Sentiment Analysis on Shopee Product Reviews using Support Vector Machine

Foryasen, Blix D.
College of Computing and Information Technologies
National University – Manila
Manila, Philippines
foryasenbd@students.national-u.edu.ph

Mapanao, Jessy Cassandra M.
College of Computing and Information Technologies
National University – Manila
Manila, Philippines
mapanaojm@students.national-u.edu.ph

Lim, Jerico C.
College of Computing and Information Technologies
National University – Manila
Manila, Philippines
limjc@students.national-u.edu.ph

***Abstract*–The burgeoning e-commerce industry in the Philippines has led to a surge in online product reviews, providing a rich source of consumer sentiment. This research investigates the application of a Support Vector Machine (SVM) model to analyze sentiment in Filipino and English product reviews from Shopee. To enhance the model's accuracy, a robust preprocessing pipeline was implemented, including profanity filtering, n-gram extraction, and letter repetition reduction. The SVM model, fine-tuned through GridSearchCV, demonstrated superior performance in classifying sentiment, outperforming baseline models such as Logistic Regression and Naive Bayes by 2.1% in accuracy. While the model achieved promising results, limitations such as the lack of comprehensive Filipino lemmatization tools and challenges in handling complex sentiment expressions were identified. Future research should explore advanced NLP techniques such as BERT to address these limitations and further improve the accuracy of sentiment analysis in multilingual e-commerce contexts.**

***Keywords:* sentiment analysis, support vector machine, e-commerce, natural language processing.**

I. INTRODUCTION

The rise of e-commerce platforms has significantly transformed purchasing behaviors among Filipino consumers, allowing them to shop conveniently from their smartphones. Over the past five years, the Philippines has emerged as the world's second-fastest-growing e-commerce market, reflecting the country's rapid adoption of digital retail channels [1]. Shopee, the Philippines' leading e-commerce platform, has become a primary marketplace offering various products, including electronics, fashion, lifestyle goods, and more. However, despite the convenience and range of options that online shopping provides, challenges still need to be addressed, such as discrepancies with the descriptive information and actual product, poor quality of sold goods, and limited after-sales support [2].

Product reviews are vital in a consumer's purchasing intention, especially in the computer category, where most purchases tend to be significant financial investments. Although e-commerce platforms capture vast customer opinions, much of this data remains underutilized due to its unstructured nature [3]. This limitation leads to an incomplete picture of customer satisfaction and product issues, impacting businesses' ability to leverage this data effectively. Furthermore, consumers' reliance on feedback from other customers is affected when the available reviews lack clear organization or proper sentiment analysis, potentially influencing their confidence in purchasing decisions.

The use of sentiment analysis allows for the analysis of available data, capturing customer sentiment that can contribute to the objectives of the customers and enterprises [4]. This method identifies and extracts subjective information from large volumes of unstructured data by combining data mining techniques, machine learning, natural language processing, and information retrieval techniques [5]. This study employed a Support Vector Machine (SVM) algorithm to assess the sentiment of product reviews, effectively categorizing them to reveal consumer attitudes. Applying sentiment analysis to these reviews provides a comprehensive understanding of consumer experience and product information, enabling consumers to make informed purchasing decisions while guiding businesses to improve product quality and customer satisfaction

II. REVIEW OF RELATED WORKS

Sentiment analysis, a vital branch of Natural Language Processing (NLP), is a process that extracts emotional tones from textual data. In e-commerce, it plays a crucial role in analyzing user opinions expressed in product reviews on platforms like Shopee. Understanding these opinions allow businesses to gauge customer satisfaction, identify product strengths and weaknesses, and improve customer experience.

Customer insights gained through sentiment analysis unveil customers' emotions towards products, revealing valuable insights into satisfaction levels and areas for improvement [6; 7; 8]. By understanding customer sentiment, businesses can refine their marketing strategies to target customers' better needs [6]. Additionally, analyzing reviews enables businesses to identify recurring negative sentiments, allowing them to address potential product issues [6]. Besides SVM, sentiment analysis also utilizes various machine learning algorithms. Notable examples include Naive Bayes, which offers efficient classification but may perform less accurately with complex datasets [9], and logistic regression, which provides interpretable results but might not capture intricate sentiment patterns [9].

