Recursive Journal of Informatics
Vol. 3, No. 2, September 2025
p-ISSN 2963-5551 https://journal.unnes.ac.id/journals/index.php/rji e-ISSN 2986-6588

Sentiment Analysis of Jobstreet Application Reviews on Google Play Store Using Support Vector Machine Algorithm with Adaptive Synthetic

Febryan Surya Shantika[1], Zaenal Abidin[2]
[1,2]Computer Science Department, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Semarang, Indonesia

Abstract.
Purpose: This research aims to test the performance result of the Support Vector Machine (SVM) classification algorithm using the help of Adaptive Synthetic (ADASYN) oversampling to analyze sentiment in Jobstreet application reviews on the Google Play Store. Sentiment analysis is a significant method to understand the market needs and application improvement.
Methods/Study design/approach: The dataset originates from Google Play reviews gained using the scrapping method, comprising 5,174 reviews with 11 attributes. The process begins with data scrapping, data labeling, and data preprocessing, including casefolding, tokenizing, filtering, and stemming using Python programs. The data is then weighted and split using an 80:20 ratio. Then applying oversampling ADASYN on a clean dataset before using SVM classification to produce the performance result.
Result/Findings: Both scenarios are conducted on SVM classification to classify the dataset. The evaluation results indicate that using SVM classification without ADASYN produces an accuracy result of 89.08%. Other scenarios by using SVM classification with the ADASYN sampling approach produce an accuracy result of 89.95%. The performance in accuracy result by using the ADASYN sampling approach on SVM classification shows an increasing result of 0.87%.
Novelty/Originality/Value: This study employs two result scenarios of SVM classification by using the ADASYN sampling approach. It contributes to the literature by demonstrating the usability of the ADASYN oversampling approach to optimalize the SVM classification result used for sentiment analysis in Jobstreet application reviews on the Google Play Store.

Keywords: Sentiment Analysis, Imbalanced Dataset, Adaptive Synthetic, Support Vector Machine.

Received August 2024 / **Revised** September 2025 / **Accepted** September 2025

INTRODUCTION

Technology is one of the factors that drives a significant change in various aspects of industry and society, including the job search process [1]. The process of delivering job information has developed along with the development of the internet as a digital information provider. Various well-known companies announce vacancies provided through the company's official website. Job seekers can also send application documents via email if provided by the company. In addition, several applications have been developed to serve as a channel for job vacancy information for companies that do not have their own official websites or want to spread vacancies more widely. Jobstreet is an application that distributes job vacancy information on Google Play Store. Jobstreet was developed in 1997 and started operating in Indonesia in 2006 [2]. Job applicants can browse suitable job applications at the touch of a finger through their Android device. Through the rating and review feature available on the Google Play Store, the Jobstreet app has received 312,000 reviews, mostly positive reviews with an average rating of
4.5 stars [3].

Sentiment analysis is a classification field of study that uses an individual's opinion, feeling, or judgment in the form of text about a particular topic issue or object [4], [5]. In the field of machine learning, sentiment

[1]*Corresponding author.
Email addresses: suryaafebryan@students.unnes.ac.id (Shantika)
DOI: 10.15294/rji.v3i2.11891

analysis works as one of the natural language processing (NLP) machine learning areas in analyzing and manipulating large amounts of text efficiently using various methods and algorithms [6], [7]. In general, sentiment is separated into positive, negative, and neutral sentiment classes. In the Jobstreet application identified user reviews are dominated by positive reviews, causing data imbalance. This imbalance in the dataset can affect the analysis results as it tends to predict the majority class over the minority class, creating bias [8], [9].

Adaptive Synthetic (ADASYN) is a sampling approach that generates synthetic data from minority data to overcome data imbalance [10], [11]. The generation of synthetic data is done to equalize the amount of minority data with majority data, with the difference from other methods is that ADASYN focuses on data with a difficult learning rate [12]. The Support Vector Machine (SVM) algorithm was introduced with the initial goal of overcoming regression analysis and classification where classification is done by finding the best hyperplane as a separator between classes, the hyperplane itself is a line that becomes a separator function that separates the two classes where the distance of the line from the nearest object is called a margin [13], [14]. In the classification of Jobstreet reviews, because they are grouped into binary classes with 2 classes, namely positive and negative sentiments, the classification is also used for binary classification using a linear kernel. This research was conducted to test the accuracy of the SVM algorithm with additional oversampling techniques using the ADASYN algorithm in analyzing the sentiment of Jobstreet application reviews on the Google Play Store.

