



Sentiment Analysis of User Reviews of the Job Search App "Kita Lulus" on the Google Play Store Using Machine Learning

Astrid Ayuzi Putri, Bunga Intan*, Armanto*

¹ Ilmu Teknik, Sistem Informasi, Universitas Bina Insan, Kota Lubuklinggau, Indonesia

Email: ¹ astridacyd4@gmail.com, ^{2*} bungaintan@univbinainsan.ac.id, ^{3*} armanto0204@gmail.com

Abstract- This study aims to analyze user review sentiments for the KitaLulus job search application on the Google Play Store using Machine Learning methods. A total of 20,000 Indonesian language reviews were collected and processed through preprocessing, labeling, and count vectorizer stages. Next, five Machine Learning algorithms were compared, namely K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), Logistic Regression (LR), and Naïve Bayes (NB). The results showed that SVM provided the best performance with 86% accuracy and an F1-score of 0.85, followed by ANN and LR with 85% accuracy, while KNN and NB achieved 83% and 81% accuracy, respectively. This study contributes by presenting a multi-algorithm comparative study on a large dataset of job search application reviews in Indonesia. These findings are expected to help application developers in improving service quality and become a reference for further research in the field of sentiment analysis..

Keywords: Sentiment Analysis; Kita Lulus Application; Google Play Store; Machine Learning.

1. INTRODUCTION

The rapid growth of mobile technology has significantly transformed the way individuals search for job opportunities. Mobile applications provide users with fast, flexible, and accessible platforms to obtain employment-related information. One of the emerging job search applications in Indonesia is Kita Lulus, which offers various features such as job vacancy information, exam preparation materials, and online communities. As the number of users increases, the app receives a substantial volume of reviews and ratings on digital platforms such as the Google Play Store. These reviews not only reflect user satisfaction but also provide valuable insights into the strengths and weaknesses of the application.

User-generated reviews are often unstructured and expressed in natural language, making it difficult for developers and stakeholders to manually extract meaningful information. Sentiment analysis, a subfield of natural language processing (NLP), provides a systematic approach to identifying and classifying opinions in text data as positive, negative, or neutral. This technique has been widely applied in e-commerce, tourism, and social media studies, but its application in analyzing job search platforms, particularly local applications like Kita Lulus, remains underexplored.

Applying sentiment analysis to user reviews on the Google Play Store enables developers and decision-makers to understand user perceptions, identify recurring issues, and improve service quality. Machine learning methods such as Support Vector Machines (SVM), Naïve Bayes, and Logistic Regression have proven effective in text classification tasks and are suitable for analyzing sentiment polarity. By leveraging these algorithms, researchers can process large amounts of textual data and provide evidence-based recommendations for app development.

Therefore, this study aims to analyze user reviews of the Kita Lulus job search app on the Google Play Store using machine learning-based sentiment analysis. The research seeks to classify reviews into sentiment categories, evaluate the performance of different machine learning algorithms in sentiment classification, and provide insights for improving the app's features and user experience. The findings of this study are expected to contribute both practically, by helping developers enhance application quality, and academically, by enriching the body of knowledge on sentiment analysis of mobile job search applications. Advances in digital technology have transformed the way people obtain information, including in job searches. Fierce competition in the Indonesian labor market is driving the growing need for effective job search platforms. One popular app is KitaLulus, with over 3 million active users by 2023. Sentiment analysis of user reviews can provide important insights into the quality of this app.[1][2]. Several previous studies have conducted sentiment analysis on various apps using machine learning algorithms. However, most studies only used limited datasets or evaluated a single algorithm[3][4]. The research gap addressed in this study is the lack of multi-algorithm comparative studies with large-scale datasets in the context of job search apps in Indonesia.[5][6]. This study aims to compare the performance of five classification algorithms: KNN, SVM, ANN, LR, and NB on a dataset of 20,000 reviews. Therefore, the main contribution of this study is to provide a benchmark for machine learning models in sentiment analysis of user review-based job search apps in Indonesia.

2. RESEARCH METHODOLOGY

The research used a quantitative approach with the following stages: (1) data collection through web scraping from the Google Play Store, (2) preprocessing (cleansing, case transformation, tokenizing, stopword removal, word normalization, and stemming), (3) labeling using TextBlob and Google Translate[7][8], (4) text representation using a count vectorizer, (5) classification using KNN, SVM, ANN, LR, and NB, and (6) evaluation using 10-fold cross-



validation, a confusion matrix, and ROC-AUC. To reduce the potential for automatic labeling errors, some of the data was manually validated by annotators. Model evaluation also considered precision, recall, and F1-score. Parameter optimization was performed using grid search to ensure optimal model performance.

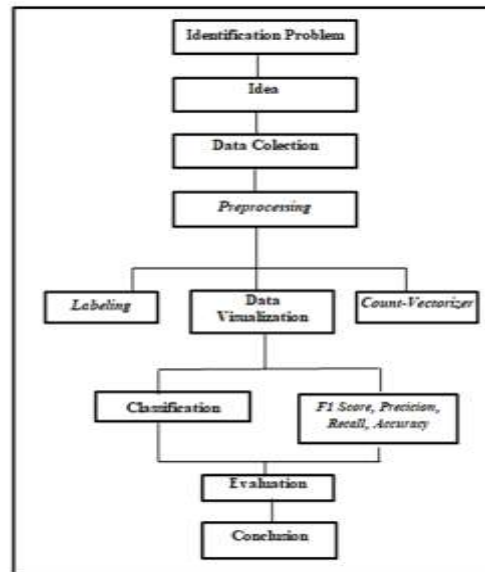


Figure 1. Framework

3. RESULT AND DISCUSSION

The classification results show that SVM excels with 86% accuracy, while ANN and LR achieve 85%, KNN 83%, and NB 81%. The classification results table indicates that SVM performs better because it is able to separate review data that tends to be linearly separable. Conversely, NB performs poorly because the distribution of words in informal texts violates the feature independence assumption[9].

