SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Intelligence Shopee Product Comparison (i-SPC) and Visualization of Product Information via Naïve Bayes Adaptation

Khyrina Airin Fariza Abu Samah[1,*], Nurul Syuhada' Abd Raub[1], Lala Septem Riza[2], Hana Fakhira Almarzuki[3], Nurazian Mior Dahalan[1], Ahmad Firdaus Ahmad Fadzil[1]

[1] Computing Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Melaka Branch, Melaka, Malaysia
[2] Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[3] Computing Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Shah Alam, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 22 June 2023
Received in revised form 5 September 2023
Accepted 29 October 2023
Available online 9 January 2024

Keywords:
E-Commerce; i-SPC, Shopee; Naïve Bayes; Data visualization

ABSTRACT

Electronic Commerce (E-Commerce) is a type of commerce that takes place online. The most used platform based on frequently visited in Malaysia is Shopee. According to a questionnaire survey of 102 respondents, 95.1% agreed that manually comparing Shopee product information takes time. Manual analyzing a group of similar products is notoriously complicated, and finding informative reviews for product purchases is becoming increasingly challenging. This study aims to obtain Shopee information from the real-time Shopee website. Hence, Intelligence Shopee Product Comparison (i-SPC), aims to design a web-based application system that compares Shopee product information from different shops using the Naïve Bayes algorithm. The user can copy and paste the chosen Shopee product link to a maximum of ten links for comparison. The i-SPC displays the information based on seven focused factors and categorizes whether the pasted link is "recommended" or "not recommended". The visualization result uses a bar chart to show four types of information: shop rating, product price, followers, and chat response. Testing phases have proven that the classifier accomplished all the research's objectives and successfully classified Shopee product information with 87.50% accuracy, which is considered "good". All test cases for the functionality test proved that the i-SPC successfully solved the problem. Therefore, it can be concluded that i-SPC overcame the problem and improved the product comparison process.

1. Introduction

As information technology advances, the Internet has become essential and beneficial in various fields, including E-Commerce, Finance, Business, Social Networking, and Marketing. Purchasing and selling products or services over the Internet is known as electronic commerce (E-Commerce) [1]. Putri & Pujani [2] pointed out that the Internet is a unified platform connecting buyers and sellers. It is crucial in promoting the growth of E-Commerce [3].

* *Corresponding author.*
E-mail address: khyrina783@uitm.edu.my

https://doi.org/10.37934/araset.37.1.179190

Budyastuti [4] stated that E-Commerce, also known as Ecom, Emmerce, or EC, is a routine business exchange. It communicates via Electronic Data Interchange, e-mail, and electronic bulletin boards. E-Commerce refers to a wide range of technologies, processes, and practices that enable transactions to be completed without using paper as a transaction medium. It has benefited many customers worldwide by allowing them to buy and sell their products on various websites available on the online platform. Customers can find everything they need under one roof using the E-Commerce platform [5-6]. Shopping is thus more convenient than the traditional method, which requires customers to manually visit each local store and search for the desired product at the best price.

Shopee, Lazada, Lelong, and Zalora are Malaysia's most popular E-Commerce platforms. According to Kiew *et al.* [7], Shopee ranks second among Malaysia's top 50 E-Commerce platforms regarding average quarterly traffic, social media followers, mobile app rankings, and employee numbers. Previous studies by Daryanto *et al.* [8] have reported that Shopee is an E-Commerce platform founded in 2009 and headquartered in Singapore under the SEA Group. Shopee operates in Singapore, Malaysia, Thailand, Taiwan, Indonesia, the Philippines, Vietnam, and Brazil [9]. Shopee is also available as a seller center in mainland China, Hong Kong, Japan, and South Korea. Shopee can be used to buy pulses, electricity tokens, and meal vouchers in addition to selling clothing, cosmetics, mobile devices, and electricity tokens [3].

Shopee does not prioritize user satisfaction with the website's interface and functional performance due to strong competition from other E-Commerce websites. When purchasing items online, users rely solely on the descriptions and images provided by the website [2]. As pointed out by Kivist [10], product information influences a website user's experience. Despite the filters provided by Shopee, the information related to each shop's specific product cannot be directly compared.

In the survey of 102 respondents who use Shopee as their online shopping platform, 95.1% agreed it is time-consuming to compare Shopee product information manually. Shopee only shows the product price, the quantity of product sold, and the product's rating. Customers must click on the product whenever they want to view additional details. Comparing multiple product prices is time-consuming and inconvenient [11-12]. It makes it difficult for customers to compare numerous similar products from different stores. Paul & Nikolaev [13] have shown that manually analyzing a group of similar products is notoriously hard.

In addition, Shopee does not have a feature that allows customers to compare product reviews from different stores. Hence, customers must perform the comparison manually. Zuhri & Maulidevi [14] stated that customers read product reviews before purchasing, significantly impacting customer decisions. 64.7% of the same survey's respondents have difficulty comparing product reviews. As more reviews circulate, it is becoming increasingly difficult to find informative reviews for product purchases [14]. 64.7% of respondents have difficulty comparing product prices. Customers are never sure if another store has a better deal, and finding the best deal is complicated. When customers receive no comparison features, the situation worsens, forcing them to spend hours searching and employing extremely complicated techniques [15].

