ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 14 (2024) No. 8, pp. 168-1778
DOI:10.33168/JSMS.2024.0811

Product Recommendation in Fashion E-Commerce using Deep Learning and Image Similarity

Gladys Patricia, Ririn Ikana Desanti

Information System Study Program, Universitas Multimedia Nusantara, Tangerang 15810, Indonesia
ririn.desanti@umn.ac.id

Abstract. This research proposes a fashion product recommendation system using content-based image retrieval (CBIR) with convolutional neural networks and the k-nearest neighbors (KNN) algorithm. Product images from an e-commerce website were collected through web scraping. Pre-trained VGG16 extracted image features. KNN identified visually similar products based on feature vectors. Cosine similarity, root mean square deviation (RMSE), and structural similarity index (SSIM) evaluated model performance. On a test set, the model achieved an average cosine similarity of 0.76 and SSIM of 0.80 between query and recommended items. The system was deployed on a website, demonstrating the feasibility of content-based visual recommendations for fashion e-commerce. This research makes a valuable contribution by developing and evaluating a domain-specific recommender system to enhance the online shopping experience.

Keywords: content-based image retrieval, convolutional neural network, image similarity, k-nearest neighbors algorithm, product recommendation.

1. Introduction

Fashion is a way to express human ideas and interests, which are represented visually through objects attached to the body. Fashion is a communicator to convey someone's identity and self-worth. Fashion is also used as an aspect to determine a person's position, class, and social status (Tajuddin, 2018). In addition, fashion is often used to seek for recognition so that a person feels they belong to certain group in society (Williams, 2018). Based on this, fashion becomes an important need for the psychological and social aspects of humans. Human needs and interests encourage the creation of business in this this field. Fashion contributed 2% of the total global Gross Domestic Product (GDP) and employed 57.8 million people in 2018 (Mair, 2018). In 2019, the global fashion business continued to grow and managed to achieve revenue of 1.3 billion USD with a total of 300 million workers (Gazzola et al., 2020).

The development of the fashion industry also demands companies to improve their business strategy to maintain their existence in the intense market competition. One of the strategies that can be carried out by a company is digital promotion. Fashion companies generally rely on physical stores to shape the shopping experience and introduce their brands to buyers (Guercini et al., 2018). However, in this modern era, companies need to expand their marketing through the use of the internet, especially e-commerce. In the case of fashion retail, e-commerce gives the freedom to choose a variety of products and provides a fast and easy shopping experience. This then adds points to the business value proposition (Sutinen et al., 2022). According to research (statista research department, 2023), the fashion e-commerce market continues to grow each year, where in 2023 the global fashion e-commerce market is forecast to reach a value over 820 billion U.S. dollars. Fashion e-commerce generally provides a product recommendation system to understand customer desires increase sales (Riyanto, 2022). When choosing fashion products, customers consider the overall design of the product. These design aspects have a major influence on a person's interest in buying (Putrisuryyana, 2021). Therefore, fashion product recommendations can be formed based on visual considerations. The recommended products in an e-commerce are expected to be accepted and liked by users to increase interest in buying. The Z fashion company e-commerce site already has a recommendation system to display a collection of visually similar products. However, this has not been working properly in displaying products. This causes customers to search for similar products according to their interests manually. Based on the existing problem, this study focuses on forming a new recommendation model based on visual similarity. This solution will be applied to corporate ecommerce so that customers have a wider selection of products and better shopping experience.

The solution offered in this study is a recommendation model based on image processing and similar image search. The result can be obtained using Content-based Image Retrieval (CBIR), as applied in the previous studies (Warongan et al., 2018; Choe et al., 2022). The image search is based on the features or characteristics of the image. Even though CBIR especially in fashion image still has limitation when it comes to visual searching in the real world due to the simultaneous availability of multiple fashion items as mentioned in the previous study (Park et al., 2019). Feature extraction is carried out with Convolutional Neural Network, namely the pre-trained model VGG16, which has been trained using millions of data from ImageNet. The image features are then processed with cosine measurements in the K-Nearest Neighbors (KNN) algorithm (Chen, 2020). K-Nearest Neighbors was chosen because it is able to get the nearest object and is often used for classification based on similar characteristics (Dewi & Dwidasmara, 2020; Cahyanti et al., 2020). Image evaluation is based on the Cosine Similarity, root mean square deviation (RMSE), and structural similarity index (SSIM) values.

