



**ANALISIS SENTIMEN APLIKASI SOLOLEARN BERDASARKAN *FEEDBACK*
PENGGUNA DI GOOGLE PLAY STORE UNTUK MENGEVALUASI
WAWASAN PENGGUNA APLIKASI MENGGUNAKAN
ALGORITMA NAÏVE BAYES**

SKRIPSI

**HANSEN KALLISTA
NIM. 2010512084**

**SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
2024**



**ANALISIS SENTIMEN APLIKASI SOLOLEARN BERDASARKAN *FEEDBACK*
PENGGUNA DI GOOGLE PLAY STORE UNTUK MENGEVALUASI
WAWASAN PENGGUNA APLIKASI MENGGUNAKAN
ALGORITMA NAÏVE BAYES**

SKRIPSI

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer

**HANSEN KALLISTA
NIM. 2010512084**

**SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
2024**

PERNYATAAN ORISINALITAS

Skripsi ini merupakan hasil karya sendiri serta semua sumber referensi yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Hansen Kallista
NIM : 2010512084
Tanggal : 25 Juli 2024

Apabila pada kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan berlaku.

Jakarta, 25 Juli 2024
Yang Menyatakan,

A handwritten signature in black ink is written over a 10,000 Indonesian postage stamp. The stamp is pink and white, featuring the Garuda Pancasila emblem and the text 'REPUBLIK INDONESIA', '10000', and 'METERAI TEMPEL'. The serial number '18BC1ALX275398904' is visible at the bottom of the stamp.

Hansen Kallista

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI
UNTUK KEPENTINGAN AKADEMIS**

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

Nama : Hansen Kallista
NIM : 2010512084
Fakultas : Ilmu Komputer
Program Studi : S1 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 karya ilmiah saya yang berjudul:

**ANALISIS SENTIMEN APLIKASI SOLOLEARN BERDASARKAN FEEDBACK
PENGGUNA DI GOOGLE PLAYSTORE UNTUK MENGEVALUASI WAWASAN
PENGGUNA APLIKASI MENGGUNAKAN ALGORITMA NAÏVE BAYES**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini, Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/peneliti dan pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Dibuat di: Jakarta
Pada Tanggal: 25 Juli 2024

Yang Menyatakan,



Hansen Kallista

LEMBAR PENGESAHAN

Tugas akhir ini diajukan oleh:

Nama : Hansen Kallista
NIM : 2010512084
Program Studi : S1 Sistem Informasi
Judul : ANALISIS SENTIMEN APLIKASI SOLOLEARN BERDASARKAN
FEEDBACK PENGGUNA DI GOOGLE PLAYSTORE UNTUK
MENGEVALUASI WAWASAN PENGGUNA APLIKASI
MENGUNAKAN ALGORITMA NAÏVE BAYES

Telah berhasil dipertahankan dihadapan 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.



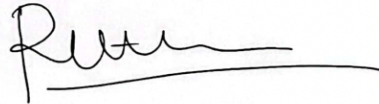
Ati Zaidiah, S.Kom., MTI.
Penguji 1



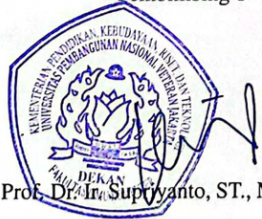
Iin Ernawati, S.Kom., M.Si.
Penguji 2



Rio Wirawan, S.Kom., M.M.S.I.
Pembimbing 1



Ruth Mariana Bunga Wadu, S.Kom., M.M.S.I.
Pembimbing 2



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM
Dekan



Anita Muliawati, S.Kom. M.TI
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 12 Juli 2024

**ANALISIS SENTIMEN APLIKASI SOLOLEARN BERDASARKAN
FEEDBACK PENGGUNA DI GOOGLE PLAY STORE UNTUK
MENGEVALUASI
WAWASAN PENGGUNA APLIKASI MENGGUNAKAN
ALGORITMA NAÏVE BAYES**