Support Vector Machine (SVM) has emerged as a powerful tool for sentiment analysis due to its ability to handle high-dimensional data and achieve strong performance [9; 10; 11]. In this context, this research explores an approach to preprocessing the data. First, profanity filtering is employed, where a predefined list of profane words is identified and replaced with a placeholder, thereby enhancing the quality of the text data and reducing noise that may affect sentiment analysis accuracy. Additionally, n-gram features are extracted from reviews, including unigrams (single words) and bigrams (two-word phrases), which can capture more nuanced sentiment than single words alone [12; 13; 14]. Furthermore, letter repetition reduction is implemented, where redundant letter repetitions are removed during text normalization. These techniques collectively aim to enhance the accuracy of SVM-based sentiment analysis for product reviews on platforms like Shopee.

III. METHODOLOGY

Implementation of the project followed most of the standard sentiment analysis pipeline from data collection to model evaluation. Deviations from the standard pipeline occurred during data pre-processing given that the analysis is multilingual. Each process is outlined and expounded as follows:

A. Data Collection

The data involved reviews scraped from Shopee, an e-commerce platform popular in Southeast Asia. The dataset, uploaded by an anonymous user on December 2, 2023, at Hugging Face, an open-source machine learning community, consists of 40,000 Filipino buyers' computer product reviews from the platform. Reviews contain a mix of Filipino and English languages, each labelled with their respective sentiment, with ‘0’ denoting a bad review and ‘1’ representing a good review. In total, there exist 20,000 bad reviews and 20,000 good reviews.

B. Data Pre-Processing

To ensure the quality, consistency, and readability of input data and help the model interpret sentiment accurately, several techniques for data cleaning, transformation, and feature extraction were employed to analyze Shopee computer product reviews, ensuring the data was suitable for training a classification model. With the dataset containing a mix of English and Filipino text and respective slang and profanities, the pre-processing pipeline is outlined as follows:

Initial Sanity Check

The dataset contained no null values and 1 duplicated value. Any null or duplicated values were dropped as they would not contribute to sentiment classification.

Profanity Filtering

Many reviews included offensive language or local slang that could influence sentiment without necessarily reflecting a negative opinion. A list of Filipino profanities served as a reference list for words to be replaced with a placeholder (“[PROFANITY]”), ensuring that context rather than specific words influenced sentiment classification.

Lowercasing

All text was converted to lowercase, minimizing the impact of capitalization on word distinctions and ensuring case consistency.

Stop Words Removal

To reduce noise, stopwords in both English and Filipino are removed. A custom stopword list combined English stopwords from NLTK and a curated list of Filipino stopwords. This step reduced the dimensionality of the data by focusing on content-carrying terms.

Reducing Repeated Characters

Filipino text, especially online reviews, often contain repeated characters for emphasis (e.g., “haaappy” instead of “happy”). By reducing consecutive repeating characters to their base form, the intended sentiment remains without inflating the term frequency.

Special Character, URL, and User Mention Removal

Non-alphabetic characters, such as punctuation, emojis, and numbers, were removed as they are typically irrelevant to sentiment and add unnecessary complexity. Moreover, URLs, email addresses, and usernames were also removed from the text.

Text Normalization

Words were standardized to a consistent format, addressing potential inconsistencies in spelling, grammar, and colloquial expressions across the dataset.

Lemmatization

Lemmatization reduces each word to its base or root form. This technique helped generalize words with the same meaning, minimizing vocabulary size and improving model interpretability. However, existing lemmatization does not support Tagalog words, thus limiting the model’s capability and accuracy for multilingual analysis.

TF-IDF Vectorization

The TF-IDF vectorizer extracted unigrams and bigrams to capture single words and word pairs that might indicate sentiment. Parameters were set to filter out terms appearing in over 90% or under 5% of the reviews, which helped remove both overly common and rare words. The custom stopword list was integrated into the TF-IDF vectorizer to refine the feature space further, ensuring these words were excluded during feature extraction. Such parameters helped to cover Tagalog words and phrases.

C. Experimental Setup

Google Colab was used as the primary IDE for the end-to-end methodology. Visual Studio Code served as a backup for running local hyperparameter optimization. To match the libraries used in Google Colab, a new Python environment was initialized containing the primary libraries matplotlib 3.8.0, pandas 2.2.2, scikit-learn 1.5.2, and nltk 3.8.1.

