Enhanced Online product similarity classification using description and Images

MSc Research Project
Data Analytics

Sureshkumar Durairaj
Student ID: x21178933

School of Computing
National College of Ireland

Supervisor: Ms. Qurrat Ul Ain

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Sureshkumar Durairaj
Student ID:	x21178933
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Ms. Qurrat Ul Ain
Submission Due Date:	14/08/2023
Project Title:	Enhanced Online product similarity classification using description and Images
Word Count:	9500
Page Count:	21

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:	
Date:	13th August 2023

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):	

Enhanced Online product similarity classification using description and Images

Sureshkumar Durairaj

21178933

MSc Data Analytics

National College of Ireland

Abstract

Online consumers and retailers share a multitude of data with e-commerce platforms, which is part of the world's expanding data fission. Magnanimous research has been on the rise especially concerning the detection of product similarities, this invites a clamour towards a possible expanse for research breakthroughs in product similarity identification. Similar products may have modest variances in their textual descriptions on e-commerce sites, but such descriptions may not be as noticeable as the accompanying images . The standard practice among e-commerce platforms in the past was to rely either on text-based comparisons or image-based techniques. The research of fusing title descriptions and image descriptions, however, is expanding as a result of technical developments. In order to identify a list of identical e-commerce products, this research effort uses deep learning algorithms in conjunction with product titles and photos . The method uses a concatenation of ecomBERT + TF-IDF vectorizer for text modelling and ResNet50 v2 with deep layers and ResNet50 for modelling images. Accurately identifying identical products is the aim of experiments involving picture augmentation, embedding, and language processing. The model used in the study, which serves as the basis for evaluation, is based on a mix of eComBERT and ResNet50v2. Cross-validation scores are used to gauge the correctness of the model, while computing time is used to gauge efficiency . The results of this study will analyze the implementation results and include a comparison study utilizing actual data from a well-known e-commerce business such as shopee.com

Keywords— product similarity match,e-Commerce,TF-IDF,e-ComBERT,ResNet-50v2

1 Introduction

The advent of digital era has resulted in the culmination of the online commerce on a large scale to an extent that we have started consuming the e-commerce services on daily basis. With the exponential growth of online retail platforms, consumers now have access to an overwhelming choices in all business acclaims. Is it not true that we all derive pleasure from securing the most exceptional deals on a wide range of e-commerce products?. Almost the entirety of online commerce have implied multiple strategies to help improve the product matching criterion for enabling the best offers and best prices for the products which they love to purchase with a consequent efforts in improving the customer experience and personalized recommendations as well. Here we have synthesized a systematic and unprecedented approach of combining the image and text data sets for building a system for identifying the product similarity match between multiple products . The proposed mechanism is an attempt to improvise the existing system by comparing the possibilities of combination of image and text based search algorithms using the advanced machine learning techniques inferred from the possibilities highlighted in (Kejriwal et al. 2021) . In order to validate the hypothesis of our proposed system we are using the images dataset used from the works of (Hari Krishnan 2021) as shown in the figure 1 The upcoming sections in this documentation gives us an elaborate picture of the proposed systems.

1.1 Research Background

In the pursuit of cost savings, astute online shoppers are keen to avoid overpaying when a better deal for the same product is available. E-commerce giants across different regions worldwide employ the concept of e-commerce product similarity matching. While some e-commerce platforms offer used items, inaccuracies can arise in the updated information provided by sellers, even due to human errors. To ensure more precise product matching for market classification and pattern descriptions, it becomes essential. This knowledge empowers e-commerce companies to take proactive measures in delivering high-quality consumer experiences. The primary attributes considered for comparison typically encompass various formats, including pictures, text, audio, and video. When it comes to product details, text and photos emerge as the common formats employed similar to the works of (ibid.). Product comparisons are conducted based on factors such as the product title, characteristics like color, structure, type, and product photos. Prominent global e-commerce companies invest substantial time and resources to continually innovate in these areas, leveraging the wealth of available data on their platforms. Furthermore , there has been additional investigation on the similar problem context as per (Sanaullah Shariff 2022) which is a prominent baseline for the proposed system .