Sentiment analysis has already become one of the topics known in machine learning research. Several studies have been explored to further expand and improve this topic using many techniques. One of the research conducted by Irmanda and Astriratma [15] testing SVM classification on children to adult rhyme that considered multiclass dataset, by using one against all methods emphasized the usage of SVM classification in multiclass scenario by converting it into binary class.

Setiawan and Kaburuan [16] trying to analyze the Korlantas application using SVM classification with 4 split data scenarios. The main goal of this research is to compare the result of SVM classification between all split data scenarios to know the optimal data ratio when classifying the Korlantas application on Google Play Store using SVM classification.

Kharisma and Ernawati [17] used SVM classification with a sampling approach in their study. They compared both scenarios, either undersampling or oversampling with SVM classification.

Hidayat, et al. [18] conducted sentiment analysis on the Airbnb dataset using an oversampling approach to SVM classification. The research was conducted to compare the result for oversampling using SMOTE and ADASYN to SVM classification with radial basis function kernel.

Magnolia, et al. [19] also researched to explore SVM classification combined with some known sampling methods on Twitter comments. This research applies 3 scenarios of max features on Term Frequency Inverse Document Frequency of records (TF-IDF) steps. The main goal of this research is to know which sampling approach and max features scenarios perform better when analyzing sentiment from Twitter application's comments.

From various studies and research conducted previously, this paper research main goal is to know the performance result of SVM classification when using ADASYN and without using ADASYN. The dataset used in this research is Jobstreet application reviews on the Google Play Store gained using the scrapping method. The main purpose is to further emphasize that the oversampling method using ADASYN can improve classification results by taking care of imbalanced data.

METHODS

To acquire the dataset needed in this research, the implementation of the Knowledge Discovery in Database (KDD) method is conducted. KDD refers to a set of processes to obtain information from a large database, by performing data mining using specific algorithms to extract relevant information from the data [20], [21]. The implementation process for this research is illustrated in Figure 1.

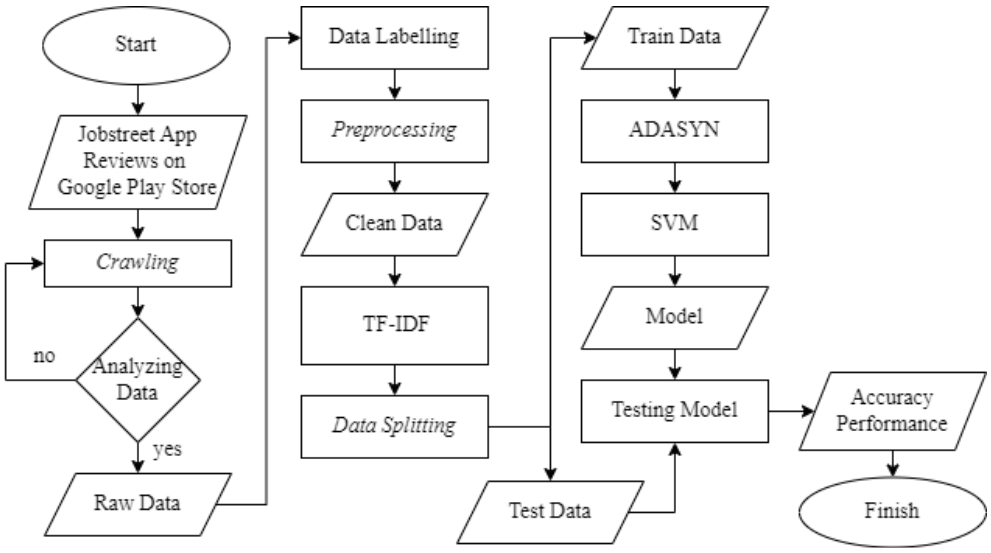

Figure 1. Research model

Dataset

In this research, the Jobstreet application reviews dataset is sourced by using the crawling method. This method works by scrapping Jobstreet Reviews from the Google Play Store computationally with Python programs. Python library using Google Play Scraper running through Jupyter Notebook. The program has been set to a 5,000 count and only scrapped reviews from Indonesian servers using Indonesian languages. From the scrapping steps dataset gained a total of 5,174 data reviews with 11 attributes. Attributes of the dataset are shown in Table *1*.