Based on the problem discussed, a web-based application using Python was implemented using the obtained data from the Shopee website via web scraping. The system uses bar charts and the NB algorithm as the classification algorithm to visualize the comparison of Shopee product information from different shops. Previous classification algorithm research has shown that NB is superior to Decision Tree and K-Nearest Neighbors [16-18]. As stated by Mican & Sitar-Taut [19], customers actively seek product information and ratings from online stores. On Shopee, product information includes vouchers, followers, chat response rate, and reviews. Hence, the following seven factors

were considered in this study: product rating, product price, product quantity, followers, chat response rate, active vouchers, and reviews. It lets users view product comparisons based on the selected factors in visualization form. Therefore, users can compare and visualize Shopee product information across shops. This paper is organized as follows: Section 1 begins with a brief introduction. Section 2 explains the methodology, followed by Section 3 with result and discussion. Finally, Section 4 concludes the study and briefly discusses potential future improvements.

2. Methodology

The research methodology employed in the design and development of this study was divided into several sections. In order to support the methodology, explanations and justifications for the research design and data collection techniques are provided. It is divided into several phases to achieve the research's goal of developing a web-based application written in Python Flask.

2.1 System Design

The system design of the web-based application that meets the research's requirements is described one by one. The system design uses use case diagrams, flowcharts, and user interfaces. A use case diagram is used to collect system requirements, both internal and external. Interactions between the system and the user are required to complete the system's task. A flowchart is a tool for analyzing the process involved in a system. Figure 1 depicts the overall system flow for this study.

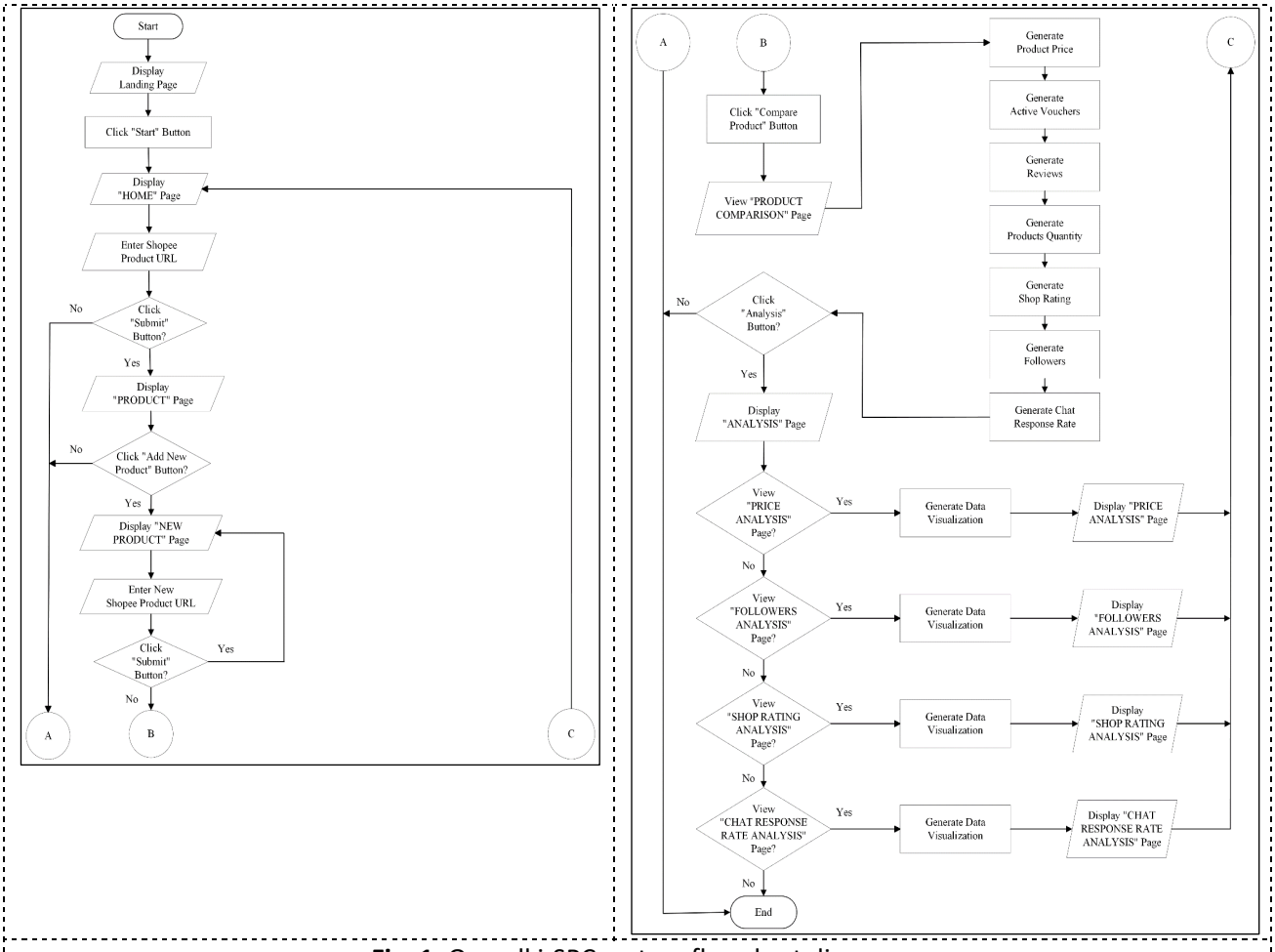


Fig. 1. Overall i-SPC system flowchart diagram

It starts by directing the users to the landing page, where they can click the "Start" button to go to the "HOME" page. On the "HOME" page, users can enter the Shopee product URL to be tracked. The system will track the product and redirect users to the "PRODUCT" page if the users click the "Submit" button. The product details are displayed, and if users want to compare the product, the "Add New Product" button can be clicked. Users will then be redirected to the "NEW PRODUCT" page, where the users can enter the new comparable product URL from other stores and click the "Submit" button to add the product.