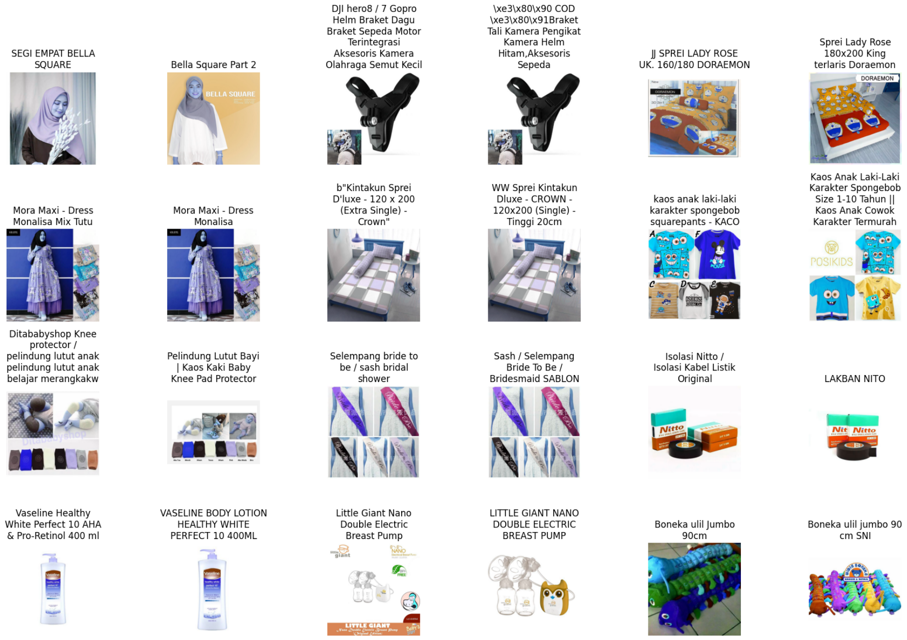

SEGI EMPAT BELLA SQUARE
Bella Square Part 2
DJI hero8 / 7 Gopro Helm Braket Dagu Braket Sepeda Motor Terintegrasi Aksesoris Kamera Olahraga Semut Kecil
\xe3\x80\x90 COD \xe3\x80\x91Braket Tali Kamera Pengikat Kamera Helm Hitam,Aksesoris Sepeda
JJ SPREI LADY ROSE UK. 160/180 DORAEMON
Sprei Lady Rose 180x200 King terlaris Doraemon
Mora Maxi - Dress Monalisa Mix Tutu
Mora Maxi - Dress Monalisa
b"Kintakun Sprei D'luxe - 120 x 200 (Extra Single) - Crown"
WW Sprei Kintakun Dluxe - CROWN - 120x200 (Single) - Tinggi 20cm
kaos anak laki-laki karakter spongebob squarepants - KACO
Kaos Anak Laki-Laki Karakter Spongebob Size 1-10 Tahun || Kaos Anak Cowok Karakter Termurah
Ditababyshop Knee protector / pelindung lutut anak pelindung lutut anak belajar merangkakw
Pelindung Lutut Bayi | Kaos Kaki Baby Knee Pad Protector
Selempang bride to be / sash bridal shower
Sash / Selempang Bride To Be / Bridesmaid SABLON
Isolasi Nitto / Isolasi Kabel Listik Original
LAKBAN NITO
Vaseline Healthy White Perfect 10 AHA & Pro-Retinol 400 ml
VASELINE BODY LOTION HEALTHY WHITE PERFECT 10 400ML
Little Giant Nano Double Electric Breast Pump
LITTLE GIANT NANO DOUBLE ELECTRIC BREAST PUMP
Boneka ulil Jumbo 90cm
Boneka ulil jumbo 90 cm SNI

Figure 1: Sample Data set form shopee.com

1.2 Motivation

In the current economic downturn resulting from the pandemic and ongoing recession, buyers are increasingly selective, opting for new or used items only when they come at a better price. Engaging with buyers is no more a timing issue; it requires active effort to foster engagement. For e-commerce shoppers, finding similar products that offer the best combination of price and quality has become essential. With the widespread availability of resources like internet connectivity and abundant increase in the usage of smart phone across the globe, e-commerce giants such as Amazon and Shopee are experiencing a significant surge in website traffic. However, increased online product searches do not enunciate successful conversions into sales unless the items are offered at the best possible prices. Personally, I recall missing out on some excellent deals for sports goods on an e-commerce platform, as similar items were listed at higher prices due to variations in product photographs, titles, and descriptions provided by different sellers. This context of concern has therefore led me to reflect on the necessity of closely examining the parallels between e-commerce items and the algorithms and models used to detect their similarity.

Additional research, specifically the work of (Rajpoot and Vishwakarma 2023) in this field, throws a wide perspective of the comparative description between varied recommendation system on the likes of similarity techniques in the e-commerce context.The emphasis on the business outcome is quite an important aspect , the focus on customer benefits holds greater significance.As an e-Commerce customer, My own experience and along with similar other shared experiences serves as evidence of the problems that arise when search results fail to display similar products in close proximity, leading to missed opportunities for the best bargains.

1.3 Research Question

The availability of vacuum especially in the area of product similarity match in e commerce platform has therefore led tot the basis of this research which is driven by the following research questions ,

Question 1: How well the proposed system improves the model response time in identifying the similar products matching in the e-Commerce platform?

Question 2: Can the proposed model with a new combination of image and text based methodology improve the overall accuracy of product similarity match in comparison to the existing systems?

The inclusion this question in this report is therefore to signify a means of improving the product similarity of related products in the e-commerce context therefore enabling the user experience on the big picture.

