National College of Ireland

Product Matching for E-commerce Platform based on Text and Image Similarity using Deep Neural Network Architecture

MSc Research Project
Data Analytics

Zahra Fathima Sanaullah Shariff
Student ID: x20221207

School of Computing
National College of Ireland

Supervisor: Christian Horn

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Zahra Fathima Sanaullah Shariff
Student ID:	x20221207
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Christian Horn
Submission Due Date:	15/08/2022
Project Title:	Product Matching for E-commerce Platform based on Text and Image Similarity using Deep Neural Network Architecture
Word Count:	7516
Page Count:	24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	S. Zahra Fathima
Date:	14th August 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Product Matching for E-commerce Platform based on Text and Image Similarity using Deep Neural Network Architecture

Zahra Fathima Sanaullah Shariff
x20221207

Abstract

Many businesses are concentrating on their e-commerce nowadays since it helps businesses sell their goods over a wider range and makes it easier to identify their clients and their needs. We all experience situations like taking time to locate favourites products and waiting in line at the register in a store on a regular basis. Also over the online platforms, many sellers are selling similar product by adding the different description of product, tags and titles. In such situation, it becomes very difficult for the e-commerce players to identify the identical products. Therefore, identification of similar product is a necessary task and still an open area of research also automation is another requirement for the advancement of society and the economy since it saves time while also being more dependable than manual processes. Deep Learning has made enormous strides recently, including successes in object detection and image categorization. Matching similar products based on the input image and text which may be tag or label can be recognized by utilizing the convolutional neural network. In this research work, three pre-trained deep convolutional models (MobileNet-V2, VGG-19, ResNet-50) are executed on image and text data to predict the most similar images with each model. Cosine similarity, Levenshtein distance, and custom metric score are some of the performance measure technique has been utilized to identify the most optimal model for similarity detection based on text and image. According to the study, the Mobilenet model is the finest ideal model for dealing with such time-consuming chores and aiding in growing a digital strategic plan.

1 Introduction

With the recent introduction and advancement of various connected technologies such as the internet, several offline tasks have been widely shifted online. Due to such advancement, most offline vendors, sellers, and buyers have opted for the online mode of approach. In today's time, the online transaction on e-commerce portals have been pacing which is also intensifying the competition among various e-commerce portals. Generally, the customers look for attractive pricing, quality, and better service offering. With the shift and acceptance of the people to buy products from e-commerce portals, numerous products are being added to the offerings. Moreover, the general way of driving buyers to the e-commerce portal is by offering various attractive incentives. Although, there are several other internal parameters which shall be considered. These e-commerce portals

must tune the product-based experience to acquire an appropriate customer lead. From a customer perspective, when similar products are available in different portals, they primarily consider the pricing among various e-commerce portals. After pricing comes the quality of the product if any variation in the products under different portals. The customers assess the quality through the product reviews. In certain scenarios, both the pricing and quality may be comparable which may lead to checking the other factor. This may be the service quality of the portals as well. The overall perspective of the customer inclines towards compares, negotiating, and buying behaviour.

The product attributes such as the name of the product, title, and specification have a certain role in driving the potential leads to the e-commerce portal. Therefore, to obtain competitional gain, the e-commerce portals or sellers have shown various efforts in both the advertisement and technological aspects. One such effort is to identify and compare identical products among different portals. Through this, the sellers can understand the potential lead perspective and add the product to the portal with minimal changes to the product attributes. In addition to these, matching products can also be suggested to potential buyers. These attributes of the products can be either in text format or in image format. Although, the text and the image may be similar but not the same. Even in these scenarios, the text can be certainly paraphrased. Furthermore, the text-based search may not be accurate for matching as every product text attribute may be slightly or completely different. To overcome this, the image-based attribute comes which depicts the physical characteristics of the product. Incorporating the visual-based matching of the product can enhance the user experience and develop potential leads for the sellers. Although to match certain products in terms of text or visual, there are various parameters to analyze. As the consequence, it will be a tedious task for the e-commerce sellers to manually sort the identical products and match them. Therefore, we have proposed a robust and autonomous way to identify and match a similar product by implementing Transfer learning algorithm.

By implementing the product match similarity function in the e-commerce portals, the outcome can be considered rewarding for the sellers and provide a convenient user experience to the potential buyers. In addition to these, implementing such a function can drastically reduce the product search time, checkout time and cart abandoning rate. The customers are also provided with a wide range of similar options for their products to buy. Moreover, by implementing machine learning or deep learning algorithms, the accuracy of the output and the experience are also enhanced. For the implementation of the task, an algorithm for text identification and image classification is considered. Primarily, we obtain the set of data from online repositories such as Kaggle. Then the information from this dataset is sorted and feature engineered. After the requisites, the data is trained by implementing certain deep neural network algorithms. The similarity based on text and images is identified by calculating the Cosin Similarity for image and Levenshtein Distance for Text. Based on these two metrics an overall score of similarity has been calculated called as the custom metric. The model having the highest custom metric score will be considered as most effective model for similarity detection. The outcome of these metrics shows the possibility of this model being implemented in real-world applications of e-commerce portals.

1.1 Research Question

RQ1: To what extent the similar products can be identified with text and image of product using the Deep Learning Algorithms?

Sub RQ: Which Deep learning Model can correctly identify the similarity among the products using image and text data? What are the metrics can be utilized for identification of similarity between the two products?

1.2 Research Objectives

Obj1: Identify the Similar product with respect to images using MobileNet, ResNet-50 and VGG-19.
Obj2: Calculate the Cosine similarity and Levenshtein distance with respect to Image and text features vectors for identification of best model.

2 Literature Review

In this section, we will discuss the literature review by researchers on various approaches for product matching for e-commerce platforms based on text and image similarity. This section is further followed into sub-sections namely, Study on Machine Learning Approach for Product Matching using Image and Text, Study on Product Matching using Deep Learning Approach and Study on Hybrid and Pre-trained approaches for Product Matching and text similarity.

2.1 Study on Machine Learning Approach for Product Matching using Image and Text

In a study by Aisha and Monira, a novel machine learning approach for product matching in e-commerce was proposed Alabdullatif and Aloud (2021). The primary aim of this study was to offer accurate product matching and categorization. This paper combined the machine learning algorithm and deep learning algorithm to form a hybrid and robust model. Through this model, the product matched and categorized will be available on a single platform. In addition, the product comparison and analysis will also be enhanced. This novel approach was named as Arabic E-commerce Products Matching (AraProdMatch) framework. It is built up of three processes namely, the data collection phase, feature extraction phase, and product matching based on related clusters. In the data collection stage, the information is scraped from various online stores and obtained as unstructured data. Then these data are pre-processed, and the data features are extracted. Once done, the proposed algorithm is implemented in the data. Due to the unstructured data, the machine learning algorithm is combined with the artificial neural network. After the implementation, the model is evaluated through certain metrics. This model showed promising results. Although, the paper suggested including the image classifier in the future scope of work. In another paper, Cherednichenko et al. (2020) studied the development of the key attributes for product matching based on the item's image tag comparison. In today's e-commerce, people tend to intensively compare the product on different attributes. To serve the purpose, there are various search and filter algorithms. But this algorithm tends to fail as they have certain fallacies. Therefore, in

this study both the image and text attributes were considered for the product matching. Through analysis of both attributes, the crucial feature would be attained, and a similar product would be categorized. The process for this approach consists of data sources and collection, data pre-processing and crowdsourcing, data sampling, and attainment of product key attributes. For the study, the data was sourced from the eBay trading platform. It consisted of two different datasets with items and tags. With various data visualization, different correlations with data were evaluated. This approach achieved an accuracy of 0.74, precision of 0.88 and recall of 0.89. In future scope of work, the paper suggested to implement larger sample and considered combining keywords.

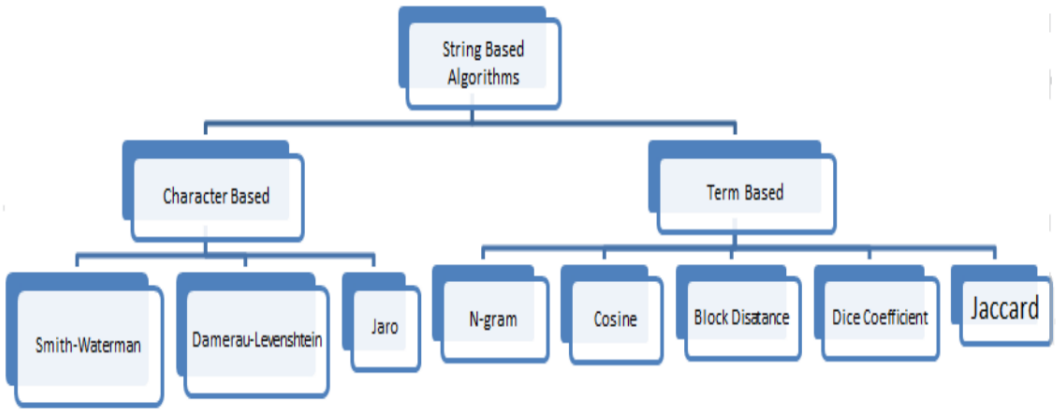

Figure 1: Text Similarity algorithms for Text data

Ko in the paper proposed an approach for product matching through multimodal image and text similarity matching KO (2021). This paper studied the different machine learning implementations for product matching under different perspectives such as vendor product feed and vendor store page. To accurately match the product, the multimodal image and text similarity index approach was considered. This paper was motivated to solve the challenge faced by APPRL, a performance-based influencer marketing firm. The challenge was the inaccurate product match between the vendor's product feed and the vendor store page. During product matching, there are certain crucial parameters which are the title of the product, description of the product, attributes of the product, category of the product and photos of the product. Generally, in the previous approach, the conventional machine learning approaches were considered by implementing the rule-based methods and manually extracting features. For the implementation, in this study, an early fusion supervised Siamese multimodal network is considered. Furthermore, the outcome of the early fusion and late fusion was also discussed in the paper. On the other hand, Pawlowski proposed a machine learning-based product classification for e-commerce Pawłowski (2021). Product classification based on user preferences has various advantages for both sellers and buyers. For the sellers, accurate product matching at their vendor feed will help appropriate inventory management and increase sales through the store page. Considering the buyers will help them in making an appropriate decision for buying a product and presenting their similar product to their preferences. This paper discussed various machine learning frameworks and algorithms which can be

