Universiti Teknologi MARA
Cawangan Melaka

In partnership with
Tadulako University

i-JaMCSIIX
2023

EXTENDED ABSTRACT BOOK

Publication Date: 30 March 2024
ISBN: 978-967-15337-0-3

https://jamcsiix.uitm.edu.my

i-JaMCSIIX 2023

INTERNATIONAL JASIN MULTIMEDIA & COMPUTER SCIENCE INVENTION AND INNOVATION EXHIBITION (I-JaMCSIIX) 2023

EXTENDED ABSTRACT

ISBN: 978-967-15337-0-3
i-JaMCSIIX
Universiti Teknologi MARA Cawangan Melaka Kampus Jasin 77300, Merlimau, Melaka

Web: https://jamcsiix.uitm.edu.my

In partnership with
Tadulako University

ORGANIZING COMMITTEE

PATRON	PM DR ISMADI MD BADARUDIN
ADVISOR I	TS DR JAMALUDDIN HJ JASMIS
ADVISOR II	DATO' DR MOHD HAJAR HASROL JONO
PROGRAM DIRECTOR	**DR. NUR SUHAILAYANI SUHAIMI**
DEPUTY DIRECTOR	TS DR NURUL HIDAYAH BINTI MAT ZAIN
SECRETARY I	**ANIS SHOBIRIN ABDULLAH SANI**
SECRETARY II	FAIQAH HAFIDZAH HALIM
TREASURER I	**SITI AISYAH ABD KADIR**
TREASURER II	UMMU MARDHIAH JALIL
PUBLICATION	**NURBAITY BINTI SABRI** DR. SITI FEIRUSZ AHMAD FESOL DR. AHMAD FIRDAUS BIN AHMAD FADZIL SITI NURAMALINA BINTI JOHARI ROSNIZA ROSLAN
JURY	**Ts DR. ALYA GEOGIANA BUJA** NORBAHIYAH AWANG Ts. DR. NOR AFIRDAUS ZAINAL ABIDIN DR. RAIHAH AMINUDDIN NOOR AFNI DERAMAN
REGISTRATION	**SITI FAIRUS BINTI FUZI** BUSHRA BINTI ABDUL HALIM NORDIANAH BINTI JUSOH@HUSSAIN AINON SYAZANA BINTI AB HAMID SITI NURSYAHIRA BINTI ZAINUDIN
SYSTEM	**FADILAH EZLINA SHAHBUDIN** HAJAR IZZATI MOHD GHAZALLI FADHLINA IZZAH SAMAN NOR AZIDA MOHAMED NOH
INVITATION AND PROMOTION	**SHAHITUL BADARIAH SULAIMAN** IZNI SYAMSINA SAARI NOR ADILA KEDIN

	ADI HAKIM BIN TALIB
	MOHD AMIRUL BIN ATAN
MULTIMEDIA	**Ts. NURUL NAJWA ABDUL RAHID@ABDUL RASHID**
	NOOR ASHITAH ABU OTHMAN
	ANWAR FARHAN ZOLKEPLAY
AWARD	**ANITA BINTI MOHD YASIN**
	NURUL EMYZA ZAHIDI
	FATIMAH HASHIM
	SITI RAMIZAH JAMA
	DR NURUL HUDA NIK ZULKIFLI
	MARIATHY BINTI KARIM
CERTIFICATE	**KHAIRUL NURMAZIANNA ISMAIL**
	NUR NABILAH ABU MANGSHOR
	ZUHRI ARAFAH ZULKIFLI
	HAZRATI ZAINI
INTERNATIONAL RELATIONS	**Ts. DR. SITI RAHAYU ABDUL AZIZ**
	ALBIN LEMUEL KUSHAN
	SHAHADAN SAAD
LIAISON OFFICER	**SYAFNIDAR ABDUL HALIM**
	AJK WAKIL UNTAD
SPONSORSHIP	**ANIS AMILAH SHARI**
	MOHD RAHMAT MOHD NOORDIN
	DR YUZAIMI YUNUS
	DR SURYAEFIZA KARJANTO
SECRETARIAT & APPRECIATION BANQUET	**RAIHANA MD SAIDI**
	NUR SYUHADA BINTI MUHAMMAT PAZIL
	ANIS AFIQAH SHARIP
	SITI MAISARAH MD ZAIN
	HAZWA HANIM MOHAMED HAMZAH

UNTAD'S COMMITTEE FOR I-JAMCSIIX 2023:

PROF. IR. MARSETYO, M.AG., PH.D.

PROF. I WAYAN SUDARSANA, S.SI., M.SI.

PROF. JUNAIDI, S.SI., M.SI., PH.D.

