# PRODUCT QUALITY CLASSIFICATION ON SHOPEE THROUGH USER REVIEW SENTIMENTS

EDGAR AARON A. GO AND RODOLFO C. CAMACLANG III

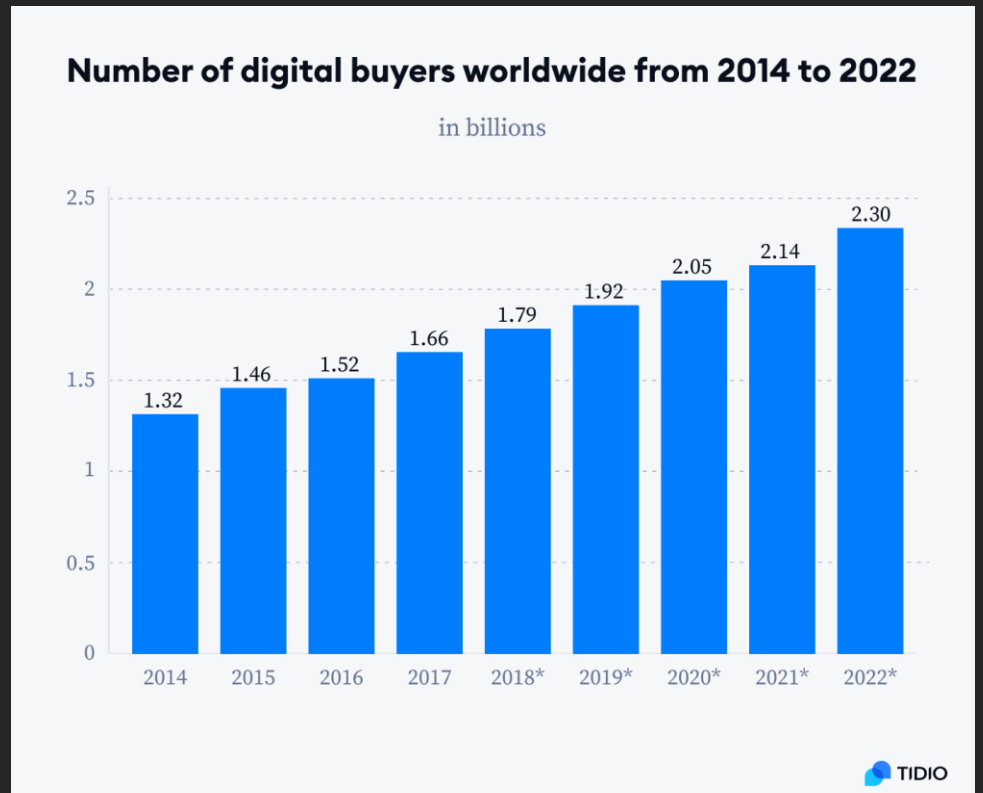
# E-COMMERCE

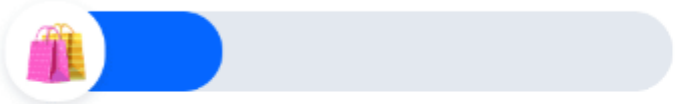
*the conduct of sale, purchase, transfer, or exchange of products, services, and information through electronic means and technologies*

There are over **2.14 billion online shoppers** in the world

**+4.4%** increase between 2020/2021

**27% of global** population shop online 🛒🌎



20% of all retail sales **happen online**

By 2024 **global online sales** are estimated to reach more than $7 trillion 😯💸

There are **more than 12 million** online stores in the world

Image Source:
https://www.tidio.com/blog/online-shopping-statistics/

# FRAUD

*any invitation, request, notification, or offer, that is spread across the Internet to obtain a victim's information or money, or otherwise obtain a financial benefit through deceptive mean*

C. Cross, R. Smith, and K. Richards, "Challenges of responding to online fraud victimization in Australia," Trends and Issues in Crime and Criminal Justice, vol. 474, 2014.
A. Abdallah, M. Maarof, and A. Zainal, "Fraud detection system: A survey," Journal of Network and Computer Applications, vol. 68, pp. 90–133, 2016

# FRAUD

*any invitation, request, notification, or offer, that is spread across the Internet to obtain a victim's information or money, or otherwise obtain a financial benefit through deceptive mean*

## Classifications of E-COMMERCE FRAUD according to IC3:

1. Non-delivery of goods
2. Misrepresentation of items
3. Poor quality control and negligence
4. Triangulation
5. Fee staking
6. Selling of black-market goods
7. Multiple bidding or shill bidding

C. Cross, R. Smith, and K. Richards, "Challenges of responding to online fraud victimization in Australia," Trends and Issues in Crime and Criminal Justice, vol. 474, 2014.
A. Abdallah, M. Maarof, and A. Zainal, "Fraud detection system: A survey," Journal of Network and Computer Applications, vol. 68, pp. 90–133, 2016

# USER REVIEWS


Customer reviews are an effective way to learn about a product's quality and features.