1.4 Research Objectives

- The pre-eminent agenda of this proposed system is to design , orchestrate and build a model for identifying the product similarity matching using the combination of image processing methodologies and text based processing methodology using ecommerce data provisioned from a e-Commerce giant such as shopee.com .

- When making online purchases or even when shopping at physical outlets, customers often compare items to similar products for a fair assessment. While comparing items becomes easier when there are fewer factors to consider, the era of extensive data availability has resulted in comprehensive information being provided for each product offered online. In the retail and commercial sectors, multiple sellers can offer the same items, leading to variations in the rich information provided, including product titles, characteristics, and pictures. This measure can enable e-commerce firms to influence consumer behavior, improve product discoverability, increase sales, establish recommendation systems, and enhance the overall shopping experience by presenting related items.

- E-commerce platforms need a product similarity matching strategy to understand the similarities between the items available on their own platform or on third-party platforms as discussed in the works of (Jain and Hegade 2021) and also considering the cross platform compatible matching capabilities from the likes of (Li et al. 2020) in the e-Commerce context. Further more the analysis has led to the fact of viewing this on a big picture with a generic context specifically the work of (Zheng et al. 2022) in understanding the self learning mechanism of understanding the similarity classification between different labelled categories of data combining both the imagery and textual description.

1.5 Contribution

There have been continuous improvements in the product similarity matching and detection across all the e-Commerce giants, it is obvious that there exists a vacuum to be brimmed for an efficient and scalable methodology on this problem context , which reflects in the fact that it is still an earnest research hypothesis for exploration by e-Commerce and many other business in general . The usage of text and imaged based solutions are available in abundance in the public domains addressing the solutions for the e-Commerce sector; provisioning an ordained combination of Deep Learning and Machine Learning methodologies where in the main intent of this project is to design and develop a deep learning model which will provide a refined similarity classification which aids the customer to search the right product with best offers at the right time enabling the highest level of user using the modernized image and text based classification approach involving deep learning neural networks by treating them independently.This ultimately yields towards better business for the e-Commerce industries.Notable progress in this specific business context in the use case of (Hari Krishnan 2021) which uses the combination of ResNet18 with Siamese twins and TF-IDF vectorisation approach which is further strengthened by the works of (Sanaullah Shariff 2022) in which they have used multiple pre-trained CNN based image processing models by comparing the efficiency of methodologies such as Mobile-Net-V2, VGG-19 and ResNet-50 which are run on the image data sets from e-Commerce using cross validation strategies based on similarity score measures.The distinctive nature of the proposed system is by leveraging the conjunctive approach of the image and text based data sets and processing them by using the individual product labels and images adjoined together to discover comparable goods that have been uploaded in the client side. Allowing the speculation in the building of an effective customer recommendation system going forward .

2 Literature Review

With the advent of the online commercial users it has become quite evident for the necessity of an eminent algorithm which would drive the enhancement of retrieving the impromptu list of products matching the appropriate label description . In the aspect of searching for synthesizing an efficient model , have began the analysis based on the works of (Hari Krishnan 2021) , whereas we have used the same dataset used for the methodology proposed . Furthermore the underlying objective of this work is to identify the similar matching products with the appropriate label description therefore resulting in the enhanced viewing experience for the e-commerce user. As an ardent online shopper myself , often times have been placed in a schema of complexity when we are to look for a particular products in any e-commerce platform yielding a overwhelming range of related products matching the description , this opens the door for further refinement in the product matching thereby enriching the possibility of identifying the relevant offers available for the customers at the right time . The primary modification made to this current research is to include the product item meta heuristics as the prime facet in discovering the match similarity along with the Textual description and Image data set . Also , we have been inspired from certain techniques from the works of (Sanaullah Shariff 2022) . Unlike the former (Hari Krishnan 2021) , here they are considering the similarity measurements such as cosine and jaccard to determine the similarities and dis-similarities which is bench marked by the cross validation score .

2.1 Analysis and discussion on the current work

The current research analysis begins by examining the work of (ibid.) ,who utilized a data set from Shopee.com available on Kaggle for academic purposes. They employed CNN models, specifically TF-IDF (Term Frequency Inverse Document Frequency) Vectorizer, to determine the significance of key aspects in text data, represented as a corpus. The TF-IDF technique helps identify word frequency in the product titles. Additionally, they utilized ResNet-18, a Residual Neural Network with 18 layers, and ResNet-50, a deeper model with 50 layers pretrained on ImageNet, to handle complex image processing tasks related to various objects like computer accessories, fashion, electronics, lifestyle, etc. Furthermore, the researchers used Siamese ResNet-50 to measure the similarity scale between two objects. This method compares the final layers of two identical neural networks to determine the similarity. They performed comparative analysis experiments to assess the performance of different combinations of research methods, with ResNet-18 + TFIDF emerging as the optimal model with a Cross Validation score of approximately 0.73.