For model training and testing, the dataset went through an 80-20 split of training and testing set, respectively. Train data consists of 16032 bad reviews and 15967 good reviews, while test data has 3968 bad reviews and 4032 good reviews. Upon completing the preprocessing steps, the data contained 19431 features, denoting high-dimensionality of the whole dataset.

To find the most optimal model, hyperparameter tuning occurred revolving around a parameter grid consisting of the regularization parameter C, the kernel to be used, and gamma or the kernel coefficient. 5-fold cross-validation took place to find the best-performing model in terms of accuracy.

D. Algorithm

Support Vector Machines (SVM)

Support Vector Machines (SVMs) are supervised algorithms used for classification and regression tasks that identify a hyperplane to optimally separate labeled data points into positive and negative classes. By utilizing structural risk minimization and the kernel trick, SVMs can transform non-linear sample spaces into linear ones, effectively handling small datasets while ensuring strong adaptability and generalization capabilities [15; 16; 17]. The algorithm's performance is often comparable to other advanced methods, enhancing its appeal in machine learning [18; 19]. Below is how the hyperplane is mathematically represented and the hyperplane formula in a two-dimensional space [20].

$$\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \cdots + \beta_n * x_n = 0$$

Equation 1: *Hyperplane Representation*

$$\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 = 0$$

Equation 2: *Hyperplane in a two-dimensional space*

Where:

- β denotes each parameter corresponding to the various dimensions in our space.
- β_0 serves as the intercept.
- β_1 corresponds to the first axis, and so forth.
- X is a point that fulfills the equation above, it indicates that the point lies on the hyperplane.

E. Training Procedure

Grid Search Cross Validation (GridSearchCV) aided in model training through hyperparameter tuning. With the predefined hyperparameter grid and baseline model passed as parameters, the preprocessed training set is fitted into GridSearchCV to perform an exhaustive search of all possible combinations. For each combination, 5-fold cross-validation occurs to reduce the risk of overfitting and evaluate the average score of each combination. Upon completion of all combinations, GridSearchCV identifies the set of hyperparameters that returns the highest performance scores, which serves as the final model for training and testing.

F. Evaluation Metrics

For evaluation and comparison of the baseline models, accuracy, f1-score, precision, and recall are considered. While accuracy measures the correctness of a model's predictions, it is important to analyze how correct the model is in classifying bad reviews or good reviews. Precision focuses on how much the model correctly identifies the target label while recall shows how much a model identifies positive instances from the total positive cases. F1-score provides a balanced measure of both recall and precision. A model might have high accuracy, however, there are instances where the same model generally performs better in classifying one class compared to another. By comparing the evaluation metrics for each baseline model, the occurrence of accuracy illusion is reduced.

Upon selecting the best model, hyperparameter tuning and cross-validation through GridSearchCV prioritized accuracy, specifically training accuracy and testing accuracy. Doing so ensures that the model performs optimally in all instances of the dataset which limits overfitting for the SVM model.

G. Baselines and Comparative Models

Logistic Regression and Naive-Bayes, specifically MultinomialNB, served as baseline and comparative models for Support Vector Machines. To prove that SVM works better as compared to the baseline models [9], all models were fitted with their default parameters and compared based on the evaluation metrics. Even without hyperparameter tuning, SVM performed slightly better in all evaluation metrics, showing a minimal advantage in consistently classifying bad and good reviews.

IV. RESULTS AND DISCUSSION

Reviews from the dataset were preprocessed following the data-preprocessing steps stated in the methodology before being fitted into the models. Every model undertook iterative comparison and optimization based on the evaluation metrics to select the best-performing model. The best-performing model is then tested through a new set of data scraped from reviews from featured Shopee Computer Products.

A. Baseline and Comparative Models

Table 1. Classification Report for Base Models

Model	Sentiment	Prec	Rec	F1	Acc
Naïve-Bayes	Bad	0.95	0.98	0.96	0.9613
	Good	0.98	0.95	0.96	
Logistic Regression	Bad	0.96	0.97	0.97	0.9654
	Good	0.97	0.96	0.97	
Support Vector Machine	Bad	0.96	0.98	0.97	0.9693
	Good	0.98	0.96	0.97	

As observed in Table 1, the SVM model has the highest accuracy (0.9693 or 96.93%), followed closely by Logistic Regression (0.9654 or 96.54%) and Multinomial Naive Bayes (0.9613 or 96.13%). However, Logistic Regression and SVM models have slightly higher F1 scores than Naive Bayes for both classes, making them more robust for sentiment analysis. Moreover, SVM and Logistic Regression consistently achieve higher Precision and Recall than Naive Bayes, indicating their strength in distinguishing between classes without sacrificing either metric.

