

SKRIPSI



**ANALISIS SENTIMEN APLIKASI SAYURBOX BERDASARKAN
ULASAN DI GOOGLE PLAYSTORE MENGGUNAKAN NAÏVE BAYES
CLASSIFIER**

AQSHAL WIN MAHARA

NIM. 2010512118

**PROGRAM STUDI S1 SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN
JAKARTA
2024**

**ANALISIS SENTIMEN APLIKASI SAYURBOX BERDASARKAN
ULASAN DI GOOGLE PLAY STORE MENGGUNAKAN NAÏVE BAYES
CLASSIFIER**

Aqshal Win Mahara

ABSTRAK

Perdagangan elektronik, atau *e-commerce*, adalah aktivitas jual atau beli melalui internet yang meliputi pembelian, penjualan barang, layanan, dan informasi secara *online*. Pertumbuhan *e-commerce* dipicu oleh kemajuan teknologi dan internet, yang memunculkan sebuah berbagai perusahaan *e-commerce*, termasuk *Sayurbox*, sebuah platform *e-commerce* yang berfokus pada buah dan sayur – sayuran, menghadapi tantangan dalam menjaga peringkatnya yang menurun di pasar yang kompetitif. Evaluasi dilakukan melalui analisis sentimen pengguna dengan metode Naïve Bayes Classifier terhadap ulasan di *Google Play Store*. Data dikumpulkan dari Juni 2024 menggunakan cara scraping menggunakan python dan mengambil data sebanyak 1000 ulasan review dengan jenis data *public*. Data melalui tahapan scraping sebanyak 1000 data lalu tahapan preprocessing (*Cleaning, Case Folding, Tokenizing, Normalisasi, Stopword Removal, Stemming*), pembobotan kata menggunakan TF-IDF, dan pembagian data. Naïve Bayes menghasilkan accuracy 86%, precision 94.4%, recall 85%, F1-Score 89.4% dengan rasio data latih dan uji 80:20. Selain hasil klasifikasi, terdapat hasil visualisasi wordcloud beserta grafik frekuensi kata terbanyak yang digunakan untuk melakukan identifikasi sentimen positif dan negatif. Setelah melakukan visualisasi terdapat nilai positif pada barang dengan bobot 370 dan untuk kata kirim terdapat hasil negatif dengan bobot 125. Luaran lainnya yaitu dilakukan pembuatan design ui website sederhana menggunakan figma untuk melakukan visualisasi hasil.

Kata Kunci : Analisis Sentimen, *Google Play Store*, *Naïve Bayes Classifier*, *Sayurbox*

**SENTIMENT ANALYSIS OF THE SAYURBOX APPLICATION BASED
ON REVIEWS ON GOOGLE PLAY STORE USING NAÏVE BAYES
CLASSIFIER**

Aqshal Win Mahara

ABSTRACT

Electronic *commerce*, or *e-commerce*, is the activity of buying or selling via the internet which includes buying, selling goods, services and information online. The growth of e-commerce is fueled by advances in technology and the internet, which has given rise to various e-commerce companies, including *Sayurbox*, an e-commerce platform that focuses on fruit and vegetables, facing challenges in maintaining its declining ranking in a competitive market. Evaluation was carried out through user sentiment analysis using the Naïve Bayes Classifier method on reviews on the *Google Play Store*. Data was collected from June 2024 using scraping using Python and taking data from 1000 reviews with public data type. The data goes through scraping stages of 1000 data then preprocessing stages (*Cleaning, Case Folding, Tokenizing, Normalization, Stopword Removal, Stemming*), word weighting using TF-IDF, and data division. Naïve Bayes produces 86% accuracy, 94.4% precision, 85% recall, 89.4% F1-Score with a training and test data ratio of 80:20. Apart from the classification results, there are wordcloud visualization results along with a graph of the frequency of the most words which are used to identify positive and negative sentiments. After doing the visualization, there was a positive value for the item with a weight of 370 and for the word send there was a negative result with a weight of 125. Another outcome was the creation of a simple website UI design using Figma to visualize the results.

Keywords: Sentiment Analysis, *Google Play Store*, *Naïve Bayes Classifier*, *Sayurbox*

KATA PENGANTAR

Puji dan syukur dipanjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunianya sehingga Proposal Tugas Akhir (Seminar Teknologi Informasi) ini dapat diselesaikan dengan baik. Proposal Tugas Akhir ini dibuat guna memenuhi persyaratan untuk menyusun Tugas Akhir/Skripsi.