2. Related Previous Research

Several previous studies have been used as references for the formation of model recommendations and image processing using various methods. This time, the research was conducted using previously used methods but applied to different objects and case studies. The study uses the image object of the Z fashion product and aims to provide recommendations based on the visual similarities of the product. The image search process was performed using the content-based Image Retrieval (CBIR) technique, which was also used in the previous research (Warongan et al., 2018; Choe et al., 2022). The technique is applied to the field of fashion, with input and output in the form of product images as the basis for similar product recommendation techniques. Specific to fashion image retrieval, it is quite challenging since it requires searching for exact items accurately from massive collections of fashion products (Park et al., 2019). The image data processing process was carried out using CNN's VGG16 architecture for the extraction of image features, as in the study (Belaid & Loudini, 2020). Next, the search for images that have visual similarities is done with the measurement of distance data using the cosine method. The selection of the cosine method was based on studies that compared the method with other measurements such as Pearson Correlation Coefficient, correlation Distance, Manhattan distance, and Euclidean Distance. Based on the research, cosine is the most accurate calculation method. The novelty of this study also lies in the use of KNN to find the nearest images. KNN is commonly used as an algorithm for classifying images and recommending tabular data, as described in previous studies (Kaesmetan & Overbeek, 2022; Suharyadi & Kusnadi, 2019). In this study, KNN will be used as an algorithm to find the nearest image neighbors based on cosine calculations. The performance of the model will be calculated and analyzed using the formulas Cosine Similarity, Root Mean Squared Error (RMSE), and Structure similarity index (SSIM). Cosine Similarity is used to see the percentage of image similarities according to research (Chen, 2020), while RMSE and SSIM are selected based on previous research (Sara et al., 2019). For this research, the recommendation model as the result of KNN algorithm and cosine calculation, will subsequently be implemented and evaluated via a website.

3. Methodology

The methodology of this study includes several key steps as illustrated in Fig. 1.

Fig. 1: Illustration of the proposed methodology.

This study starts with the collection of product images from Z e-commerce using scraping method. The data used in this study amounted to 1918 images which were divided into two main lines, namely woman's and men's products. Women's products consist of shoes, sandals, bags, clutches, wallets, and other accessories with a total of 908 products. Men's products consist of shoes, sandals, bags, clutches, wallets, belts, and other accessories with a total of 1010 products. The scrapping process is carried out for approximately one month.

The second type of data used in this study is tabular data which is presented in table form containing products description. Tabular data is used in the deployment process and testing process of recommendation model. This data is formed by disguising the original data and generating random data. In addition to taking picture data, dummy data is also created which contained information from each product as supporting data. Subsequently, data preparation is conducted to make all image data the same size. Data manipulation is carried out to convert data into 3 types, namely normal, contour, and grayscale data. Data preparation also includes feature extraction for model building. Each type of image input is built into the K-Nearest Neighbors and compared to get the best input type. The result of recommendation model is deployed to a website for testing purposes. Every process is performed using Python programming language.

The feature extraction process is carried out with the help of a pre-trained model of VGG16. This model is available in the Keras library and can be called into a worksheet via code. The VGG16 used is the result of training on millions of images from the imageNet database. The structure and arrangement of the model is used as an image feature extractor for this research.

4. Results and Discussion

4.1. Dataset

The dataset used in this study is divided into two kinds, namely image data and table data. Image data represents a product in a visual form that can be seen by the human eye. The unique product image is 1918 and obtained by scraping data from the official product Z website. The scraping process was carried out in January 2023, all products taken are up to date until January 2023. Table data is presented in tabular form and contains a description of the products. Tabular data is not the main component to build recommendation model, but it supports the process of deploying and model testing. This data is formed by generating random data. The description of each variable in the tabular data is provided as follows:

- **Product Code**: A unique code that represents a product.
- **Product Name**: The name of the product being sold.
- **Label**: Product line, consisting of men and women. Men are denoted by 'M', while women are denoted by 'W'.
- **Department**: The first hierarchy department in the distribution of product types, consisting of footwear, bags, belts and accessories.
- **Class**: The second hierarchy in the division of product types, consists of shoes, sandals, bags, accessories, clutches, belts and wallets.
- **Category**: The third order hierarchy that reduces product classes consists of 29 different product categories.
- **Color**: Product color
- **Price**: Product price
- **Image Name**: Product image name
- **Main Material**: The main material that makes up the product.
- **Other Material**: Additional material making up the product.
- **Image Path**: Directory where images are stored.

