

**PERBANDINGAN SENTIMEN PADA ULASAN APLIKASI  
PERJALANAN ONLINE DENGAN MENGGUNAKAN  
ALGORITMA K-NEAREST NEIGHBORS DAN NAÏVE BAYES**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi SI Informatika



disusun oleh

**AGENG NANDANG PAMBUDI**

**20.11.3661**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2025**

**PERBANDINGAN SENTIMEN PADA ULASAN APLIKASI  
PERJALANAN ONLINE DENGAN MENGGUNAKAN  
ALGORITMA K-NEAREST NEIGHBORS DAN NAÏVE BAYES**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi S1 Informatika



disusun oleh  
**AGENG NANDANG PAMBUDI**  
**20.11.3661**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2025**

## HALAMAN PERSETUJUAN



## HALAMAN PENGESAHAN

### SKRIPSI

#### PERBANDINGAN SENTIMEN PADA ULASAN APLIKASI PERJALANAN ONLINE DENGAN MENGGUNAKAN ALGORITMA K- NEAREST NEIGHBORS DAN NAÏVE BAYES

yang disusun dan diajukan oleh

**Ageng Nandang Pambudi**

20.11.3661

Telah dipertahankan di depan Dewan Pengaji  
pada tanggal 14 Maret 2023

Sesuai Dewan Pengaji

Nama Pengaji

Tanda Tangan

Supriatni, A.Md., S.Pd., M.Kom.  
NIK. 1903022239

Moch Farid Fauzi, S.Kom., M.Kom.  
NIK. 1903022284

Kusnawi, S.Kom., M.Eng.  
NIK. 1903022112

Skrripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 14 Maret 2023

DEKAN FAKULTAS ILMU KOMPUTER



Haniff Al Fatta, S.Kom., M.Kom.  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ageng Nandang Pambudi  
NIM : 20.11.5661

Menyatakan bahwa Skripsi dengan judul berikut:

**Perbandingan Sentimen Pada Ulasan Aplikasi Perjalanan Online Dengan Menggunakan Algoritma K-Nearest Neighbors Dan Naïve Bayes**

Dosen Pembimbing : Kusnawi, S.kom., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencahutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 14 Maret 2025

Yang Menyatakan,



Ageng Nandang Pambudi

## HALAMAN PERSEMBAHAN

Alhamdulillah, Segala Puji Bagi ALLAH SWT. Kepada-Nya kita memohon, memuji dan meminta perlindungan, serta petunjuk Kepada-Nya. Kita berlindung kepada ALLAH SWT dari kejahatan diri kita dan keburukan amal-amal kita. Dengan ini penulis mempersembahkan skripsi ini kepada semua pihak yang telah membantu dan memberikan motivasi serta dukungan yang besar sehingga skripsi ini dapat diselesaikan dengan sebaik baiknya.

Melalui skripsi ini saya persembahkan terimakasih serta syukur kepada:

1. Rasa syukur kepada Allah SWT yang telah memberikan rahmat serta karunia untuk terus mengucapkan syukur dan sabar dalam menyelesaikan tugas akhir skripsi.
2. Kepada orang tua penulis Bapak Wahyu Murtono, Ibu Sutinem yang selalu memberikan doa, cinta dan dukungan yang tidak pernah berhenti diberikan kepada penulis.
3. Bapak Kusnawi, S.Kom., M.Eng. selaku dosen pembimbing yang selalu memberikan saran dan masukan kepada penulis dalam menyelesaikan skripsi ini.
4. Bapak/Ibu dosen yang telah memberikan ilmu dan pendidikan selama berkuliah di Amikom.
5. Terima kasih kepada Canfidelia yang selalu mensupport serta memberikan doa, cinta dan kasih sayang penulis sampai dengan tugas akhir skripsi ini. Walaupun dalam proses penulisan skripsi ini banyak hal yang terlewat dan harus dikorbankan. Semoga kedepannya banyak hal baik dan menyenangkan yang harus bisa dilewati bersama.
6. Staf UPT ME dan teman teman SS UPT ME yang selalu memberikan support dalam mencapai kesuksesan.
7. Teman teman Kontrakan yang menjadi pemberi dukungan dalam mencapai kesuksesan yang penuh suka dan duka. Terima kasih banyak telah

membantu banyak hal yang tidak diketahui sampai dengan sekarang penulis bisa menyelesaikan tugas akhir skripsi ini.