ELISA SESA, S.SI., M.SI., PH.D.

MUKRIM, M.ED., PH.D.

ZARKIANI HASYIM, S.PD., M.ED.

DR. HJ. ANI SUSANTI, M.SI.

DR. ISKANDAR, M.HUM.

DR. IR. ROIS., MP.

SYARIFUL ANAM, S.SI., M.SI., PH.D.

DR. NAHARUDDIN, S.PD, M.SI.

DR. DRG. ELLI YANE BANGKELE, M.KES.

HERMAN, SKM., M.MED.ED.

DR. IR. SAMLIOK NDOBE, M.SI.

DR. RAHMAT BAKRI, S.H., M.H.

DR. HAERUL ANAM, SE., M.SI.

DR. IR. BAKRI, S.T., PG. DIPL. ENG., M.PHIL.

DR. IR. MUHAMMAD YAZDI PUSADAN, S.KOM., M.ENG.

IR. SYAIFUL HENDRA, S.KOM., M.KOM.

RIZANA FAUZI S.T., M.T.

MOHAMMAD FAJRI, S.SI., M.SI.

NURUL FISKIA GAMAYANTI, S.SI., M.SI.

DR. NUR'ENI, S.SI., M.SI.

IMAN SETIAWAN, S.SI., M.SI.

FADJRIYANI, S.SI., M.SI.

LIST OF SPONSORS

External Company Sponsors

Klinik Dr Jamaluddin
Klinik Mawar Jasin
Nasi Ayam Ala Cina Zul
ADS Oasis Enterprise
Noorys Enterprise
Che Ramli bin Che Ismail
Beria Maju Enterprise
Rintiz rezeki
H&K food cafe
HS Gerak Wawasan

Individual Sponsors

En. Muhammad Hanif bin Abdul Aziz
Nor Suhaida binti Karjanto

Table of Contents

Universiti Teknologi MARA Cawangan Melaka
i-JaMCSIIX 2023
International Jasin Multimedia & Computer Science Invention and Innovation Exhibition
In Partnership with: Tadulako University

CRC – Clothing Review Classification using Sentiment Analysis

Nur Suhailayani Suhaimi[1], Anis Afiqah Sharip[2], Norazam Arbin[3], and Azyan Izzati[4]

[1,2,3,4] Universiti Teknologi MARA, Malaysia

suhailayani@uitm.edu.my, anis588@uitm.edu.my, noraz574@uitm.edu.my, azyanizzati19@gmail.com

***Abstract*—Safehse is a clothing brand established in 2020. Safehse clothing highlights the essence of South Korean streetwear and puts it under the spotlight. The customer text review is one of the features that Safehse have to play a role in helping the customer to make their purchasing decision. However, Safehse encountered a few problems with the current process of identifying genuine or fake reviews, time-consuming to classify positive and negative from the customer text review, and misleading or misunderstanding reviews that need to be clarified for the customers. In order to reduce and minimize the problem, Safehse needs to use the classification of text reviews by using sentiment analysis. This research project aims to develop a system to classify text reviews for Safehse, which can identify positive and negative reviews. The sentiment analysis technique used for this text review classification is supervised machine learning that anticipates occurrences by combining what it has learned from prior and current data with labels. The outcome of this text review classification is to display the categorized reviews with the calculated tokenization. As for the result of this project, the system will display the categorized review with the classification of positive and negative. The project will also display the genuine or fake review categories with the percentage of criteria from the data training. In average, more 75 % of sample data are correctly classified based on their pre-defined classes and more 55 % of data precisely categorized into fake or genuine label. For future enhancement of this project, a mobile application feature can be added to ease the text classification process quickly.**

Keywords—text mining, sentiment analysis, classification

I. INTRODUCTION

Safehse is a clothing brand, established in 2020. Safehse clothing highlights the essence of South Korean streetwear and putting it under the spotlight. Safehse keep their product's quality high and maintain it at a reasonable price. In addition, everyone can look trendy, stylish and feelcomfortable through their product. As for this project, it mainly focused on product text review for Safehse. The text review is oneof the features that Safehse website have. The text review plays the role to help customer to make their purchasing decision. Previously, text review was a section where every customer can write a review after purchasing the product. After that, the text review will be displayed onthe product page with no classification making it hard for customers to make their purchasing. Consumers have long relied on reviews from others before making purchasing decisions. In fact, Faridani (2021) the Forbes Councils. Member reveals that 82% of the consumers reading online reviews before buying a new product and more than 63% indicating they are more likely to buy from a site that has reviews. Due to reviews are written in free form on online platforms, the emotion of the users is difficult to explain. Certain words used in reviews have uncertain value of negative or positive opinion. There are reviews that are nicely written purposely to increase the sales even if the product isnot as good as its reviews. The value of ‘good’, ‘slightly good’ and ‘moderately good’ could have different reviews from different users.