implemented for accurate classification of the product. Dataset was obtained from a polish e-commerce page for this study. Different types of classifiers were considered for the implementation. These were evaluated based on certain evaluation metrics. On the final stance, three classifiers were considered based on the performance which was Linear SVC, Ridge Classifier and Random Forest Classifier. These results showed that upon implementing these classifiers there was a subtle improvement in the outcome. Although the limitation of this study was the lack of correlation between the helper module and the searching or browsing module.

Peter in the paper studied the approach to detecting similar products by utilizing machine learning algorithms Peter (2020). With the pacing shift toward online shopping, there is an increase in online shops which leads to an increase in the products online. Manually sorting the products may be a tedious task which can lead to the inaccurate output. Therefore, this paper proposes an approach to overcome certain challenges and experiences in detecting similar products. In the study, the approach of gradient boosting was utilized which is a type of machine learning. The dataset was obtained from certain online repositories which consisted of dog food category information. Furthermore, the Siamese neural network was also implemented to enhance the accuracy of the output. For the text embedding, various pre-trained models were considered such as word2vec, fast text and more. Moreover, the paper stated that with the increase in the dataset samples, there is an increase in accuracy. In the bottom line, the study of the evaluation of these algorithms showed that the gradient boosting algorithm had a better output. Also, Khan et al. (2022) proposed an approach of recommending products based on visual similarities utilizing machine learning. Initially, the product recommenders were based on text classifications. These worked on the general keywords and preferences of the users and thereby recommended the product. Although the accuracy of these was limited as it depended on certain textual attributes of the product. Furthermore, the manual interference also inhibited the accuracy of the output. Therefore, to overcome the challenge, the implementation of the machine learning algorithms was considered. Through this the model accuracy was enhanced. This paper studied various conventional machine learning algorithms and evaluated the algorithms. Furthermore, the visual similarity approach was also evaluated. In addition to these, several advantages and disadvantages were also discussed in the paper. In the final stance, the paper also discussed the future scope of work.

A similar approach of product similarity matching for food retail using machine learning was proposed by Kerek (2020). The motivation for this study evolves from the need for a similar product suggestion in food retail when a product is out of stock irrespective of the product brand. Therefore, the paper analyzed and compared the best machine learning algorithm for the purpose. The dataset was sourced from matspar repositories. Various parameters were considered such as name, description, nutritional values, weight, category, filter, and brand. Then the data were pre-processed, and different libraries were considered. The similarity values were calculated using these parameters. The similarity value was evaluated using the Jaccard similarity. For the implementation, the model considered were regularized logistic regression, random forest, and SVM. For the evaluation, various metrics were considered such as accuracy, precision, recall, and f-1 score. Additionally, the model was hyper-tuned to enhance the output. The ROC curve and AUC curve were also considered. In another paper, Ristoski also proposed the machine learning approach for product matching and categorization Ristoski et al. (2018). Through this

study, the paper aimed to improvise the product suggestions and advertisement accuracy. The model proposed will harness the semantically structured data and categorize the product utilizing the product descriptions or the textual attributes. The model process consists of the following steps namely feature extractions, calculating the similarity feature vectors and classification. The types of feature extractions include dictionary-based, conditional random field (CRF), CRF with text embedding and image feature extraction model. Then the model was trained using different classifiers which are Random Forest, Support Vector Machine, Naïve Bayes, and Logistic Regression. For the dataset, Yahoo's Gemini product ads data was considered. Then the model was evaluated utilizing the PRF score. The result of the proposed approach was promising, and the model could be implemented in real-world applications as it would enhance the user experience. Similarly, Borst et al. (2020) proposed the CNN-based model for product similarity detection and classification for the product data. Furthermore, transformers were implemented for the extraction of semantic features. Ahsan proposed an approach for visually compatible home décor recommendations utilizing the object detection technique and the product matching technique Ahsan et al. (2021). In this model, when the image of a room is fed, it detects the objects in the room. Then by utilizing the image retrieval technique the products like the ones extracted from the image are suggested in the Catalogue. On the other hand, Rahul proposed the product recommendations using textual similarity-based learning models Shrivastava and Sisodia (2019). The study considered the BoW and TF-IDF algorithms which are text vectorization methods. Although this model was on par with other existing algorithms, it had certain upper hands regarding the accuracy scores and performances. The text classification through BoW (Bag of Words) method is shown in Figure 2.

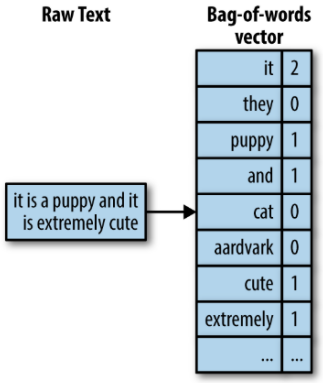

Figure 2: Text Classification through BoW

Kuppili suggested a mechanic-based similarity measure for text classification in the Machine Learning aspect Kuppili et al. (2020). To determine the similarity, the paper considered the Maxwell-Boltzmann method which obtained different characteristics and attributes of the text descriptions. Therefore the proposed approach was known as Maxwell-Boltzmann Similarity Measures (MBSM). This proposed approach consisted of different variants of the KNN algorithm. Furthermore, the method was thoroughly compared with the other algorithms. In the final interpretation, the MBSM model achieved higher performance. On the other hand, Omid did a comparative study on the text similarity in the vector space model Shahmirzadi et al. (2019). The vector space model

included the TF-IDF, topic models and neural models. Various approaches and methods were studied in the different papers. Through the analysis, advantages, disadvantages, and approachability were scrutinized in the study. In addition, the words are classified in different variants based on similarity and retardness as shown in **??**. The paper stated that each study had different fallacies which had to be overcome for the real-world application.

2.2 Study on Product Matching using Deep Learning Approach

In a paper by Gupte et al. (2021), a deep neural network-based approach for multimodal product matching and category mapping was researched. The study was considered for both text and image attributes of the product. Multiple retail samples were collected for the implementation of the study. For the training, Siamese neural network was combined with ResNet architecture and the transformer. On this note, the weighted multi-modal approach was proposed in the study. Previously, the authors of this paper studied only the analysis of the textual attributes. Before implementing the dataset, contextual embedding was done. During the evaluation, it was stated that the proposed model outperformed the conventional models. On the other hand, Jeremy did a detailed comparison of the frameworks for the product matching algorithms Foxcroft (2021). Various machine learning algorithms and deep learning algorithms were discussed and evaluated. These algorithms were evaluated based on different evaluation metrics such as accuracy, precision, recall, f-1 score, AUC, ROC curve and more. Several concepts and viewpoints were also discussed in the paper. The thorough analysis of each algorithm for implementation was evaluated and suggested for the future scope of work. In another paper, Ralph proposed the supervised contrastive learning approach for product matching Peeters and Bizer (2022). The proposed approach is a robust and novel method in computer vision and information retrieval. This method is applied in product matching for the comparison of various products from several e-commerce pages. Utilizing the contrastive learning algorithm, the transformer encoder is pre-trained. For the implementation, three different datasets were considered which are the Abt-Buy dataset, the Amazon-Google dataset, and the WDC computer dataset. The performance of the model in these datasets was promising. Furthermore, the author stated the future scope of work for the enhancement of the model through data augmentation.

Mario proposed a deep learning approach for image similarity in product matching Rivas-Sánchez et al. (2017). It was duly noticed that the model with the analysis of textual attributes was limited to a certain range of performance. Although to enhance the product matching, the researchers considered the visual attributes. To extract the important features of the images, the content-based image retrieval (CBIR) method is utilized. Through this different complex, features are extracted with very large samples. In addition to these, the paper also analyzed different deep learning algorithms for the implementation. Similarly, Huang et al. (2021) proposed a metric learning-based vision transformer for product matching. Through this deep learning algorithm, the crucial features are extracted from the image and the image embedding is learned from the implementation. To achieve intra-class compactness and inter-class dispersion, the concept of arc face loss was implemented here. For the evaluation, the paper compared the proposed model with the Siamese neural network considering various evaluation metrics such as precision, recall, f-1 score, accuracy and more.