Online product reviews can help influence the buying decisions of potential customers.

Feedback from real consumers can shed light on the way a product or service performs in real life

Reviews can be helpful for learning about problems with a given product or shop. Knowing about potential problems ahead of time can help you avoid them in the future

Image Source: https://www.zoho.com/commerce/insights/importance-of-product-reviews.html

# THE PROBLEM

In the rapidly expanding realm of e-commerce, the proliferation of e-commerce fraud targeting unsuspecting consumers poses a significant challenge. Leveraging sentiment analysis on user reviews presents a promising approach to discerning the authenticity of online offerings through customer satisfaction.

# OBJECTIVES

The general objective of the study is to develop a Chrome browser extension that uses TF-IDF to create vectors from preprocessed user reviews to train a logistic regression model. This is used to classify the quality of a product accurately from Shopee through sentiment analysis of the reviews

# OBJECTIVES

The general objective of the study is to develop a Chrome browser extension that uses Word2Vec to create vectors to be used to train a logistic regression model. This is used to classify a product accurately from an ecommerce platform in its legitimacy through sentiment analysis from user reviews.

SPECIFIC OBJECTIVES:

1. Annotate training data manually given a set of criteria
2. Preprocess and create vectors using TF-IDF from the annotated data to train a logistic regression model
3. Develop an application to perform API requests to retrieve user reviews with text from a Shopee product page
4. Develop a Chrome extension to automate product classification based on the results of the model; and
5. Evaluate the performance of the model and extension given the evaluation metrics.

# METHODOLOGY – TECHNOLOGIES USED

PYTHON


SCIKIT-LEARN

- LOGISTIC REGRESSION
- TF-IDF VECTORIZER
- METRICS

FLASK

# PROPOSED METHODOLOGY – MODEL TRAINING PROCESS

DATA GATHERING

DATA ANNOTATION

DATA PREPROCESSING

MODEL TRAINING AND SEMANTIC ANALYSIS

‘text’
user review text

‘label’
star rating
accompanying review

## SHOPEE REVIEWS DATA SET (TWO COLUMNS)

https://huggingface.co/datasets/scaredmeow/shopee-reviews-tl-stars/viewer/default/train

| | text | label |
|---|---|---|
| 2 | A cute niya umiilaw nman siya. I will order again para r | 4 |
| 3 | First time order sa shop nato at di ko inakalang Ang Ga | 4 |
| 4 | Madaling masira yung sa may sinisintasan nya. Wala r | 2 |
| 5 | Maganda naman sya mas malaki lang sa ineexpect ko | 4 |
| 6 | All goods na tester q 1. 2v ngalang pro ok Naman char | 4 |
| 7 | hindi sya anti rad mas lalong masisira mata nyo. blurre | 0 |
| 8 | Malamig yung hangin kahit hindi malamig na tubig o wa | 4 |
| 9 | sakto lang, cute sya pero hindi ganon katagal yung bat | 2 |

*13,000 total reviews for training*
*2,500 total reviews for testing*
*(two separate csv files)*

DATA GATHERING

DATA ANNOTATION

DATA PREPROCESSING

MODEL TRAINING AND SEMANTIC ANALYSIS

In the process of annotating the training dataset, two respondents, which are both university students who have a proficiency in the English and Tagalog language, along with the researcher independently provided annotations for each data point as inaccuracies may exist from the ratings. Reviews are annotated based a criteria.

The annotations from the respondents and the researcher were compared, and the majority decision was chosen as the final annotation for each data point. A total of 6,632 positive reviews and 6,368 negative reviews were annotated in the final dataset.

