skenario. Skenario yang dibuat terdiri dari penggunaan library Vader dan Roberta untuk pelabelan data. Lalu, terdapat pengecekan EDA dimana data yang didapatkan imbalanced, sehingga perlu dilakukan skenario pengecekan hasil akurasi

undersampling dan oversampling. Adapun, pengecekan skenario nilai tetangga  $k=1$  hingga  $k=25$  untuk kedua skenario tersebut. Hasil penelitian menunjukkan bahwa akurasi tertinggi diperoleh oleh model klasifikasi menggunakan library Roberta dengan metode oversampling dengan model KNN nilai  $k=6$ , yaitu 92% dan process time 1,14 detik. Model library Roberta oversampling  $k=6$  juga memberikan nilai F-1 Score yang baik, yaitu untuk kelas negatif 93% dan untuk kelas positif, 91%. Evaluasi confusion matrix menunjukkan tingkat kebenaran prediksi yang tinggi, yaitu 94% untuk kelas negatif dan 90% untuk kelas positif. Oleh karena itu, model klasifikasi menggunakan menggunakan library Roberta dengan metode oversampling nilai  $k=6$  dipilih untuk diterapkan pada proses visualisasi.

3. Proses visualisasi ulasan aplikasi OTA dilakukan dengan menyajikan hasil sentimen, grafik untuk setiap kelas sentimen, wordcloud pada dataset yang diupload, dan tabel kelas sentimen. Pada proses visualisasi dibuat dengan menggunakan html (web). Dengan mengimport model yang telah dibuat berupa berkas (.pkl) lalu mengimport flask dan dimasukkan ke dalam html.

## 5.2 Saran

Berikut beberapa saran yang dapat diberikan untuk pengembangan penelitian ini selanjutnya:

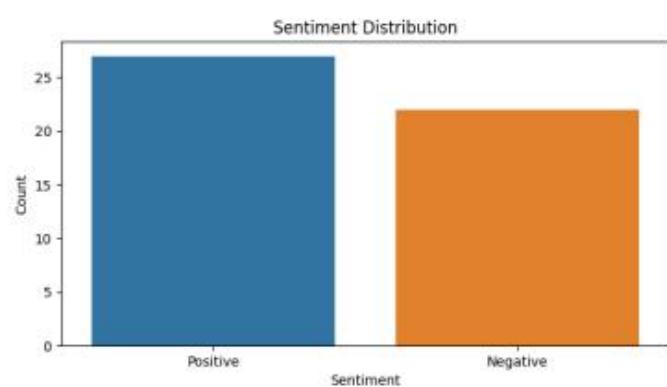
1. Memperluas kemampuan sistem untuk mengatasi ulasan dengan kata-kata singkatan dan kesalahan pengejaan (typo), sehingga dapat meningkatkan akurasi dalam mengolah data yang tidak standar.

2. Menambahkan kata tidak baku seperti “gak” saat preprocessing, sehingga dapat meningkatkan akurasi dalam pemberian label sentiment.
3. Menambah kelas sentimen netral ketika library sudah menyediakannya di kemudian hari.
4. Menyempurnakan fitur yang dikeluhkan oleh pengguna pada aplikasi OTA dan menjaga fitur yang telah baik oleh developer aplikasi OTA. Berdasarkan data penelitian, kata sentimen negatif ulasan yang sering muncul, yaitu refund, bayar, war, dan limit paylater. Maka dari itu, pihak developer perlu melakukan improvisasi terkait fitur tersebut. Sedangkan, kata sentimen positif ulasan yang sering muncul, yaitu kemudahan, kelancaran, kecepatan, promo, dan transaksi, sehingga fitur – fitur terkait perlu dijaga agar pengguna puas dan terbantu saat menggunakan aplikasi OTA.

Dengan mengimplementasikan saran-saran di atas, diharapkan penelitian ini dapat memberikan kontribusi yang lebih besar dalam analisis sentimen pada ulasan aplikasi OTA baik di Google Play Store atau App store, serta meningkatkan kualitas dan keakuratan model klasifikasi yang digunakan.

## LAMPIRAN

Pengujian 50 data pada sistem sentimen menggunakan data yang telah melewati preprocessing.