4.2. Data Preparation

The data that has been collected is then processed through the data preparation stage. This stage is divided into 4 processes, namely handling missing values in tabular data, image cropping, image manipulation, and feature extraction. The data preparation process is carried out using Python.

Missing value handling begins with checking the table to find out which column has a null value and how many rows of data are missing. Figure 2 shows the checking process using the is NA and sum functions. The isna() function is used to check for missing values in each row and column. This function returns true when it encounters a null value and returns false when the row has data. The sum() function is used to add up each true value returned by isna(), so that the number of missing values can be known along with the column it came from.

```
desc.isna().sum()
✓ 0.1s                                   Python

product_code      0
product_name      0
label             0
department        0
class            17
category         17
color             0
price             0
image_name        0
main_material     0
other_material    0
image_path        0
dtype: int64
```

Fig. 2: Checking Missing Data.

After knowing that the missing values are in the class and category columns with relatively small amounts, the process of replacing the missing values is carried out by analyzing product images. Through product images, category and class values will be determined manually. Therefore, each image of the row that has a missing value is displayed. Figure 3 shows each image with null values, there are 17 different product images to identify. Every missing category and class are determined manually but imputed by looping function in Python. The prepared data is then transferred to the database for deployment purposes.

Fig. 3: Products with Missing Values.

The second phase of data preparation is image cropping. All images are originally in 10: ratio, meanwhile the VGG16 receives input in 1:1 ratio. Preparing the image is done by cropping the image automatically through looping. The image is cropped to avoid the resizing process which has an impact on object shrinkage. The crop is calculated from each side, such as left, right, top and bottom resulting in a pixel size of 760x760 with a 1:1 ratio. The image is then saved back to local storage in a different folder.

The third phase of data preparation is image manipulation. Manipulation is done by changing the original image using two types of filters, namely grayscale and contour. Each image is processed with a filter from ImageFilter library, then saved back to the local directory for processing. Converting an image to a contour is done with the ImageFilter.CONTOUR() functions. The contour filter is used to change the image object so that the outline and texture remain. The process of converting an image to grayscale is done using the convert function, namely Image.convert('L'). Based on visuals, this data manipulation is able to highlight the texture and shape of objects, as well as automatically remove background noise to make it plain. The results of data manipulation are shown in figure 4. Normal, grayscale, and contour image will be compared to see the best recommendation results.

Fig. 4: Image Manipulation Results.

The last phase of data preparation is feature extraction. The feature extraction process is carried out with pre-trained model VGG16. Through this extraction, the original array is changed so that it has a certain value that accentuates the shape, edge, texture, color, contrast, and other components that make up the image. These results make objects in the image recognizable. One feature extractor can be used for three different types of images. Each type of image is processed, then the extracted results are stored in a local directory using the pickle format. Extracted features can be called and reused as long as there's no change in the data. The features that have been extracted are the main components in the formation of the K-Nearest Neighbor (KNN) model.

4.3. Modeling and Evaluation

The process of forming the model and evaluation in this study is divided into two stages: search for the best type of input and formation of the final model.

At the first stage, three kinds of input image that have been extracted are then formed into different KNN models. Establishment of this model aims to compare the results of recommendations from each type of input and to get the best results. The KNN model formed in this study produces 13 closest data (n_neighbors) with a cosine matrix. One value returned by KNN is a product that is 100% the same as the input (distance = 0). The value of n_neighbors can be changed according to the needs of the recommendations. Figure 5 show the input and prediction results for each type of image:

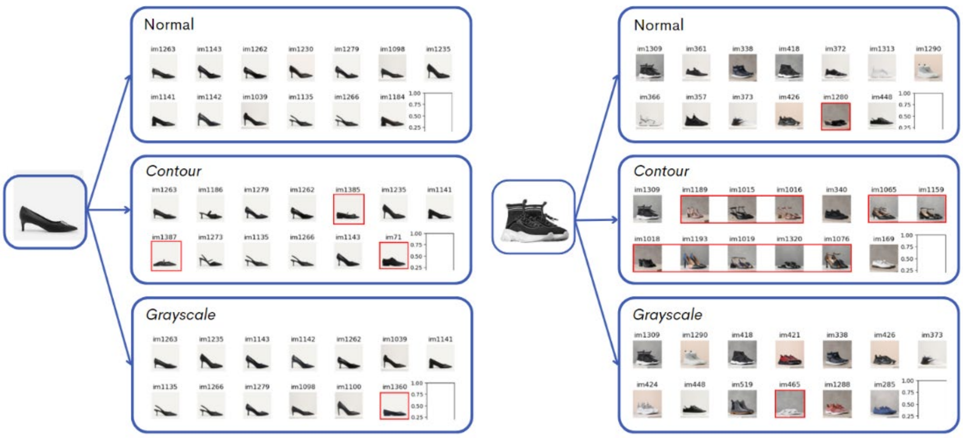

Fig. 5: Recommendation Results based on Different Inputs.

Evaluation can be run by calling any of the cosine, RMSE, and SSIM functions. In addition, dissimilar products (the one that is marked in red) are counted manually. Figure 6 contains the results of evaluation of each model. All types of input are generally able to provide good image recommendations, but normal data supports the model to produce better output.

Jenis	Cosine	RMSE	SSIM	Visual err	Line err
Normal	**0.80**	50.09	0.81	**3**	**13**
Kontur	0.73	**48.50**	0.81	17	15
Grayscale	0.77	50.22	0.81	6	27

Fig. 6: Evaluation Result based on Different Inputs.

The second model is built to separate between man and woman’s product. The previous model is built based on all the combined data. There’s another problem with the model, image similarity-based recommendations cannot guarantee that the results are in accordance with the product line. This means that the model is likely to issue women's products in men's recommendations and vice versa. This problem is caused by the image-based recommendation formation process that does not receive any input other than the features of the image. The process of forming male and female models is carried out with the same process and stages as the previous model. The change lies in the image data used in each process. Image data needs to be separated between women and men, then processed into different models. Image searches will be carried out separately, male products will be searched from the men's line set, while women's products will be searched from the women's line data set. This separation guarantees the accuracy of recommendations based on product lines and shortens feature extraction time. Figure 7 show the results from each man and woman recommendation model.

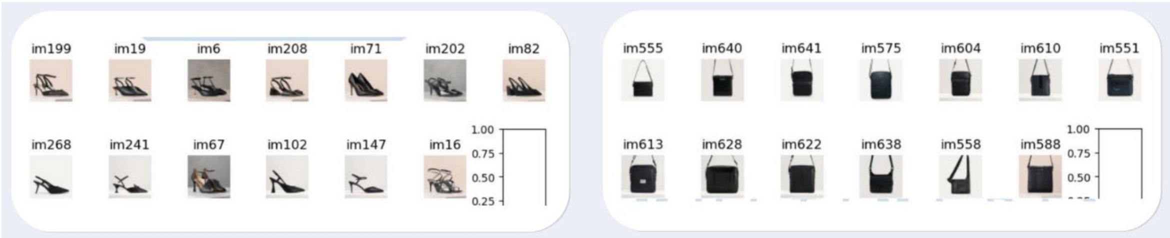

Fig. 7: Woman and Man Recommendation Result.

Recommendations from this model are then reevaluated using cosine, RMSE, and SSIM. The women's products consist of 19 categories, therefore it is iterated 19 times with different images and each represents a certain category. Men's products consist of 20 categories, thus the evaluation iteration is carried out 20 times. The woman’s evaluation results show that the similarity from cosine metric is 0.73, the RMSE error value is 52.69, and the SSIM value is 0.80. The evaluation results for the male model obtained an average cosine value of 0.79, an RMSE of 47.81, and an SSIM value of 0.80. This evaluation value means that the product recommendation results have an average similarity of 73% and 70% based on cosine calculations, both also have a structural similarity of 80%.

4.4. Deployment

The final models are deployed to the website for testing purposes. The website will call an API which is built using Flask. The API runs as a server providing data from the recommendation model to the front end of the website. Flask works as the communicator between website and the model, it passes input image to the model and give the recommendation result back to website in JSON format. Figure 8 shows similar products results.