II. MATERIALS

A. Dataset

This experiment used 6 sets of datasets from Safehse range from year 2019 until 2021. Three years (2 sets per year) of data comprises of all the reviews from customers regarding their clothes sold online. The data were divided into two categories which are 70 % of training data and remaining 30 % for testing.

B. User Requirements

The user requirements for this research consists of six (6) use cases which are 'Reviews Collection', 'Pre-processing Text Review', 'Tokenization', 'Classify Review', 'Generate Review Categories', 'Generate Review Visualization' and 'View Genuine or Fake Review'.

III. METHODS

Classifying text reviews of Safehse customer satisfaction in buying clothes using sentiment analysis involves the accomplishment of these task:

a. Data Collection: Gather a dataset of text reviews from customers who have purchased clothes. You can obtain this data from various sources such as e-commerce websites, social media platforms, or survey responses.

b. Data Pre-processing: Clean the text data by removing any irrelevant information like special characters, punctuation, and numbers. Normalize the text by converting it to lowercase and removing stop-words (commonly used words that do not carry significant meaning like "the," "is," etc.) Tokenize the text, splitting it into individual words or phrases (tokens).

c. Feature Extraction: Convert the text into numerical features that can be understood by machine learning algorithms. Common techniques include - Bag of Words (BoW); Represent each document as a vector where each dimension corresponds to the frequency of a term in the document. TF-IDF (Term Frequency-Inverse Document Frequency): Weigh the importance of each term in the document relative to its frequency across all documents in the corpus. Word Embeddings (e.g., Word2Vec, GloVe): Represent words as dense vectors in a continuous vector space, capturing semantic relationships between words.

d. Model Selection: Run on a LTSM and SVM machine learning or deep learning model for sentiment analysis. The categorizing includes positive reviews and negative reviews. All neutral reviews smoothed into positive category.

e. Training the Model: Split the dataset into training and testing sets to evaluate the performance of the model. Train the chosen model on the training set using the extracted features. Fine-tune the model parameters if necessary to improve performance.

f. Evaluation: Evaluate the trained model on the testing set using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. Analyze the model's performance and iterate on the process if needed, such as trying different pre-processing techniques, feature extraction methods, or model architectures.

g. Deployment: Once satisfied with the model's performance, deploy it to classify new text reviews of customer satisfaction in buying clothes. Integrate the model into Safehse application to automate the sentiment analysis process for incoming customer reviews.

h. Monitoring and Maintenance: Regularly monitor the performance of the deployed model and retrain it periodically with new data to maintain its accuracy and relevance over time. Address any issues or drift in performance that may arise due to changes in the data distribution or customer preferences.

IV. RESULTS AND FINDINGS

Accuracy, precision, and F-score are all metrics used to evaluate the performance of sentiment analysis models, but they measure different aspects of the model's performance. Accuracy measures the proportion of correctly classified instances out of the total number of instances. Precision gives an indication of how many of the positively classified instances are actually positive. In the context of sentiment analysis, precision tells us the proportion of correctly identified positive reviews out of all reviews predicted as positive. The F-score balances both precision and recall, providing a single metric that captures the model's overall performance. It's particularly useful when you want to find a balance between precision and recall, especially in scenarios where you cannot prioritize one over the other. In sentiment analysis, a high precision indicates that when the model predicts a sentiment (positive or negative), it's likely correct. A high recall indicates that the model is capturing most of the positive or negative sentiments present in the data. The F-score combines these two metrics, giving a comprehensive evaluation of the model's performance. The acceptable level of accuracy for sentiment analysis in clothing reviews can vary depending on several factors, including the specific application, the desired level of confidence, and the distribution of sentiments in the dataset. However, generally speaking, a high level of accuracy for sentiment analysis in clothing reviews would typically be considered to be above 80%. If the dataset contains a significant imbalance in sentiment distribution (e.g., there are far more positive reviews than negative reviews or vice versa), achieving high accuracy can be more challenging. In such cases, other metrics like precision, recall, and F-score become more important to consider alongside accuracy. Table 1 shows the accuracy results for all six datasets of reviews in Safehse data warehouse.