8. Terimakasih kepada sahabat saya ulin yang selalu membantu perbaiki eror pada kodingan daya dan selalu menemani ngopi saya.
9. Kepada teman teman "SUBEDU" yang hampir setiap saat memberikan segala yang mereka bisa kepada penulis sehingga dapat menyelesaikan tugas akhir skripsi.
10. Yang terakhir terima kasih kepada anak laki-laki ganteng yang selalu memiliki pemikiran akan hal hal kecil yang sulit dimengerti, sang penulis sebuah karya tulis ini, diri saya sendiri, Ageng Nandang Pambudi. Seorang laki-laki yang berusia 23 tahun saat menyelesaikan tugas akhir skripsi. Terima kasih kepada diri saya sendiri yang tetap mengupayakan semua hal dan terus berusaha memperbaiki segala hal terkait tugas akhir skripsi ini. "Jangan malu karena kau bodoh, namun malulah kalau kau tak berubah". "Jadikan masa lalu atau hal yang terlewat sebagai bahan evaluasi dan motivasi diri".

## KATA PENGANTAR

Puji syukur kehadiran ALLAH SWT karena berkat rahmat serta hidayahnya penulis diberikan kesehatan dan kekuatan serta iman yang kuat sehingga dapat menyelesaikan skripsi yang berjudul "PERBANDINGAN SENTIMEN PADA ULASAN APLIKASI PERJALANAN ONLINE DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS DAN NAÏVE BAYES". Oleh karena itu penulis mengucapkan syukur karena dapat menyelesaikan skripsi ini. Skripsi ini diajukan kepada Program Studi S1 Informatika, Fakultas Ilmu Komputer, serta Universitas AMIKOM Yogyakarta.

Penulis dalam menyelesaikan skripsi mendapatkan beberapa hambatan dalam berbagai hal, namun banyak pihak yang selalu membantu sehingga skripsi ini dapat diselesaikan dengan baik. Dengan begitu penulis hendak mengungkapkan terima kasih terhadap:

1. Bapak prof. Dr. M Suyanto, M.M selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradya D, M.Kom., selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
4. Bapak Kusnawi, S.kom., M.Eng. selaku dosen pembimbing yang selalu memberikan saran dan arahan dalam proses penulisan skripsi ini.
5. Bapak Murtono dan Ibu Sutinem selaku orang tua yang selalu memberikan doa dan motivasi kepada penulis.