Table 1. Description of each data attribute

Attributes	Description
Review ID	Identification data of the review.
User name	Name of the reviewer.
User image	Profile photo of the reviewer.
Content	Review from the user.
Score	Rating value for the application by the user.
Thumbs up count	A number of other users approved or appreciated the user review.
Review created version	The version of the application when the user review was given.
At	The time and date when the review was given.
Reply content	Replies to user reviews from the application developer.
Replied at	The time and date when the developer's reply was given.
App version	The version of the application.

From this dataset attributes, only content and score attributes will be used for this research. Then, labelling data was conducted to split the dataset into binary class types with positive and negative value. The content attributes will be changed into comment attributes, and score attributes will be the parameter for applying value. Reviews with a score of 1 to 3 are labeled with “NEGATIF”, while scores 4 to 5 are labeled with “POSITIF”. Reviews with a rating value of 3 will be considered as reviews with negative sentiments. This is because the tendency of rating 3 reduces the trustworthiness of the product, and negative sentiment is more weighted in the neutral class [22], [23].

Preprocessing Data

Text preprocessing is the steps of processing raw text data into clean text data that can be learned by machines before the classification model-building process [17], [24]. The text preprocessing process is used to remove the noise so that clean text is obtained and easily learned by machine algorithms. Data preprocessing steps consist of case folding, tokenizing, filtering, and stemming.

a. Case folding

The Casefolding stage converts every word in the entire dataset to lowercase. This conversion is to generalize the words in the text to reduce the cases of re-detecting the same word due to different

upper and lower case letters in the word. In addition, characters in words other than alphabets and numbers will also be removed to create clean text. Unnecessary characters in the text are characters such as punctuation marks or symbols. Case folding is applied so that the data used is consistent [13], [17].

b. Tokenizing
The tokenizing stage works by dividing text and sentences into smaller parts that are word by word. Each of these word fragments is called a token. The separation of each word is done based on the space that separates each word. Word separation aims to facilitate text analysis because the word fragments represent an array of words owned by sentences in the dataset text [13].

c. Filtering
In the filtering stage, words that are considered stopwords will be ignored. Stopwords are words that have almost no semantic meaning so they will have no effect as a reference for the model to learn. By removing stopwords from the text, the analysis can be centered on important and relevant words only [13], [24].

d. Stemming
The stemming process aims to reprocess each word into its root or basic form. This process removes affix words such as prefixes and suffixes contained in the word. The stemming process is done to avoid word duplication due to words that have similar meanings but have different affixes [13].

Weighting and Splitting Data

Data weighting by Term Frequency Inverse Document Frequency of records (TF-IDF) is weighting the dataset to determine the relevance of each word in the text. TF-IDF is commonly used to find how often a word appears in the dataset text so that the relevance of the word in the text is known [13], [25]. After weighting, the dataset will be split into train data and test data with an 80:20 ratio, 80% for training data and 20% for testing data.

Sampling Approach

The sampling approach in this research will use an oversampling technique called Adaptive Synthetic (ADASYN). ADASYN is a sampling approach that generates synthetic data from minority data to overcome data imbalance [10], [11]. This imbalance is one of the problems in the classification process because it has a comparison of the amount of data with a large enough difference, data that has more numbers is called the majority class, while data that has fewer numbers is called the minority class. The oversampling technique is used to equalize the data class by increasing the number of minority class samples, the duplication of the minority class is usually done randomly [10]. This research uses ADASYN to handle the data imbalance in Jobstreet application reviews before running the classification test.