All three models demonstrate high performance across metrics, with accuracy values around 96–97%. This suggests that the dataset is likely well-prepared, and the features derived are relevant for sentiment classification. Given the minimal differences in metrics, the choice of model could be influenced by considerations like computational efficiency and model interpretability.

However, SVM's marginally higher accuracy and balanced performance across Precision, Recall, and F1 Scores proved to be the better choice for sentiment analysis compared to the other two models, proving past research on Naive Bayes and Logistic Regression's performance concerns on complex and high-dimensional datasets [4, 5, 6]. SVM, as the best baseline model, undertook further hyperparameter tuning. Logistic Regression might be considered should computational efficiency be a concern, as it generally requires less computational power than SVM

B. Hyperparameter Tuning of SVM Models

Table 2. Parameter Grid for GridSearchCV

Parameter	Values
C	0.1, 1, 5, 10, 100
Kernel	'linear', 'rbf'
Gamma	1, 0.1, 0.01, 0.001, 'scale', 'auto'

Given the computational power needed for training SVM, the parameter grid for hyperparameter tuning focused more on tweaking the C, kernel, and gamma values, yielding 70 combinations. Kernel is limited to linear and radial basis function (rbf) which serves as the default kernel for the baseline model to compare which kernel generally performs better for sentiment analysis.

Upon fitting the parameter grid and training data with 5-fold cross-validation, the following are the best three models per kernel.

Table 3. Top 3 Best Models per SVM kernel after GridSearchCV

C	Gamma	Train Acc	Test Acc	Acc Diff
rbf				
5	0.1	0.9901	0.9691	0.0210
5	'scale'	1.0000	0.9691	0.0309
5	1	1.0000	0.9690	0.0310
linear				
1	1	0.9894	0.9684	0.0210
1	0.1	0.9894	0.9684	0.0210
1	0.01	0.9894	0.9684	0.0210

The GridSearchCV results suggest that RBF and linear kernels are strong choices for sentiment analysis. The slight advantage of the RBF kernel with carefully tuned C values and gamma values indicates its suitability for capturing non-linear relationships, while the linear kernel offers a simpler alternative with comparable performance. The finding deviates from results on past research data citing that rhe linear kernel performs better compared to RBF when using high dimension data [21]. Given that maximum performance, reduced overfitting, and test accuracy are the primary goals, the RBF kernel with C = 5 and gamma = 0.1 was selected as the best model.

Table 4. *Classification Report for Base Models and Best Model*

Model	Sentiment	Prec	Rec	F1	Acc
Naïve-Bayes	Bad	0.95	0.98	0.96	0.9613
	Good	0.98	0.95	0.96	
Logistic Regression	Bad	0.96	0.97	0.97	0.9654
	Good	0.97	0.96	0.97	
Support Vector Machine	Bad	0.96	0.98	0.97	0.9693
	Good	0.98	0.96	0.97	
Best SVM	Bad	0.97	0.97	0.97	0.9714
	Good	0.97	0.97	0.97	

The best SVM model achieves near-optimal metrics across all evaluation categories. The Precision, Recall, and F1 Score for both "Bad" and "Good" sentiment classes are uniformly identical at 0.97, indicating that the model performs equally well in identifying both positive and negative sentiments. Along with boasting the highest accuracy and balanced performance across classes, the model is neither overfitting nor underfitting the data and is a reliable choice for analyzing customer feedback.

C. Evaluation of Best Model with New Data

To test the reliability of the model in encountering new data, reviews were scraped from the computer products category in Shopee. Five different products were chosen for model testing, selecting two reviews each, totaling 10 reviews. Each unique review is transformed and preprocessed before model prediction.