Yogyakarta, 14 Maret 2025

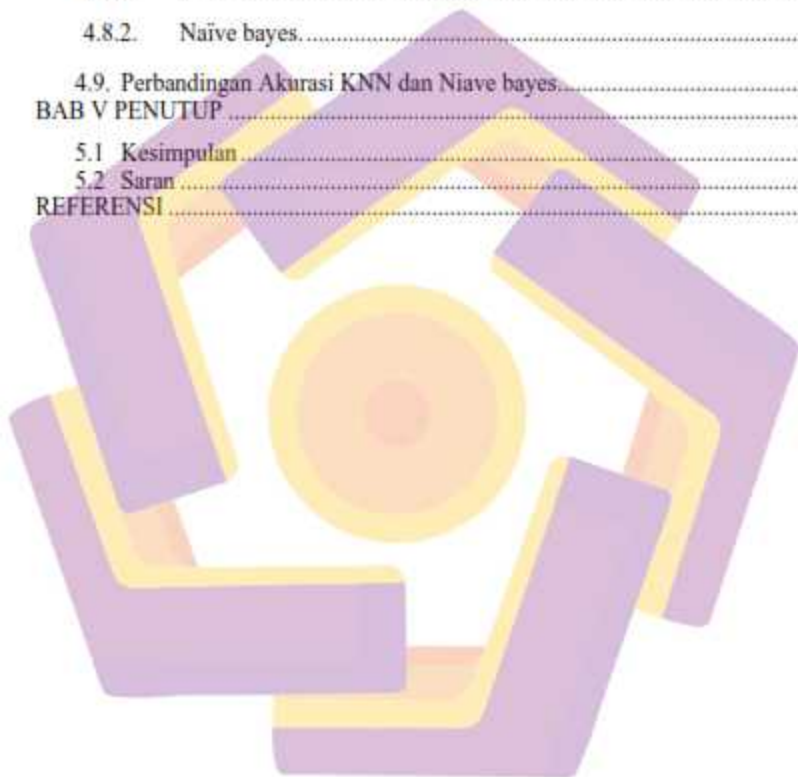
Ageng Nandang Pambudi

## DAFTAR ISI

|                                           |      |
|-------------------------------------------|------|
| HALAMAN JUDUL .....                       | i    |
| HALAMAN PERSETUJUAN.....                  | ii   |
| HALAMAN PENGESAHAN .....                  | ii   |
| HALAMAN PERNYATAAN KEASLIAN SKRIPSI ..... | iii  |
| HALAMAN PERSEMBAHAN .....                 | iv   |
| KATA PENGANTAR .....                      | vii  |
| DAFTAR ISI.....                           | viii |
| DAFTAR TABEL.....                         | xi   |
| DAFTAR GAMBAR.....                        | xiii |
| INTISARI .....                            | xiv  |
| ABSTRACT.....                             | xv   |
| BAB I PENDAHULUAN.....                    | 1    |
| 1.1 Latar Belakang .....                  | 1    |
| 1.2 Rumusan Masalah.....                  | 2    |
| 1.3 Batasan Masalah .....                 | 3    |
| 1.4 Tujuan Penelitian .....               | 3    |
| 1.5 Manfaat Penelitian .....              | 3    |
| 1.6 Sistematika Penulisan .....           | 4    |
| BAB II TINJAUAN PUSTAKA .....             | 6    |
| 2.1 Studi Literatur.....                  | 6    |
| 2.2 Dasar Teori .....                     | 17   |
| 2.2.1 Analisis Sentimen .....             | 17   |
| 2.2.2 Traveloka .....                     | 17   |
| 2.2.3 Klasifikasi .....                   | 18   |
| 2.2.4 Python .....                        | 18   |
| 2.2.5 Google Colaboraty .....             | 18   |
| 2.2.6 Web Scrapping .....                 | 19   |
| 2.2.7 Labeling .....                      | 19   |
| 2.2.8 Pre-Processing .....                | 19   |

|                                   |                                                            |    |
|-----------------------------------|------------------------------------------------------------|----|
| 2.2.9                             | Pembobotan TF-IDF .....                                    | 20 |
| 2.2.10                            | <i>K-NEAREST NEIGHBOR</i> .....                            | 22 |
| 2.2.11                            | <i>Naïve Bayes</i> .....                                   | 23 |
| 2.2.11                            | Confusion Matrix .....                                     | 24 |
| BAB III METODE PENELITIAN .....   |                                                            | 26 |
| 3.1                               | Objek Penelitian .....                                     | 26 |
| 3.2                               | Alur Penelitian .....                                      | 26 |
| 3.2.1                             | Pengumpulan Data .....                                     | 28 |
| 3.2.2                             | Labelling Data .....                                       | 28 |
| 3.2.3                             | Preprocessing Data .....                                   | 31 |
| 3.2.4                             | Visualization .....                                        | 32 |
| 3.2.5                             | Splitting Data .....                                       | 32 |
| 3.2.6                             | TF-IDF .....                                               | 33 |
| 3.2.7                             | Pelatihan dengan K-Nearest Neighbors Dan Naïve Bayes ..... | 33 |
| 3.2.8                             | Evaluasi .....                                             | 34 |
| 3.3                               | Alat dan Bahan .....                                       | 34 |
| BAB IV HASIL DAN PEMBAHASAN ..... |                                                            | 36 |
| 4.1.                              | Pengambilan Dataset .....                                  | 36 |
| 4.2.                              | Pelabelan Data .....                                       | 38 |
| 4.2.1                             | Pelabelan Data Rating .....                                | 38 |
| 4.2.2                             | Pelabelan Data Teks .....                                  | 40 |
| 4.3.                              | Pre-Processing .....                                       | 46 |
| 4.3.1                             | Case Folding .....                                         | 46 |
| 4.3.2                             | Data Cleaning .....                                        | 48 |
| 4.3.3                             | Stopword Removal .....                                     | 49 |
| 4.3.4                             | Stemming .....                                             | 51 |
| 4.3.5                             | Tokenizing .....                                           | 53 |
| 4.4.                              | Visualitation .....                                        | 54 |
| 4.5.                              | Spliting Data .....                                        | 56 |