Dalam penyusunan persyaratan proposal ini hingga selesai, terdapat banyak banturn dari banyak pihak. Demikian, segala rasa hormat dan terima kasih disampaikan setinggi – tingginya kepada :

1. Tuhan Yang Maha Esa atas rahmat dan karunia yang tak terhingga.
2. Kedua orang Tua yang telah memberikan dukungannya sepenuh hati.
3. Pak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM, selaku Dekan Fakultas Ilmu Komputer UPNVJ.
4. Ibu Anita Muliawati, S.Kom. M.TI, S.Pd, M.Kom, selaku ketua Program Studi S1 Sistem Informasi.
5. Ibu Ati Zaidah, S.kom, M.T.I, selaku Dosen Pembimbing 1.
6. Pak Muhammad Panji Muslim., S.Pd., M.Kom selaku Dosen Pembimbing 2.
7. Kepada teman – teman yang membantu saya dalam penyusunan penulisan Skripsi ini Vincentius L, M. Fahrizal, Hanifa W, Raditya R, Nugroho A.
8. Seluruh pihak yang tidak dapat disebutkan satu – persatu,

Masih banyak kekurangan yang terdapat di Proposal Tugas Akhir ini, baik dari segi isi maupun penulisan, mengingatkan kurangnya pengetahuan dan pengalaman serta hakikat sebagai manusia yang selalu salah. Oleh karena itu, saran dan kritik yang membangun akan sangat berarti.

Jakarta, Juli 2024



Penulis

LEMBAR PERSETUJUAN

LEMBAR PERSETUJUAN

Yang bertanda tangan di bawah ini:

Nama : Aqshal Win Mahara

NIM : 2010512118

Program Studi : S1 Sistem Informasi

Judul Skripsi/TA : Analisis Sentimen Aplikasi Sayurbox Berdasarkan Ulasan
Di Google Playstore Menggunakan Naive Bayes Classifier

Dinyatakan telah memenuhi syarat dan menyetujui untuk mengikuti ujian sidang skripsi.

Jakarta, 25 Juni 2024

Menyetujui,

Dosen Pembimbing I,



Ati Zaidiah, S.Kom, M.T.I.

Dosen Pembimbing II,



Muhammad Panji Muslim,
S.Pd., M.Kom.

Mengetahui,

Ketua Program Studi,



Anita Muliawati, S.Kom., M.T.I.

LEMBAR ORISINALITAS

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Aqshal Win Mahara
NIM : 2010512118
Fakultas : Ilmu Komputer
Program Studi : S-1 Sistem Informasi

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 20 Juli 2024

Yang Menyatakan,



Aqshal Win Mahara

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Aqshal Win Mahara
NIM : 2010512118
Fakultas : Ilmu Komputer
Program Studi : S-I Sistem Informasi

Demi pembangunan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Eksklusif (*Non-exclusive Royalty Free Right*) atas skripsi saya yang berjudul "**Analisis Sentimen Aplikasi Sayurbox Berdasarkan Ulasan Di Google Playstore Menggunakan Naive Bayes Classifier**". Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (basis data), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Dibuat di : Jakarta

Pada Tanggal : 20 Juli 2024

Yang Menyatakan,



Aqshal Win Mahara

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Nama : Aqshal Win Mahara
NIM : 2010512118
Studi : S1 Sistem Informasi
Judul Skripsi/TA : Analisis Sentimen Aplikasi Sayurbox Berdasarkan Ulasan Di Google Playstore Menggunakan Metode Naïve Bayes Classifier

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Ati Zaidiah, S.Kom, MTL.

Dosen Pembimbing I



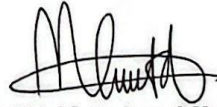
Muhammad Panji Muslim, S.Pd., M.Kom.

Dosen Pembimbing II



Theresa Wati, S.Kom., MTL.

Dosen Penguji I



Nur Hafifah Matondang, S.Kom, MM

Dosen Penguji II



Prof. Dr. Ir. Suprivanto, ST., M.Sc., IPM

Dekan FIK UPN Veteran Jakarta



Anita Muliawati, S.Kom., MTL.