2.3 Study on Hybrid and Pre-trained approaches for Product Matching

In a paper by Kertkeidkachorn and Ichise. (2019), a model based on an ensemble pre-trained learning algorithm was proposed for product matching. This novel approach was named as PMap. For the implementation, the paper considered mining the web of Html embedded product data (MWPD). The process for this model includes pre-processing, fine-tuning pre-trained models, and implementing the ensemble model. As a subset, the WDC dataset was initially utilized for the study. Furthermore, fine-tuning the concept of transfer learning was implemented. Here, the ensemble model utilized was RoBERTa which showed a promising result. The outcome of the models was evaluated based on the f-1 score. On the other hand, Roman and Mnich (2021) experimented the bilingual transformers for product matching. The paper utilized the pre-trained transformers which were fine-tuned and implemented for the product matching. It aimed the product categorization in both the languages which are English and Polish. The models implemented were mBERT and XLM-RoBERTa for the task. Furthermore, to test the performance the author here created a custom dataset and implemented it on the trained model. The algorithm considered in this paper were outperforming other conventional algorithms and this could be implemented in real-world scenarios. Łukasik, Michałowski, Kowalski and Gandomi (2021) proposed another approach for text-based product matching with incomplete and inconsistent product descriptions. The paper aimed for an accurate assortment optimization in the retail industry by interpreting the competitor's organization. To build a model for this challenge, the paper utilized the text-mining approach to match and categorize the product even in the case of incomplete and inconsistent product descriptions. Furthermore, the paper implemented the real-world dataset here the evaluate the true performance of the model.

Shah in the paper proposed an approach for product matching and categorization utilizing neural network-based extreme classification and similarity models Shah et al. (2018). The paper is motivated to improve the e-commerce user experience by making the search effective, search engine optimization, providing product reviews and product price estimation. Therefore, the authors in this paper considered two different algorithms which are classification based on shallow neural network and similarity based on deep Siamese network. During the evaluation of the models, it was stated that the proposed model outperformed the conventional models by five percent in terms of accuracy. On the other hand, Juan proposed a deep cross-platform product matching in e-commerce Li et al. (2020). This paper considered the unique and robust neural matching model which will overcome the challenges faced by the conventional models. On evaluation, the model's performance stated it could be implemented in a real-world application.

2.4 Research Summary

After studying all the prior methods for product matching, it has been found that, in most of the studies, the similarity matching among the products has been found either on the image basis or either based on textual data. For image similarity, features of images are extracted using the deep learning models and algorithms such as K-nearest neighbour, K-means clustering has been applied to identify the euclidean distance among these feature vectors, images having the minimum euclidean distance are considered as similar. On

the other hand, for text similarity many common approaches such as Jaccard Coefficient, TF-IDF and Bag of Words (BoW) methods has been used highly. However, a combined study on images and text based similarity between the two products are rarely found. Therefore, in this research a combined approach of text and image will be applied for product matching, where the cosine similarity is calculated for images and levenshteing distance has been calculated among the textual data for similarity detection.

3 Methodology

Nowadays, various companies are focusing on online businesses which supports them in selling their products throughout an extended range and ease the process of recognizing their customers and their requirements. Customers are trusting in buying products through online methods because it provides various similar products according to the need. In this method, a customer can input the image, text, or both of their need and as result, they are offered a large variety of similar products from different brands. A more perfect and relevant system which can show various similar products attracts a large number of customers. In this work, a method is proposed which can predict similar images of the product based on the input image and text which may be a label or tag. In this research work, the main objective is to identify the best deep learning-based model which can predict the most similar matching product based on the image and text or both. Data collecting, duplicate elimination, data pre-processing, data analysis and exploration, model training, and evaluation are enveloped in the uniform flow diagram to accomplish this task. Each process is inscribed in an extended depth in further sections. The framework for Product detection similarity is shown in Figure 3.

3.1 Data Collection

To identify the similar products based on image and textual description, the first step was to identify and collect the correct dataset. Therefore, shopee product matching dataset is collected from the Kaggle website *Shopee - Price Match Guarantee* (n.d.). This dataset consists a set of products having the images along with the label description. There are 32,412 unique images are available in the dataset along with their image title in a csv file. However, along with the image title the dataset also contains the other information such as posting_id, image_phash and label group. All images in the dataset are found to be in png format. On exploring further, it has been found that the dataset has 28,735 unique image_phash values and 11,014 unique label groups are available. The Description of textual data along with attributes are shown in Figure 4.

	posting_id	image	image_phash	title	label_group
0	train_129225211	0000a68812bc7e98c42888dfb1c07da0.jpg	94974f937d4c2433	Paper Bag Victoria Secret	249114794
1	train_3386243561	00039780dfc94d01db8676fe789ecd05.jpg	af3f9460c2838f0f	Double Tape 3M VHB 12 mm x 4,5 m ORIGINAL / DO...	2937985045
2	train_2288590299	000a190fdd715a2a36faed16e2c65df7.jpg	b94cb00ed3e50f78	Maling TTS Canned Pork Luncheon Meat 397 gr	2395904891
3	train_2406599165	00117e4fc239b1b641ff08340b429633.jpg	8514fc58eafea283	Daster Batik Lengan pendek - Motif Acak / Camp...	4093212188
4	train_3369186413	00136d1cf4ededе0203f32f05f660588.jpg	a6f319f924ad708c	Nescafe \xc3\x89clair Latte 220ml	3648931069
5	train_2464356923	0013e7355ffc5ff8fb1ccad3e42d92fe.jpg	bbd097a7870f4a50	CELANA WANITA (BB 45-84 KG)Harem wanita (bisa...	2660605217

Figure 4: Sample Description of Text data

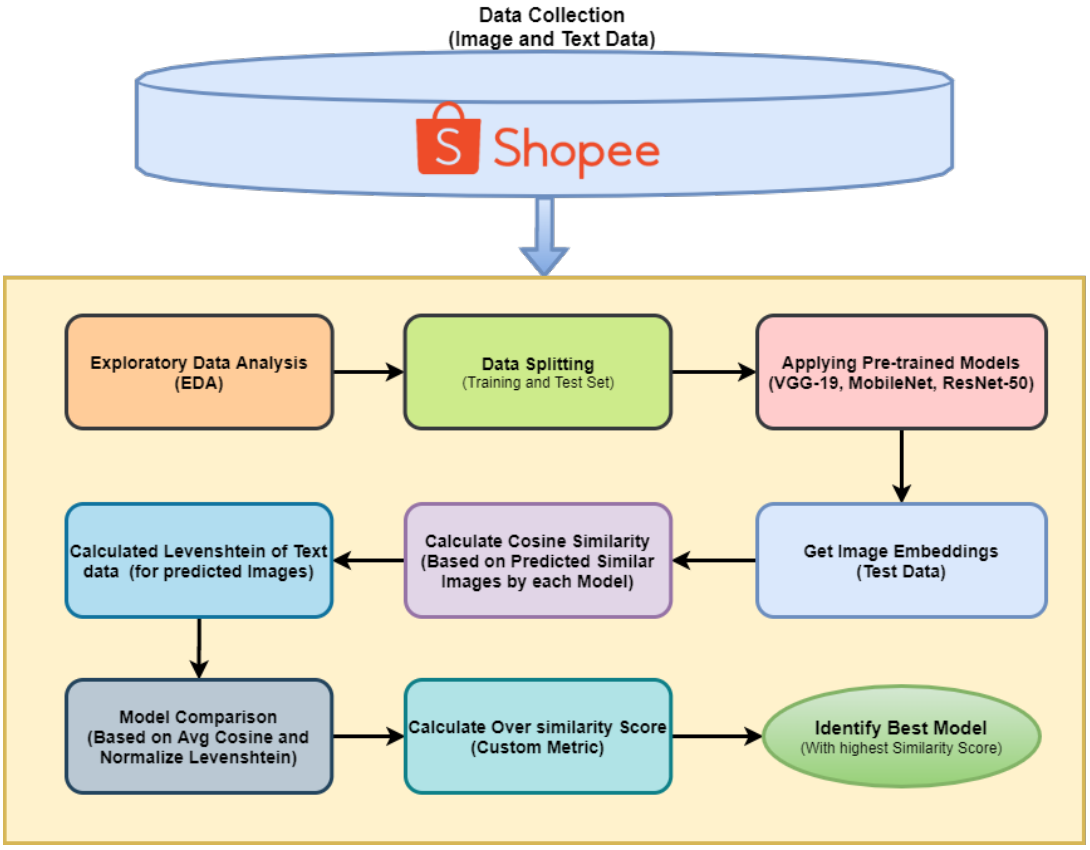

Figure 3: Methodology for Product Similarity Detection

3.2 Data Pre-Processing

After the collection of data from a legitimate source pre-processing step is performed on the available raw data. Pre-processing of the data is required to get the optimal results from deep learning model. Resizing the images is the first step that has been followed in this approach, where each image has been resized to 256 X 256 using tensorflow library, as deep learning model trains faster on the small images. For a larger input image, the neural network model needs to learn a lot depending on the number of pixels and pixel size. Therefore, for reducing the training overhead, resizing images is considering as one of efficient data pre-processing method. In order to extract the image embedding, the images are pre-processed and augmented to reduce the biasing of the models and to make them more generalized. After this step, images are converted into an array and fitted to the models for extraction of image embedding.

3.3 Exploratory Data Analysis

After the preprocessing of the data, data analysis and visualizations are performed on the data. This step allows getting a better insight into the data that helps in selecting the correct hyper-parameters of the algorithm. In this task, first, the visualization of the data is performed based on the label group where the same label group images are visualized as shown in Figure 5.