Further discussion shows that these results are consistent with similar international research, where SVM is often reported to excel in text classification with large datasets[3][10]. The practical implication of these findings is that application developers can utilize SVM models to more accurately assess user satisfaction. Furthermore, these results can serve as a basis for further research in developing sentiment-based recommendation systems.



	username	score	at	comment	
14826	Pengguna Google	5	2024-12-02 14:16:04		netral
14805	Pengguna Google	5	2024-12-02 13:27:27	sangat bagus aplikasi ini untuk mencari pekerjaan	
19482	nasukin.tubungubetari	5	2024-12-02 11:50:39	Karya profesional HR MAGASO Kitchen 2016 HRD K.	
1780	Pengguna Google	5	2024-12-02 11:04:29	bagus sekali saya suka karena sangat membantu di	
14836	Pengguna Google	5	2024-12-02 10:30:24	kemudahan dengan pemisahan data yang penting	
18001	Pengguna Google	5	2024-12-02 10:03:57		sangat memuaskan
14822	Pengguna Google	5	2024-12-02 09:30:29	terpenting bagi para hasil pencarian: semua kategori	
14854	Pengguna Google	1	2024-12-02 08:30:05	Untuk aplikasi ini, apakah error saya cari pem...	
19470	V Z	1	2024-12-02 08:12:06	Pada pencarian tidak bisa di gunakan ketika di	
19473	Vita Mustakim	5	2024-12-02 06:52:15	Aplikasi terapanya untuk orang yang sedang me...	
19879	Tubagus Muhammad Syahri	5	2024-12-02 06:49:08		Membuat
14844	Pengguna Google	5	2024-12-02 06:47:26		Recommended
14822	Pengguna Google	5	2024-12-02 05:03:14	ini kenapa tidak ada apa apa ya? Tapi...	
14821	Pengguna Google	1	2024-12-02 03:40:57	Siapa, oh kenapa ya? Nah, itu namanya k...	
14806	Pengguna Google	5	2024-12-02 01:51:13		puas
14845	Pengguna Google	5	2024-12-01 17:14:58		ok

Figure 2.Preprocessing

The goal is to maintain the clarity of the message and improve the quality and accuracy of the information. The result is a cleaner, easier to read, and more relevant text for analysis.[11]

	content	cleaning
0	hebat	hebat
1	sangat bagus aplikasi ini untuk mencari pekerj...	sangat bagus aplikasi ini untuk mencari pekerj...
2	Kerja profesional HR MAGANG Kitchen 2019 HRD K...	Kerja profesional HR MAGANG Kitchen HRD KERJA ...
3	bagus sekali saya suka,semoga semakin sukses b...	bagus sekali saya sukasemoga semakin sukses bu...
4	berintraksi dengan pembeli itu hal yang pertam...	berintraksi dengan pembeli itu hal yang pertam...

Figure 3.Cleaning

Next, the case transformation process to change or equalize letters to lowercase so that text data is more consistent[11]. The results of this letter transformation are as follows:

	content	cleaning	transform_case
0	hebat	hebat	hebat
1	sangat bagus aplikasi ini untuk mencari pekerj...	sangat bagus aplikasi ini untuk mencari pekerj...	sangat bagus aplikasi ini untuk mencari pekerj...
2	Kerja profesional HR MAGANG Kitchen 2019 HRD K...	Kerja profesional HR MAGANG Kitchen HRD KERJA ...	kerja profesional hr magang kitchen hrd kerja...
3	bagus sekali saya suka,semoga semakin sukses b...	bagus sekali saya sukasemoga semakin sukses bu...	bagus sekali saya sukasemoga semakin sukses bu...
4	berintraksi dengan pembeli itu hal yang pertam...	berintraksi dengan pembeli itu hal yang pertam...	berintraksi dengan pembeli itu hal yang pertam...
10004	Super super super	Super super super	super super super
10005	Kita lulus Oke yessssss	Kita lulus Oke yessss	Kita lulus oke yessss
10006	Semoga selalu diberi kelancaran kemudahan sel...	Semoga selalu diberi kelancaran kemudahan sel...	semoga selalu diberi kelancaran kemudahan sel...
10007	Semoga ini merupakan akhir di pejuang NPK L...	Semoga ini merupakan akhir di pejuang NPK L...	semoga ini merupakan akhir di pejuang npk lu...
10009	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	aplikasi yg bagus buat belajar persiapan cpns