Ketua Program Studi S1 Sistem Informasi

Ditetapkan di : Jakarta

Tanggal Ujian : 1 Juli 2024

DAFTAR ISI

| | |
|---------------------------------------------------------------------------------------------|-------------|
| KATA PENGANTAR..... | iii |
| LEMBAR PERSETUJUAN | iv |
| LEMBAR ORISINALITAS | v |
| LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS | vi |
| LEMBAR PENGESAHAN | vii |
| DAFTAR ISI..... | viii |
| DAFTAR GAMBAR..... | xii |
| DAFTAR TABEL | xiii |
| DAFTAR RUMUS | xiv |
| BAB 1 | 1 |
| PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah..... | 2 |
| 1.3 Tujuan Peneltian | 3 |
| 1.4 Manfaat Penelitian..... | 3 |
| 1.5 Batasan Penelitian..... | 3 |
| 1.6 Sistematika Penulisan | 4 |
| BAB I PENDAHULUAN | 4 |
| BAB II LANDASAN TEORI | 4 |
| BAB III METODOLOGI PENELITIAN | 4 |
| BAB IV METODOLOGI PENELITIAN | 4 |
| BAB V METODOLOGI PENELITIAN | 4 |
| BAB II | 5 |
| LANDASAN TEORI..... | 5 |
| 2.1 E - commerce | 5 |
| 2.2. Sayurbox..... | 5 |
| 2.3 Google Play Store..... | 5 |
| 2.4 Google Colab | 6 |
| 2.5 WordCloud | 6 |
| 2.6 Python | 6 |
| 2.7 Text mining..... | 6 |

| | |
|----------------------------------------------|-----------|
| 2.8 Analisis Sentimen | 7 |
| 2.9 Text Preprocessing | 7 |
| 2.9.1 Data Cleaning | 8 |
| 2.9.2 Case Folding | 8 |
| 2.9.3 Tokenizing | 8 |
| 2.9.4 Normalisasi..... | 8 |
| 2.9.5 Stopword Removal..... | 8 |
| 2.9.6 Stemming | 9 |
| 2.10 Pembobotan TF-IDF | 9 |
| A. Term Frequency (TF)..... | 9 |
| B. Inverse Document Frequency (IDF) | 9 |
| 2.11 Naïve Bayes Classifier..... | 9 |
| 2.12 Kelebihan Naïve Bayes | 10 |
| 2.13 Kekurangan Naïve Bayes..... | 11 |
| 2.14 Confusion Matrix..... | 11 |
| 2.15 Confusion Matrix..... | 12 |
| 2.14.1 Accuracy (Akurasi)..... | 12 |
| 2.14.2 Precision (presisi) | 12 |
| 2.14.3 Recall | 12 |
| 2.14.4 F-1 Score | 13 |
| 2.15 WorldCloud | 13 |
| 2.16 Figma..... | 13 |
| 2.17 Recommended untuk sayurbox | 13 |
| 2.18 Review penelitian terdahulu | 13 |
| BAB III..... | 17 |
| METEOROLOGI PENELITIAN | 17 |
| 3.1 Alur Penelitian..... | 17 |
| 3.2 Identifikasi masalah | 18 |
| 3.4 Studi Literatur | 18 |
| 3.4.1 Scraping Data (Pengumpulan Data) | 18 |
| 3.4.2 Pelabelan Data | 19 |
| 3.4.3 Praprocessing Data | 19 |
| a. Cleaning (Pembersihan Data)..... | 19 |
| b. Case Folding | 20 |

| | |
|--------------------------------------------------------|-----------|
| c. Tokenisasi | 20 |
| d. Normalisasi..... | 21 |
| c. Stopword Removal | 21 |
| c. Stemming | 22 |
| 3.4.6 Term Weighting (Pembobotan) | 22 |
| 3.4.7 Splitting data (belum dirapiin)..... | 22 |
| 3.5 Model Training..... | 22 |
| f. Evaluasi Performa Model..... | 22 |
| 3.6 Visualisasi Wordcloud | 23 |
| 3. Perancangan..... | 23 |
| 3.3 Spesifikasi Alat Pembuatan..... | 24 |
| 3.3.1 Perangkat Keras..... | 24 |
| 3.3.2 Perangkat Lunak..... | 24 |
| 3.4 Jadwal Kegiatan | 24 |
| BAB IV | 26 |
| Hasil dan Pembahasan..... | 26 |
| 4.1 Pengumpulan data ulasan (Scraping) | 26 |
| 4.2 Data Labeling | 26 |
| 4.3 Preprocessing..... | 27 |
| 4.3.1 Data Cleaning | 27 |
| 4.3.2 Case Folding | 28 |
| 4.3.3 Tokenizing | 28 |
| 4.3.4 Normalisasi..... | 29 |
| 4.3.5 Stopword Removal..... | 30 |
| 4.3.5 Stemming | 30 |
| 4.4 Pembobotan TF – IDF | 30 |
| 4.5 Splitting Data (pembagian data)..... | 32 |
| 4.6 Klasifikasi Naïve Bayes Classifier | 33 |
| 4.7 Evaluasi model Naïve Bayes Classifier..... | 33 |
| 4.8 Analisis hasil Klasifikasi | 35 |
| 4.8.1 Visualisasi WordCloud Sentimen Keseluruhan | 35 |
| 4.8.2 Visualisasi WordCloud Positif..... | 36 |
| 4.8.3 Visualisasi WordCloud Negatif..... | 38 |
| 4.9 Perancangan | 39 |