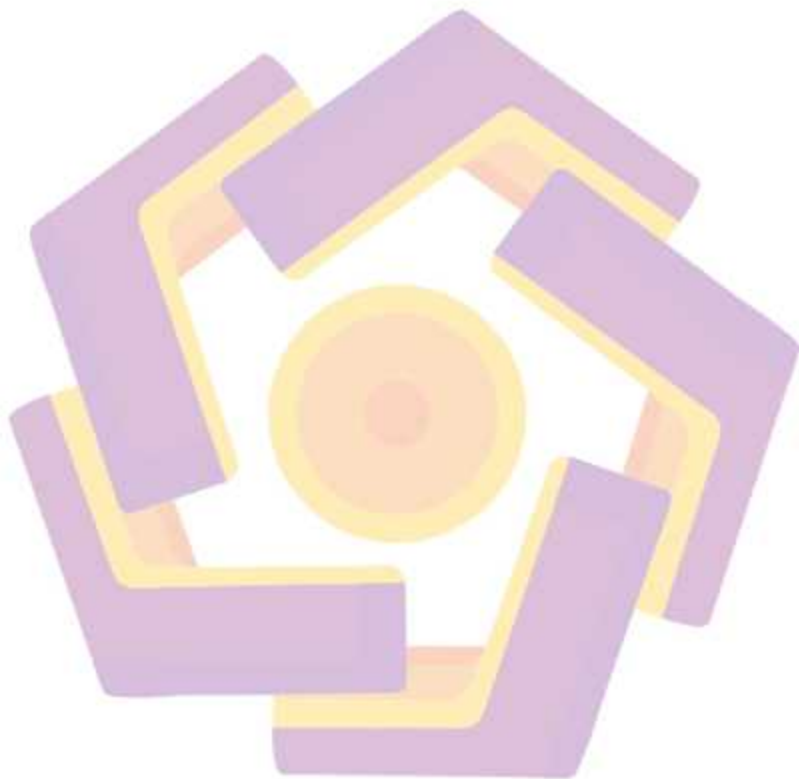
|                                                    |    |
|----------------------------------------------------|----|
| 4.6. Pembobotan TF-IDF .....                       | 57 |
| 4.7. Klasifikasi dengan KNN dan Naïve Bayes.....   | 60 |
| 4.7.1. Klasifikasi KNN.....                        | 60 |
| 4.7.2. Klasifikasi Naïve Bayes .....               | 62 |
| 4.8. Evaluasi (Confusion matrix).....              | 64 |
| 4.8.1. KNN.....                                    | 64 |
| 4.8.2. Naïve bayes.....                            | 65 |
| 4.9. Perbandingan Akurasi KNN dan Naïve bayes..... | 67 |
| BAB V PENUTUP .....                                | 70 |
| 5.1 Kesimpulan.....                                | 70 |
| 5.2 Saran .....                                    | 71 |
| REFERENSI .....                                    | 72 |