Another relevant study by (Sanaullah Shariff 2022) was considered, where they employed three deep learning methods, namely MobileNet-V2, VGG-19, and ResNet-50, trained on image and text description data similar to our research context. They also explored the use of Cosine Similarity, Levenshtein distance, and a custom metric score to improve performance. Among the methods, MobileNet achieved the highest custom metric score of 0.7969, indicating better performance for this research context. However, both studies were conducted independently without leveraging the potential benefits of combining algorithms. Therefore, in our research, we propose to use various combinations of ResNet for processing the image data set and multiple combinations of BERT for processing the text description data from Shopee.com. The goal is to design an effective model that can accurately identify similarity matches between identical or similar products.

2.2 eCommerce based Similarity classification

The conventional approaches that have been Incorporated in the product classification and match identification specifically on the lines of online commercial systems have isolated itself from the usage of BERT algorithm . Despite ebring a renowned NLP algorithm has however has its own set backs . This has been clearly described in the work of (Zhang et al. 2020) . The author in (ibid.) further expands the indispensability of a new technique since the BERT lacks in two levels of domain knowledge namely phase level and Induction level . The former could actually enhance the NLP modelling on the whole however this has the residual noise as a result of this overwhelming knowledge . They are introducing a concept such as adaptive hybrid masking . On the other hand for utilizing the product level knowledge they employ a neighbour product reconstruction strategy which leads the proposed mechanism E-BERT for the prediction of an individual product;s associated neighbour using a noiseless cross retention layer by considering the parameters such as review-based question , answering, aspect extraction, aspect sentiment classification, and product classification. They have used e commerce corpus such as the amazon data set validating their proposed system which has resulted in the comparison of various versions of BERT on the corpus data set such as RAW BERT, BERT NP , Span BERT , E- BERT DP , E-BERT AHM and E-BERT of which the proposed mechanism E-BERT, has proven to yield the maximum performance across the various downstream tasks as listed above with product classification task acquiring the maximum accuracy of 78.4% . This experimental approach has therefore motivated in our proposed system to use the E-BERT for enriched performance specifically in the context of product match identification and classification .
The product category matching has been particularly addressed in the works of (Kejriwal et al. 2021), mainly towards the enhancement of search and recommendation systems in online contexts. It introduces a carefully designed set of guidelines and methodology to acquire annotations in a cost-effective and reliable manner. This methodology is essential for evaluating various existing and future models for product category matching enabling a systematic approach to acquire annotations, the paper (ibid.) enables researchers and practitioners to make more informed decisions and compare different product category matching methods effectively. Additionally, they also present a methodology to compare solutions from two or more product category matching methods. This comparison includes both pre- and post-annotation evaluations, ensuring a comprehensive analysis of the models' performance. The authors demonstrate the effectiveness of their proposals using three widely used e-commerce product category taxonomies and multiple metrics, making their findings more generalized and applicable to real-world scenarios. Overall, the paper is valuable in its contributions, as it addresses an important problem in the context of e-commerce and digital marketplaces. The guidelines and methodology for acquiring annotations are likely to be of significant interest to researchers and practitioners working on product category matching and related tasks. The use of widely-used taxonomies and multiple metrics to validate the proposed approaches enhances the credibility and applicability of the findings. However, it's important to consider potential inferences from this work towards orchestrating an effective product category matching , therefore this has led us towards the usage of CNN / Transfer learning verctorization techniques for product category match .

The complexity of entity matching tasks is influenced by various factors, including the number of challenging corner-case pairs, the ability to generalize to unseen entities, and the size of the development set as discussed in the works of (Peeters, Der and Bizer 2023). Existing entity matching benchmarks typically focus on single dimensions, such as the amount of training data, but fail to address the crucial aspect of generalizing to unseen entities. To address this gap, a new benchmark called WDC Products is introduced. This benchmark evaluates entity matching systems across three dimensions: corner-cases, generalization to unseen entities, and development set size, using real-world data from diverse product datasets collected from multiple e-shops, each marked up with schema.org annotations. The evaluation formulates entity matching as both pair-wise and multi-class classification tasks, enabling direct comparisons between the two formulations. The benchmark is evaluated using state-of-the-art matching systems, including Ditto, HierGAT, and R-SupCon, as well as several other supervised entity matching methods: word (co-)occurrence baseline, Magellan, RoBERTa-base, and HierGAT. The experiments demonstrate that all systems struggle to varying degrees with unseen entities, and some models are more data-efficient than others. There are certain insights which can be derived from the above such as increasing number of corner-cases, highlighting the challenge of handling such cases. Deep learning-based systems, including R-SupCon, Ditto, and RoBERTa, perform relatively well, achieving F1 scores between 72.18 and 79.99 for variants with 80% corner-cases.All methods exhibit decreased performance when tested on unseen entities, but deep learning approaches outperform symbolic baselines, even for unseen products. Much