| | |
|----------------------------------------------------------|-----------|
| 4.10 Implementasi dan pengujian Prototype | 39 |
| 4.10.1 Tampilan Halaman Dashbaord..... | 40 |
| 4.10.2 Tampilan Informasi Data..... | 40 |
| 4.10.3 Tampilan Naïve Bayes Classifier..... | 41 |
| 4.10.4 Tampilan Halaman WordCloud..... | 42 |
| 4.10.5 Tampilan Halaman Sentimen Berdasarkan Aspek | 43 |
| BAB V | 46 |
| PENUTUP | 46 |
| 5.1 Kesimpulan | 46 |
| 5.2 Saran | 46 |
| Daftar Pustaka | 48 |

DAFTAR GAMBAR

| | |
|----------------------------------------------------|----|
| Gambar 3. 1 Alur Penelitian..... | 17 |
| Gambar 3. 2 Worcloud..... | 23 |
| Gambar 4. 1 Data labeling | 27 |
| Gambar 4. 2 Data Cleaning..... | 28 |
| Gambar 4. 3 Confusion Matrix | 34 |
| Gambar 4. 4 Wordcloud Keseluruhan..... | 35 |
| Gambar 4. 5 Diagram wordcloud keseluruhan | 36 |
| Gambar 4. 6 Wordcloud Positif | 36 |
| Gambar 4. 7 Grafik wordcloud positif | 37 |
| Gambar 4. 8 Wordcloud Negatif..... | 38 |
| Gambar 4. 9 Grafik Negatif | 38 |
| Gambar 4. 10 Halaman Dashboard | 40 |
| Gambar 4. 11 Tampilan Informasi Data | 41 |
| Gambar 4. 12 Tampilan Naïve Bayes Classifier..... | 41 |
| Gambar 4. 13 Tampilan Wordcloud..... | 42 |
| Gambar 4. 14 Tampilan Wordcloud Positif..... | 42 |
| Gambar 4. 15 Wordcloud Negatif..... | 43 |
| Gambar 4. 16 Tampilan Sentimen Berdasar Aspek..... | 43 |
| Gambar 4. 17 sentimen berdasarkan aspek | 44 |
| Gambar 4. 18 Upload File..... | 44 |
| Gambar 4. 19 Tampilan setelah upload file | 45 |

DAFTAR TABEL

| | |
|----------------------------------------------------------------------|----|
| Tabel 2. 1 Confusion Matrix | 12 |
| Tabel 2. 2 Penelitian Terlebih Dahulu | 14 |
| Tabel 3. 1 Content Positif Negatif..... | 19 |
| Tabel 3. 2 Sebelum dan Sesudah Cleaning | 20 |
| Tabel 3. 3 Sebelum dan Sesudah Case Folding | 20 |
| Tabel 3. 4 Sebelum dan sesudah Tokenisasi..... | 21 |
| Tabel 3. 5 Sebelum dan Sesudah Normalisasi | 21 |
| Tabel 3. 6 Sebelum dan Sesudah Stopword | 21 |
| Tabel 3. 7 Sebelum dan Sesudah Stemming | 22 |
| Tabel 3. 8 Actual | 23 |
| Tabel 3. 9 Jadwal Kegiatan | 24 |
| Tabel 4. 1 Case Folding | 28 |
| Tabel 4. 2 Tokenizing | 29 |
| Tabel 4. 3 Normalisasi | 29 |
| Tabel 4. 4 Stopword Removal..... | 30 |
| Tabel 4. 5 Stemming | 30 |
| Tabel 4. 6 Data Sampel | 31 |
| Tabel 4. 7 Hasil Perhitungan Pembobotan TF-IDF pada data sampel..... | 32 |
| Tabel 4. 8 Pembagian Data dengan rasio 80:20..... | 33 |

DAFTAR RUMUS

| | |
|----------------------------------------------------|----|
| <u>Term Frequency (TF)</u> (2.1) | 9 |
| <u>Inverse Document Frequency (IDF)</u> (2.2)..... | 9 |
| <u>Naïve Bayes Classifier</u> (2.3) | 10 |
| <u>Accuracy</u> (2.4)..... | 12 |
| <u>Precision</u> (2.5) | 12 |
| <u>Recall</u> (2.6)..... | 13 |
| <u>F-1 Score</u> (2.7) | 13 |