Figure 4.Transformasi Case

The procedure of dividing text into smaller units, known as tokens. These tokens may consist of words, phrases, or even sentences, depending on how they are utilized. The following illustrates the outcome of the tokenization process.[12]

	content	cleaning	transform_case	tokenisasi	stopword	normalisasi	stemming
0	hebat	hebat	hebat	[hebat]	[hebat]	[hebat]	[hebat]
1	sangat bagus aplikasi ini untuk mencari pekerj...	sangat bagus aplikasi ini untuk mencari pekerj...	sangat bagus aplikasi ini untuk mencari pekerj...	[sangat, bagus, aplikasi, ini, untuk, mencari, pekerj, ...]	[sangat, bagus, aplikasi, ini, untuk, mencari, pekerj, ...]	[sangat, bagus, aplikasi, ini, untuk, mencari, pekerj, ...]	[sangat, bagus, aplikasi, ini, untuk, mencari, pekerj, ...]
2	Kerja profesional HR MAGANG Kitchen 2019 HRD K...	Kerja profesional HR MAGANG Kitchen HRD KERJA ...	kerja profesional hr magang kitchen hrd kerja...	[kerja, profesional, hr, magang, kitchen, hrd, kerja, ...]	[kerja, profesional, hr, magang, kitchen, hrd, kerja, ...]	[kerja, profesional, hr, magang, kitchen, hrd, kerja, ...]	[kerja, profesional, hr, magang, kitchen, hrd, kerja, ...]
3	bagus sekali saya suka,semoga semakin sukses b...	bagus sekali saya sukasemoga semakin sukses bu...	bagus sekali saya sukasemoga semakin sukses bu...	[bagus, sekali, saya, suka, semoga, semakin, sukses, b..., ...]	[bagus, sekali, saya, suka, semoga, semakin, sukses, b..., ...]	[bagus, sekali, saya, sukasemoga, semakin, sukses, bu..., ...]	[bagus, sekali, saya, sukasemoga, semakin, sukses, bu..., ...]
4	berintraksi dengan pembeli itu hal yang pertam...	berintraksi dengan pembeli itu hal yang pertam...	berintraksi dengan pembeli itu hal yang pertam...	[berintraksi, dengan, pembeli, itu, hal, yang, pertam, ...]	[berintraksi, dengan, pembeli, itu, hal, yang, pertam, ...]	[berintraksi, dengan, pembeli, itu, hal, yang, pertam, ...]	[berintraksi, dengan, pembeli, itu, hal, yang, pertam, ...]
10004	Super super super	Super super super	super super super	[super, super, super]	[super, super, super]	[super, super, super]	[super, super, super]
10005	Kita lulus Oke yessssss	Kita lulus Oke yessss	Kita lulus oke yessss	[Kita, lulus, Oke, yessss, yessss, yessss, yessss]	[Kita, lulus, Oke, yessss, yessss, yessss, yessss]	[Kita, lulus, oke, yessss, yessss, yessss, yessss]	[Kita, lulus, oke, yessss, yessss, yessss, yessss]
10006	Semoga selalu diberi kelancaran kemudahan sel...	Semoga selalu diberi kelancaran kemudahan sel...	semoga selalu diberi kelancaran kemudahan sel...	[Semoga, selalu, diberi, kelancaran, kemudahan, sel..., ...]	[Semoga, selalu, diberi, kelancaran, kemudahan, sel..., ...]	[semoga, selalu, diberi, kelancaran, kemudahan, sel..., ...]	[semoga, selalu, diberi, kelancaran, kemudahan, sel..., ...]
10007	Semoga ini merupakan akhir di pejuang NPK L...	Semoga ini merupakan akhir di pejuang NPK L...	semoga ini merupakan akhir di pejuang npk lu...	[Semoga, ini, merupakan, akhir, di, pejuang, npk, lu..., ...]	[Semoga, ini, merupakan, akhir, di, pejuang, npk, lu..., ...]	[semoga, ini, merupakan, akhir, di, pejuang, npk, lu..., ...]	[semoga, ini, merupakan, akhir, di, pejuang, npk, lu..., ...]
10009	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	aplikasi yg bagus buat belajar persiapan cpns	[Aplikasi, yg, bagus, buat, belajar, persiapan, cpns]	[Aplikasi, yg, bagus, buat, belajar, persiapan, cpns]	[aplikasi, yg, bagus, buat, belajar, persiapan, cpns]	[aplikasi, yg, bagus, buat, belajar, persiapan, cpns]

Figure 5.Tokenizing

The application of stopwords entails the removal of non-essential words by using a CSV file that contains a list of stopwords. The objective is to enhance the quality of the text by eliminating words that disrupt the meaning or diminish the clarity of the message. The outcomes of applying stopwords are as follows:[13]