Hansen Kallista

ABSTRAK

Sentiment analysis telah berkembang menjadi salah satu bidang penelitian paling aktif dalam *Natural Language Processing*, selain itu aktivitas industri lainnya seputar analisis sentimen terhadap aplikasi juga berkembang secara pesat. Sololearn adalah sebuah aplikasi edukasi yang dikembangkan oleh perusahaan di Amerika yang memungkinkan pengguna untuk belajar bahasa pemrograman seperti HTML, CSS, JavaScript, dll, yang dilengkapi oleh fitur seperti compiler kode. Oleh karena itu, penulis ingin melakukan penelitian terhadap aplikasi Sololearn untuk melihat sentimen pengguna aplikasi tersebut serta mempelajari keunggulan serta kelemahan aplikasi. Dengan total 728 data ulasan yang diperoleh melalui *data scraping* dari Google Play Store, yang kemudian diberikan label data menggunakan sistem rating ulasan secara manual dimana jika ulasan tersebut memiliki rating 4 dan 5 maka data tersebut diberi label positif, sedangkan jika ulasan tersebut memiliki rating 3 atau lebih rendah maka data tersebut akan diberi label negatif. Data melalui tahap *pre-processing* untuk memastikan data siap untuk diproses, lalu menggunakan metode pembobotan kata TF-IDF. Setelah seluruh tahapan *pre-process* selesai, data akan digunakan untuk implementasi model dengan menggunakan algoritma *naïve bayes* dengan pembagian data 80% data training dan 20% data testing. Hasil yang diperoleh dari pembuatan model ini yaitu skor akurasi yang didapat sebesar 80%, skor presisi 64%, recall 80%, dan f1-score 71%. Hasil dari analisis sentimen data ini ialah pengguna mengharapkan adanya tambahan fitur bahasa indonesia untuk memudahkan proses pembelajaran, serta memperbaiki bug yang masih terdapat di sebagian course yang disajikan.

Kata Kunci: Analisis Sentimen, Sololearn, Naïve Bayes, Modeling Data

**SENTIMENT ANALYSIS OF THE SOLOLEARN APPLICATION BASED ON
USER FEEDBACK ON GOOGLE PLAY STORE TO EVALUATE USER
INSIGHTS
USING THE NAÏVE BAYES ALGORITHM**

Hansen Kallista

ABSTRACT

Sentiment analysis has evolved into one of the most active research fields in Natural Language Processing, and industrial activities around sentiment analysis of applications have also rapidly developed. Sololearn is an educational application developed by an American company that allows its users to learn programming languages such as HTML, CSS, JavaScript, etc., equipped with features like a code compiler. Therefore, the author wants to conduct research on the Sololearn application to observe user sentiment toward the application and to study its strengths and weaknesses. With a total of 728 review data obtained through data scraping from the Google Play Store, the data were then manually labeled using the review rating system where if the review had a rating of 4 and 5, the data was labeled positive, whereas if the review had a rating of 3 or lower, the data was labeled negative. The data went through a pre-processing stage to ensure the data was ready to be processed, then using the TF-IDF word weighting method. After all pre-process stages were completed, the data was used to implement the model using the naïve bayes algorithm with an 80% training data and 20% testing data split. The results obtained from this model creation were an accuracy score of 80%, a precision score of 64%, a recall of 80%, and an f1-score of 71%. The results of this sentiment analysis indicate that users expect the addition of an Indonesian language feature to facilitate the learning process, as well as the fixing of bugs still present in some of the courses offered.

Keywords: *Sentiment Analysis, Sololearn, Naïve Bayes, Data Modelling*

KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT. karena berkat rahmat serta karunia-nya yang telah memberikan kesehatan kepada peneliti sehingga dapat menyelesaikan skripsi atau tugas akhir yang berjudul “Analisis Sentimen Aplikasi Sololearn Berdasarkan Feedback Pengguna di Google Play Store Untuk Mengevaluasi Wawasan Pengguna Aplikasi Menggunakan Algoritma Naïve Bayes”.