Users can begin the comparison process after adding the products by clicking the "Compare Product" button. The system will then proceed to the "PRODUCT COMPARISON" page, which compares the information on Shopee products from different stores. Product price, active vouchers, reviews, product quantity, shop rating, followers, and chat response rate are the seven factors that will be considered. Users can navigate to the "ANALYSIS" page by clicking the "Analysis" button. The page will include buttons for analyzing the four factors: product price, followers, shop rating, and chat response rate.

The user interface (UI) design is the final step in the design phase before moving on to the implementation phase. UI design visualizes the proposed system's components layout where users can interact with the system directly. The UI design describes the system's features by following the planned idea in the flowchart.

2.2. Back-End Development

During development, the back end is the data access layer of the system, also known as the server-side development code, as presented in Figure 2. Starting with data preparation consists of data collection and pre-processing. It ends with the system's back end on the model deployment of NB written in Python, consisting of classifier and output. The output provides two sections, which are report data and visualization data. The following subsection discusses the data preparation process and NB classifier in more detail.

2.2.1. Data Preparation

Web scraping with BeautifulSoup and Selenium in Python was used to collect the data from https://shopee.com.my/. Scraping data can be accomplished using various methods, including Octopurse and rapid miner software. Using Python, however, is more efficient and allows users to customize the data required for this study.

The data is scraped in real time and are dependent on the users. The scraped data for one product is approximately 5 minutes long. Hence, the scraped data is limited to 10 URLs with 10 attributes: product name, price, rating, shop name, rating, quantity, followers, chat response rate, active vouchers, and reviews. The data is saved in CSV (comma-separated values) format after scraping.

Before data can be processed and used in a machine, it must be cleaned several times. The Shopee dataset has been pre-processed to remove unnecessary features, such as special symbols in the data, that add no value and only degrade the quality of the developed model. Data is cleaned before encoding to fill in the dataset's null values. Several factors, including shop rating, followers, product quantity, chat response rate, and product price, are cleaned by transforming their type from string to numerical. The data are then cleaned using feature extraction. This dimensionality reduction process converts original data into a dataset with fewer variables known as features. Data transformation is carried out using the Python library LabelEncoder from sklearn.

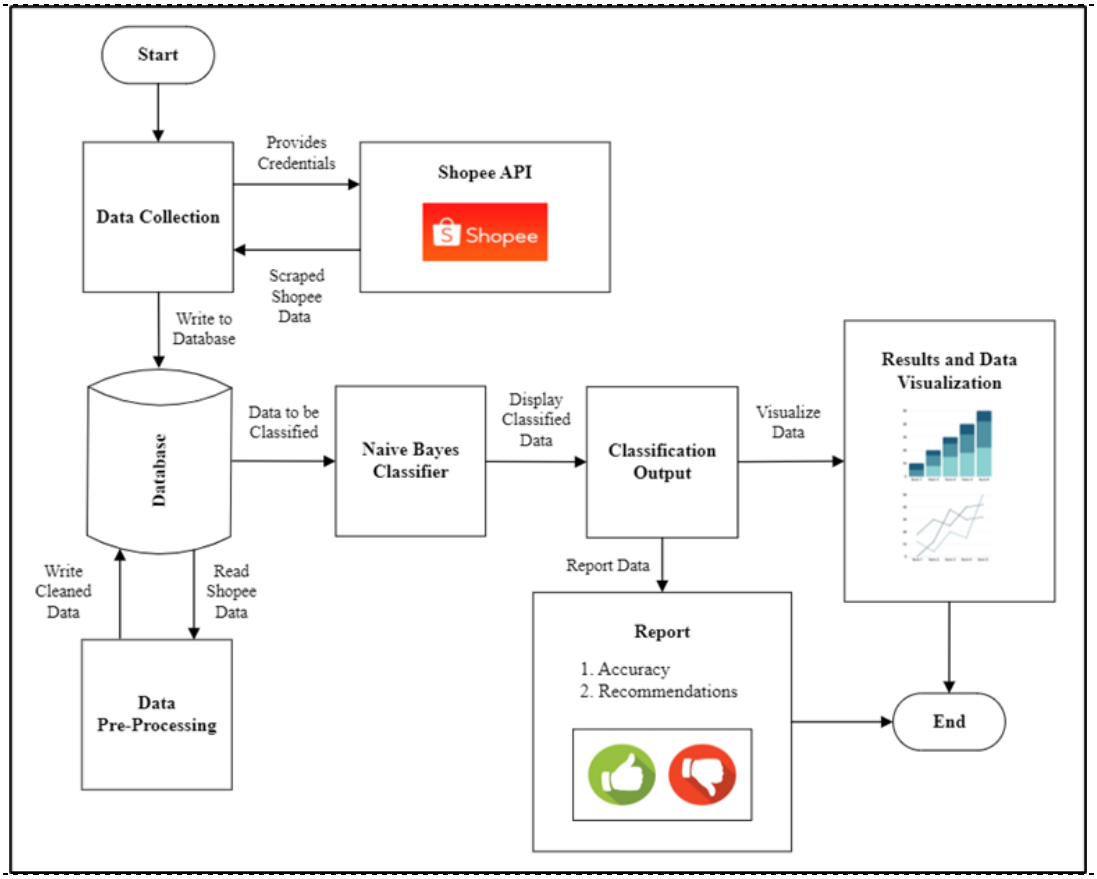


Fig. 2. Flow diagram of the algorithm design for i-SPC and visualization information

Next, a new attribute called "Category" has been added to the dataset. The attribute classifies the product as "Recommended" or "Not Recommended." If a product meets four criteria, it is classified as "Recommended": its price is lower than the average of product prices, its chat response rate is higher than the average of product chat response rates, its followers are higher than the average of product followers, and its shop rating is higher than the average of product shop ratings. The conditions are otherwise classified as "Not Recommended".