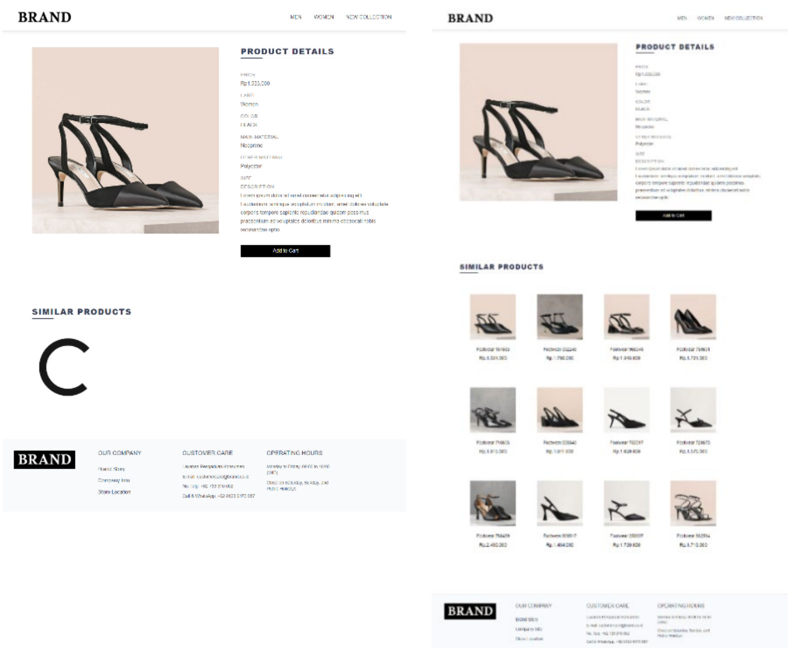

Fig. 8: Deployment Result to Website.

4.5. Discussion

This research focuses on forming a recommendation model for similar products based on the principle of content-based image retrieval using VGG16 and KNN. The CBIR principle is not used to generate data classifications based on certain labels but to obtain images with the highest similarity as a basis for recommendations. Therefore, there is no calculation of accuracy, precision, or recall, as was done in previous studies. Even though the evaluation methods are different, this research is still in line with previous research for searching for the closest image from a database.

Table. 1: Comparison of this Study with Previous Studies.

Researcher	Total Data	Tools	Evaluation Method	Result
Current research team	1918	Python	Cosine Similarity	76%
			SSIM	80%
Warongan, T. S., Sompie, S. R. U. A., & Jacobus, A.	250	Python	Recall	72.2%
			Precision	72.2%
Choe, J., Hwang, H. J., Seo, J. B., Lee, S. M., Yun, J., Kim, M. J., Jeong, J., Lee, Y., Jin, K., Park, R., Kim, J., Jeon, H., Kim, N., Yi, J., Yu, D., & Kim, B	288	Python	Accuracy	60.9%

Table 1 shows a comparison of this study with previous CBIR studies. Based on a previous study, CBIR and feature extraction methods were successfully implemented for the case of fashion products. The model in this study has the highest similar image recognition rate compared to the two previous studies. A study (Warongan et al., 2018) has a recall value and precision of image prediction suitability of 72.2%, while another research (Choe et al., 2022) has a general classification accuracy value of 60.9%. In the study, the evaluation of the female model obtained an average cosine value of 0.73 and an SSIM of 0.80. The male model has an average cosine value of 0.79 and an SSIM of 0.80. In general, the average cosine is 76% and the SSIM is 80%. According to further analysis, the quantity of comparable image data in the database can also have an impact on this number's size in addition to model performance. The more similar products in the database, the higher the possibility of getting recommendation results that have high cosine and SSIM scores. This is proven by products from the men's line, such as loafers and wallets, which have a similar cosine recommendation of 93%, as well as moccasins of 89% because they have many products with similar shapes. The evaluation value of this research can be categorized as quite good when compared with previous research. This can be because more data is used so that the model can learn the image better, as well as giving the model freedom to get the best output option. Based on the comparative evaluation, this research is proven to be in line with previous research and shows that the CBIR method is good enough to be used in different cases.

The quantity of similar image data in the database can also have an impact on the evaluation score in addition to model performance. The more similar products in the database, the higher the possibility of obtaining recommendation results with high cosine and SSIM scores. Based on the comparisons, this study is proven to be in line with previous studies and shows that the CBIR method is good enough to be used in different cases.