Classification Algorithm

This study will use the Support Vector Machine (SVM) algorithm for dataset classification. The SVM algorithm's initial goal is to overcome regression analysis and classification where classification is done by finding the best hyperplane as a separator between classes, the hyperplane itself is a line that becomes a separator function that separates the two classes where the distance between the line and the nearest object is called a margin [13], [14]. The formula to find a hyperplane can be seen in Equation 1.

$$y = f(x) = sgn(w \,.\, x + b) \tag{1}$$

Where:

y : the result of class classification predicted with the SVM algorithm
f : SVM algorithm decision function
x : input feature vector
w : word weight vector
b : bias
sgn : *signum*, function that produces +1 or -1 value based on the sign of the given argument

In this research, the Jobstreet application review is categorized as binary class data with only positive and negative reviews. The kernel will use a linear kernel to handle binary classification.

Model Evaluation

Model evaluation for this research emphasizes more on accuracy results. Conducting a method called confusion matrix evaluation will give data prediction for each class then count it into the proposed result. Accuracy measurement is used to measure the performance of the classification model of the research conducted. Accuracy in NLP is defined as the ratio of closeness between the actual data as a whole and the predictions given by the classification model. In addition to accuracy, to get information on model performance, other retrieval information can also be used, namely precision, recall, and f1 score. Precision is the degree of accuracy between the prediction model to actual information, recall is the success rate of retrieving information, and f1 score is the harmony between precision and recall. The accuracy calculation can be seen in Equation 2.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

Precision is the degree of accuracy between the prediction model to actual information. The precision calculation can be seen in Equation 3.

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

Recall is the success rate of the system in retrieving information. The recall calculation can be seen in Equation 4.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

The F1 score is the harmony between precision and recall. The f1 score calculation can be seen in Equation 5.

$$F1 - Score = \frac{2\ x\ R\ x\ P}{R+P} \tag{5}$$

In this binary classification research, confusion matrix classification produces 4 components, true positive (TP), false positive (FP), false negative (FN), and true negative (TN). These components were used to calculate the classification result in accuracy, precision, recall, and f1 score needed [26].

RESULTS AND DISCUSSION

To acquire the dataset needed in this research, the implementation of KDD is conducted. Dataset gained through a process called crawling using a Python library package named Google Play Scraper. The programs run computatically scrapping through Google Play reviews for the Jobstreet application. After this process was conducted, the obtained dataset had a total amount of 5,174 reviews with 11 attributes. Some samples from the dataset retrieved can be seen in Figure 2.

reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVer
5c902ced-640-41a6-900c-3cc0ea7ed	Iwan Iwan	https://play-lh.googleusercontent.com/a/ACg8oc...	Rekom banget	5	0	14.2.1	2024-05-12 13:05:17	None	NaT	1.
94e0dfe3-710-4189-93aa-21a4fee5b	agungs liar	https://play-lh.googleusercontent.com/a/ACg8oc...	Keamanankurang nomer pribadi tersebar kemana"	1	0	None	2024-05-12 12:15:49	Hi, saat ini aplikasi kami sudah memiliki fitu...	2024-05-13 09:11:32	!
f8ef-8cd0-460-9836-761107865	apolonius Nehe	https://play-lh.googleusercontent.com/a-/ALV-U...	Titip doa admin, supaya saya Mendapatkan peker...	5	0	14.2.1	2024-05-12 10:21:36	None	NaT	1.
1629ce4d-e17-4167-9089-da3492c8	Melankolis Sang Pengamat	https://play-lh.googleusercontent.com/a-/ALV-U...	Aplikasi penipu sudah kasih lamaran sebanyak-b...	1	0	14.2.1	2024-05-12 08:06:15	Hi, terima kasih atas reviewnya dan kami memin...	2024-05-13 09:11:53	1.
76139447-37f5-4b63-a30d-d0490d47	Andika Wiraputra	https://play-lh.googleusercontent.com/a-/ALV-U...	Nice	5	0	14.2.1	2024-05-12 07:52:46	None	NaT	1.

Figure 2. Dataset from the scrapping process

From this data, only 2 attributes are needed for the next steps, content and score attributes. The next step is labeling data, data reviews with scores 1 to 3 are labeled as “NEGATIF”, while data reviews with scores 4 and 5 are labeled as “POSITIF”. From this step, the dataset is known to have 3,873 positive reviews and 1,301 negative reviews.