<b>content</b>	<b>sentiment</b>
update malah ganggu terus lah gagal muat data lah	Positive
jelek apa war ticket setting bgtu biar org dalem aja yg dapet	Negative
app mesan paling murah dulu pake krn bnyk promo skrng udah jarang jd booking hotel langsung order call hotel suka lebih murah drpd aplikasi	Positive
moga tetap lancar dn gampang dlm beli tiket	Positive
sangat manfaat cara pakai gampang fiktur yg sedia lengkap pokok good	Positive
promo promo tambah donk min	Positive
main war tiket cukup jahat tipu	Negative
aplikasi mantap cepet mudah	Positive
bagi baik lama tahun selalu tani selalu beri baik bets pokok jamin keamana beli tiket khusus sekse selalu	Positive
gampang beli tiket mogak makin banyak diskon	Positive
fee doang yg gedeeee sistem ancoorrr	Negative
semua jadi gampang mudah lalu aplikasi percaya	Positive
sngat bantu butuh tiket utk jalan keluar kota utama mudik lebaran	Positive
alhamdulillah bagus terus smoga bnyk diskon utama customer yg mungkin sdh tahan pakai aplikasi brp tahun	Positive

content	sentiment
senang sekali bwli tiket pesawat sini selalu promo	Positive
sangat puas pakai apk respon cepat bantu	Positive
saat genting loading mlu gaesiki	Negative
payleter kurang baik bayar aplikasi bayar	Negative
sangat bantu penting traveling booking hotel tiket pesawat	Positive
apk eror terus gak fungsi	Negative
rasa rugi transaksi tiket pesawat karna jadi refun tiket malah suruh bayar biarin kena ojk	Negative
bagus kok aplikasi sangat bantu proses cepat gk ribet	Positive
total harga sblm bayar ubah selisih lumayan banyak padahal hy beda waktu mnt	Negative
buat war jelek banget allah puasa bikin gedeg kali benah	Negative
cari tiket gak pake ribet	Negative
jelek licik ga adil fix your app asap do n t corrupt	Negative
cukup puas karna responya cepat pokok recomendasi bngat lah buat traveling	Positive
beli tiket bus nyata habis saat selesai bayar online lalu uang balik hubung costumer service nya robot chat otomatis bantu hati hati kawan masalah sistem ganti rugi pwnumpang	Negative
aku suka layan makanya kasih bintang	Positive
kecewa layan atas beli tiket garuda indonesia hari hari minggu tuju gorontalo sistem adu	Negative

content	sentiment
robot feedback lama ganti refund juta harga tiket juta bapak ibu saran pindah aplikasi travels penuh layan cenderung rugi konsumen contoh istri	
saudara saudari jangan download apk sampe nih beli e tiket skrg keluar e tiket	Negative
aplikasi selalu tolong saat pergi keluar kota	Positive
apknya gak cocok buat war tiket war aja dah crash pas jam padahal biaya platfrom nya mahal gak nampung banyak customer	Positive
baru coba kasih bintang dulu	Negative
ngecrush parah banget kemaren	Negative
dah lama pake krn urus refund bikin illfill udah ngisi form disuru nunggu minggu eh bulan kemudian disuru ngisi form aplikasi saing banyaaaak klo layan kek gin mending auto uninstall aplikasi deh	Negative
tadi enak bgt bru aktifin paylatter udh dpet pake selalu lancar udh upgrade kok jd turun lg buat andal plah bgini kecewa si	Negative
baik beri tiket diskon costomer	Positive
lama enjoy aja pakai aplikasi terima kasih	Positive
parah sekarang menu reschedule cuma pajang asli gabisa semudahbl kata kendala gatau kapan selesainye ujung suruh urus sendiri nel maskapai helooooo ngapain w beli elu kalo gamaau repot ngurus payah lo uninstal fix	Negative
the best banyak diskon langgan baik tiket pesawat maupun hotel	Positive

<b>content</b>	<b>sentiment</b>
bagus kok aplikasi sangat bantu proses cepat gk ribet	Positive
dev tambah metode bayar pake aplikasi bayar gampang oakai apk uang ribet	Positive
biaya platfrom besar asli besar	Positive
fee gede sistem gak jelas makan duit haram tuh gak pernah berkah loh mana bulan puasa nih tiati	Negative
aplikasi bagus kalo diadain dong tiket pelni	Positive
akun hp tiba hilang akses biaa akses guna laptop ipad	Negative
sy daftar paylater kok uda tuju batal kurang puas	Negative
gara elo gue gak nonton konser suga	Positive