	content	cleaning	transform case	tokenisasi	stopword
0	hebat	hebat	hebat	{hebat}	{hebat}
1	sangat bagus aplikasi ini untuk mencari paket	sangat bagus aplikasi ini untuk mencari paket	sangat bagus aplikasi ini untuk mencari paket	{sangat, bagus, aplikasi, ini, untuk, mencari, paket}	{bagus, aplikasi, mencari, paket}
2	Karya profesional HR MAGANG Kitchen 2019 HRD K...	Karya profesional HR MAGANG Kitchen 2019 HRD K...	Karya profesional hr magang kitchen terd karya	{karya, profesional, hr, magang, kitchen, terd, karya}	{karya, profesional, hr, magang, kitchen, terd, karya}
3	bagus sekali saya suka semoga semakin sukses di	bagus sekali saya suka semoga semakin sukses di	bagus sekali saya suka semoga semakin sukses di	{bagus, sekali, saya, suka, semoga, semakin, sukses, di}	{bagus, suka, semoga, sukses, di}
4	berinteraksi dengan pembeli di hal yang pertam	berinteraksi dengan pembeli di hal yang pertam	berinteraksi dengan pembeli di hal yang pertam	{berinteraksi, dengan, pembeli, di, hal, yang, pertam}	{berinteraksi, pembeli, pertam}
19994	Super super super	Super super super	super super super	{super, super, super}	{super, super, super}
19995	Kita lulus Oke yessssssss	Kita lulus Oke yessss	kita lulus oke yessss	{kita, lulus, oke, yesss}	{lulus, oke, yesss}
19996	Demoga selalu diberi kelancaran kemudahan seli	Demoga selalu diberi kelancaran kemudahan seli	demoga selalu diberi kelancaran kemudahan seli	{demoga, selalu, diberi, kelancaran, kemudahan, seli}	{demoga, kelancaran, kemudahan, seli}
19997	Demoga in merupakan dtkar di pejuang NP 202	Demoga in merupakan dtkar di pejuang NP 202	demoga in merupakan dtkar di pejuang np 202	{demoga, in, merupakan, dtkar, di, pejuang, np, 202}	{demoga, dtkar, pejuang, np, 202}
19998	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	aplikasi yg bagus buat belajar persiapan cpns	{aplikasi, yg, bagus, buat, belajar, persiapan, cpns}	{aplikasi, bagus, belajar, persiapan, cpns}

Figure 6. Stopwords

Next, the procedure of converting slang or non-standard words into a more formal version, utilizing a CSV file that contains normalization rules[14]. The primary aim of this step is to enhance the consistency and uniformity of word usage, as well as to ensure that the text adheres to formal language standards. The following presents the results of the word normalization process.[15]

	content	cleaning	transform case	tokenisasi	stopword	normalisasi
0	hebat	hebat	hebat	{hebat}	{hebat}	hebat
1	sangat bagus aplikasi ini untuk mencari paket	sangat bagus aplikasi ini untuk mencari paket	sangat bagus aplikasi ini untuk mencari paket	{sangat, bagus, aplikasi, ini, untuk, mencari, paket}	{bagus, aplikasi, mencari, paket}	sangat bagus aplikasi ini untuk mencari paket
2	Karya profesional HR MAGANG Kitchen 2019 HRD K...	Karya profesional HR MAGANG Kitchen 2019 HRD K...	Karya profesional hr magang kitchen terd karya	{karya, profesional, hr, magang, kitchen, terd, karya}	{karya, profesional, hr, magang, kitchen, terd, karya}	Karya profesional hr magang kitchen terd karya
3	bagus sekali saya suka semoga semakin sukses di	bagus sekali saya suka semoga semakin sukses di	bagus sekali saya suka semoga semakin sukses di	{bagus, sekali, saya, suka, semoga, semakin, sukses, di}	{bagus, suka, semoga, sukses, di}	bagus sekali saya suka semoga semakin sukses di
4	berinteraksi dengan pembeli di hal yang pertam	berinteraksi dengan pembeli di hal yang pertam	berinteraksi dengan pembeli di hal yang pertam	{berinteraksi, dengan, pembeli, di, hal, yang, pertam}	{berinteraksi, pembeli, pertam}	berinteraksi dengan pembeli di hal yang pertam
19994	Super super super	Super super super	super super super	{super, super, super}	{super, super, super}	super super super
19995	Kita lulus Oke yessssssss	Kita lulus Oke yessss	kita lulus oke yessss	{kita, lulus, oke, yesss}	{lulus, oke, yesss}	Kita lulus oke yessss
19996	Demoga selalu diberi kelancaran kemudahan seli	Demoga selalu diberi kelancaran kemudahan seli	demoga selalu diberi kelancaran kemudahan seli	{demoga, selalu, diberi, kelancaran, kemudahan, seli}	{demoga, kelancaran, kemudahan, seli}	demoga selalu diberi kelancaran kemudahan seli
19997	Demoga in merupakan dtkar di pejuang NP 202	Demoga in merupakan dtkar di pejuang NP 202	demoga in merupakan dtkar di pejuang np 202	{demoga, in, merupakan, dtkar, di, pejuang, np, 202}	{demoga, dtkar, pejuang, np, 202}	demoga in merupakan dtkar di pejuang np 202
19998	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	aplikasi yg bagus buat belajar persiapan cpns	{aplikasi, yg, bagus, buat, belajar, persiapan, cpns}	{aplikasi, bagus, belajar, persiapan, cpns}	aplikasi yg bagus buat belajar persiapan cpns

Figure 7. Normalization

Then, the stemming process is carried out using the Sastrawi library, which aims to change words into their basic form. Here are the results of the stemming process:[16]