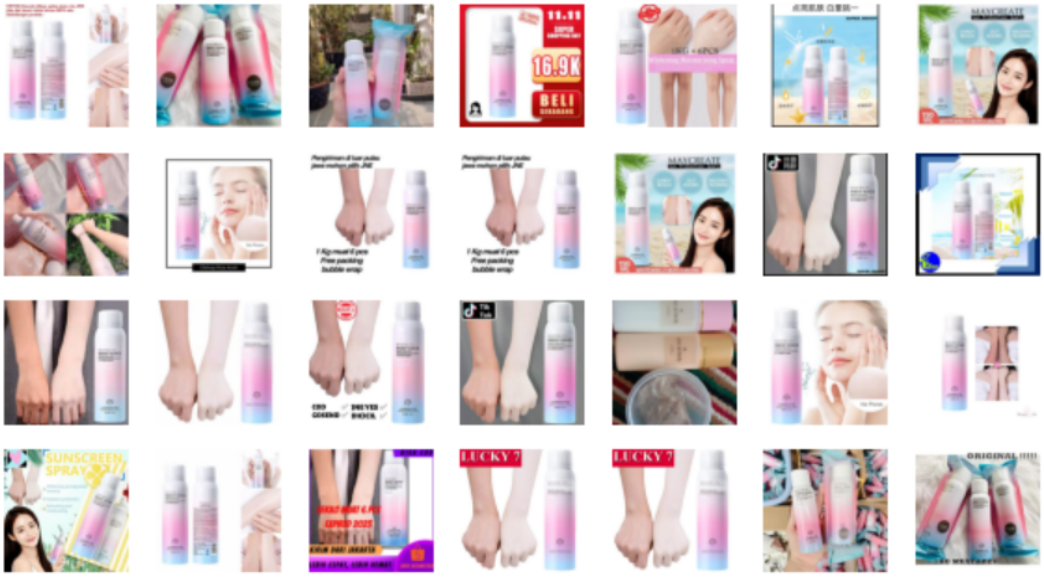

Figure 5: Visualization based on same label group

In order to identify the images with respect to their title, images has been plotted based on product title shown in Figure 6. Same title can have a set of multiple images.

Figure 6: Visualization of Images based on Titles

Images can be categorized into multiple label groups. Therefore, each label group can have the multiple set of images. Bar graph shown in Figure 7 represent the Top-15 label groups based on their image count. After analysing the graph, it has been found that highest number of images in label group is found to be 52.

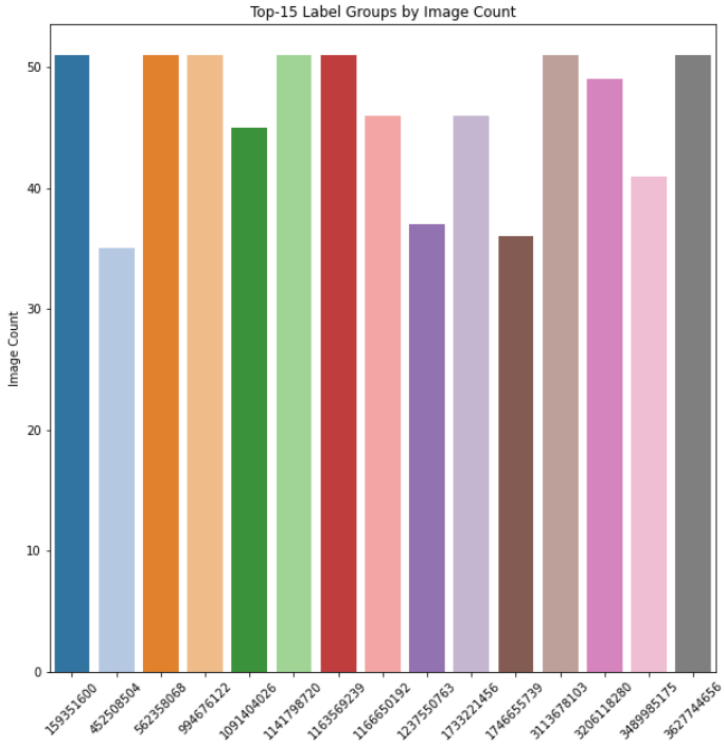

Figure 7: Top 15 Label Groups based on Image Count

3.4 Model Training and Testing

For the training and testing of the model, the first data is divided into training and test data in the ratio of 60:40, where both image data and text data are splitted in the same ratio. This training data is used to train the model while test data is used for testing purposes. Evaluation of the model can be executed only on test data. In this task three pre-trained deep learning models are executed which are Vgg-19, Mobilenet and ResNet-50. In the training of these models, the transfer learning concepts are integrated. In this method, deep convolutional models are trained on large dataset such as imagenet and these weights are stored and can be used further with our dataset.

3.5 Evaluation of Model Performance

Identification of best model with respect to optimal product matching is difficult and crucial task, as there is no standard metric available to evaluate the performance of model for similarity identification. Therefore, in order to assess each implemented model a custom metric is defined, which calculates the overall similarity of recommended product. The similarity of products can be measured by cosine similarity and Levenstein distance. So, in this task, the cosine similarity of the predicted images is calculated and for text similarly, Levenshtein distance is calculated. Based on input image, each model has predicted 5 similar images, where their cosine similarity is calculated. In order to generalize the results, the Average cosine similarity of the recommended similar 5 images are calculated. After calculation of Cosine similarity, the title of recommended products are extracted and the levenshtein distance among the text features is calculated with respect to the title of input image. After calculation of levenshtein distance of 5 different images, the

results are normalized by considering the average levenshtein distance of these 5 products. Model with Maximum Average Cosine similarity and minimum normalized Levenshtein Distance will be considered as most optimal model. Based on these 2 metrics, a custom metric is developed which calculates the overall similarity of product based on input image and text. In order to calculate the overall similarity score, the weight is assigned to each metric in a way that 0.6 is assigned to cosine similarity and 0.4 for the normalized Levenshtein distance because the image of the product will be same throughout all the regions of the world but the text may differ on the basis of language constraints, therefore, a weighted sum of both are considered for evaluation of models. The formula for calculating the custom metric is as follow.

$$CustomMetric = (Avg.CosineSimilarity \times 0.6) + (Avg.LevenshteinDistance \times 0.4)$$

Custom Metric informs about the overall similarity of product based on image and text data with respect to test image.

4 Design Specification

Taking decisions and evaluating patterns of behavior both rely heavily on deep learning techniques. Three pre-trained deep learning models are thus used to anticipate matched goods. One of these is a transfer learning method, which uses a weight that has already been taught to yield precise results. CNN is a deep learning system created to replicate the operation of the occipital lobe in the cerebral cortex. It performs image recognition by treating the picture as a grid with several rows and columns comprised of 0s and 1s. It first organizes the array in a linear way, after which it begins analyzing each piece of a picture by giving it a score. In essence, it is giving every part of the photo more significance, which will aid in identification in the future.

4.1 VGG-19

There are 19 layers in total, 16 of which are convolutional and three of which are fully connected. A convolutional neural network is what this is. Its database is Image Net, a model trained on approximately 1 million images. Image Net can categorise images into over 1000 different categories, including cars, bikes, vehicles, tools, animal types, numbers, and more. It has a good photo category and requires an image size of 224x224 to process. The concept of transfer learning is developed by using this model with pre-trained weights. Transfer learning, as the name implies, involves the transfer of knowledge from one model to another. It works by using one model as the foundation for another, which saves a significant amount of computing power and time because you don't have to start from scratch every time you train a model. Many models include a pre-trained model that can be used as a starting point, with the results being transferable to a new model for improved accuracy and less processing. For instance, if you have a trained model that recognises animals based on their eyes, you can use that model to train a new model that recognises animals based on their ears or noses. VGG-19's architecture is depicted in Figure 3. In this study, we used the transfer learning technique and two CNNs, VGG19 and ResNet15, as pre-trained models. The architecture of VGG-19 is shown in Figure 8.

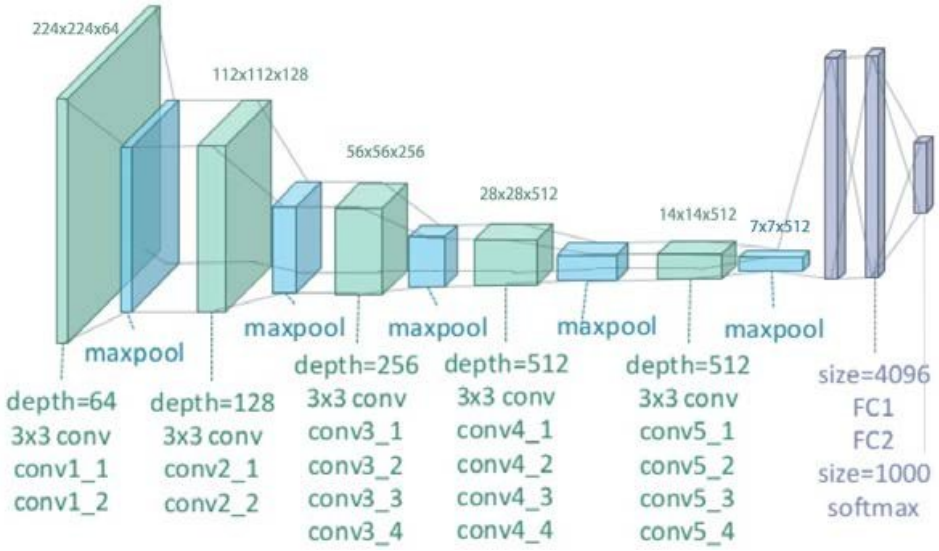

Figure 8: VGG-19 Architecture

4.2 MobileNet-V2

MobileNet is an efficient and portable CNN architecture in real-world applications. MobileNet essentially replaces the traditional convolutions used in older architectures with depth-resolvable convolutions to create a lighter model. Model authors can trade off speed and small size for latency or accuracy by using two new global meta-parameters introduced by MobileNet width multiplier and resolution gain. Deep separable convolutional layers are used to construct MobileNet. Each depth separable convolution layer is composed of depth convolution and point convolution. MobileNet has 28 layers if depth and point convolutions are counted separately. We can reduce the 4.2 million parameters that make up a typical MobileNet by fine-tuning the width factor meta-parameters. MobileNet-V2 Architecture is shown in Figure 9