The next step is preprocessing the dataset. Datasets gained from the previous process still have many objects that could hinder the classification result. Preprocessing is needed to create a clean dataset for the classification process. Four steps in text preprocessing need to be carried out, casefolding, tokenizing, filtering, and stemming. This whole step will create a clean dataset for the next step. Some comparison samples from the raw dataset into a clean dataset after preprocessing can be seen in Table 2.

Table 2. Dataset Comparison

Num	Raw Dataset	Clean Dataset
1	Banyak notifikasinya anjirr	notifikasi anjirr
2	Baru coba langsung kasih bintang 5	coba langsung kasih bintang 5
3	Terimakasih jobstreet,disini saya menemukan pekerjaan yg terbaikðŸ‘□ðŸ‘□	terimakasih jobstreetdisini temu kerja yg baik
4	Lowongan kerja tipu2 semua...ala negara konoha	lowong kerja tipu2 semuaala negara konoha
5	Mempermudah mendapatkan informasi lowongan pekerjaan. Terima kasih	mudah informasi lowong kerja terima kasih

After the preprocessing process is complete, the data will be converted into strings for comments and categories for values. The dataset was reduced to 4,139 reviews with 3,090 positive reviews and 1,049 negative reviews. This happens because preprocessing removes duplicate reviews to better clean the dataset. The dataset is then weighted in the TF-IDF process and then divided into two parts training data and test data. Split data ratio will use 80 percent for training data, and 20 percent for test data.

In one scenario, ADASYN will be applied before classification. The purpose is to balance the minority class by creating synthetic data. Before oversampling with ADASYN, it is known that there are 3,090 positive reviews and 1,049 negative reviews. In this dataset, negative reviews are categorized as a minority class. ADASYN works by balancing out between the majority class and minority class as can be seen in Figure 3.

```
Before Oversampling, counts of label 'POSITIF': 3090
Before Oversampling, counts of label 'NEGATIF': 1049

After Oversampling, the shape of train_X: (6145, 4202)
After Oversampling, the shape of train_y: (6145,)

After Oversampling, counts of label 'POSITIF': 3090
After Oversampling, counts of label 'NEGATIF': 3055
```

Figure 3. ADASYN program result

In the next step, the SVM classification method is applied to both scenarios, without ADASYN and using ADASYN. The SVM kernel used in this research is linear kernel since the dataset used is binary class. When running SVM classification without ADASYN, the result can be seen in Table 3.

Table 3. SVM classification without ADASYN

Accuracy	Precision	Recall	F1 score
89.08%	81.73%	71.03%	76%

Without using ADASYN, SVM classification for this Jobstreet reviews dataset results in an accuracy of 89.08%, precision of 81.73%, recall of 71.03%, and f1 score of 76%. In other scenarios, ADASYN is applied first before running classification using SVM. The classification result after oversampling using ADASYN can be seen in Table 4.

Table 4. SVM classification with ADASYN

Accuracy	Precision	Recall	F1 score
89.95%	77.20%	83.33%	80.15%

In this scenario, when applying oversampling using ADASYN, SVM classification shows result accuracy of 89.95%, precision of 77.20%, recall of 83.33%, and the F1 score of 80.15%. The results of SVM classification research with ADASYN show an increase in accuracy of 0.87%, an increase in recall of 12.30%, and an increase in the F1 score of 4.15%. Meanwhile, precision decreased by 4.53%.

Discussion

Several previous studies have shown an increase in the accuracy value of the classification algorithm after a sampling approach using ADASYN. The test results in this study show the accuracy value of the SVM algorithm before using ADASYN is 89.08 percent, while after using ADASYN the accuracy value is 89.95 percent. This shows the accuracy value of the SVM algorithm model has increased by 0.87 percent. In addition, the increase in recall and f1 score values shows that the SVM algorithm is better at predicting minority classes. This is because the evaluation of the f1 score value compares precision and recall so that it can be used to measure how well the classification algorithm performs [19]. The increase in accuracy value obtained in this study is fairly small compared to previous studies, with an increase of only 0.87 percent. However, based on the collection of research results, it still shows that ADASYN is quite effective in handling data imbalance and increasing the accuracy of the SVM algorithm even though it is used in different cases and datasets.