Table 4. Evaluation data and respective predictions

Review	Star Rating	Prediction
kvpal ka ba boss?! 3200mhz inorder ko tapos 2133mhz ibibigay mo?	3	Bad
A bit disappointed na unsealed na yung item, pero Ok na rin dahil gumana ana ang PC	4	Bad
Tang*na mo seller na budol ako di maayus puny*ta ka	1	Bad
pede na po hehe	5	Bad
Sayang excited pa naman ako Gamitin kaso may damage pala.	1	Bad
Delivered safe and working, but does not look durable. Plug cable too thin	3	Good
Putik damage ayaw na palitan. Kahit may unboxing video damage item kitang kita.	1	Bad
All working. All power **Pingkot yung box hindi iningatan. All box matters	4	Good
Sablay mag pisowifi ka nalang kesa dito	1	Bad
akala ko sira hindi gumana ang coinslot sinubukan ko lang ni reflash success gumana ng maayos maganda ang performance sana tumagal	5	Bad

Despite the model's high performance, for predicting sentiments, inconsistencies occur in predicting real-life reviews, especially those invoking contradicting statements, lacking context, or sarcasm. The misclassification may be attributed to lacking lemmatization and language support for Filipino words, leading the model to miss the whole context of the text. Overall, the model performed well in analyzing the sentiments of new input data.

V. CONCLUSION

Given the rapid growth of the e-commerce market in the Philippines, consumers rely heavily on reviews when making costly purchasing decisions, especially with electronics. However, unstructured reviews, which can be subjective and varied in sentiment, make it challenging for businesses and other consumers to draw clear insights. This study aimed to apply an optimized Support Vector Machine (SVM) model for sentiment analysis, effectively categorizing them to better understand customer satisfaction and product quality issues and enhance the preprocessing techniques to improve sentiment classification accuracy.

The approach addressed challenges to multilingual sentiment analysis, particularly linguistic challenges unique to Filipino and English text, through advanced data preprocessing techniques. Profanity filtering, n-gram extraction, and letter repetition reduction maximized the quality of the text data. The results showed that the SVM model with RBF kernel achieved high performance, significantly outperforming baseline models such as Logistic Regression and Naive Bayes. Through GridSearchCV hyperparameter tuning, the model demonstrated a balanced accuracy and reliability in classifying sentiments, even when faced with high-dimensional data. The findings underscore the SVM model's adaptability and robustness, making it an effective tool for sentiment analysis in e-commerce, especially within linguistically diverse markets, for targeted product quality and customer service improvements through data-driven decisions.

However, despite its strengths, the study faced several limitations. Notably, there exists lacked comprehensive lemmatization tools for Filipino, limiting the model's ability to capture nuanced local expressions fully. Additionally, the model occasionally misclassified complex sentiments, such as sarcasm or mixed sentiments, revealing areas where further linguistic fine-tuning could be beneficial, suggesting room for improvement in terms of contextual understanding.

Future work could explore fine-tuning advanced NLP models like BERT to capture better Filipino-English sentiment and expand the dataset to include other Southeast Asian languages. Integrating Filipino-specific lemmatization tools could further improve classification accuracy and contextual interpretation. Additional research could also focus on handling complex sentiments like sarcasm and applying this model to other e-commerce categories with diverse product types and customer demographics.

Support Vector Machines provides a replicable, scalable, and baseline approach to analyzing sentiment within multilingual e-commerce data, enhancing the quality of consumer insights for both customers and businesses. By demonstrating the power of SVM for sentiment classification in a real-world context, this study opens the door for broader applications across diverse markets, contributing to the ongoing evolution of customer feedback analysis in e-commerce.

REFERENCES

[1] Balita, "Topic: E-commerce in the Philippines," *Statista*, Oct. 02, 2024. https://www.statista.com/topics/6539/e-commerce-in-the-philippines/#:~:text=E%2Dcommerce%20landscape%20in%20the,its%20home%20and%20living%20items

[2] P. Ji, H.-Y. Zhang, and J.-Q. Wang, "A Fuzzy Decision Support Model With Sentiment Analysis for Items Comparison in e-Commerce: The Case Study of http://PConline.com," Oct. 01, 2019. https://ieeexplore.ieee.org/document/8515098.

[3] W. Muhammad, M. Mushtaq, K. N. Junejo, and M. Y. Khan, "Sentiment Analysis of Product Reviews in the Absence of Labelled Data using Supervised Learning Approaches," *ResearchGate*, 2020, doi: 10.22452/mjcs.vol33no2.3.