There are 11 columns in the final dataset, including the seven factors used in the NB classifier: product name, product rating, product price, shop name, shop rating, product quantity, followers, chat response rate, vouchers, reviews, and category. Before using data visualization, the data is saved in a CSV file that serves as a database. Following the pre-processing step, the final dataset is ready to be used in the machine learning algorithm for classification.

2.2.2. Naïve Bayes Classifier

Machine learning (ML) is a subset of artificial intelligence (AI) that relies on knowledge derived from training data based on known facts [20]. Its goal is to create algorithms capable of learning and building statistical models for data analysis and prediction [21]. Rincy & Gupta [22] reported that the NB algorithm was chosen as the ML algorithm to build the model because it quickly learns and performs classification. Mishra *et al.* [23] pointed out that NB has been successfully used in various classification task issue fields, including image and pattern recognition, weather forecasting, and medical diagnosis. The model learns from the training set's pre-labeled data and then performs classification on the dataset, mapping it to one of several predefined categorical classes. NB is a probabilistic classifier that uses Bayes' theorem to compute the probability of an event based on the

probabilistic joint distribution of other events [24]. The equation for Bayes' Theorem is represented as Eq. (1), and the equation parameter is explained in Table 1.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

Table 1
Equation parameter

Parameter's Name	Explanation
Prior probability of A, $P(A)$	Probability of hypothesis A being true regardless of the data
Prior probability of B, $P(B)$	Probability of the data regardless of the hypothesis
Posterior probability of A given B, $P(A\|B)$	Probability of hypothesis A given the data B
Posterior probability of B given A, $P(B\|A)$	Probability of data given that hypothesis A was true

NB predicts each product's category based on seven factors: product rating, product price, product quantity, followers, chat response rate, active vouchers, and reviews. The data is split into training and testing segments with a 60:40 split, with 60% for training and 40% for testing. The test set ratio for this study is 0.40. The train set should not be larger than the test set because there might not be enough data to train. The random state serves as the seed for the random number generator. This study's random state is 109. After the splitting process, the response for the test dataset is predicted. The model's accuracy prediction for the test dataset was expressed as a percentage of accurate predictions out of all predictions made.

2.3. Front-End Development

Front-end development is client-side development that focuses on user interfaces in web browsers. Front-end development uses languages like HTML, CSS, and JavaScript to communicate with Flask's back-end framework. Python's data visualization tools and the Flask web framework are used to create custom plots and charts within the context of a Python web application, combining the power of front-end and back-end development. It has specifics for each interface available for user interaction: the new product page, the product comparison page, the product price analysis page, the followers analysis page, the shop rating analysis page, and the chat response rate analysis page.

3. Result and Discussion

This section presents and discusses the analysis results performed on real-world data and the testing on the accuracy NB classifier model, functionality and usability testing.

3.1 Accuracy Testing of Naïve Bayes Classifier Model

The accuracy testing of the NB Classifier Model is obtained from simple Python code. Figure 3 shows the accuracy testing results for the NB model. The accuracy score is 87.5% when expressed in percentage form. This score indicates that the classification result is 87.5% correct. The model classified seven of the eight correct results as "Recommended" or "Not Recommended." In the confusion matrix, the "Not Recommended" class is represented as 0, while the "Recommended" class

is represented as 1. Vujović [25] reported that Anthony Ladson provided a model performance table based on the efficiency coefficient. In the case of model performance validation, the efficiency coefficient values describe the classification as follows: E≥0.93 means “Excellent”, 0.8≤E<0.93 means “Good”, 0.6≤E<0.8 means “Satisfactory”, 0.3 ≤E<0.6 means “Transient”, E<0.3 means “Bad”. Hence, the accuracy of 87.5% indicates that the performance of the NB classifier is good.

```
Model accuracy score: 0.8750

Confusion matrix

 [[7 0]
 [1 0]]

True Positives(TP) =  7

True Negatives(TN) =  0

False Positives(FP) =  0

False Negatives(FN) =  1

              precision    recall  f1-score   support

           0       0.00      0.00      0.00         1
           1       0.88      1.00      0.93         7

    accuracy                           0.88         8
   macro avg       0.44      0.50      0.47         8
weighted avg       0.77      0.88      0.82         8
```

Fig. 3. Result of accuracy testing for Naïve Bayes model

3.2 Functionality Testing

Testing is required to ensure that all system features function correctly and that any unusual system behavior is quickly identified and corrected. Functional testing aims to review each function of the visualization application and see how well the specifications complement each other by providing appropriate input and comparing the output to the functional requirements stated in the previous chapter. This test employs test cases derived from system specifications. The results demonstrate that the system can redirect to the correct page and display the appropriate information. The bar chart visualizations were also tested to ensure the user could view the selected visualization.