## DAFTAR PUSTAKA

- Afuan, L. (2018). Sentimen Analisis Di Twitter (Studi Kasus: Resepsi Pernikahan Putri Presiden RI Ke-7). *Semnasteknomedia Online*, 6(1), 1-6.
- Aisha Alfani W. P. R., Fahrur Rozi, F. S. (2021). Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(1), 155–160.
- Alasmari, S. F., Dahab, M. 2017. Sentiment Detection, Recognition, and Aspect Identification. *International Journal of Computer Applications*, Vol. 177, No. 2, hal. 31-38.
- Amrullah, A. Z., Anas, A. S., & Hidayat, M. A. J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal BITe*, 2(1), 40–44. <https://doi.org/10.30812/bite.v2i1.804>
- Andriana, H., Hilab, S. S., & ... (2023). Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter terhadap KTT G20 di Indonesia. *JURIKOM (Jurnal ...)*, 10(1), 60–67. <https://doi.org/10.30865/jurikom.v10i1.5427>
- Aqlan, A. A. Q., Manjula, B., Naik, R. L. 2019. *A Study of Sentiment Analysis: Concepts, Techniques, and Challenges. Proceedings of International Conference on Computational Intelligence and Data Engineering*, hal. 147-162.
- Cahyo, J. (2022). *Analisis Sentimen Pengguna Twitter Terhadap Penerapan Pembelajaran Tatap Muka Pada Masa Pandemi Covid-19 Menggunakan ....* [http://repository.upnjatim.ac.id/id/eprint/6923%0Ahttp://repository.upnjatim.ac.id/6923/1/18082010034\\_Cover.pdf](http://repository.upnjatim.ac.id/id/eprint/6923%0Ahttp://repository.upnjatim.ac.id/6923/1/18082010034_Cover.pdf)
- Chapman, P et al. (2000). CRISP-DM 1.0: Step-by-Step Data Mining Guide.
- Cici, H., & Hwan, S. Y. (2023). *Mitigating Class Imbalance in Sentiment Analysis through GPT-3-Generated Synthetic Sentences*. <https://doi.org/10.3390/app13179766>
- Erfina, A., Basryah, E. S., Saepulrohman, A., & Lestari, D. (2020). Analisis Sentimen Aplikasi Pembelajaran Online di Play Store pada Masa Pandemi Covid-19 Menggunakan Algoritma Support Vector Machine. Seminar Nasional Informatika (SEMNASIF), 1(1), 145-152.
- Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*, 6(1), 64–69.
- Emarketer. (2018). *Global Digital Travel Sales 2018 Ease of Mobile Payment Options Boost Online Bookings*.
- Etaiwi, W & Naymat, G. (2017). The Impact of applying Different Preprocessing Steps on Review Spam Detection. *Procedia Computer Science* 113 (2017) 273-279. <https://www.sciencedirect.com/science/article/pii/S1877050917317787>
- Fajri, K., Jey, H. L., & Timothy, B. (2020). *In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing (AACLIJCNLP 2020)*.
- Feldman, R., & Sanger. J. (2007). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press,