DATA GATHERING

DATA ANNOTATION

**DATA PREPROCESSING**

MODEL TRAINING AND SEMANTIC ANALYSIS

**Data Preprocessing Steps:**

- **Case conversion: The reviews are transformed into lowercase.**
- **Removal of non-alphanumeric characters: Alphanumeric characters are retained. The removed characters are replaced with a single space.**
- **Removal of whitespace: Tabs, newlines, and multiple spaces are replaced into a single space.**
- **Removal of stop words: Words such as "the", "are", and "is" are removed from the list of tokens. A Tagalog stop word list was also used.**
- **Tokenization: TF-IDF is used to convert the reviews into tokens with vector representation.**

DATA GATHERING

DATA ANNOTATION

DATA PREPROCESSING

**MODEL TRAINING AND SEMANTIC ANALYSIS**

According to the survey of Shahid et al. (2022), the logistic regression model has the highest mean accuracy from other binary classification models. This model was used for this paper along with a TF-IDF vectorizer.

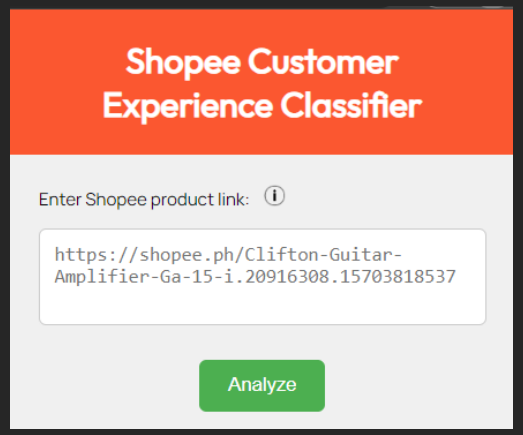
Different vectorization techniques and the support vector machine model was also used to compare performances of the different combinations.

The preprocessed data will be used to train the model and the testing data set will be used to determine several evaluation metrics of the model, such as accuracy, precision, and recall.

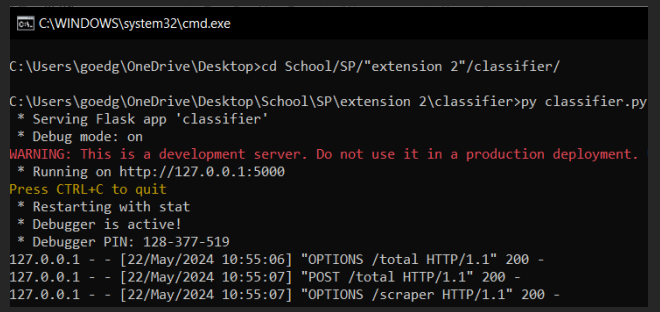
Once trained, the Chrome extension  will gather reviews off a product page in Shopee. These will be preprocessed and classified through the model.

M. Shahid, A. Chaudhary, and . K. A. Gupta D., "Review based rating prediction using machine learning techniques," 2022 11th International Conference on System Modeling Advancement in Research Trends (SMART), pp. 118–122, 2022