[4] S. Zad, M. Heidari, J. Jones, and O. Uzuner, "A survey on Concept-Level Sentiment Analysis Techniques of Textual data," *IEEE Conference Publication | IEEE Xplore*, May 10, 2021. https://ieeexplore.ieee.org/document/9454169

[5] S. Ainin, A. Feizollah, N. B. Anuar, and N. A. Abdullah, "Sentiment analyses of multilingual tweets on halal tourism," *ScienceDirect*, 2020, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211973620300258

[6] Rahul, Raj, and Monika, "Sentiment Analysis on Product Reviews," *SemanticScholar*, doi: 10.1109/ICCCIS48478.2019.8974527.

[7] Bose, Dey, Roy, and Sarddar, "Sentiment Analysis on Online Product Reviews," *ResearchGate*, 2019, doi: 10.1007/978-981-13-7166-0_56.

[8] Y. Jiang, H. Wang, and T. Yi, "Evaluation of Product Reviews Based on Text Sentiment Analysis," *ACM Digital Library*, May 2021, doi: 10.1145/3469213.3470379.

[9] Brownfield and Zhou, "Sentiment Analysis of Amazon Product Reviews," *SemanticScholar*, 2020, doi: 10.1007/978-3-030-63319-6_68.

[10] He, Zhou, and Zhao, "Exploring E-Commerce Product Experience Based on Fusion Sentiment Analysis Method," *IEEE Xplore*, 2022. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9919154

[11] Dubey and Jain, "Sentiment analysis of keenly intellective smart phone product review utilizing SVM Classification technique," *IEEE Conference Publication | IEEE Xplore*, Jul. 01, 2019. https://ieeexplore.ieee.org/document/8944795

[12] Trueman, Jayaraman, Cambria, and Ananthakrishnan, "An N-gram-Based BERT model for Sentiment Classification Using Movie Reviews," *ResearchGate*, 2022, doi: 10.1109/AIDE57180.2022.10060044.

[13] A. Zakaria and M. Siallagan, "Predicting Customer Satisfaction through Sentiment Analysis on Online Review," *International Journal of Current Science Research and Review*, Jan. 24, 2023. https://ijcsrr.org/wp-content/uploads/2023/01/56-24-2023.pdf

[14] Purbaya, Rakhmadani, Arum, and Nasifah, "Implementation of n-gram Methodology to Analyze Sentiment Reviews for Indonesian Chips Purchases in Shopee E-Marketplace," *Rek Ayas a Siste M Dan Tek N Ol Og I Inf Orm Asi*, 2023. https://www.researchgate.net/publication/371281311_Implementation_of_n-gram_Methodology_to_Analyze_Sentiment_Reviews_for_Indonesian_Chips_Purchases_in_Shopee_E-Marketplace

[15] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *ScienceDirect*, 2020, [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0925231220307153

[16] Q. Wang, "Support Vector machine algorithm in machine learning," *2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, Jun. 2022, doi: 10.1109/icaica54878.2022.9844516.

[17] Švec, "Vector," *Definitions*, Feb. 2020, doi: 10.32388/7uyp6e.

[18] Z. Jun, "The development and application of Support Vector Machine," *Journal of Physics Conference Series*, vol. 1748, no. 5, p. 052006, Jan. 2021, doi: 10.1088/1742-6596/1748/5/052006.

[19] R. A. Berk, "Support vector machines," in *Springer series in statistics*, 2008, pp. 1–28. doi: 10.1007/978-0-387-77501-2_7.

[20] Sirohi, "Support Vector Machine (Detailed Explanation) - towards data science," *Medium*, Dec. 11, 2021. [Online]. Available: https://towardsdatascience.com/support-vector-machine-support-vector-classifier-maximal-margin-classifier-22648a38ad9c

[21] Prastyo, Ardiyanto, and Hidayat, "Indonesian Sentiment Analysis: An Experimental Study of Four Kernel Functions on SVM Algorithm with TF-IDF," *IEEE Conference Publication | IEEE Xplore*, Oct. 26, 2020. https://ieeexplore.ieee.org/document/9325685?fbclid=IwZXh0bgNhZW0CMTEAAR2e3ghps8cV5vGnqkXjPN6atSYS-pMyxIu6m_5mVm-Ku57n3fxWqjFZ7R0_aem_nBmZrcolCiuqBNdPphXOzw