	content	cleaning	transform case	tokenisasi
0	hebat	hebat	hebat	{hebat}
1	sangat bagus aplikasi ini untuk mencari paket	sangat bagus aplikasi ini untuk mencari paket	sangat bagus aplikasi ini untuk mencari paket	{sangat, bagus, aplikasi, ini, untuk, mencari, paket}
2	Karya profesional HR MAGANG Kitchen 2019 HRD K...	Karya profesional HR MAGANG Kitchen 2019 HRD K...	Karya profesional hr magang kitchen terd karya	{karya, profesional, hr, magang, kitchen, terd, karya}
3	bagus sekali saya suka semoga semakin sukses di	bagus sekali saya suka semoga semakin sukses di	bagus sekali saya suka semoga semakin sukses di	{bagus, sekali, saya, suka, semoga, semakin, sukses, di}
4	berinteraksi dengan pembeli di hal yang pertam	berinteraksi dengan pembeli di hal yang pertam	berinteraksi dengan pembeli di hal yang pertam	{berinteraksi, dengan, pembeli, di, hal, yang, pertam}
19994	Super super super	Super super super	super super super	{super, super, super}
19995	Kita lulus Oke yessssssss	Kita lulus Oke yessss	kita lulus oke yessss	{kita, lulus, oke, yesss}
19996	Demoga selalu diberi kelancaran kemudahan seli	Demoga selalu diberi kelancaran kemudahan seli	demoga selalu diberi kelancaran kemudahan seli	{demoga, selalu, diberi, kelancaran, kemudahan, seli}
19997	Demoga in merupakan dtkar di pejuang NP 202	Demoga in merupakan dtkar di pejuang NP 202	demoga in merupakan dtkar di pejuang np 202	{demoga, in, merupakan, dtkar, di, pejuang, np, 202}
19998	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	aplikasi yg bagus buat belajar persiapan cpns	{aplikasi, yg, bagus, buat, belajar, persiapan, cpns}

Figure 8. Stemming

Next, the labeling stage is the process of determining the sentiment label, whether it is positive, neutral, or negative. This process is carried out using the TextBlob library. Because TextBlob cannot understand text in Indonesian, the GoogleTrans library is also needed to translate the text into English [17] Results of the translation process

id	comment	comment	comment	comment	comment	comment	comment	comment	comment
1	target tugas aplikasi ini untuk mencari peluang...	target tugas aplikasi ini untuk mencari peluang...	target tugas aplikasi ini untuk mencari peluang...	target tugas aplikasi ini untuk mencari peluang...	target tugas aplikasi ini untuk mencari peluang...	target tugas aplikasi ini untuk mencari peluang...	target tugas aplikasi ini untuk mencari peluang...	target tugas aplikasi ini untuk mencari peluang...	This application is very good for finding work...
2	kerja profesional HR di GIGAH Kitchen 2019 HR...	kerja profesional HR di GIGAH Kitchen 2019 HR...	kerja profesional HR di GIGAH Kitchen 2019 HR...	kerja profesional HR di GIGAH Kitchen 2019 HR...	kerja profesional HR di GIGAH Kitchen 2019 HR...	kerja profesional HR di GIGAH Kitchen 2019 HR...	kerja profesional HR di GIGAH Kitchen 2019 HR...	kerja profesional HR di GIGAH Kitchen 2019 HR...	professional work by internships kitchen HR wor...
3	bagus sekali rasanya...	bagus sekali rasanya...	bagus sekali rasanya...	bagus sekali rasanya...	bagus sekali rasanya...	bagus sekali rasanya...	bagus sekali rasanya...	bagus sekali rasanya...	Very good, I like it, hopefully more success f...
4	berinteraksi dengan pembeli itu hal yang penting...	berinteraksi dengan pembeli itu hal yang penting...	berinteraksi dengan pembeli itu hal yang penting...	berinteraksi dengan pembeli itu hal yang penting...	berinteraksi dengan pembeli itu hal yang penting...	berinteraksi dengan pembeli itu hal yang penting...	berinteraksi dengan pembeli itu hal yang penting...	berinteraksi dengan pembeli itu hal yang penting...	Interacting with buying is the first thing for...
14277	Super super super	Super super super	Super super super	Super super super	Super super super	Super super super	Super super super	Super super super	super super super
14278	Kita bisa The yesssssss	Kita bisa The yesssssss	Kita bisa The yesssssss	Kita bisa The yesssssss	Kita bisa The yesssssss	Kita bisa The yesssssss	Kita bisa The yesssssss	Kita bisa The yesssssss	we pass okay yesssss
14279	Semoga selalu dilah...	Semoga selalu dilah...	Semoga selalu dilah...	Semoga selalu dilah...	Semoga selalu dilah...	Semoga selalu dilah...	Semoga selalu dilah...	Semoga selalu dilah...	I hope it always goes smoothly so that we can ...
14280	Semoga in persiapan akhir...	Semoga in persiapan akhir...	Semoga in persiapan akhir...	Semoga in persiapan akhir...	Semoga in persiapan akhir...	Semoga in persiapan akhir...	Semoga in persiapan akhir...	Semoga in persiapan akhir...	I hope that with all my efforts, I will pass a...
14281	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	Aplikasi yg bagus buat belajar persiapan CPNS	A good application for teaching CPNS preparation

Figure 9.Labeling

After the text is translated into English, sentiment labels are assigned using TextBlob. This process involves the TranslateToEnglish(text) function to translate, and the 'stemming' column is translated into English with the polarity result measuring the level of emotional sentiment (a value of 0 indicates positive sentiment)[17]. Here is the result of labeling using TextBlob

```

0          great 0.800000 positif
1  This application is very good for finding work... 0.671667 positif
2  professional work hr internship kitchen HR wor... 0.100000 positif
3  Very good, I like it, hopefully more success f... 0.552500 positif
4  Interacting with buying is the first thing for... 0.250000 positif
...
14277      super super super 0.333333 positif
14278      we pass okay yesssss 0.500000 positif
14279  I hope it always goes smoothly so that we can ... 0.400000 positif
14280  I hope that with all my efforts, I will pass a... 0.000000 netral
14281  A good application for teaching CPNS preparation 0.700000 positif

[14281 rows x 10 columns]
```

Figure 10.Labeling

The results of the sentiment composition from the labeling process are as follows:

```

TextBlob
positif    10175
netral     2927
negatif    1179
Name: count, dtype: int64
```