Smaller data size pose a challenge for algorithms to calibrate and acquire higher accuracy , resulting in unreliable accuracy rates up to 78%. R-SupCon demonstrates a 3-6% performance improvement over its pair-wise counterpart, indicating its suitability for multi-class matching to recognize known products as described in (Peeters, Der and Bizer 2023).WDC Products serves as a comprehensive benchmark for entity matching, offering evaluation along three crucial dimensions and supporting both pair-wise and multi-class formulations. It presents a challenging evaluation for advanced matching systems, especially concerning the generalization to unseen entities, a dimension that previous benchmarks failed to address adequately. The inclusion of the seen/unseen dimension enhances the benchmark's utility for evaluating entity matching systems in real-world scenarios. Thee experimental results from these have considered the specific data sets from single domain perception , this limitation can be eliminated and applied upon diverse category of data sets similar to the data set we have considered .

The essence of culinary artistry has opened up a world of countless possibilities. As food enthusiasts explore various online platforms to find the most delightful recipes, they yearn for an interactive system that optimizes their cooking journey. To achieve this, the culinary community is increasingly turning to the power of advanced recipe recommendation systems. In a groundbreaking article (Abluton 2022), the focus is on leveraging cutting-edge multi-modal capabilities to revolutionize how users discover recipes. The experimental approach involves user engagement, where individuals can upload images of the dishes they wish to explore further. Using sophisticated image processing algorithms like Yolo, the system provides them with visually similar recipes. This visual recommendation and search approach prove remarkably effective, particularly for smaller to medium-scale cooking platforms. The heart of this technique lies in employing KNN (K-Nearest Neighbors) to search for similar images within the vast recipe dataset. Interestingly, many culinary platforms offer a plethora of multi-modal features that remain largely untapped. An article by (Mehta et al. 2022) delves into a similar approach, emphasizing the significance of incorporating fashion BERT (Bidirectional Encoder Representations from Transformers) and CNN (Convolutional Neural Network) based models. Addressing a common challenge faced in the world of cooking, another e-commerce journal (Hendriksen 2022) highlights the mismatch between textual recipe descriptions and corresponding images. The authors propose a novel method called CTIR (Category-to Image Retrieval) and FGTIR (Fine-Grained Text-to-Image Retrieval) utilizing BERT for text classification and match detection. By leveraging these techniques, the study evaluates and compares performance across multiple culinary categories online.

Intriguingly, the research community finds that BERT is an ideal candidate for culinary research due to its enhanced ability to access multi-modal capabilities, as corroborated by (Tracz et al. 2020). Additionally, an article (Le and Hinneburg 2022) concentrates on product similarity in e-commerce, ensuring that the recommended recipes are not only popular among users but also closely related to the initial choice. They employ a classification model for entity match and various image processing methods like Siamese CNN and Ditto model, achieving an impressive F1 score of approximately 87. A pivotal aspect in the culinary research context is recipe matching, which leads to further exploration in (Peeters and Bizer 2022). This article explores the use of a pre-trained transformer-based model in a supervised learning environment. By employing source-aware sampling, the authors significantly boost the model's performance. Notably, AptBuy reaches an impressive F1 score of about 94%, while Amazon-Google achieves approximately 79%. Although these techniques are most effective for smaller datasets, the use of F1-score as a performance measure remains valuable and insightful for the culinary community.

2.3 Evolution of Image based Similarity Classification

In the semantics of E commerce platform , learning with few labeled data has always been an existing issue, therefore a need for the construction of a semi supervised learning heavily reliant on the similarities such as semantic similarity and instance similarity is required . This is expected to yield a consistency regularisation using various augmented views of same instance to same class prediction In the work of (Zheng et al. 2022) they have proposed a system using semantic similarity match . In their experiments they have discovered that the sim match would improve the performance of a semi supervised model orchestrated using 400 epochs with the sim match yielding to 67.2 % and up to 74.4% with the top most accuracy with 1% and 10 % labeled data sets arrived and extracted from imagenet . They have used CIFAR-10 and CIFAR-100 data set sim match using WRN28-2, and WRN28-8 respectively followed by ImageNet containing 1k data samples . After which they have also performed SimMatch on the large-scale ImageNet-1k dataset to show the the superiority. The base model is constructed using

ResNet-50 for processing the images on the above described dataset and used a standard SGD optimizer with Nesterov momentum. Here in they have used the method , InfoNCE SwAV SimMatch with Top-1 dataset yielding an accuracy of about 53.5%, 49.7% and 61.7% respectively . These experimental results from (Zheng et al. 2022) has therefore paved way for the usage of Resnet 50 for our research context mainly because of its similarity in the e-commerce business domain.