Figure 4 shows the “New Product” page, where users can enter the URLs of new comparable Shopee products to be compared, up to a maximum of 10 URLs. For this example, we use 10 URLs of the product “Adidas Adilette Slides Slipper Selipar Men Women Unisex”. Once complete, the menu, as in Figure 5, depicts the “Product Comparison” page. It provides a comprehensive product comparison overview. Users can compare similar Shopee products from different stores based on seven criteria: product rating, product price, product quantity, followers, chat response rate, active vouchers, and reviews. It displays, together with the additional four columns: product name, shop name, shop rating and category display.

The category column contains the product recommendations. If a product meets four criteria, it is classified as “Recommended”: it has a lower price than the average product price, a higher chat response rate than the average product chat response rates, a higher number of followers than the average product followers, and a higher shop rating than the average product shop ratings. Otherwise, the conditions are classified as “Not Recommended.” These recommendations assist users in locating relevant products.

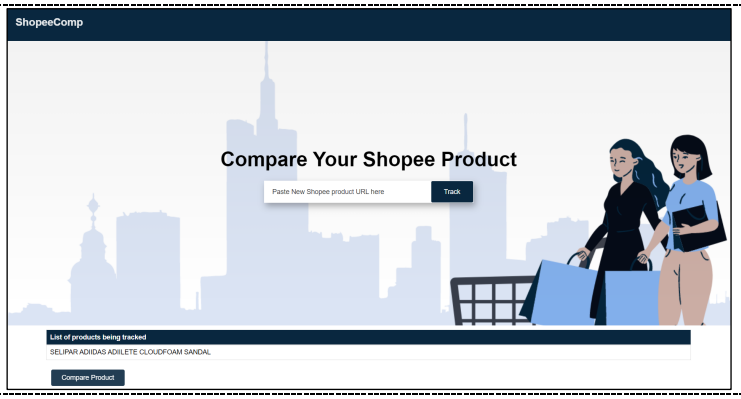
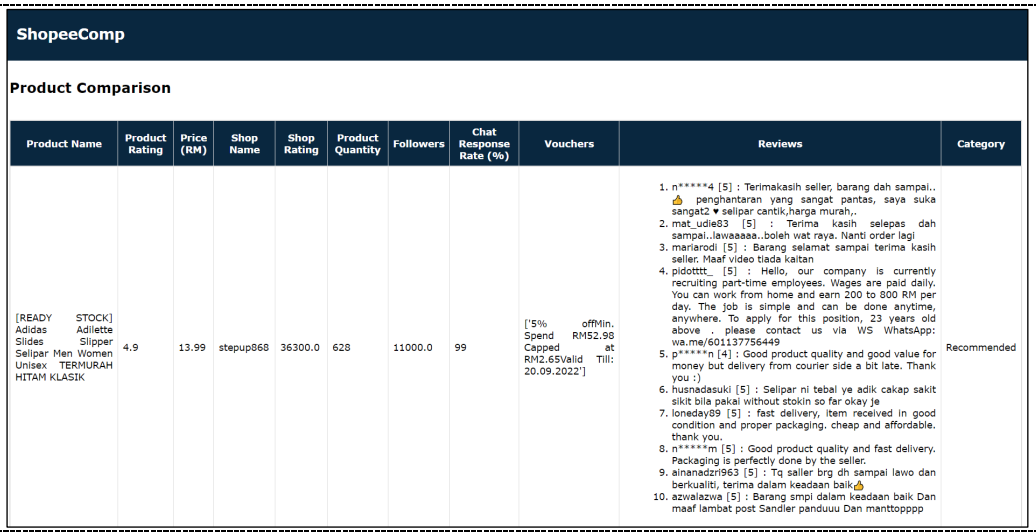


Fig. 4. View the "New Product" page



Product Name	Product Rating	Price (RM)	Shop Name	Shop Rating	Product Quantity	Followers	Chat Response Rate (%)	Vouchers	Reviews	Category
[READY STOCK] Adidas Adilette Slides Slipper Selipar Men Women Unisex TERMURAH HITAM KLASIK	4.9	13.99	stepup868	36300.0	628	11000.0	99	['5% offMin. Spend RM52.98 Capped at RM2.65Valid Till: 20.09.2022']	1. n*****4 [5] : Terimakasih seller, barang dah sampai.. penghantaran yang sangat pantas, saya suka sangat2 ♥ selipar cantik,harga murah,. 2. mat_udie83 [5] : Terima kasih selepas dah sampai..lawaaaaa..boleh wat raya. Nanti order lagi 3. mariarodi [5] : Barang selamat sampai terima kasih seller. Maaf video tiada kaitan 4. pidottt_ [5] : Hello, our company is currently recruiting part-time employees. Wages are paid daily. You can work from home and earn 200 to 800 RM per day. The job is simple and can be done anytime, anywhere. To apply for this position, 23 years old above . please contact us via WS WhatsApp: wa.me/601137756449 5. p*****n [4] : Good product quality and good value for money but delivery from courier side a bit late. Thank you :) 6. husnadasuki [5] : Selipar ni tebal ye adik cakap sakit sikit bila pakai without stokin so far okay je 7. loneday89 [5] : fast delivery, item received in good condition and proper packaging. cheap and affordable. thank you. 8. n*****m [5] : Good product quality and fast delivery. Packaging is perfectly done by the seller. 9. ainanadzri963 [5] : Tq saller brg dh sampai lawo dan berkualiti, terima dalam keadaan baik 10. azwalazwa [5] : Barang smpi dalam keadaan baik Dan maaf lambat post Sandler panduuu Dan manttopppp	Recommended

Fig. 5. View the "Product Comparison" page

The analysis of four Shopee product factors is visualized using a bar chart. It begins with the price analysis in Figure 6, which gives users an overview of product price differences. It has been sorted descending to help them observe which shop has the highest and lowest prices. Based on the example used, the shop name "uwalkmy.os" offer the highest price of RM99 and the shop name "kasutmurahlagimurah" shows the lowest price of RM10.9.