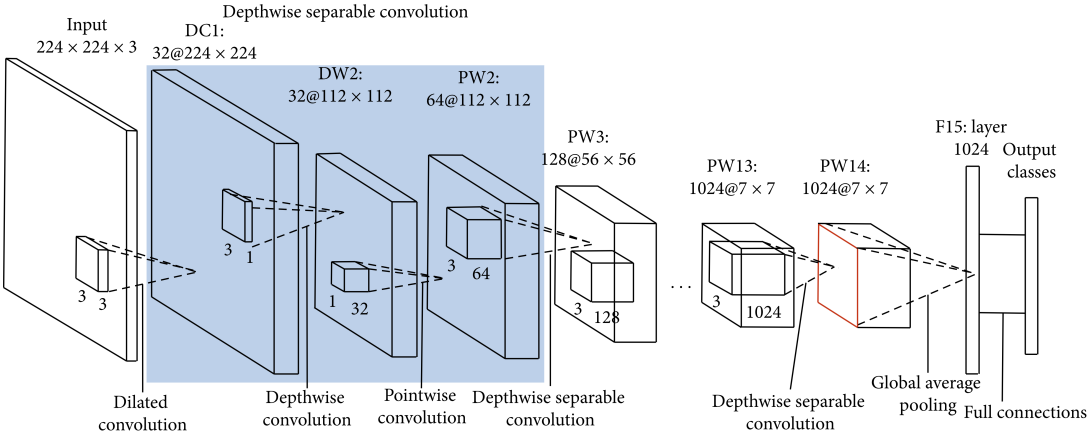

Figure 9: Architecture of MobileNet-V2

4.3 ResNet-50

ResNet, also known as residual network, is a popular neural network that solves the problem of training deep neural networks. Since the invention of ResNet, neural networks with more than 140 layers have become relatively easy to train. Prior to ResNet, the gradients vanished after each successive layer because they were inversely related. This means that after a certain number of layers, performance reaches saturation. ResNet accomplishes this through the use of identity connectivity, a technique that addresses the issue of vanishing gradients. ResNet50's architecture is depicted in Figure 10.

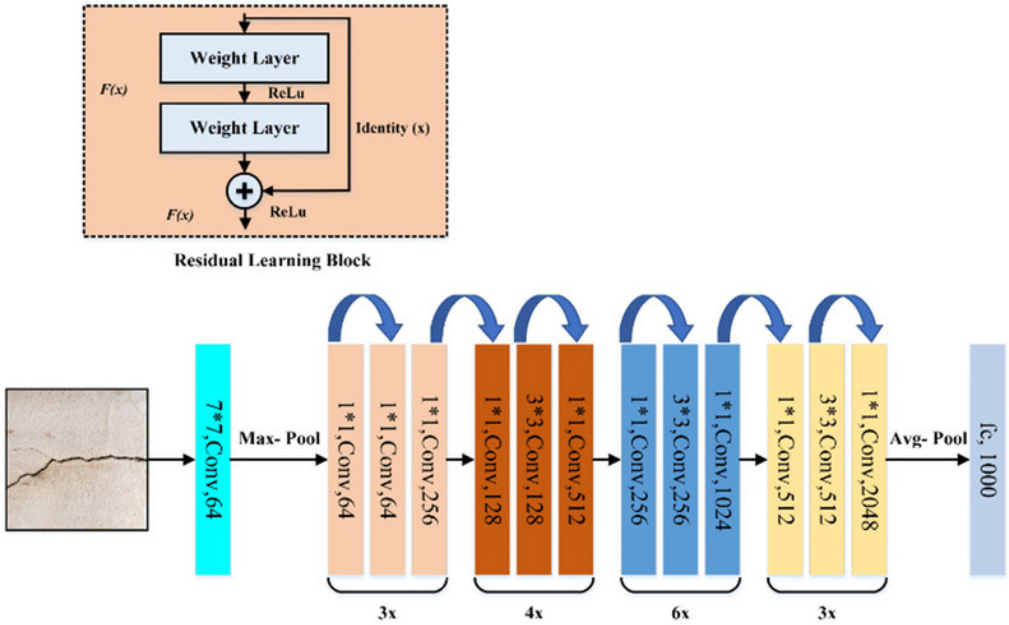

Figure 10: Architecture of ResNet-50

5 Implementations

In this task three deep learning-based algorithms are executed which are Vgg-19, Mobilenet, and Resnet-50, and the superlative model can be carried out whose weighted custom metric value will be the highest. This metric is based on the logic that the cosine similarity for similar images should be more and Levenshtein distance for the text should be less and the weight is assigned to each metric in a way that 0.6 is assigned to cosine similarity and 0.4 for the normalized Levenshtein distance because the image of the product will be same throughout all the regions of the world but the text may differ on the basis of language constraints, therefore, a weighted sum of both are considered for evaluation of models. These all models are incorporated with categorical cross entropy as the loss function and adam optimizer is used for optimizing the performance of the model. Each model is trained over the same data and then tested over the same data and for each model, the metric is calculated. During the implementation, various libraries are utilized which are NumPy, scipy, plotly, sklearn, OpenCV, TensorFlow, Keras, tqdm, etc. Since the proposed process involves CNN, therefore, google collab is used for the training of the models as it provides free GPU services. Here python language is used for the programming purpose. The following specification was required in order to implement

the model.

System Configuration	
Platform	Google-Colab
RAM	12GB
CPU Cores	4
Disk Space	15GB
GPU	Nvidia P100
Programming Language	Python
Python Libraries	Numpy, pandas, CV2, Levenshtein, TensorFlow

6 Evaluation

Here the main objective is to identify similar images based on the input image and the text label to the input image. This objective is achieved by executing pre-trained deep learning models such as VGG-19, Mobilenet, and ResNet-50, therefore it is necessary to assess each executed model based on the cosine similarity, normalized Levenshtein distance, and custom metric. It is a recommendation task, more specifically a multi-class recommendation task so calculating cosine similarity, Levenshtein distance, and a custom metric on test data to evaluate various pre-trained deep learning models. After the training of each model on training data, each model is assessed on test data and all metrics are intended. The model which achieves the highest value of the custom metric is selected as the finest model for recommending the matching products. Bar plots are plotted to visualize the comparison among the implemented models. In this research, the experimentation has been performed over the 3 different test images.

6.1 Experimentation 1 / Test Data (Image 1)

Models such as VGG-19, ResNet-50 and Mobilenet has been implemented and image embedding is extracted using these models by using the train data. Each image is processed through the model and image embeddings are concatenated in the list. Furthermore, based on test image, 5 most similar images are predicted using these models and their cosine similarities are calculated with the respect to the test image. Later, the average of all the obtained cosine similarity has been calculated. The average cosine similairy obtained through each model is shown in Figure 11. The horizontal line among the bar graphs shown in Figure 11, represents the average cosine similarity of all the recommended images.

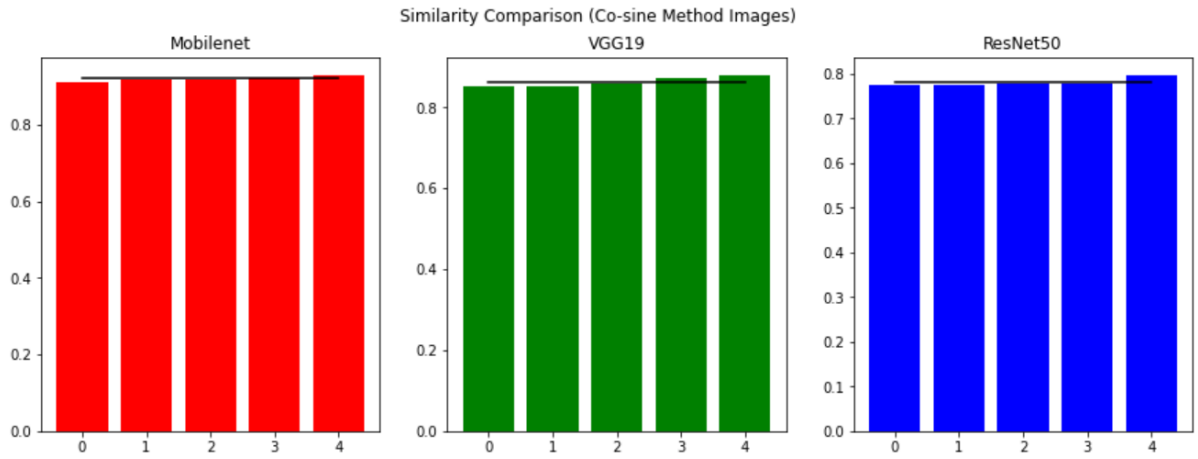

Figure 11: Cosine similarity comparison of Different Models for Recommended images based on Test Image-1

For Mobilenet Model the average cosine distance observed is 0.9205. For VGG-19 the average cosine similarity score is found to be 0.8628. Where for ResNet50 the average cosine similarity is found to be 0.7800. Later based on the recommended 5 images, the text data is extracted and their levenshtein distance is calculated with respect to test image-1. Average normalized Levenshtein distance is calculated for VGG-19 is found to be 0.378, followed by MobileNet where normalized levenshtein distance is calculated values is 0.380 and for ResNet-50 model the normalized Levenshtein distance is 0.462. The Comparative analysis of models based on the Levenshtein distance is shown in Figure 12.