## DAFTAR TABEL

|                                                                                                           |    |
|-----------------------------------------------------------------------------------------------------------|----|
| Tabel 4. 1 script Scrapping dan filtering .....                                                           | 36 |
| Tabel 4. 2 dataset .....                                                                                  | 37 |
| Tabel 4. 3 <i>script</i> pelabelan <i>score</i> atau <i>rating</i> .....                                  | 38 |
| Tabel 4. 4 Hasil pelabelan text menggunakan <i>rating</i> .....                                           | 39 |
| Tabel 4. 5 Script pelabelan text menggunakan <i>Lexicon</i> .....                                         | 40 |
| Tabel 4. 6 Proses translate ulasan dari Bahasa Indonesia menjadi Bahasa Inggris .....                     | 40 |
| Tabel 4. 7 Hasil Translate .....                                                                          | 41 |
| Tabel 4. 8 menampilkan hasil <i>labeling</i> menggunakan <i>lexicon</i> .....                             | 42 |
| Tabel 4. 9 Hasil <i>labeling</i> Data teks ( <i>content</i> ) dan <i>rating</i> ( <i>score</i> ) .....    | 43 |
| Tabel 4. 10 Script Menampilkan Jumlah Data Tiap Kelas dari Hasil Pelabelan .....                          | 45 |
| Tabel 4. 11 Script case foolding .....                                                                    | 46 |
| Tabel 4. 12 Hasil <i>Casefolding</i> .....                                                                | 47 |
| Tabel 4. 13 Script <i>cleaning</i> .....                                                                  | 48 |
| Tabel 4. 14 Hasil <i>Cleaning</i> .....                                                                   | 48 |
| Tabel 4. 15 Script Pada Tahap <i>Stopword Removal</i> .....                                               | 49 |
| Tabel 4. 16 Hasil <i>stopword removal</i> .....                                                           | 50 |
| Tabel 4. 17 script proses <i>stemming</i> .....                                                           | 51 |
| Tabel 4. 18 Hasil <i>stemming</i> .....                                                                   | 52 |
| Tabel 4. 19 script proses <i>tokenizing</i> .....                                                         | 53 |
| Tabel 4. 20 Hasil <i>Tokenizing</i> .....                                                                 | 53 |
| Tabel 4. 21 Script Pada Tahap <i>Visualization</i> Data Ulasan Pelabelan Menggunakan <i>Lexicon</i> ..... | 55 |
| Tabel 4. 22 Script Pada Tahap <i>Visualization</i> Data Ulasan Pelabelan Menggunakan <i>Rating</i> .....  | 55 |
| Tabel 4. 23 Script Pada Tahap <i>splitting data</i> .....                                                 | 56 |
| Tabel 4. 24 Script Proses <i>TF-IDF</i> .....                                                             | 57 |
| Tabel 4. 25 Hasil Pemisahan <i>Term</i> .....                                                             | 58 |
| Tabel 4. 26 Hasil <i>TF</i> Normalisasi .....                                                             | 58 |
| Tabel 4. 27 Hasil penjumlahan <i>term</i> .....                                                           | 59 |
| Tabel 4. 28 Hasil dari perhitungan nilai <i>IDF</i> .....                                                 | 59 |
| Tabel 4. 29 Hasil Perhitungan <i>TF-IDF</i> .....                                                         | 59 |
| Tabel 4. 30 Script Klasifikasi Menggunakan Metode <i>KNN</i> .....                                        | 60 |
| Tabel 4. 31 Script Hasil Akurasi Metode <i>KNN</i> .....                                                  | 61 |
| Tabel 4. 32 Script Klasifikasi Menggunakan Metode <i>Naïve Bayes</i> .....                                | 62 |
| Tabel 4. 33 Script Hasil Akurasi Metode <i>Naïve Bayes</i> .....                                          | 62 |
| Tabel 4. 34 Perbandingan Akurasi Algoritma Klasifikasi pada Pembagian Tertentu .....                      | 63 |
| Tabel 4. 35 Script <i>Confusion Matrix knn</i> .....                                                      | 64 |
| Tabel 4. 36 <i>Confusion Matrix knn</i> Data Labeling Menggunakan <i>Lexicon</i> .....                    | 64 |
| Tabel 4. 37 <i>Confusion Matrix knn</i> Data Labeling Menggunakan <i>Rating</i> .....                     | 65 |
| Tabel 4. 38 Script <i>Confusion Matrix Naïve bayes</i> .....                                              | 66 |

|                                                                                         |    |
|-----------------------------------------------------------------------------------------|----|
| Tabel 4. 39 <i>Confusion Matrix Naïve bayes Data Labeling Menggunakan Lexicon</i> ..... | 66 |
| Tabel 4. 40 <i>Confusion Matrix Naïve bayes Data Labeling Menggunakan Rating</i> .....  | 66 |
| Tabel 4. 41 <i>Script Perbandingan Akurasi Algoritma KNN dan Naïve Bayes</i> .....      | 67 |