Dalam pelaksanaan proses penulisan proposal, peneliti menerima dukungan, semangat, dan bimbingan dari berbagai pihak. Dengan ini, peneliti ingin menyampaikan rasa terima kasih kepada:

1. Ibu Anita Muliawati, S.Kom., MTI. selaku Ketua Program Studi S1 Sistem Informasi.
2. Bapak Rio Wirawan, S.Kom., M.M.S.I. selaku Dosen Pembimbing pertama saya atas bimbingan dan saran selama proses penulisan penelitian.
3. Ibu Ruth Mariana Bunga Wadu, S.Kom., M.M.S.I. selaku Dosen Pembimbing kedua saya atas bimbingan dan saran selama proses penulisan penelitian.
4. Orang tua, saudara, dan keluarga atas doa, motivasi, dan arahan yang telah diberikan selama ini, dan.
5. Teman-teman satu jurusan S1 Sistem Informasi atas semangat dan kerja samanya.

Peneliti menyadari bahwa isi dari penulisan proposal penelitian ini tidak luput dari kekurangan. Dengan ini, peneliti mengharapkan saran dan kritikan sebagai masukan agar penelitian ini dapat berkembang demi kepentingan penelitian yang akan datang.

Jakarta, 20 Juni 2024

Peneliti

DAFTAR ISI

| | |
|--|------------|
| PERNYATAAN ORISINALITAS | i |
| PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS | ii |
| LEMBAR PENGESAHAN | iii |
| ABSTRAK | iv |
| ABSTRACT | v |
| KATA PENGANTAR | vi |
| DAFTAR ISI | vii |
| DAFTAR GAMBAR | x |
| DAFTAR TABEL | xi |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah | 2 |
| 1.3 Tujuan Penelitian..... | 2 |
| 1.4 Manfaat Penelitian..... | 3 |
| 1.5 Batasan Masalah..... | 3 |
| 1.6 Luaran Penelitian..... | 3 |
| 1.7 Sistematika Penulisan..... | 4 |
| BAB II TINJAUAN PUSTAKA | 5 |
| 2.1 Landasan Teori | 5 |
| 2.1.1 Analisis Sentimen | 5 |
| 2.1.2 Sololearn..... | 5 |
| 2.1.3 <i>Text Mining</i> | 6 |
| 2.1.4 <i>Web Scraping</i> | 6 |
| 2.1.5 <i>Data Pre-processing</i> | 6 |
| 2.1.6 <i>Data Labelling</i> | 7 |
| 2.1.7 <i>Feature Extraction</i> | 7 |
| 2.1.8 <i>Wordcloud</i> | 8 |
| 2.1.9 <i>Naïve Bayes Classifier</i> | 8 |
| 2.1.10 <i>Confusion Matrix</i> | 8 |

| | | |
|----------------|---|-----------|
| 2.1.11 | Evaluasi Performa Model | 9 |
| 2.1.12 | <i>K-fold Cross Validation</i> | 10 |
| 2.2 | Kajian Literatur | 10 |
| BAB III | METODOLOGI PENELITIAN..... | 14 |
| 3.1 | Tempat dan Waktu Penelitian | 14 |
| 3.1.1 | Tempat Penelitian | 14 |
| 3.1.2 | Waktu Penelitian | 14 |
| 3.2 | Diagram Alur Penelitian | 14 |
| 3.2.1 | Perumusan Masalah | 16 |
| 3.2.2 | Studi Literatur | 16 |
| 3.2.3 | Pengumpulan Data..... | 16 |
| 3.2.4 | <i>Data Labeling</i> | 16 |
| 3.2.5 | <i>Data Pre-processing</i> | 16 |
| 3.2.6 | <i>Feature Extraction</i> | 17 |
| 3.2.7 | <i>Split Data</i> | 17 |
| 3.2.8 | <i>Modeling</i> | 17 |
| 3.2.9 | <i>Model Performance Evaluation</i> | 17 |
| 3.2.10 | Analisa Hasil..... | 17 |
| 3.3 | Alat Pendukung Penelitian | 18 |
| 3.3.1 | Perangkat Keras | 18 |
| 3.3.2 | Perangkat Lunak | 18 |
| 3.4 | Jadwal Penelitian | 18 |
| BAB IV | HASIL DAN PEMBAHASAN..... | 20 |
| 4.1 | Hasil Penelitian..... | 20 |
| 4.1.1 | Prosedur Pengumpulan Data..... | 20 |
| 4.1.2 | Cleaning Data | 21 |
| 4.1.3 | <i>Normalize</i> | 22 |
| 4.1.4 | <i>Stopword</i> | 23 |
| 4.1.5 | <i>Tokenizing</i> | 25 |
| 4.1.6 | <i>Stemming</i> | 26 |
| 4.1.7 | <i>Labeling</i> | 27 |
| 4.1.8 | <i>Visualization</i> | 28 |
| 4.1.9 | <i>TF-IDF</i> | 30 |