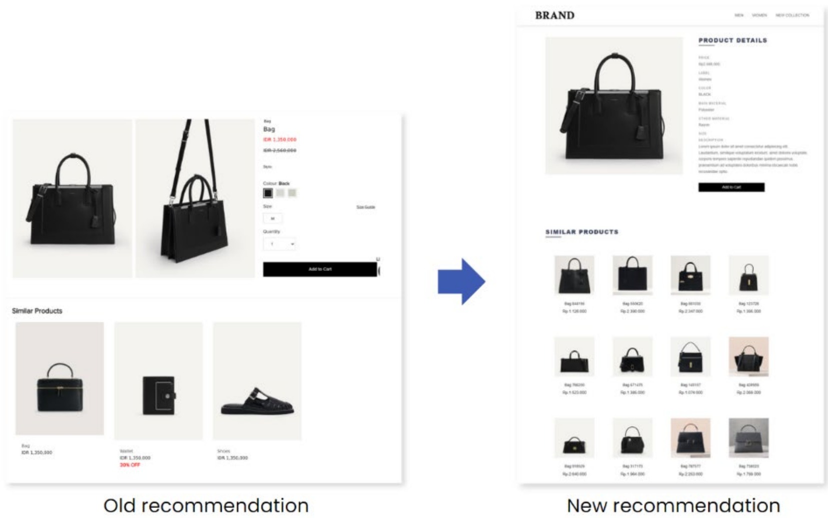

Fig. 9: Comparison of Old and New Recommendation.

In addition, another evaluation is made to compare the current model with the old model. From figure 9, bag recommendations have been improved into 12 bag products with similar visuals. The final validation was carried out on 20 random users in the Jakarta and Bogor city areas to compare the results of the previous and latest recommendations. Respondents were asked to fill in between a scale of 1 (strongly disagree) and 5 (strongly agree) for each aspect or statement. Table 2 shows the statements and their average score.

Table. 2: The Result of Evaluation by Random Users.

Statement	Score
The new recommendation is more in line with the product being viewed	4.9
The new recommendation is more in line with my expectation as an e-commerce visitor	4.7
The new recommendation is better at displaying product collections	4.8
The new recommendation is more helpful in finding similar products	4.6
The new recommendation increases user interest to view and explore the product collections offered	4.5
The new recommendation increases interest in buying products offered	4.5

Based on the table above, respondents agree that the results of the latest recommendations have a higher suitability for the products they are looking for or are interested in. Apart from that, the results of the latest recommendations are also considered to be more capable of introducing the Z product collection and increasing user interest in exploring these products. Respondents were also asked to provide a general rating in the range of 1 (very poor) to 5 (very good) for previous and most recent recommendations. Previous recommendations received an average score of 2.5, while the latest recommendations received a score of 4.5. A score of 2.5 can be categorized as a poor recommendation, while the recommendations from the results of this research are categorized as a very good score. In general, the survey results provide additional validation that the latest recommendations perform better than previous recommendations.

5. Conclusion

This research presented a fashion product recommendation system using content-based image retrieval techniques tailored for the e-commerce domain. Evaluation on a dataset of real-world fashion images showed the model can effectively identify visually similar items with high cosine similarity and SSIM.

The system was successfully deployed to a website, confirming the value of content-based recommendations for creating more relevant and engaging online shopping experiences. A key advantage of this approach is utilizing visual product appearance rather than reliance on metadata or user ratings. This provides a more universal model of perceived similarity. The research makes a contribution by demonstrating a domain-specific application of CBIR and deep learning-based feature extraction for the understudied area of fashion e-commerce. Further work can build upon these findings by exploring additional factors like product categories, testing on larger datasets, and comparing recommendation performance to metadata-based techniques. Overall, this research makes progress in bringing content-based recommender systems into the real-world fashion retail context to enhance personalization and sales.

References

Belaid, O.N, & Loudini, M. (2020). Classification of Brain Tumor by Combination of Pre-Trained VGG16 CNN. *J. Inf. Technol. Manag., 12(2), 13-25.* https://doi.org/10.22059/jitm.2020.75788.

Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, *1*(2), 39–43. https://doi.org/10.33096/ijodas.v1i2.13.

Chen, Y. (2020). Exploring the Impact of Similarity Model to Identify the Most Similar Image from a Large Image Database. *Journal of Physics: Conference Series*, *1693*(1), 1–7. https://doi.org/10.1088/1742-6596/1693/1/012139.