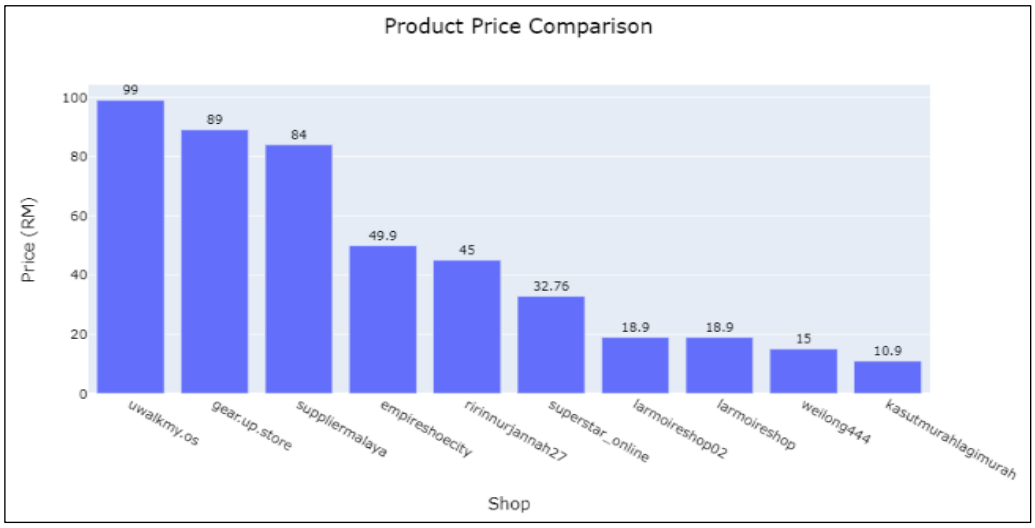


Fig. 6. Bar chart of product price analysis

Figure 7 indicates the follower's analysis in a bar chart. It gives users an overview of followers' differences. It has been sorted descending to help them determine which shop has the highest number of followers. Based on the example, the shop name "larmoireshop" has the highest of followers, 109,600 and the shop name "ririnnurjannah27" shows the lowest number of followers, 209.

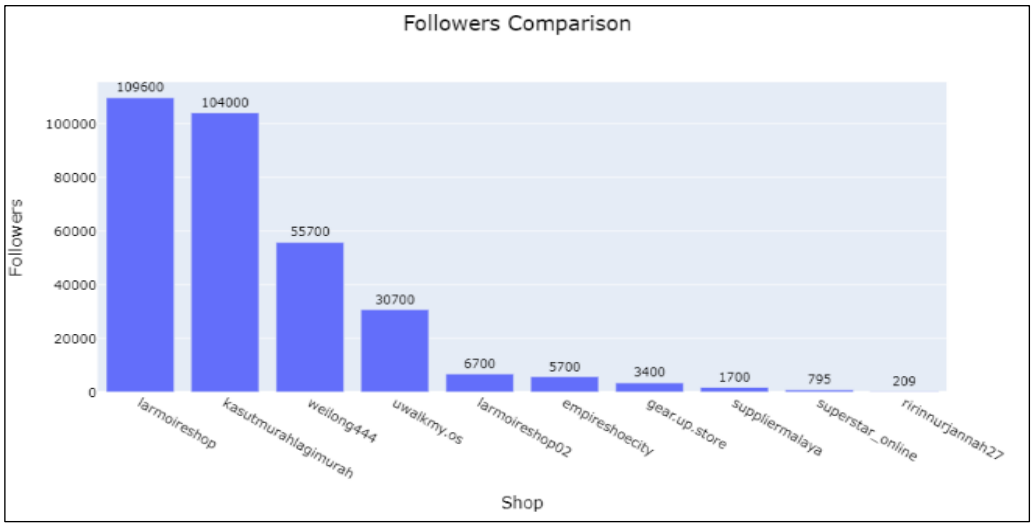


Fig. 7. Bar chart of followers analysis

The shop rating analysis uses a bar chart depicted in Figure 8. It shows information to users with an overview of shop rating differences. With that, it has been sorted descending to help them determine which shop has the highest shop rating. Based on the example, the shop name "weilong444" has the highest shop rating, 92,900 and the shop name "ririnnurjannah27" shows the lowest, 1300.