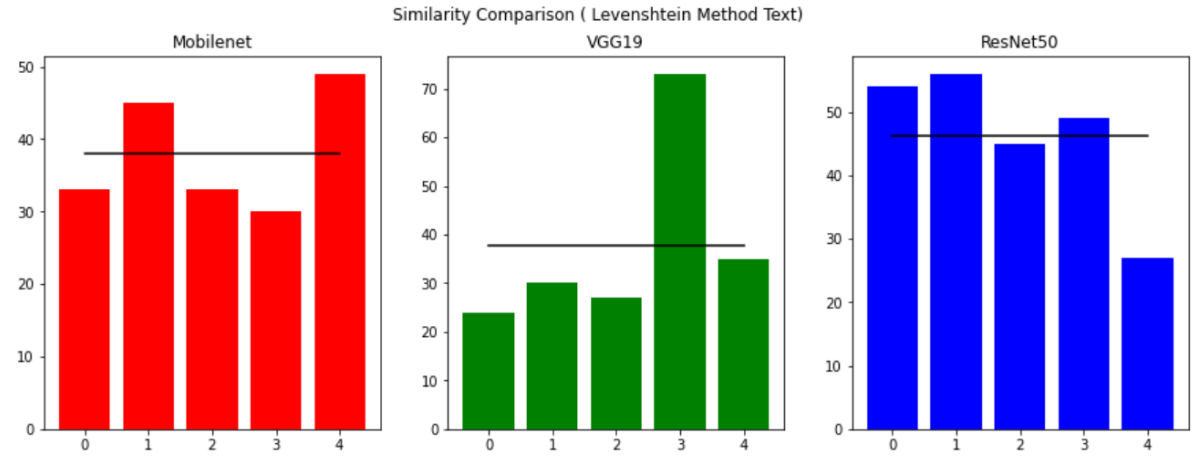

Figure 12: Levenshtein Distance comparison of Model Recommended images based on Test Image-1

Model having the Highest Average Cosine Similarity and minimum normalized levenshtein distance can be considered as the most optimal model. In order to make this calcultion much easier, based on the Cosine Similarity and Levenshtein Distance an overall similarity score has been calculated called as custom metric. The custom metric along with Average Cosine and Levenshtein distance for each model with respect test image-1 is shown in Figure 13.

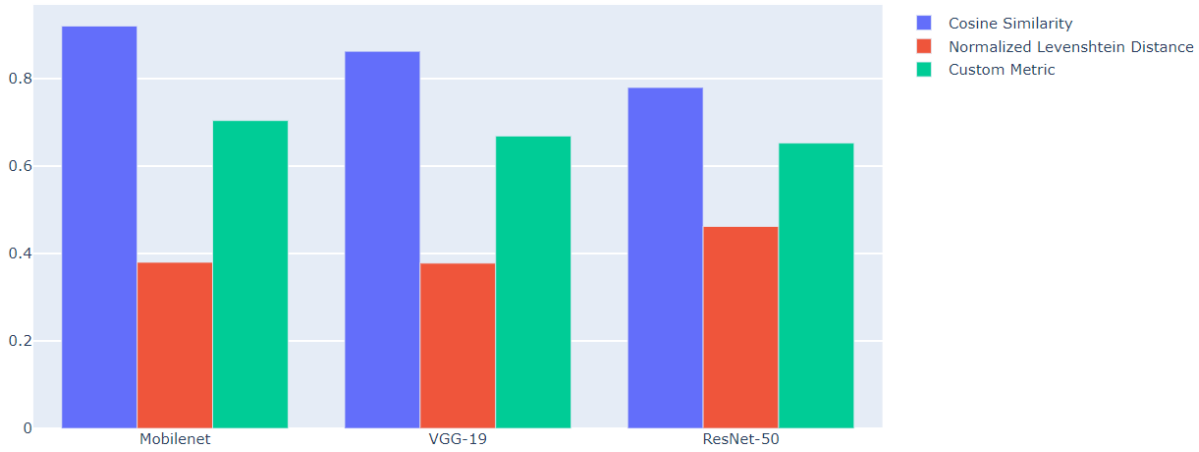

Figure 13: Cosine similarity, Normalized Levenshtein distance and Custom metric comparison of Different Models based on Test Image-1

After analysing the values of bar chart shown in Figure 13 it has been found that Custom metric for MobileNet model is found to be 0.7043. Where, for VGG-19 model the custom metric score was found to be 0.6689 and for ResNet-50 model the calculated custom metric score is 0.6528 and after following the similar steps mentioned above all three metrics were calculated for this model. Model having the high custom metric score will be considered as most optimal model for similarity detection for test image-1, MobileNet model have shown the best results.

6.2 Experimentation 2 / Test Data (Image 2)

In the second experiment, a second image from the test data is considered and five similar products are predicted by each model using the same process utilized in the Experiment 1. For each model, all three metrics have been calculated including cosine similarity and Levenshtein distance and custom metric. The average cosine similarity of all the three models are shown in Figure 14.

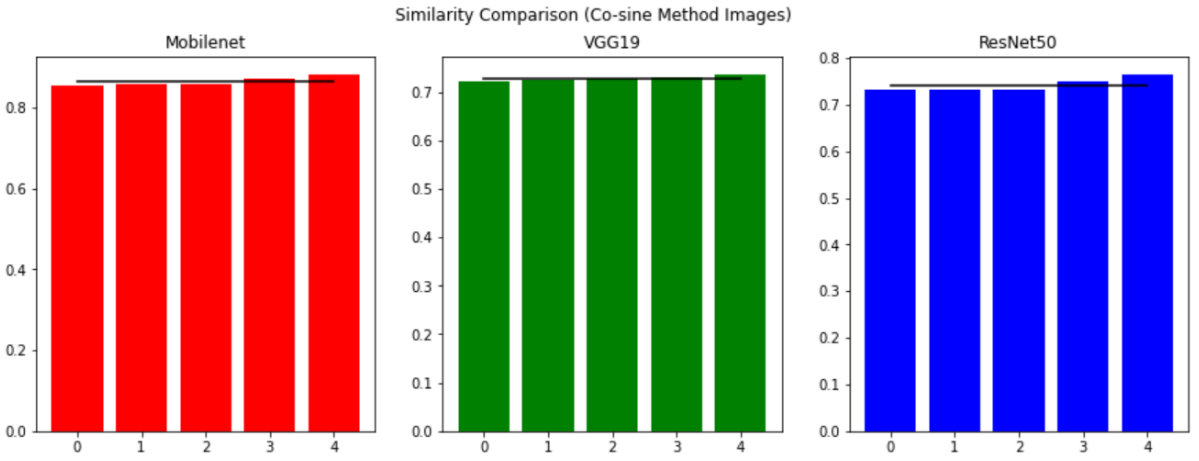

Figure 14: Cosine similarity comparison of Different Models for Recommended images based on Test Image-2

After analysing the graph shown in Figure 14, it has been found that average cosine similarity score obtained using MobileNet model, VGG-19 and ResNet-50 is 0.8640, 0.7286 and 0.7418. Where the maximum Average Cosine similarity score is obtained using MobileNet-V2 architecture. After calculating the cosine similarity the Normalized Levenshtein distance for the recommended images has been calculated. The Levenshtein distance obtained from each model with respect to test image-2 is shown in Figure 15.

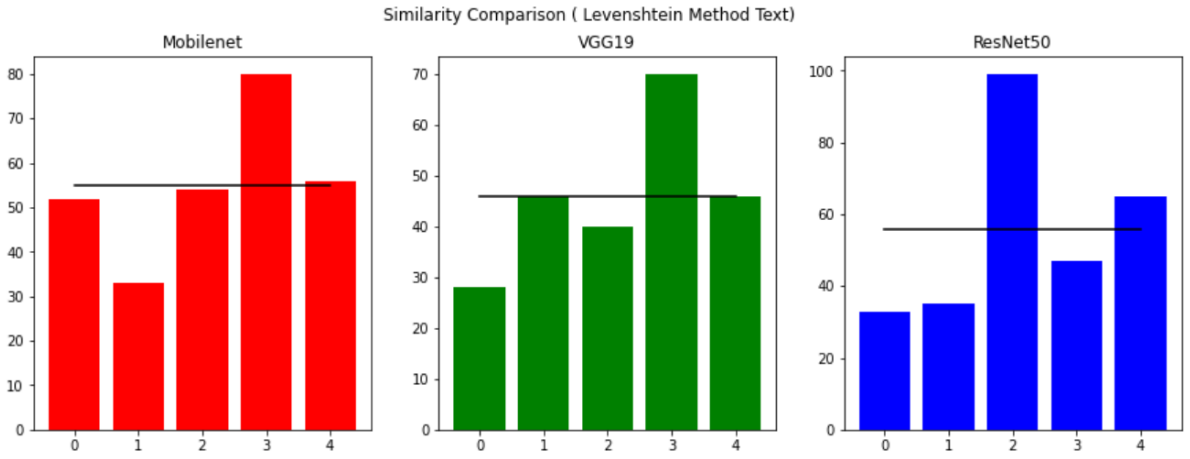

Figure 15: Levenshtein Distance comparison of Model Recommended images based on Test Image-2

The Normalized Levenshtein distance obtained from MobileNet, VGG-19 and ResNet-50 are 0.550, 0.460 and 0.558. Where the minimum Levenshtein distance is obtained using VGG-19 architecture. Based on the calculation of Average Cosine similarity and Normalized Levenshtein distance, the overall score can be calculated, called as the custom metric. The custom metric score for each model obtained is shown in Figure 16.