Figure 11.Labeling TextBlob

Visualization is a graphical representation stage of information in data that aims to facilitate understanding more intuitively and effectively. Data visualization is done using pie charts and word clouds[9][18]. The results of sentiment representation with pie charts are as follows:

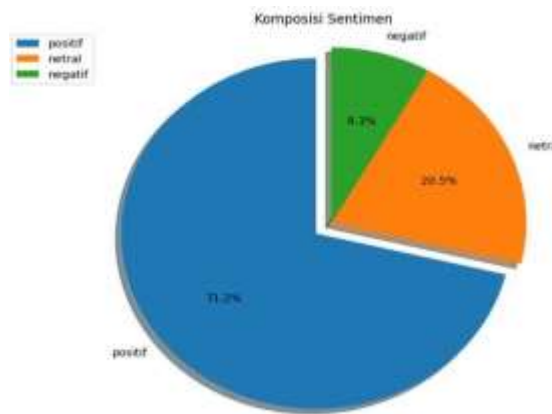


Figure 22. Visualisasi Pie

The next visualization is using a wordcloud, which depicts the words that appear most frequently in user review data for the kitalulus application[19].



Figure 13. Wordcloud Indonesian



Figure 15. Wordclaude English

The next step is Count-Vectorizer transforms Bag of Words (BoW) into a vector representation by extracting words from documents and counting the frequency of occurrence of each word. Each document is represented by a vector of the same size as the number of comprehensions, where entries indicate the number of occurrences of each word[20]. In this study, classification was performed by applying the K-Nearest Neighbors algorithm, Support Vector Machine, Artificial Neural Network, Naïve Bayes, and Logistic Regression. The data was partitioned into training and testing sets using the k-fold 10 Cross-Validation method, where one subset served as the test data, while the other nine subsets were used as training data[21]. 10 times to ensure a consistent and representative evaluation.

Table 1. K-Nearest Neighbors Classification

Sentiment	Precision	Recall	F1 Score	Accuracy	ROC AUC	CV K-10
Negative	77%	36%	49%			
Neutral	62%	81%	71%	83%	93%	83%
Positive	91%	89%	90%			

Table 2. Support Vector Machine Classification

<i>Sentiment</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	<i>Accuracy</i>	<i>ROC AUC</i>	<i>CV K-10</i>
<i>Negative</i>	67%	47%	55%	86%	92%	83%
<i>Neutral</i>	71%	85%	77%			
<i>Positive</i>	92%	90%	91%			

Table 3. Artificial Neural Network Classification

<i>Sentiment</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	<i>Accuracy</i>	<i>ROC AUC</i>	<i>CV K-10</i>
<i>Negative</i>	68%	45%	55%	85%	92%	83%
<i>Neutral</i>	69%	85%	76%			
<i>Positive</i>	92%	90%	91%			

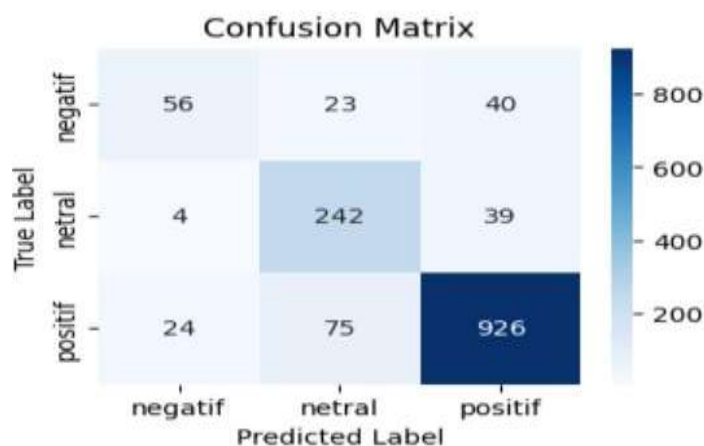
Table 4. Logistic Regression Classification

<i>Sentiment</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	<i>Accuracy</i>	<i>ROC AUC</i>	<i>CV K-10</i>
<i>Negative</i>	72%	45%	56%	85%	92%	83%
<i>Neutral</i>	69%	82%	75%			
<i>Positive</i>	92%	91%	91%			

Table 5. Naïve Bayes Classification

<i>Sentiment</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	<i>Accuracy</i>	<i>ROC AUC</i>	<i>CV K-10</i>
<i>Negative</i>	55%	45%	49%	81%	91%	83%
<i>Neutral</i>	70%	61%	65%			
<i>Positive</i>	87%	91%	89%			

The assessment of the classification outcomes includes 10-fold Cross Validation, Confusion Matrix, and ROC-AUC graphs. This offers a thorough insight into the performance of the model using K-Nearest Neighbors, Support Vector Machine, Artificial Neural Network, Naïve Bayes, and Logistic Regression[22]. The assessment outcomes of the five algorithms are displayed as follows:


Figure 16. Confusion ANN

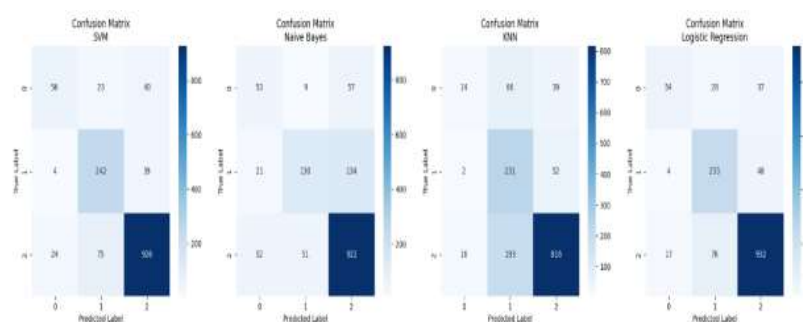


Figure 17. Confusion SVM,KNN,LR,NB

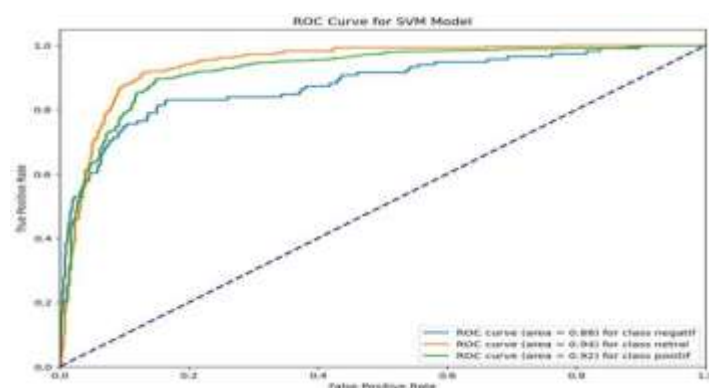


Figure 3.ROC AUC ANN

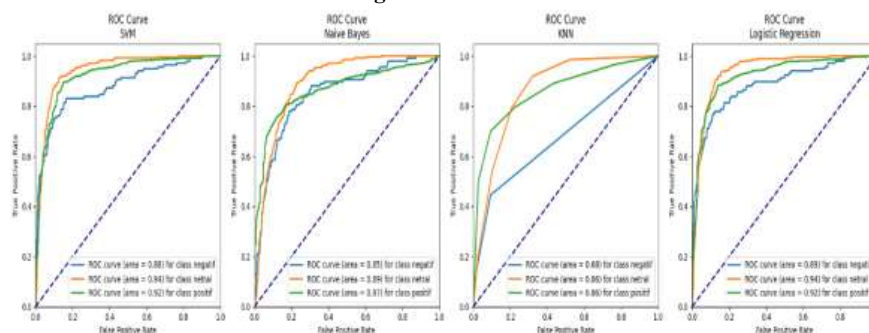


Figure 48.SVM,KNN,LR,NB

The issue addressed in this study is the examination of user reviews for the KitaLulus job search application utilizing the K-Nearest Neighbors, Support Vector Machine, Artificial Neural Network, Naïve Bayes, and Logistic Regression algorithms. The purpose of this study is to compare the performance of these algorithms in classification[13]. The initial step of this study is data collection (crawling) using google_play_scraper with the keyword "KitaLulus application" from January 2021 to November 2024, with a total of 20,000 user reviews. After the data is collected, the next stage is preprocessing which produces clean data, which can be seen in Table 6.

Table 6. Data after preprocessing

Review	Preprocessing Results
KitaLulus Application	20.000

After preprocessing, the next stage is labeling using the TextBlob library. The composition results of the labeling can be seen in table 7.

**Table 7** TextBlob sentiment labels

Review	Positive	Neutral	Negative
KitaLulus Application	10.175	2.927	1.179

The final stage is the application of classification using the K-Nearest Neighbors algorithm, Support Vector Machine, Artificial Neural Network (ANN), Logistic Regression and Naïve Bayes, by implementing the Cross-Validation k-10 technique which produces an evaluation of the average value, evaluation of the Confusion Matrix and evaluation of the ROC-AUC can be seen as follows:

Table 8. Classification weighted avg

Classification Model	Accuracy	Precision	Recall	F1-Score
KNN	83%	84%	83%	83%
SVM	86%	86%	86%	85%
ANN	85%	86%	85%	85%
LR	85%	86%	85%	85%
NB	81%	81%	81%	81%

4. CONCLUSION

This study concluded that SVM was the best algorithm for sentiment analysis of KitaLulus app reviews, with an accuracy of 86% and an F1-score of 0.85. The main contribution of this study is the presentation of a comparative study of five machine learning algorithms on a large dataset of job search app reviews in Indonesia, which can serve as a reference for further studies. However, this study has limitations: the automatic labeling process is potentially noisy, and the dataset is limited to the Google Play Store. For future research, the use of deep learning models such as IndoBERT and the expansion of the dataset to include multiple platforms and languages are recommended.