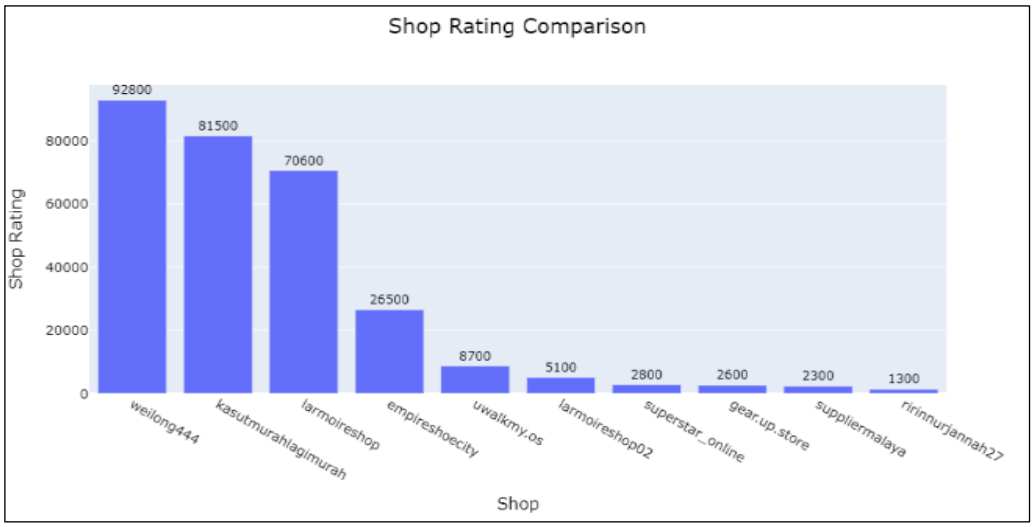


Fig. 8. Bar chart of shop rating analysis

As shown in Figure 9, the chat response rate analysis is depicted in a bar chart form. It gives users an overview of the chat response rate percentage and allows them to see which shops have the highest rate. All the data visualizations are interactive visualization. The user can see the graph

information when hovering over the graph and using Plotly's function to download the desired graph and change the size of the graph. Based on the example, the shop name "ririnnurjannah27" has the highest chat rating, 90% and the shop name "suppliermalaya" shows the lowest chat rating, 16%.

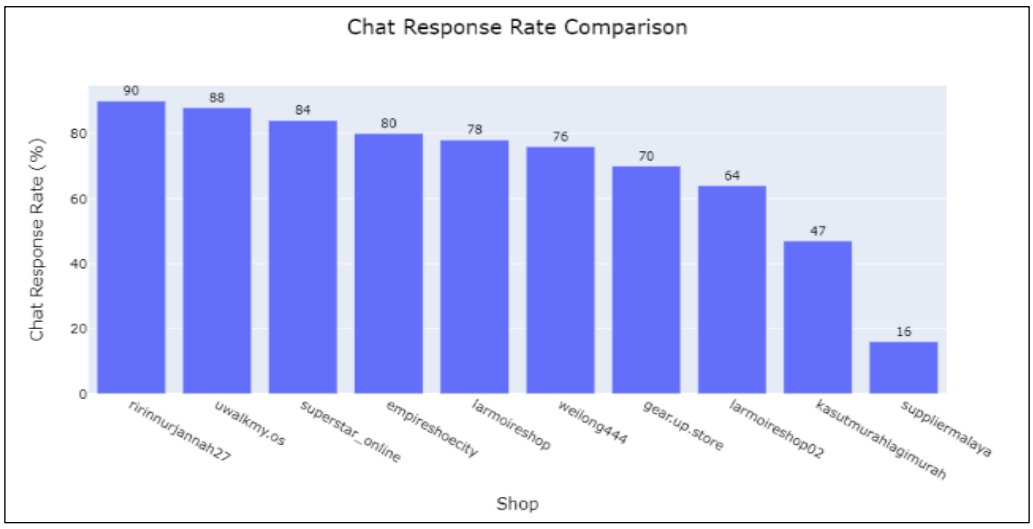


Fig. 9. Bar chart of chat response rate analysis

4. Conclusions

A web application titled "Intelligence Shopee Product Comparison (i-SPC) and Visualization of Product Information via Naïve Bayes Adaptation" was developed to visualize the comparison of Shopee product information from various stores. The information obtained from the system application can assist the user in making decisions by assisting them in understanding the differences between the same product from various stores. Furthermore, the system application's various visualizations make it simple for the user to better understand each shop's product. Shopee users can use this app to help users get a fast overview of Shopee product information comparison and assist them by recommending which products are better to buy. For further study, a larger sample size could also be obtained for future research by increasing the amount of data collected by speeding up the system's web scraping process.

Acknowledgement

This research was funded by a grant from Universiti Teknologi MARA Cawangan Melaka (TEJA Grant 2023 GDT 2023/1-14).

References

[1] Elveny, M., S. M. Hardi, I. Jaya, and P. Gundari. "Web-based E-Commerce Products Grouping." In *Journal of Physics: Conference Series*, vol. 1898, no. 1, p. 012018. IOP Publishing, 2021. https://doi.org/0.1088/1742-6596/1898/1/012018.

[2] Putri, Wilda Kurnia, and Vera Pujani. "The influence of system quality, information quality, e-service quality and perceived value on Shopee consumer loyalty in Padang City." *The International Technology Management Review* 8, no. 1 (2019): 10-15. https://doi.org/10.2991/itmr.b.190417.002.

[3] Fadhillah, Auliya, Yuniman Zebua, and Yudi Prayoga. "Analysis of Information quality, trust and satisfaction on customer participation (Case study on customer online shop shopee in Rantauprapat)." *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)* 2, no. 2 (2021): 3039-3051. https://doi.org/10.33258/birci.v4i2.2010.