# METHODOLOGY – FLOWCHART



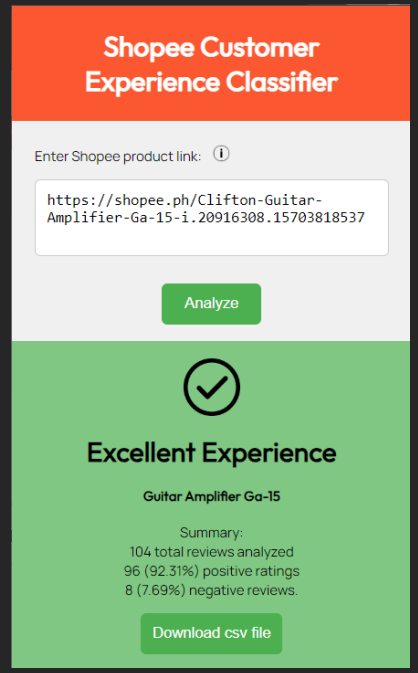

CHROME EXTENSION




FLASK SERVER

API REQUESTS TO RETRIEVE REVIEWS FROM URL

REVIEWS ARE PREPROCESSED AND CLASSIFIED THROUGH THE MODEL AND SENDS RESULT TO EXTENSION




EXTENSION DISPLAYS RESULT

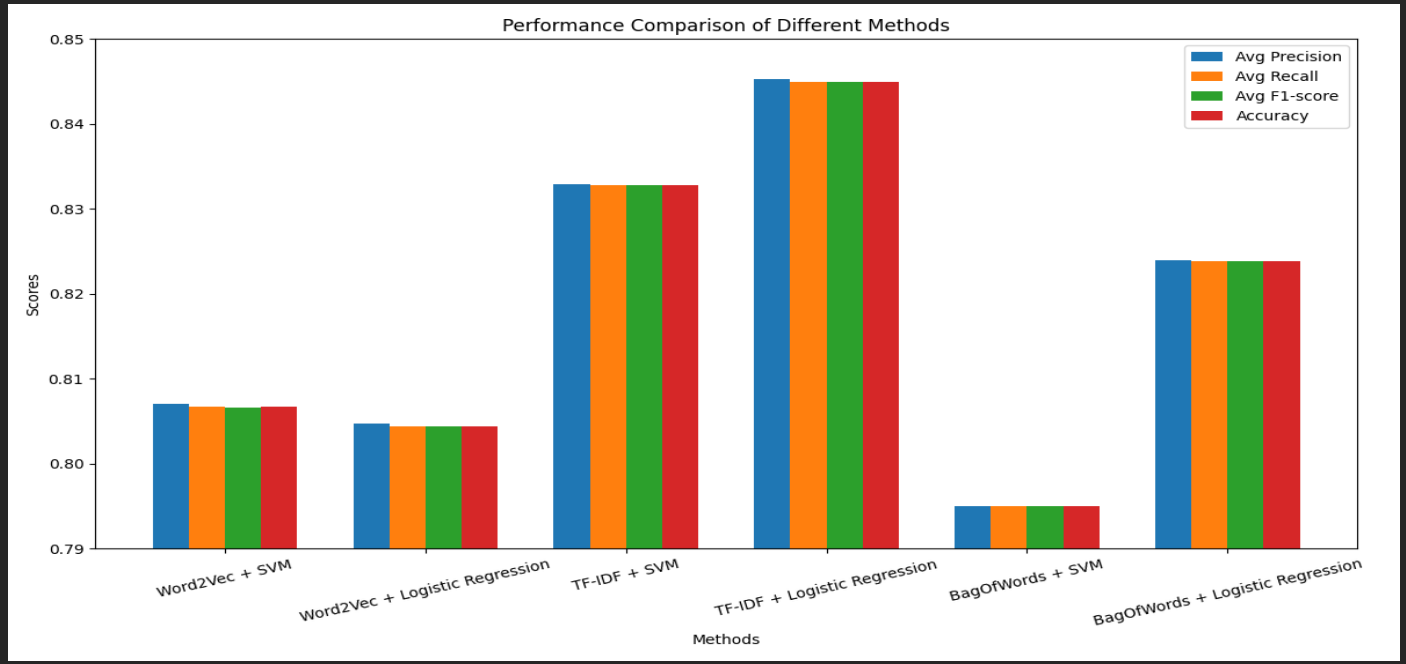
| | |
|---|---|
| performance ok product quality ok ok nmn medjo buzz kaso ok nmn | 1 |
| best feature okay sya pag wala kana talagang magamit product quality it really so | 1 |
| wala akong video maganda yung amp solid malakas maganda bou yung tunong la | 1 |
| pde nag ccmula mag aral mag electric guitar mabagal lang shipping panget cour | 0 |
| maganda kaso basag top right lower right | 0 |
| the speaker terminal is loose because the rivet broke fixed with a glue | 0 |
| best feature not good product quality not good performance not good buzz sayan | 0 |
| product quality amazing best feature excellent performance wondeful | 1 |
| best feature good performance good product quality good good amp | 1 |
| product quality excellent ganda product malinis yung tunog | 1 |
| performance good best feature good product quality good | 1 |
| performance best feature product quality | 1 |
| best feature ok product quality ok performance ok good | 1 |
| all goods po boss sulit yung boss what you see is what you get tlga | 1 |

CSV FILE OF RESULT CAN BE DOWNLOADED

# RESULTS AND DISCUSSION – PERFORMANCE EVALUATION

**TABLE I:** Performance Metrics for Different Methods

| Method | Avg Precision | Avg Recall | Avg F1-score | Accuracy |
|---|---|---|---|---|
| W2V + SVM | 0.80701 | 0.80667 | 0.80661 | 0.80667 |
| W2V + LR | 0.80470 | 0.80444 | 0.80440 | 0.80444 |
| TFIDF + SVM | 0.83293 | 0.83278 | 0.83276 | 0.83278 |
| TFIDF + LR | 0.84523 | 0.84500 | 0.84497 | 0.84500 |
| BoW + SVM | 0.79502 | 0.79500 | 0.79500 | 0.79500 |
| BoW + LR | 0.82400 | 0.82389 | 0.82387 | 0.82389 |