The Residual Network is considered to be one of the effective deep learning methodology when it comes to product similarity match detection . This article on the Residual Network (Hanif and Bilal 2020) entails further broader understanding on the same . Considering the shortcomings of the ResNet , one of the most cons is the presence of the deep layer stochastic gradient , when there are no explicit regularization . The modernization and recent advancements in the Residual Network has effected in the convergence of gradient flow of the parameters. Also , (ibid.) states that the Residual Network is highly perfromant when compared to AlexNet,GoogleNet,VggNet and MobileNet respectively . Since these ResNets are basically RNN with unique identity mappings. It uses a custom based CoRN on CIFAR data-set for image processing . This shows that out of all the methodologies ResNet with about 110 layers deep neural network on CIFAR-10 datset produces only 1.7% error rate. From the above study and analysis and also considering (Hari Krishnan 2021) , the best way is to go forward with the combinative comparison approach by using ResNEt50 versions along with BERT for Vectorization discusses in the section 2.4 below .

2.4 Textual Description based Similarity Classification

Since we are dealing with online merchant based similarity match detection based on Image as well as textual description . It is important to begin with BERT classification - Bidirectional Encoder Representations , This article (Tracz et al. 2020) deals withe the implementation of BERT based methodology for product matching and offer matching for an e-commerce data-set by applying a transformer based deep learning architecture with appropriate sampling technique which significantly improves the performance for a varied list of e-commerce. The product matching problem is address by using the zero-shot strategy in order to avoid the heftiness of retraining the model . The consideration of offer matching vs matching and non-matching product is carried out . This is improved by adjusting the network parameters to minimize the triplet loss objectives . The usage of context specific BERT such as eCOmBERT when used in combination with using Category Random(BH) , Category Hard(CH) and Batch hard(BH) strategies respectively yielding an accuracy of 93% when used with CH . The article (Abolghasemi, Verberne and Azzopardi 2022) uses QBD in which the seed document is used as the query and the objective is to retrieve the documents relevant to the search seed. The multi-task optimization is proved to develop the performance thereby increasing the ranking performance . These experimental analysis are intended to effect an early risk prediction of the user in an attempt to enhance the user experience which can be included as our future scope.

With the introduction of advanced NLP tasks introduced by contextual bindings a new form of BERT called prod2BERT is trained to generate representations via masked sessions . In this article (Bianchi, Yu and Tagliabue 2020) they have compared the performance of prod2BERT and prod2vec in terms of accuracy metrics . They have subjected the considered dataset through various experimentation such as Next event prediction , intent prediction using multiple methods such as zero encoding , Three byte encoding , concatenation, linear weighted combination of all hidden layers and fine-tune for both prod2BERT and prod2vec . Though the experimental results were effective , however they have not considered image dataset rather only text based dataset and methods for evaluating the text vectorisation is considered . Consequently , the essence of an effective BERT is inferred from the findings of (ibid.) .

2.5 Final Discussion and Inference

From the above analysis on various research articles on the existing methods , it can be inferred that for our research we can use ResNet methodologies for image processing based similarity match detection and BERT for text based vectorised approach and use the performance measure such as Cross Validation score , F1-Score , Cosine Similarity , Levenshtien distance and few other metrics as evident from the comparison of the literature works as shown in the table 1.

Comparison of the Previous Works			
Author	Techniques Used	Results	Run Time
(Hari Krishnan 2021)	ResNet-18 TFIDF ResNet-50 ResNet-50+TFIDF Siamese ResNet50	0.652 0.613 0.663 0.734 0.712	01:45:00 00:19:00 00:42:00 01:52:00 00:52:25
(Sanaullah Shariff 2022)	Mobilenet VGG19 ResNet50	0.9205 0.8628 0.7800	8h
(Abolghasemi, Verberne and Azzopardi 2022)	SPECTER SPECTER w/ HF BM25 BM25optimized BERT MTFT-BERT	83.6% 83.4% 75.4% 76.26% 85.2% 86.2%	2h
(Tracz et al. 2020)	BOW eComBERT-NFT eComBERT	0.8016 0.6656 0.8873	2h

Table 1: Highlights from the Previous Work

3 Research Methodology

This paraphrase primarily focuses on describing the methodology, models, and possible assumptions used to develop an effective algorithm for determining similarity classification. It also elaborates on the details of the framework considered for this study. To simplify matters, a five-phased KDD (Knowledge Discovery Database) approach is being adopted as shown in the Figure 2. This framework is essential to ensure a refined and comprehensive approach to address the problem successfully, providing a roadmap for the entire process. The framework is specifically designed to extract data in a detailed manner to effectively solve the problem. It outlines the process, starting from data extraction to finalizing the similarity classifying model, enabling successful resolution by evaluating the results. The proposed mechanism involves two stages of research to build the similarity classification model and evaluate the results iteratively for effectiveness. In the first stage, ResNet-50V2 (an RNN-based Deep Learning Neural Network for Image-based similarity classification) is used, and the processed results are stored in the Knowledge Base (KB). The second stage extends the classification and analysis by subjecting the results to the Textual Description-based similarity classification algorithm, e-BERT (BERT)+Tf-IDF. The similarities obtained from both classification strategies are compiled and integrated to produce a refined model, enabling customers to identify matching and non-matching products and offers for a given product. The following sections describe the details of each step involved in the process and, therefore, the outcomes of the proposed system.