Several possibilities can be proposed in this research regarding the insignificant increase in accuracy value. The first possibility is that the preprocessing stage is still not clean enough to process the words in the dataset. This is related to the daily writing method made by users when giving reviews. Processing of slang words, subject and object words, and removal of stopwords or common sentences in the preprocessing stage affect the accuracy of the classification algorithm [27]. The second possibility lies in the accuracy results provided by the original data or data that does not use ADASYN. Although the accuracy is high, it cannot be fully trusted given the imbalance in the dataset. The imbalance in this dataset can affect the condition of the accuracy value because the classification algorithm tends to predict the majority class compared to the minority class, creating a bias due to the neglect of the minority class that can result in misclassification of the minority class [8], [9]. The increase in recall and f1 score in this study after using ADASYN shows the algorithm's sensitivity to minority classes so that the accuracy value can be more trusted.

CONCLUSION

This research was conducted to observe the effectiveness of ADASYN when used in the SVM algorithm classification model for data imbalance problems in the implementation of sentiment analysis of the Jobstreet application review dataset on the Google Play Store. After implementation, the results of the research conducted show an increase in the performance of the SVM classification algorithm when using ADASYN. In testing SVM performance, the accuracy value is 89.08%, while testing the SVM algorithm using ADASYN, the accuracy value is 89.95%. A comparison of the two analysis results shows an increase in accuracy value of 0.87% after ADASYN is used on the dataset. This shows that the ADASYN oversampling technique can improve the performance results of the SVM classification algorithm accuracy value in sentiment analysis research of Jobstreet application user reviews.

REFERENCES

[1] K. Fauziah and A. Triyanto, “The influence of Jobstreet.com toward the fulfillment of job vacancy information needs,” in Library Philosophy and Practice (e-journal), 2018. https://digitalcommons.unl.edu/libphilprac/2284

[2] Jobstreet, “Tentang Jobstreet,” 2024. https://www.jobstreet.co.id/id/about

[3] Google, “Jobstreet: Info Lowongan Kerja,” 2024. https://play.google.com/store/apps/details?id=com.jobstreet.jobstreet&hl=id&gl=id

[4] R. Obiedat, R. Qaddoura, A. M. Al-Zoubi, L. Al-Qaisi, O. Harfoushi, M. Alrefai, and H. Faris, “Sentiment Analysis of Customers’ Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution,” in Institute of Electrical and Electronics Engineers (IEEE), 2022. https://doi.org/10.1109/ACCESS.2022.3149482

[5] F. M. Sarimole and Kudrat, "Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," in Jurnal Sains Dan Teknologi, 2024, vol. 5, no. 3, pp. 783–790. https://doi.org/10.55338/saintek.v5i1.2702

[6] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," in Journal of the American Medical Informatics Association, 2011, vol. 18, no. 5, pp. 544-551. https://doi.org/10.1136/amiajnl-2011-000464

[7] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," in Multimedia Tools and Applications, 2023, vol. 82, no. 3, pp. 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

[8] C. Kaope and Y. Pristyanto, "The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance," in MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer, 2023, vol. 22, no. 2, pp. 227–238. https://doi.org/10.30812/matrik.v22i2.2515

[9] R. A. Nurdian, M. Ridwan, and A. Yusuf, "Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru," in Jurnal Teknik Informatika Dan Sistem Informasi, 2022, vol. 8, no. 1, pp. 24-32. https://doi.org/10.28932/jutisi.v8i1.4004

[10] F. S. Dhitama and F. A. Bachtiar, "Penentuan Kelayakan Debitur Menggunakan Metode Decision Tree C4.5 Dan Oversampling Adaptive Synthetic (ADASYN)," in Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2020, vol. 4, no. 10, pp. 3712–3721. https://doi.org/10.37148/bios.v3i1.36

[11] Y. Pristyanto, A. F. Nugraha, I. Pratama, A. Dahlan, and L. A. Wirasakti, "Dual Approach to Handling Imbalanced Class in Datasets Using Oversampling and Ensemble Learning Techniques," in Institute of Electrical and Electronics Engineers (IEEE), 2021. https://doi.org/10.1109/IMCOM51814.2021.9377420

[12] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning," in Institute of Electrical and Electronics Engineers (IEEE), 2008, pp. 1322-1328.