# RESULTS AND DISCUSSION – USABILITY SURVEY

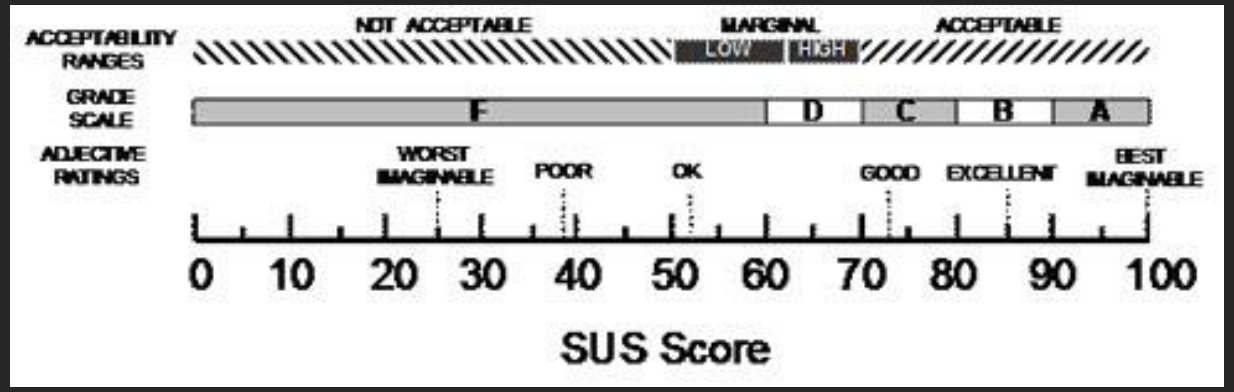
## SYSTEM USABILITY SCALE SURVEY

1. I think that I would like to use this extension frequently.
2. I found the extension unnecessarily complex.
3. I thought the extension was easy to use.
4. I think that I would need the support of a technical person to be able to use this extension.
5. I found the various functions in this extension were well integrated.
6. I thought there was too much inconsistency in this extension.
7. I would imagine that most people would learn to use this extension very quickly.
8. I found the extension very cumbersome to use.
9. I felt very confident using the extension.
10. I needed to learn a lot of things before I could get going with this extension

# RESULTS AND DISCUSSION – USABILITY SURVEY

| Respondent | SUS Score |
|---|---|
| 1 | 70 |
| 2 | 90 |
| 3 | 60 |
| 4 | 80 |
| 5 | 90 |
| 6 | 70 |
| 7 | 85 |
| 8 | 70 |
| 9 | 72.5 |
| 10 | 60 |
| 11 | 75 |
| 12 | 75 |
| 13 | 85 |
| 14 | 85 |
| 15 | 80 |
| 16 | 80 |
| **Avg SUS Score** | **76.71875** |

**TABLE II:** User Scores and SUS Score




A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," Journal of User Experience, 2009

# CONCLUSION

The main objective of this study is to develop a Chrome browser extension that uses TF-IDF to create vectors to be used to train a logistic regression model in order to classify a product accurately from Shopee in its legitimacy through sentiment analysis from user reviews. Based on the findings, the model is able to outperform different combinations of text vectorization methods and classifiers and it is able to achieve an accuracy of 84.5%, average precision of 84.52%, average recall of 84.5% and an average F1-score of 84.49%.

In terms of usability, the system usability scale (SUS) surveys revealed that users found the extension generally easy to use, well-integrated, and not overly complex. The SUS scores averaged to 76.72, which falls between the ”good” and ”excellent” categories, indicating that the extension is well- received by users and meets usability standards

# REFERENCES


https://www.zoho.com/commerce/insights/importance-of-product-reviews.html

https://www.tidio.com/blog/online-shopping-statistics/

C. Cross, R. Smith, and K. Richards, “Challenges of responding to online fraud victimization in Australia,” Trends and Issues in Crime and Criminal Justice, vol. 474, 2014.

A. Abdallah, M. Maarof, and A. Zainal, “Fraud detection system: A survey,” Journal of Network and Computer Applications, vol. 68, pp. 90–133, 2016

M. Shahid, A. Chaudhary, and . K. A. Gupta D., “Review based rating prediction using machine learning techniques,” 2022 11th International Conference on System Modeling Advancement in Research Trends (SMART), pp. 118–122, 2022

A. Bangor, P. Kortum, and J. Miller, “Determining what individual sus scores mean: Adding an adjective rating scale,” Journal of User Experience, 2009