[4] Budyastuti, Triyani. "Factors that Affect the Use of the Online Payment System on E-Commerce (Case Study on Users of Shopee. CO. ID)." In *Information and Knowledge Management*, vol. 9, no. 5. 2019. https://doi.org/10.7176/IKM.
[5] Černikovaitė, Miglė, Žaneta Karazijienė, Lina Bivainienė, and Valdas Dambrava. "Assessing customer preferences for shopping centers: Effects of functional and communication factors." *Sustainability* 13, no. 6 (2021): 3254. https://doi.org/10.3390/su13063254.
[6] Isa, Khairunesa, J. M. Shah, Sarala Palpanadan, and F. Isa. "Malaysians' popular online shopping websites during movement control order (Mco)." *International Journal of Advanced Trends in Computer Science and Engineering* 9, no. 2 (2020): 2154-2158. https://doi.org/10.30534/ijatcse/2020/190922020.
[7] CHING, KIEW CHEE, ZUHA ROSUFILA ABU HASAN, and Nadia Abu Hasan. "Factors influencing consumers in using shopee for online purchase intention in east coast Malaysia." *Universiti Malaysia Terengganu Journal of Undergraduate Research* 3, no. 1 (2021): 45-56.
[8] Daryanto, Wiwiek Mardawiyah, Putri Tresna Meiliawati, and Tesha Syarafina. "The Financial Impact of Conventional Retail Industry Due to Online Based Shopping Model in Indonesia (Case Study of PT Matahari Departement Store Tbk and PT Shopee International Indonesia Period of 2013-2018)." *South East Asia Journal of Contemporary Business, Economics and Law* 21, no. 2 (2020): 26-37.
[9] Nursyirwan, Vivi Iswanti, and Sasmita Sari Ardaninggar. "The factor analysis that influence the student purchase intention in shopee E-commerce." *Economics and accounting journal* 3, no. 2 (2020): 118-129.
[10] Kivistö, Reetta. "E-commerce product information supporting user experience." Master's thesis, 2020.
[11] Saric, Dragana, and Marian Mikolasik. "The impact of electronic services on traditional services." *Developments in Information & Knowledge Management for Business Applications: Volume 2* (2021): 305-334. https://doi.org/10.1007/978-3-030-76632-0_20.
[12] Menghwar, Parmanand. "A Product Pricing Comparison Model and Data Visualization for Online Retailers." (2019).
[13] Paul, Himangshu, and Alexander Nikolaev. "Fake review detection on online E-commerce platforms: a systematic literature review." *Data Mining and Knowledge Discovery* 35, no. 5 (2021): 1830-1881. https://doi.org/10.1007/s10618-021-00772-6.
[14] Zuhri, Hamdi Ahmad, and Nur Ulfa Maulidevi. "Product Review Ranking in e-Commerce using Urgency Level Classification Approach." *Jurnal Online Informatika* (2020): 212-216. https://doi.org/10.15575/join. join.v5i2.612.
[15] Iyer, Srinivasan B. "RISING TREND IN DIGITAL CONSUMERS." (2020).
[16] Maliha, Shanjida Khan, Romana Rahman Ema, Simanta Kumar Ghosh, Helal Ahmed, Md Rafsun Jony Mollick, and Tajul Islam. "Cancer disease prediction using naive bayes, K-nearest neighbor and J48 algorithm." In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-7. IEEE, 2019. https://doi.org/10.1109/ICCCNT45670.2019.8944686.
[17] Yadav, Kirtika, and Reema Thareja. "Comparing the performance of naive bayes and decision tree classification using R." *International Journal of Intelligent Systems and Applications* 11, no. 12 (2019): 11. https://doi.org/10.5815/ijisa.2019.12.02.
[18] Slamet, Cepy, Rian Andrian, Dian Sa'adillah Maylawati, Suhendar, Wahyudin Darmalaksana, and Muhammad Ali Ramdhani. "Web scraping and Naïve Bayes classification for job search engine." In *IOP Conference Series: Materials Science and Engineering*, vol. 288, p. 012038. IOP Publishing, 2018. https://doi.org/10.1088/1757-899X/288/1/012038.
[19] Mican, Daniel, and Dan-Andrei Sitar-Taut. "Analysis of the factors impacting the online shopping decision-making process." *Studia Universitatis Babes-Bolyai Oeconomica* 65, no. 1 (2020): 54-66. https://doi.org/10.2478/subboec-2020-0004 ANALYSIS.
[20] Fujii, Gaku, Koichi Hamada, Fuyuki Ishikawa, Satoshi Masuda, Mineo Matsuya, Tomoyuki Myojin, Yasuharu Nishi et al. "Guidelines for quality assurance of machine learning-based artificial intelligence." *International journal of software engineering and knowledge engineering* 30, no. 11n12 (2020): 1589-1606. https://doi.org/10.1142/S0218194020400227.
[21] Chibani, Siwar, and François-Xavier Coudert. "Machine learning approaches for the prediction of materials properties." *Apl Materials* 8, no. 8 (2020). https://doi.org/10.1063/5.0018384.
[22] Gupta, Roopam. "An efficient feature subset selection approach for machine learning." *Multimedia tools and applications* 80, no. 8 (2021): 12737-12830. https://doi.org/10.1007/s11042-020-10011-7.
[23] Mishra, Smriti, Ranjan Kumar, Sanjay Kumar Tiwari, and Priya Ranjan. "Machine learning approaches in the diagnosis of infectious diseases: a review." *Bulletin of Electrical Engineering and Informatics* 11, no. 6 (2022): 3509-3520. https://doi.org/10.11591/eei.v11i6.4225.
[24] da Silva, Paula Renatha Nunes, Hossam A. Gabbar, Petrônio Vieira Junior, and Carlos Tavares da Costa Junior. "A new methodology for multiple incipient fault diagnosis in transmission lines using QTA and Naïve Bayes

classifier." *International journal of electrical power & energy systems* 103 (2018): 326-346. https://doi.org/10.1016/j.ijepes.2018.05.036.

[25] Vujović, Ž. "Classification model evaluation metrics." *International Journal of Advanced Computer Science and Applications* 12, no. 6 (2021): 599-606. https://doi.org/10.14569/IJACSA.2021.0120670.