REFERENCES

- [1] V. M. Hidayah, "Pemodelan Teknologi dalam Aplikasi KitaLulus untuk Lowongan Pekerjaan Menggunakan Metode Technology Acceptance Model (TAM)," vol. 5, no. 3, pp. 2801–2812, 2024.
- [2] I. Iwandini, A. Triayudi, and G. Soepriyono, "Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor," *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 543–550, 2023, doi: 10.47065/josh.v4i2.2937.
- [3] H. Imaduddin, F. Y. A'la, and Y. S. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 113–117, 2023, doi: 10.14569/IJACSA.2023.0140813.
- [4] S. Tabinda Kokab, S. Asghar, and S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," *Array*, vol. 14, no. April, p. 100157, 2022, doi: 10.1016/j.array.2022.100157.
- [5] A. G. Tando and M. I. Irawan, "Analisis Dinamika Harga Saham yang Dipengaruhi oleh Analisis Sentimen di Media Sosial Menggunakan Algoritma Support Vector Machine," *J. Sains dan Seni ITS*, vol. 12, no. 1, 2023, doi: 10.12962/j23373520.v12i1.107080.
- [6] A. E. Sari, S. Widowati, and K. M. Lhaksana, "Klasifikasi Ulasan Pengguna Aplikasi Mandiri Online di Google Play Store dengan Menggunakan Metode Information Gain dan Naive Bayes Classifier," A. E., Widowati, S., & Lhaksana, K. M. (2019). Klasifikasi Ulasan Pengguna Aplikasi Mandiri Online di Goog," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 9143–9157, 2019, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/viewFile/9933/9790>
- [7] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback," *Appl. Sci.*, vol. 15, no. 1, pp. 1–28, 2025, doi: 10.3390/app15010172.
- [8] N. P. I. Maharani, A. Purwarianti, Y. Yustiwani, and F. C. Rochim, "Domain-Specific Language Model Post-Training for Indonesian Financial NLP," *Proc. Int. Conf. Electr. Eng. Informatics*, 2023, doi: 10.1109/ICEEI59426.2023.10346625.
- [9] S. A. R. Rizaldi, S. Alam, and I. Kurniawan, "Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 109–117, 2023, doi: 10.55123/storage.v2i3.2334.
- [10] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of





- Indonesian reviews using fine-tuning IndoBERT and R-CNN,” *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 348–354, 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [11] D. Iskandar Mulyana and N. Lutfianti, “Penerapan Sentimen Analisis Dengan Algoritma SVM Dalam Tanggapan Netizen Terhadap Berita Resesi 2023,” *Sisfotenika*, vol. 13, no. 1, pp. 53–64, 2023.
- [12] D. T. Wisudawati, “Analisis Sentimen Terhadap Dampak COVID-19 Pada Perfoma E-Commerce di Indonesia Menggunakan SVM,” *Univ. Muhammadiyah Semarang*, pp. 1–146, 2020.
- [13] I. Kurniawan, A. Lia Hananto, S. Shofia Hilabi, A. Hananto, B. Priyatna, and A. Yuniar Rahman, “Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 731–740, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [14] M. Parmitha, “ANALISIS SENTIMEN OPINI PUBLIK MENGENAI CALON PRESIDEN 2024 PADA SOSIAL MEDIA TWITTER MENGGUNAKAN MACHINE LEARNING ANALISIS SENTIMEN OPINI PUBLIK MENGENAI CALON PRESIDEN 2024 PADA SOSIAL MEDIA TWITTER MENGGUNAKAN,” 2024.
- [15] E. P. A. Akhmad, “Analisis Sentimen Ulasan Aplikasi DLU Ferry Pada Google Play Store Menggunakan Bidirectional Encoder Representations from Transformers,” *J. Apl. Pelayaran Dan Kepelabuhanan*, vol. 13, no. 2, pp. 104–112, 2023, doi: 10.30649/japk.v13i2.94.
- [16] A. Muhammadin and I. A. Sobari, “Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma Svm Dan Nbc,” *Reputasi J. Rekayasa Perangkat Lunak*, vol. 2, no. 2, pp. 85–91, 2021, doi: 10.31294/reputasi.v2i2.785.
- [17] A. Baita, Y. Pristyanto, N. Cahyono, P. Covid-, K. N. N. Akurasi, and K. Kunci, “ANALISIS SENTIMEN MENGENAI VAKSIN SINO VAC MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN) Abstraksi Keywords :,” vol. 4, no. 2, pp. 42–46, 2021.
- [18] A. Wibowo, “Analisa Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis Pada PT. Telkominfra,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2292–2304, 2022, doi: 10.35957/jatisi.v9i3.2737.
- [19] M. Kholilullah, M. Martanto, and U. Hayati, “Analisis Sentimen Pengguna Twitter(X) Tentang Piala Dunia Usia 17 Menggunakan Metode Naive Bayes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 392–398, 2024, doi: 10.36040/jati.v8i1.8378.
- [20] S. Khomsah and A. S. Aribowo, “Model Text-Preprocessing Komentar Youtube Dalam Bahasa IndonesiaText-Preprocessing Model Youtube Comments in Indonesian,” vol. 4, no. 4, pp. 650–650, 2020.
- [21] M. R. Fauzan, H. Oktafia, L. Wijaya, and J. Karman, “KENAIKAN HARGA BBM DI MEDIA SOSIAL TWITTER,” vol. 1, no. 1, pp. 82–89, 2023.
- [22] D. Pradana, M. Luthfi Alghifari, M. Farhan Juna, and D. Palaguna, “Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network,” *Indones. J. Data Sci.*, vol. 3, no. 2, pp. 55–60, 2022, doi: 10.56705/ijodas.v3i2.35.
- [23] K. S. Nugroho, A. Y. Sukmadewa, H. Wuswilahaken Dw, F. A. Bachtiar, and N. Yudistira, “BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews,” *ACM Int. Conf. Proceeding Ser.*, pp. 258–264, 2021, doi: 10.1145/3479645.3479679.
- [24] Mesran, M., Syahrizal, M., Sarwandi, S., Aripin, S., Utomo, D. P., & Karim, A. (2024, April). A comparison of the performance of data mining classification algorithms on medical datasets with the application of data normalization. In AIP Conference Proceedings (Vol. 3048, No. 1, p. 020047). AIP Publishing LLC.