| | | |
|-----------------------------|---|-----------|
| 4.1.10 | <i>Data Training Model</i> | 30 |
| 4.1.11 | Hasil <i>Training Model</i> | 31 |
| 4.1.12 | Hasil Model <i>Naïve Bayes Classifier</i> | 31 |
| 4.1.13 | Hasil <i>K-fold Cross Validation</i> | 33 |
| 4.1.14 | Hasil <i>Dashboard Visualisasi Data</i> | 33 |
| BAB V | PENUTUP | 34 |
| 5.1 | Kesimpulan..... | 34 |
| 5.2 | Saran | 34 |
| DAFTAR PUSTAKA | | 36 |
| LAMPIRAN | | 38 |

DAFTAR GAMBAR

| | | |
|-------------|---|----|
| Gambar 3.1 | Diagram Alur Penelitian | 15 |
| Gambar 4.1 | Script Cleaning Data..... | 22 |
| Gambar 4.2 | <i>Script Normalize</i> | 23 |
| Gambar 4.3 | <i>Script Stopword</i> | 24 |
| Gambar 4.4 | <i>Script Tokenizing</i> | 26 |
| Gambar 4.5 | <i>Script Stemming</i> | 27 |
| Gambar 4.6 | <i>Script Labeling</i> | 28 |
| Gambar 4.7 | Distribusi Sentim..... | 29 |
| Gambar 4.8 | Visualisasi <i>Wordcloud</i> Positif..... | 29 |
| Gambar 4.9 | Visualisasi <i>Wordcloud</i> Negatif..... | 30 |
| Gambar 4.10 | Hasil TF-IDF | 27 |
| Gambar 4.11 | <i>Confusion Matrix</i> | 32 |
| Gambar 4.12 | <i>Dashboard</i> | 33 |

DAFTAR TABEL

| | | |
|------------|---|----|
| Tabel 2.1 | <i>Confusion Matrix</i> | 8 |
| Tabel 2.2 | Kajian Penelitian Terdahulu | 10 |
| Tabel 3.1 | Penjadwalan Kegiatan Penelitian | 19 |
| Tabel 4.1 | Sampel Data Playstore (SoloLearn) | 21 |
| Tabel 4.2 | Hasil Cleaning Data | 22 |
| Tabel 4.3 | Hasil <i>Normalize Data</i> | 23 |
| Tabel 4.4 | Hasil <i>Stopword Data</i> | 24 |
| Tabel 4.5 | <i>List Stopword</i> | 24 |
| Tabel 4.6 | Hasil <i>Tokenizing Data</i> | 26 |
| Tabel 4.7 | Hasil <i>Stemming Data</i> | 27 |
| Tabel 4.8 | Hasil <i>Labeling Data</i> | 28 |
| Tabel 4.9 | Sampel <i>data preprocessing</i> | 31 |
| Tabel 4.10 | <i>Test Result Naïve Bayes Classifier</i> | 32 |