Table 1. Accuracy Result for Reviews Classification

Dataset	Accuracy Results for Reviews 2019-2021		
	Description	***F-Score (%)***	***Precision (%)***
2019A	Reviews for Jan-Jun2019	71.5	80.3
2019B	Reviews for Jul-dec2019	74.2	86.4
2020A	Reviews for Jan-Jun2020	81.3	88.9
2020B	Reviews for Jul-dec2020	84.9	90.4
2021A	Reviews for Jan-Jun2021	88.6	91.3
2021B	Reviews for Jul-dec2021	85.0	89.9

In average, an F-score of 81% suggests that the model achieves a good balance between precision and recall. The F1-score is the harmonic mean of precision and recall, so achieving 81% indicates that the model performs well in both correctly identifying positive instances (precision) and capturing most of the actual positive instances in the data (recall). All precision above 80 % means that out of all instances predicted as positive (whether positive reviews or sentiments), 80% of them are actually positive. This indicates that the model is relatively conservative in classifying instances as positive, ensuring that the majority of positive predictions are accurate. In many sentiment analysis applications, achieving an F-score of 81% and a precision of 80% would be considered quite satisfactory. However, the interpretation of these metrics can also depend on the specific requirements and constraints of the application. Ultimately, the acceptable level of accuracy for sentiment analysis in clothing reviews should be determined based on a thorough understanding of the application's requirements, the characteristics of the dataset, and the trade-offs between different evaluation metrics.

V. CONCLUSIONS

Sentiment analysis helps in understanding the sentiment behind customer reviews. By analyzing the language used in reviews, businesses can categorize them as positive, negative, or neutral. This understanding allows businesses to gauge overall customer satisfaction with their products, services, or shopping experience. Insights derived from sentiment analysis can inform product development decisions. Positive feedback can highlight features or aspects of products that customers particularly enjoy, guiding future product enhancements or expansions. Conversely, negative feedback can signal areas for improvement, helping companies prioritize issues to address in subsequent product iterations. Sentiment analysis allows businesses to enhance the overall customer experience. By identifying pain points or areas of dissatisfaction, companies can take corrective actions to improve customer service, streamline the ordering process, or optimize website usability. Addressing these concerns can lead to higher customer satisfaction and loyalty. Understanding how customers perceive competing products or services can provide valuable insights into market positioning, strengths, and weaknesses. This information can inform competitive strategies and help businesses differentiate themselves in the market. By continuously tracking customer feedback, businesses can promptly respond to emerging issues, engage with customers, and mitigate negative publicity. This proactive approach demonstrates responsiveness and commitment to customer satisfaction. Overall, while there is no universally "good" accuracy threshold for sentiment analysis, an F-score of 81% and a precision of 80% indicate that the model performs well and can be considered quite satisfactory in many practical applications.

REFERENCES

[1] A. Abdul Aziz & A. Starkey (2020). Predicting supervise machine learning performances for sentiment analysis using contextual-based approaches. IEEE Access, 8, 17722–17733. https://doi.org/10.1109/access.2019.2958702

[2] J. R. Alharbi & W. S. Alhalabi (2020). Hybrid approach for sentiment analysis of Twitter posts using a dictionary-based approach and Fuzzy Logic Methods. International Journal on Semantic Web and Information Systems, 16(1), 116– 145. https://doi.org/10.4018/ijswis.2020010106

[3] A. Alsaeedi & M. Zubair (2019). A study on sentiment analysis techniques of Twitter data. International Journal of Advanced Computer Science and Applications, 10(2). https://doi.org/10.14569/ijacsa.2019.0100248

[4] A. A. Alsolamy, M. A. Siddiqui & I. H. Khan (2019). A corpus based approach to build Arabic sentiment lexicon. International Journal of Information Engineering and Electronic Business, 11(6), 16–23. https://doi.org/10.5815/ijieeb.2019.06.03

[5] M. E. Basiri & A. Kabiri (2018). Words Are Important. ACM Transactions on Asian and Low-Resource Language Information Processing, 17(4), 1–18. https://doi.org/10.1145/3195633

[6] M. Birjali, M. Kasri & A. Beni-Hssane (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and Trends. Knowledge-Based Systems, 226, 107134. https://doi.org/10.1016/j.knosys.2021.107134

[7] V. Bonta, N. Kumaresh, & N. Janardhan (2019). A comprehensive study on lexicon based approaches for sentiment analysis. Asian Journal of Computer Science and Technology, 8(S2), 1–6. https://doi.org/10.51983/ajcst-2019.8.s2.2037

i-JaMCSIIX 2023

PUBLISHED BY:
i-JaMCSIIX
Universiti Teknologi MARA Cawangan Melaka
Kampus Jasin
77300 Merlimau, Melaka

Tel: 062645000
Email: jamcsiix@uitm.edu.my
Web: https://jamcsiix.uitm.edu.my/