3.1 Data Collection

The dataset used for this analysis is sourced from the Shopee online commercial business organization and is publicly available for academic research and Kaggle competitions since 2020. The dataset is not directly provided as an image and text set; instead, researchers are provisioned with a link to facilitate the data extraction from the Shopee Website. The considered dataset consists of approximately

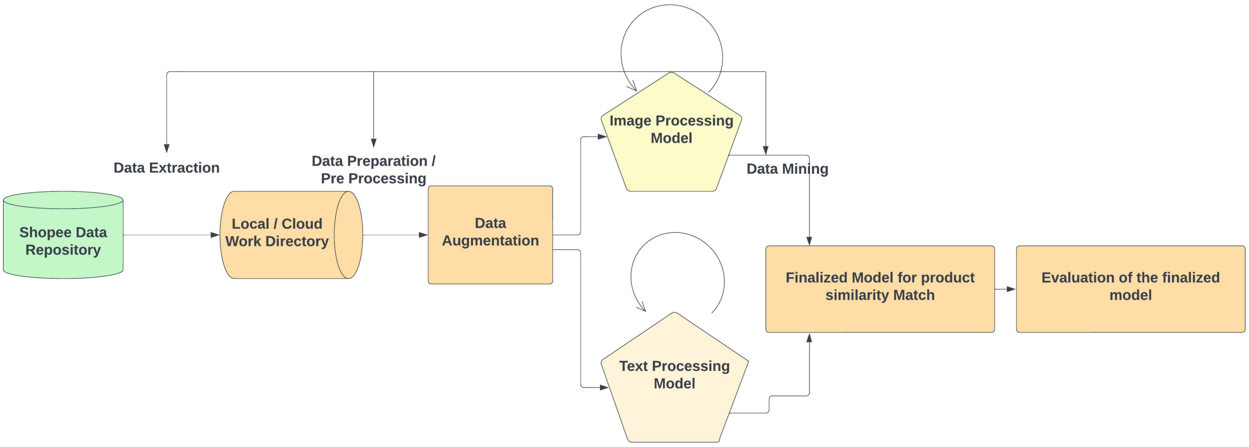
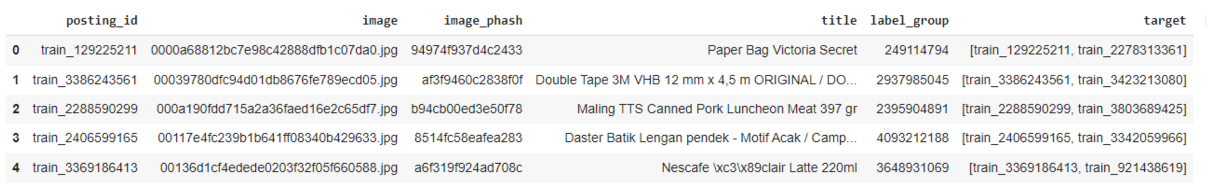

Data Extraction
Data Preparation / Pre Processing
Image Processing Model
Data Mining
Shopee Data Repository
Local / Cloud Work Directory
Data Augmentation
Finalized Model for product similarity Match
Evaluation of the finalized model
Text Processing Model

Figure 2: Processing Steps in Knowledge Discovery Database

32,400 images [1] and textual descriptions of online commercial products from various product categories. The analysis performed on these datasets yields a result set containing identical products among the listed items, which are subjected to several iterations of training and testing. The following Figure 3 illustrates the key attributes of the datasets, such as product title, product image, and image phash. Additionally, attributes like Label Group and PostingId should also be taken into consideration. A sample of the dataset, showing how the input dataset appears, is depicted in the figure. The next step involves analyzing the detailed characteristics of the image and text datasets separately. Subsequently, the datasets should be classified into two main categorical types: "Match Detected" and "No-Match Detected," by associating them with different groups.



	posting_id	image	image_phash	title	label_group	target
0	train_129225211	0000a68812bc7e98c42888dfb1c07da0.jpg	94974f937d4c2433	Paper Bag Victoria Secret	249114794	[train_129225211, train_2278313361]
1	train_3386243561	00039780dfc94d01db8676fe789ecd05.jpg	af3f9460c2838f0f	Double Tape 3M VHB 12 mm x 4,5 m ORIGINAL / DO...	2937985045	[train_3386243561, train_3423213080]
2	train_2288590299	000a190fdd715a2a36faed16e2c65df7.jpg	b94cb00ed3e50f78	Maling TTS Canned Pork Luncheon Meat 397 gr	2395904891	[train_2288590299, train_3803689425]
3	train_2406599165	00117e4fc239b1b641ff08340b429633.jpg	8514fc58eafea283	Daster Batik Lengan pendek - Motif Acak / Camp...	4093212188	[train_2406599165, train_3342059966]
4	train_3369186413	00136d1cf4edede0203f32f05f660588.jpg	a6f319f924ad708c	Nescafe \xc3\x89clair Latte 220ml	3648931069	[train_3369186413, train_921438619]

Figure 3: Data-Set - Data-frame Snapshot

3.2 Data Pre-Processing

The shopee dataset in consideration has two different types of data , image data and text data. Whereas the image data contains about 34253 images , this needs to be compared against each other along with the label description in the csv files for these images.Therefore processing all these images would be a daunting task for even modern computers.The given Figure 4 below shows the plot of image count per label in the product titles category .