## DAFTAR GAMBAR

|                                                                                                                                     |    |
|-------------------------------------------------------------------------------------------------------------------------------------|----|
| Gambar 4. 1 Jumlah Data Tiap Kelas dari Hasil Pelabelan .....                                                                       | 46 |
| Gambar 4. 2 Menampilkan hasil proses cleaning .....                                                                                 | 47 |
| Gambar 4. 3 Menampilkan hasil proses casefolding .....                                                                              | 48 |
| Gambar 4. 4 Menampilkan hasil proses stopword removal .....                                                                         | 50 |
| Gambar 4. 5 Menampilkan Hasil Proses <i>Stemming</i> .....                                                                          | 52 |
| Gambar 4. 6 Menampilkan hasil Tahap <i>Tokenizing</i> .....                                                                         | 53 |
| Gambar 4. 7 Hasil <i>Visualization</i> Data Ulasan <i>Labeling</i> Menggunakan <i>Lexicon</i> .....                                 | 55 |
| Gambar 4. 8 Hasil <i>Visualization</i> Data Ulasan <i>Labeling</i> Menggunakan <i>Rating</i> .....                                  | 56 |
| Gambar 4. 9 Hasil Akurasi Metode <i>KNN</i> .....                                                                                   | 61 |
| Gambar 4. 10 Hasil Akurasi Metode <i>Naïve bayes</i> .....                                                                          | 63 |
| Gambar 4. 11 Hasil <i>Convusion Matrix</i> Metode <i>KNN</i> .....                                                                  | 65 |
| Gambar 4. 12 Hasil <i>Convusion Matrix</i> Metode <i>Naïve bayes</i> .....                                                          | 67 |
| Gambar 4. 13 Hasil Perbandingan Akurasi Algoritma <i>KNN</i> dan <i>Naïve Bayes</i> pada Pelabelan Menggunakan <i>Lexicon</i> ..... | 68 |
| Gambar 4. 14 Hasil Perbandingan Akurasi Algoritma <i>KNN</i> dan <i>Naïve Bayes</i> pada Pelabelan Berdasarkan <i>Rating</i> .....  | 69 |



## INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi Traveloka di Google Play Store dengan menggunakan dua algoritma klasifikasi, yaitu K-Nearest Neighbors (KNN) dan Naïve Bayes. Data yang digunakan dalam penelitian ini terdiri dari 3000 ulasan yang diambil melalui metode web scraping. Ulasan tersebut kemudian dilabeli menjadi tiga kategori: positif, negatif, dan netral. Proses analisis melibatkan tahapan preprocessing, termasuk cleaning data, case folding, stopword removal, stemming, dan tokenizing. Selanjutnya, data dibagi menjadi data pelatihan dan pengujian dengan rasio 70:30 untuk evaluasi model.

Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes memberikan akurasi yang lebih tinggi dibandingkan dengan KNN, dengan nilai akurasi masing-masing sebesar 84,44% untuk data label berdasarkan rating dan 74,00% untuk data label menggunakan lexicon. Sementara itu, KNN menghasilkan akurasi 82,67% untuk data label berdasarkan rating dan 71,33% untuk data label menggunakan lexicon. Penelitian ini memberikan wawasan yang berguna bagi pengembang aplikasi Traveloka dalam memahami sentimen pengguna dan meningkatkan kualitas layanan berdasarkan umpan balik yang diterima. Diharapkan penelitian selanjutnya dapat memperbaiki klasifikasi sentimen netral dan menerapkan metode machine learning yang lebih beragam untuk analisis yang lebih mendalam.

**Kata kunci:** Analisis Sentimen, Traveloka, KNN, Naïve Bayes, Ulasan Pengguna.

## **ABSTRACT**

*This research aims to analyze the sentiment of user reviews of the Traveloka application on the Google Play Store using two classification algorithms, namely K-Nearest Neighbors (KNN) and Naïve Bayes. The data used in this research consists of 3000 reviews taken through the web scraping method. The reviews were then labeled into three categories: positive, negative, and neutral. The analysis process involved preprocessing steps, including data cleaning, case folding, stopword removal, stemming, and tokenizing. Next, the data was divided into training and testing data with a ratio of 70:30 for model evaluation.*

*The results show that the Naïve Bayes algorithm provides higher accuracy compared to KNN, with accuracy values of 84.44% for rating-based label data and 74.00% for lexicon-based label data, respectively. Meanwhile, KNN produces 82.67% accuracy for label data based on rating and 71.33% for label data using lexicon. This research provides useful insights for Traveloka app developers in understanding user sentiment and improving service quality based on the feedback received. It is hoped that future research can improve the classification of neutral sentiment and apply more diverse machine learning methods for more in-depth analysis.*

**Keyword:** Sentiment Analysis, Traveloka, KNN, Naïve Bayes, User Reviews.