[13] F. M. Rizky, Jondri, and K. M. Lhaksmana, "Twitter Sentiment Analysis of Kanjuruhan Disaster using Word2Vec and Support Vector Machine," in Building of Informatics, Technology and Science (BITS), 2023, vol. 5, no. 1, pp. 219-227. https://doi.org/10.47065/bits.v5i1.3612

[14] D. Ismafillah, T. Rohana, and Y. Cahyana, "Implementasi Model Support Vector Machine dan Logistic Regression Untuk Memprediksi Penyakit Stroke," in JURIKOM (Jurnal Riset Komputer), 2023, vol. 10, no. 1, pp. 248–256. https://doi.org/10.30865/jurikom.v10i1.5478

[15] H. N. Irmanda and R. Astriratma, "Klasifikasi Jenis Pantun dengan Metode Support Vector Machines (SVM)," in Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 2020, vol. 4, no. 5, pp. 915–922.

[16] N. R. Setiawan and E. R. Kaburuan, "Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM," in Jurnal Sisfokom (Sistem Informasi Dan Komputer), 2023, vol. 12, no. 1, pp. 105–116. https://doi.org/10.32736/sisfokom.v12i1.1614

[17] A. Kharisma and I. Ernawati, "Sentimen Analisis Opini Masyarakat Jakarta Pada Kinerja Pemerintah Jakarta Terhadap Isu Tenggelamnya Jakarta Menggunakan Algoritma Support Vector Machine," in Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA), 2023, pp. 488–497.

[18] W. Hidayat, M. Ardiansyah, and A. Setyanto, "Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb," in Edumatic: Jurnal Pendidikan Informatika, 2021, vol. 5, no. 1, pp. 11–20. https://doi.org/10.29408/edumatic.v5i1.3125

[19] C. Magnolia, A. Nurhopipah, and B. A. Kusuma, "Penanganan Imbalanced Dataset untuk Klasifikasi Komentar Program Kampus Merdeka Pada Aplikasi Twitter," in Edu Komputika Journal, 2022, vol. 9, no. 2, pp. 105–113.

[20] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," in AI Magazine, 1996, vol. 17, no. 3, pp. 37–54. https://doi.org/10.37148/bios.v3i1.36

[21] M. Jaiswal and D. Patel, "Data Mining Techniques and Knowledge Discovery Database," in International Journal of Research and Analytical Reviews, 2015, vol. 2, no. 1, pp. 248-259.

[22] E. Bigne, C. Ruiz, C. Perez-Cabañero, and A. Cuenca, "Are customer star ratings and sentiments aligned? A deep learning study of the customer service experience in tourism destinations," in Service Business, 2023, vol. 17, no. 1, pp. 281–314. https://doi.org/10.1007/s11628-023-00524-0

[23] D. Jabbour, “3-Star Reviews Result In A -70% Decrease In Trust [Data Study],” 2023. https://gofishdigital.com/blog/3-star-reviews-result-in-70-decrease-in-trust-data-study/
[24] Ismatullah, F. Fauzi, and I. M. Nur, “Adaptive Synthetic Support Vector Machine Multiclass untuk mengklasifikasikan Imbalance data pada Sentimen kenaikan Bahan Bakar Minyak,” in Seminar Nasional Sains Data, 2023, pp. 304–312.
[25] F. Fauzi, Ismatullah, and I. M. Nur, “Unbalanced multiclass classification with adaptive synthetic multinomial Naive Bayes approach,” in Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Srodowiska, 2023, vol. 13, no. 3, pp. 64–70. https://doi.org/10.35784/iapgos.3740
[26] D. Y. Utami, E. Nurlelah, and F. N. Hasan, “Comparison of Neural Network Algorithms, Naive Bayes and Logistic Regression to predict diabetes,” in JITE (Journal of Informatics and Telecommunication Engineering, 2021, vol. 5, no. 1, pp. 53–64. https://doi.org/10.31289/jite.v5i1.5201
[27] S. Khomsah and A. S. Aribowo, “Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia,” in Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 2020, vol. 4, no. 4, pp. 648–654.