- Firdaus, M. F. El, Nurfaizah, N., & Sarmini, S. (2022). Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor. *JURIKOM (Jurnal Riset Komputer)*, 9(5), 1329. <https://doi.org/10.30865/jurikom.v9i5.4774>
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- He, H., Zhang, W., & Zhang, S. (2018). *A novel ensemble method for credit scoring: adaption of different imbalance ratios*. *Expert Systems With Applications*. doi:10.1016/j.eswa.2018.01.012
- Haryanto. (2017). Text Mining dan Sentimen Analisis Twitter pada Gerakan LGBT. *INTUISI: Jurnal Psikologi Ilmiah*, 9(1), 18–25.
- Hutto, C.J. and Gilbert, E. (2014). *VADER: A Parsimonious Rule-based Model for Eighth International AAAI Conference on Weblogs and Social Media*, 18. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109>
- Ilmawan, L. B., & Winarko, E. (2015). Aplikasi Mobile untuk Analisis Sentimen pada Google Play. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 9(1), 53. <https://doi.org/10.22146/ijccs.6640>
- Kannan & Gurusamy. (2015). Preprocessing Techniques for Text Mining. [https://www.researchgate.net/publication/273127322\\_Preprocessing\\_Techniques\\_for\\_Text\\_Mining](https://www.researchgate.net/publication/273127322_Preprocessing_Techniques_for_Text_Mining)
- Katadata (2022). Survei Populix: Traveloka Aplikasi Perjalanan Online Paling Banyak Disukai Konsumen. <https://databoks.katadata.co.id/datapublish/2022/12/16/survei-populix-traveloka-aplikasi-perjalanan-online-paling-banyak-disukai-konsumen>
- Kotsiantis, S. B., D. Kanellopoulos dan P. E. Pintelas. 2006. *Data Preprocessing for Supervised Learning*. *International Journal of Computer Science* Volume 1 Number 2. Patras: University of Patras.
- Kusrini dan Lutfhi, E.T., 2009. Algoritma Data Mining, Andi Publishing, Yogyakarta.
- Lidya, K.,S, Opim, S., & Syahril, E. (2015). Sentiment Analysis Pada Text Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (K-NN). Seminar Nasional Teknologi Informasi dan Komunikasi 2015, Yogyakarta.
- M. Bramer, *Principles of Data Mining*, Springer, 2007.
- Maharani, L. (2022). Evaluasi User Experience pada Aplikasi Pemesanan Tiket Pesawat Traveloka dan Tiket.com Menggunakan User Experience Questionnaire (UEQ). *Universitas Islam Negeri Syarif Hidayatullah Jakarta*. <https://repository.uinjkt.ac.id/dspace/handle/123456789/65499>
- Manning, C, Raghavan, P, dan Schutze, H. 2009. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press. Dapat diunduh dari: <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>.
- Marisa, F., Maukar, A. L., Farhan, A., Widodo, E. A., Sa'adah, I., & Dasilva, R. T. L. (2021). Pengukuran Tingkat Kematangan Kopi Arabika Menggunakan Algoritma K-Nearest Neighbour. *JIMP: Jurnal Informatika Merdeka Pasuruan*, 6(3), 1-5. <http://dx.doi.org/10.37438/jimp.v6i3.346>

- Obviously.ai. (2022). Cara Mengetahui Apakah Model Machine Learning Anda Memiliki Performa Baik. <https://www.obviously.ai/post/machine-learning-model-performance>.
- Prasetyo Adhi, & Wita, N. (2022). Analisis Minat Pengguna OTA Tiket.com di Indonesia Menggunakan Model Pendekatan Modifikasi UTAUT2. *Jurnal Manajemen Dan Keuangan*, 11(1), 36–54.
- Rahayu, S., MZ, Y., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 98–106. <https://doi.org/10.29408/edumatic.v6i1.5433>
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50. <https://doi.org/10.35314/isi.v3i1.335>
- Sepri, D. (2020). Penerapan Algoritma Naïve Bayase Untuk Analisis Kepuasan Penggunaan Aplikasi Bank. *JOSYC (Journal of Computer System and Informatics)* 2(1), 135-139.
- Sugianto, C. A., & Apandi, T. H. (2018). Pengaruh Tokenisasi Kata N-Grams Spam SMS Menggunakan Support Vector Machine [Preprint]. INA-Rxiv. <https://doi.org/10.31227/osf.io/vjc7k>
- Sumarlin (2015). Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM. *Jurnal Sistem Informasi Bisnis* (01) 2015. <http://ejournal.undip.ac.id/index.php/jsinbis>
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan. *Edumatic : Jurnal Pendidikan Informatika*, 4(1), 37–46. <https://doi.org/10.29408/edumatic.v4i1.2081>
- Tiket.com. (2023). Tiket.com – Tentang Kami. <https://www.tiket.com/info/about-us>.
- Traveloka. (2021). Traveloka - Tentang Kami. <https://www.traveloka.com/id-id/about-us>.
- Wahyudi, R., & Kusumawardana, G. (2021). Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *Jurnal Informatika*, 8(2), 200–207. <https://doi.org/10.31294/ji.v8i2.9681>