Choe, J., Hwang, H. J., Seo, J. B., Lee, S. M., Yun, J., Kim, M. J., Jeong, J., Lee, Y., Jin, K., Park, R., Kim, J., Jeon, H., Kim, N., Yi, J., Yu, D., & Kim, B. (2022). Content-based Image Retrieval by Using Deep Learning for Interstitial Lung Disease Diagnosis with Chest CT. *Radiology*, *302*(1), 187–197. https://doi.org/10.1148/radiol.2021204164.

Dewi, A. M. S. I., & Dwidasmara, I. B. G. (2020). Implementation Of The K-Nearest Neighbor (KNN) Algorithm For Classification Of Obesity Levels. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, *9*(2), 277. https://doi.org/10.24843/jlk.2020.v09.i02.p15.

Gazzola, P., Pavione, E., Pezzetti, R., & Grechi, D. (2020). Trends in the fashion industry. The perception of sustainability and circular economy: A gender/generation quantitative approach. *Sustainability (Switzerland)*, *12*(7), 1–19. https://doi.org/10.3390/su12072809.

Guercini, S., Bernal, P. M., & Prentice, C. (2018). New marketing in fashion e-commerce. *Journal of Global Fashion Marketing*, *9*(1), 1–8. https://doi.org/10.1080/20932685.2018.1407018.

Kaesmetan, Y. R., & Overbeek, M. V. (2022). Digital Image Processing using Texture Features Extraction of Local Seeds in Nekbaun Village with Color Moment, Gray Level Co Occurance Matrix, and K-Nearest Neighbors. *Ultim. J. Tek. Inform., 13(2), 81-88.* https://doi: 10.31937/ti.v13i2.2038.

Mair, C. (2018). *The Psychology of Fashion*. Routledge.

Park, S., Shin, M., Ham, S., Choe, S., & Kang, Y. (2019). Study on Fashion Image Retrieval Methods for Efficient Fashion Visual Search. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, USA.* https://doi.org/10.1109/CVPRW.2019.00042.

Putrisuryyana, A. P. E. (2021). *Analisis Faktor-Faktor yang Mempengaruhi Nilai Konsumen Terhadap Niat Beli: Studi Calon Konsumen Merek Lokal Kind Denim* [Universitas Multimedia Nusantara]. https://kc.umn.ac.id/16975/

Riyanto, D. Y. (2022). Deeper Understanding of Customer Needs Utilizing Customer Relationship Management. *Journal of Applied Management and Business (JAMB)*, *3*(1), 1–10. https://doi.org/10.37802/jamb.v3i1.241.

Sara, U., Akter, M., & Uddin, M. S. (2019). Image Quality Assessment through FSIM, SSIM, MSE, and PSNR-A Comparative Study. *J. Comput. Commun., 07(03), 8-18.* https://doi: 10.4236/jcc.2019.73002.

Suharyadi, J., & Kusnadi, A. (2019). Design and Development of Job Recommendation System Based on Two Dominants on Psychotest Results using KNN Algorithm. *Int. J. New Media Technol., 5(2), 116-120.* https://doi: 10.31937/ijnmt.v5i2.954.

Sutinen, U. M., Saarijärvi, H., & Yrjölä, M. (2022). Shop at your own risk? Consumer activities in fashion e-commerce. *International Journal of Consumer Studies*, *46*(4), 1299–1318. https://doi.org/10.1111/ijcs.12759.

Statista Research Department. (2023, 5 June). Fashion E-Commerce Worldwide – Statistics & Facts. Last accessed on June 30, 2023, https://www.statista.com/topics/9288/fashion-e-commerce-worldwide/#topicOverview.

Tajuddin, F. N. (2018). Cultural and Social Identity in Clothing Matters “Different Cultures, Different Meanings.” *European Journal of Behavioral Sciences*, *1*(4), 21–25. https://doi.org/10.33422/ejbs.2018.07.67.

Warongan, T. S., Sompie, S. R. U. A., & Jacobus, A. (2018). Penerapan Metode Content-Based Image Retrieval untuk Pengenalan Jenis Bunga. *Jurnal Teknik Informatika*, *13*(3), 1–5. https://ejournal.unsrat.ac.id/index.php/informatika/article/view/28070/27542.

Williams, D. (2018). Fashion Design as a Means to Recognize and Build Communities-in-Place. *She Ji*, *4*(1), 75–90. https://doi.org/10.1016/j.sheji.2018.02.009.