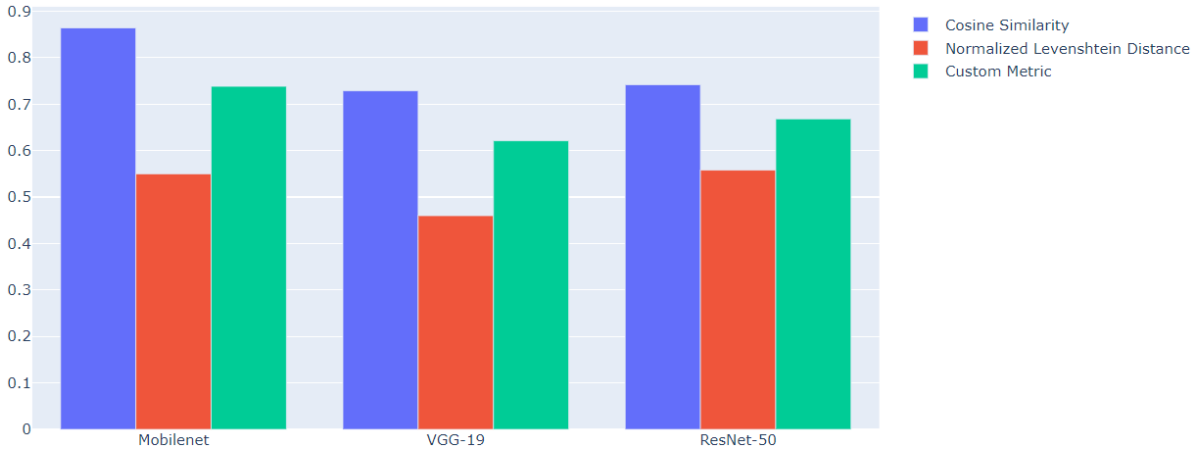

Figure 16: Cosine similarity, Normalized Levenshtein distance and Custom metric comparison of Different Models based on Test Image-2

The Custom metric score obtaied for MobileNet, VGG-19 and ResNet-50 are 0.7384, 0.6212, 0.6682. Where the highest custom metric score (overall similarity score) has been provided by MobileNet architecture, followed by ResNet-50 for Test Image-2.

6.3 Experimentation 3/ Test Data (Image 3)

In the third experiment, a third image from the test data is considered and five similar products are predicted by each model using the same process utilized in the Experiment 1 and Experiment 2. For each model, all three metrics have been calculated including cosine similarity and Levenshtein distance and custom metric. The average cosine similarity of all the three models for test image-3 are shown in Figure 17.

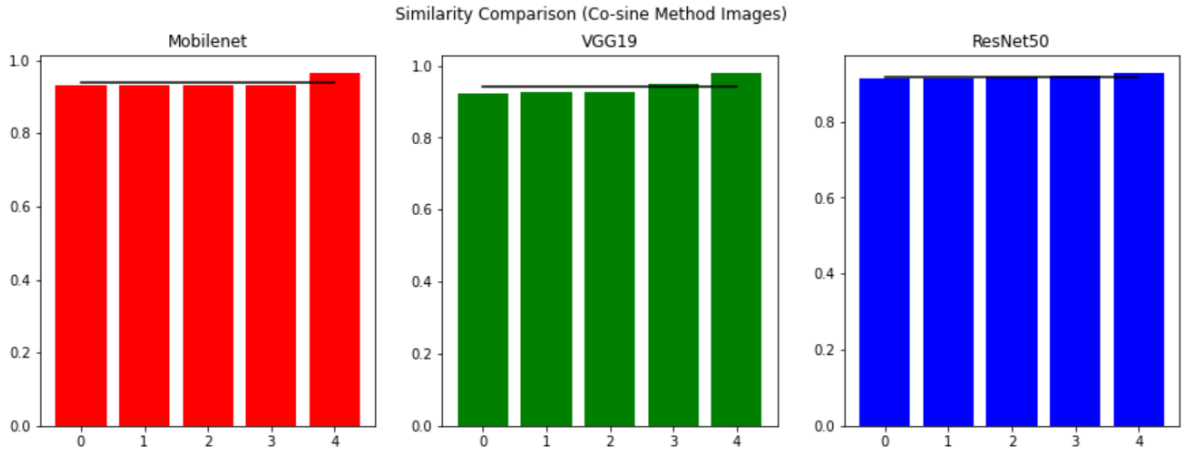

Figure 17: Cosine similarity comparison of Different Models for Recommended images based on Test Image-3

After analysing the graph shown in Figure 17, it has been found that average cosine similarity score obtained using MobileNet model, VGG-19 and ResNet-50 is 0.9375, 0.9318 and 0.9184. Where the maximum Average Cosine similarity score is obtained using MobileNet-V2 architecture. After calculating the cosine similarity the Normalized Levenshtein distance for the recommended images has been calculated. The Levenshtein distance obtained from each model with respect to test image-3 is shown in Figure 18.

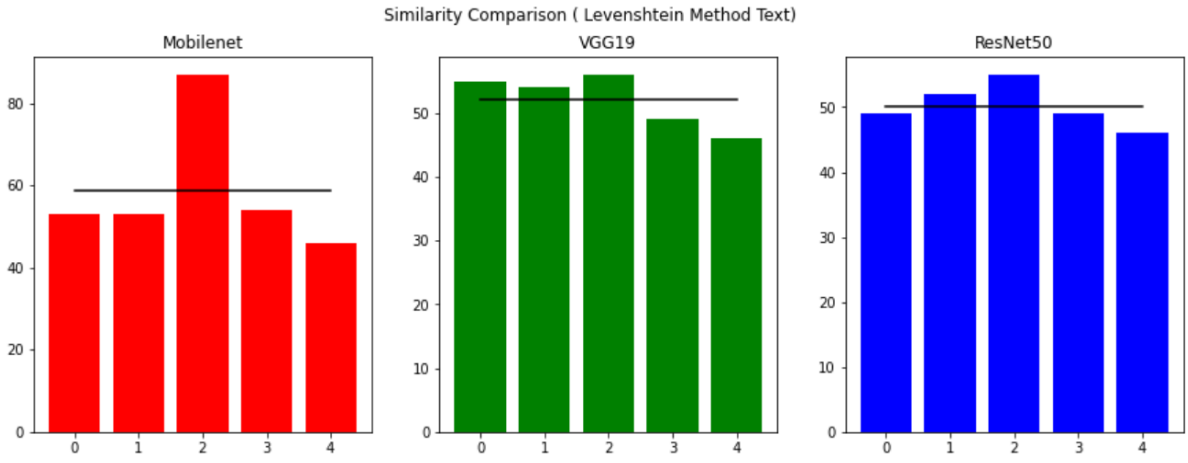

Figure 18: Levenshtein Distance comparison of Model Recommended images based on Test Image-3

The Normalized Levenshtein Distance obtained from MobileNet, VGG-19 and ResNet-

50 are 0.586, 0.520 and 0.502. Where the minimum Levenshtein distance is obtained using Resnet-50 architecture. Based on the calculation of Average Cosine similarity and Normalized Levenshtein distance, the overall score can be calculated, called as the custom metric. The custom metric score for each model obtained is shown in Figure 19.

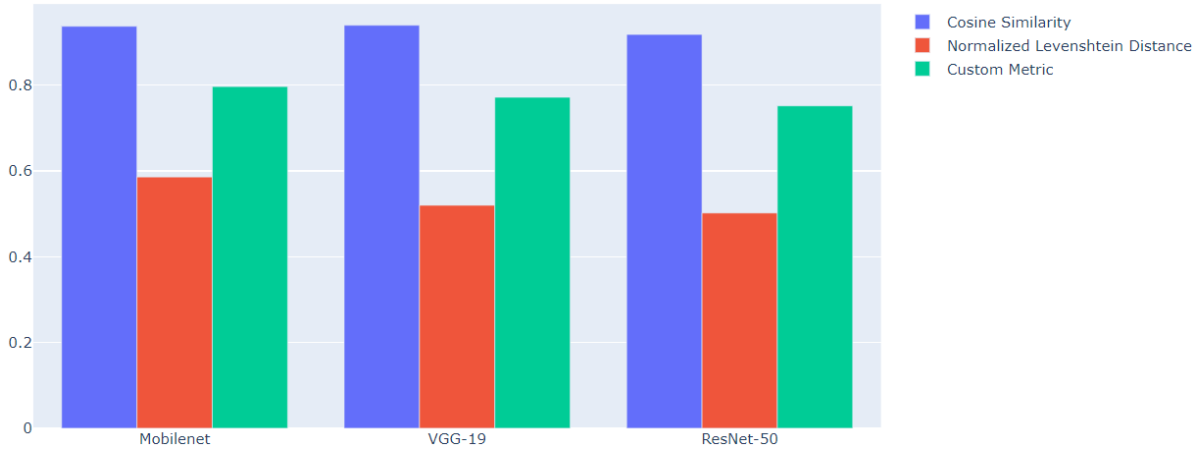

Figure 19: Cosine similarity, Normalized Levenshtein distance and Custom metric comparison of Different Models based on Test Image-3

The Custom metric score obtaied for MobileNet, VGG-19 and ResNet-50 are 0.7969, 0.7719, 0.7518. Where the highest custom metric score (overall similarity score) has been provided by MobileNet architecture, followed by ResNet-50 for Test Image-3.

6.4 Discussion

Prediction or recognition of similar matching products based on the image, text, and both is carried out by training the three different deep learning pre-trained models on the shopee data. Three sets of experiments are executed on test data for testing and evaluating the model performance for similarity prediction. Based on the experiments, the most optimal model for product similarity matching is found to MobileNet Architecture, which outperformed than the other models such as VGG-19 and ResNet-50. MobileNet is a very deep and complex convolutional-layered model, which is lightweight in nature. This model exceeds the other executed model in terms of the evaluation metrics which are cosine similarities, Levenshtein distance, and custom metrics. It is interesting to notice that the other models, Vgg-19 and ResNet-50 have performed well and output satisfying results. From the above analysis, it can be deduced that the mobilenet model is working outstanding in cosine similarity while in terms of Levenshtein distance output is somewhat nearer to other models which are ignorable. Since all models are deep and trained on ImageNet data therefore results of all models are comparable. The matching products predicted by these algorithms are almost similar but the value of the metrics such as cosine similarity is less, Levenshtein distance is higher and the custom metrics value are less than the mobilenet model. Since the size of the dataset is large hence each model takes about 8 hours time to train. The mobilenet model achieved the 0.9205, 0.8640, and 0.9375 values for the average cosine similarity while 0.380, 0.550, and 0.586 are the values for the average normalized Levenshtein distance and 0.7043, 0.7384, and 0.7969 are the score of the custom metrics for experiment 1, experiment 2, and experiment 3 respectively.

The Figure 20 shown below represents the predicted image by each implemented model. It can be analyzed from the Figure 20 that the prediction by the MobileNet model is efficient and the predicted image by this model is most similar to the input image. Since MobileNet-V2 Architecture based model achieved the most prominent results. Thus, it can be used for deployment in the real-world framework.

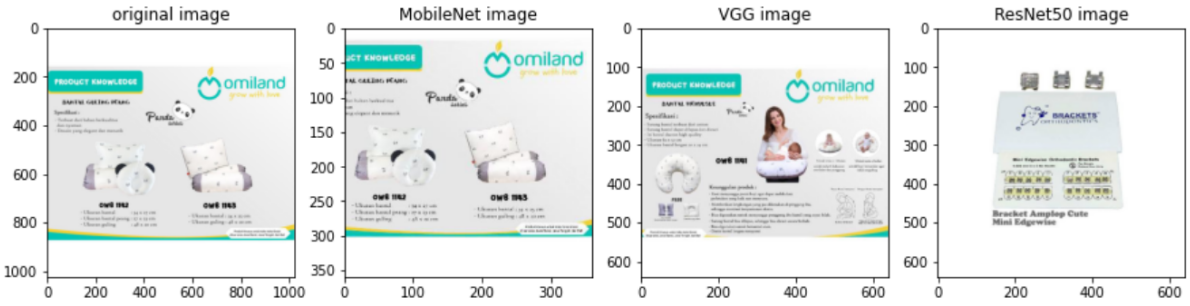

Figure 20: Similar Image Identification by each Model

7 Conclusion

Recently, it has been discovered that a wide range of business titans are turning their product marketing techniques towards online platforms as a result of improvements in a wide range of technologies and broadband services. Digital product sales allow businesses to conduct their operations more specifically, and both sellers and purchasers find this procedure to be convenient, affordable and fulfilling from a variety of perspectives. Consumers are therefore increasingly drawn to platforms that offer benefits like comparable or matching items based on the image and words given by the customers in order to meet their wants and usefulness, according to request analysis. Prediction of the most similar products or identifying the most matching products is still a stretching task. In this research project work, the first pre-trained deep convolutional models are implemented to extract the image embeddings and text embeddings of the images and the labels associated with the respective images and then to identify the surpassing model cosine similarity, Levenshtein distance, and a custom metric is calculated. In the proposed methods the pretrained Mobilenet model is able to predict most similar images or matching products based on image and text with the highest average cosine similarity, average normalized Levenshtein distance, and highest value of the custom metric score. The major aspect of implementing this model is the deep convolutional architecture and lightweight nature of the model. In this prediction and analysis work, a dataset having 32,412 images with their respective labels are used for identifying similar image purposes. This work helped in identifying the most matching products based on the image and text which contains the label or tags associated with the image and fulfills the objective of the similar product identifying the task. Due to the limited number of product images along with their description, the achieved results are somewhat restricted to their optimum results because of limited computing capabilities. In future work, a more complex matching product system can be integrated with the large size data of product images and labeled tags which will surely result in higher accuracy and better results. The time required to train these models will be shorter with greater computing resources. An accurate similarity prediction of product can save lot of time for e-commerce players to classify the

similar product and can be used for recommendation system to help the customer with selection of best choice.

References

Ahsan, U., Wang, Y., Guo, A., Tynes, K., Xu, T., Afshar, E. and Cui, X. (2021). Visually compatible home decor recommendations using object detection and product matching, pp. 214–220.

Alabdullatif, A. and Aloud, M. (2021). Araprodmatch: A machine learning approach for product matching in e-commerce, *International Journal of Computer Science and Network Security* .

Borst, J., Körner, E., Opasjumruskit, K. and Niekler, A. (2020). Language model cnn-driven similarity matching and classification for html-embedded product data.

Cherednichenko, O., Yanholenko, O. and Kanishcheva, O. (2020). Developing the key attributes for product matching based on the item's image tag comparison, *MoM-LeT+DS*.

Foxcroft, J. (2021). A comparison framework for product matching algorithms.

Gupte, K., Pang, L., Vuyyuri, H. and Pasumarty, S. (2021). Multimodal product matching and category mapping: Text+image based deep neural network, pp. 4500–4505.

Huang, L., Wei, S., Wang, F., Xie, W. and Wong, K.-C. (2021). Metric learning based vision transformer for product matching, pp. 3–13.

Kerek, H. (2020). *Product Similarity Matching for Food Retail using Machine Learning*.

Kertkeidkachorn, N. and Ichise., R. (2019). Apmap: Ensemble pre-training models for product matching.

Khan, U., Memon, F., Bhutto, M. and Arain, A. (2022). *Recommending Products Based on Visual Similarity Using Machine Learning*, pp. 261–268.

KO, E. (2021). Product matching through multimodal image and text combined similarity matching.

Kuppili, V., Biswas, M., Edla, D. R., Prasad, K. J. R. and Suri, J. S. (2020). A mechanics-based similarity measure for text classification in machine learning paradigm, *IEEE Transactions on Emerging Topics in Computational Intelligence* **4**(2): 180–200.

Li, J., Dou, Z., Zhu, Y., Zuo, X. and Wen, J.-R. (2020). Deep cross-platform product matching in e-commerce, *Information Retrieval Journal* **23**.

Pawłowski, M. (2021). Machine learning based product classification for ecommerce, *Journal of Computer Information Systems* pp. 1–10.

Peeters, R. and Bizer, C. (2022). Supervised contrastive learning for product matching.

Peter, J. (2020). Using machine learning to detect if two products are the same.

Ristoski, P., Petrovski, P., Mika, P. and Paulheim, H. (2018). A machine learning approach for product matching and categorization: Use case: Enriching product ads with semantic structured data, *Semantic Web* **9**: 1–22.

Rivas-Sánchez, M., Guerrero-Lebrero, M., Guerrero, E., Bárcena-González, G., Martel, J. and Galindo, P. (2017). Using deep learning for image similarity in product matching, Vol. 10305, pp. 281–290.

Roman, A. and Mnich, M. (2021). Test-driven development with mutation testing – an experimental study, *Software Quality Journal* **29**: 1–38.

Shah, K., Kopru, S. and Ruvini, J. (2018). Neural network based extreme classification and similarity models for product matching, pp. 8–15.

Shahmirzadi, O., Lugowski, A. and Younge, K. (2019). Text similarity in vector space models: A comparative study, *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pp. 659–666.

Shopee - Price Match Guarantee (n.d.).
URL: *https://www.kaggle.com/competitions/shopee-product-matching/data*

Shrivastava, R. and Sisodia, D. S. (2019). Product recommendations using textual similarity based learning models, *2019 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–7.
ΩŁukasik et al.

Łukasik, S., Michałowski, A., Kowalski, P. A. and Gandomi, A. (2021). *Text-Based Product Matching with Incomplete and Inconsistent Items Descriptions*, pp. 92–103.