Also the label count have been analysed across the frequency distribution, this is as shown in the Figure 5 below.We have performed some image augmentation which would saturate and de-saturate the images , each of the model considered for the research requires different pre-processing steps.However,as a common approach we have tried to reshape the images using augmentation and tokenize the words from the text file matching the title for the corresponding images.Then a simple algorithm for determining the nearest neighbours such as KNN is used and also using the cosine similarity to classify the wide range of products by similarity index arrived .

[1] https://www.kaggle.com/competitions/shopee-product-matching/data

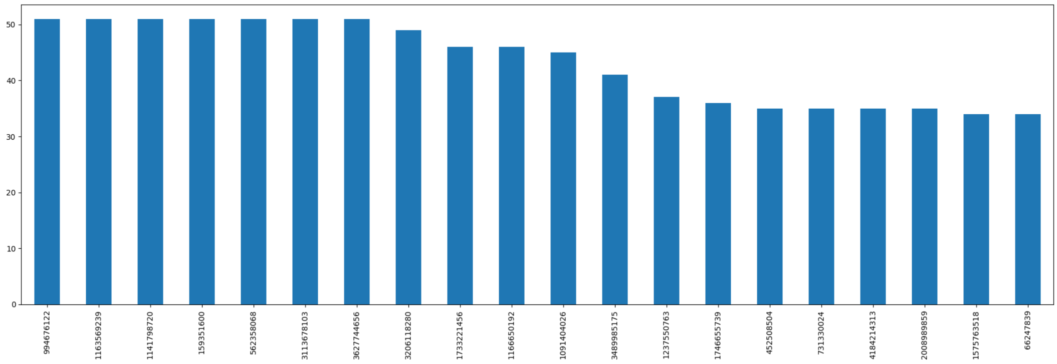


Figure 4: Image count for each label

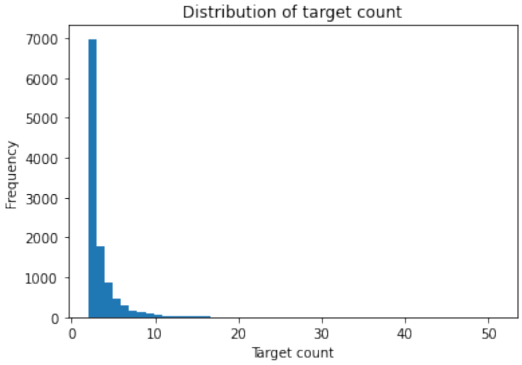


Figure 5: Target Count Frequency Distribution

3.3 Data Augmentation and Normalization

In order to perform the detailed analysis on the given image data set the image data-sets considered are normalized and augmented which is synthesized by the regularization function aiding in controlling the over-fitting of data . This is done by using the library utility form kera library in python such as ImageDataGenerator to resize the image size . Sample augmentation on the given data set is as shown in the Figure 6 below ,

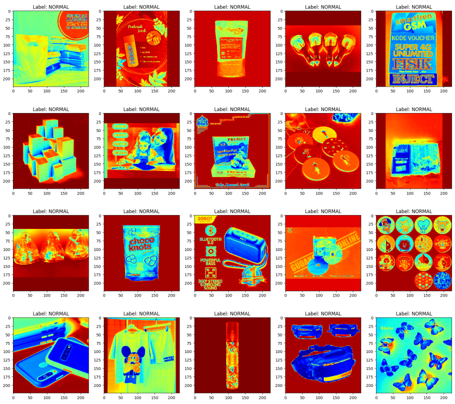


Figure 6: Image Augmentation

These normalised and augmented images in the training set are further fed in to the model for identifying the product similarity match without over-fitting the data .

4 Design Specification

The matching criterion between images and text data set pertaining to the similarity match prediction relies significantly on the proposed architecture as shown in the Figure below 7. Therefore in order to attain effective indications of matched product items, a variety of pre-trained deep learning models are employed. Building upon the proposed mechanism of using Resnet50v2 + eBERT, the concept of transfer learning is embraced. This approach involves leveraging previously computed values from one model and function to enhance the precision of results. The encompassed framework will amalgamate the application of various models developed as an evolution towards the finalized and refined model for product match.

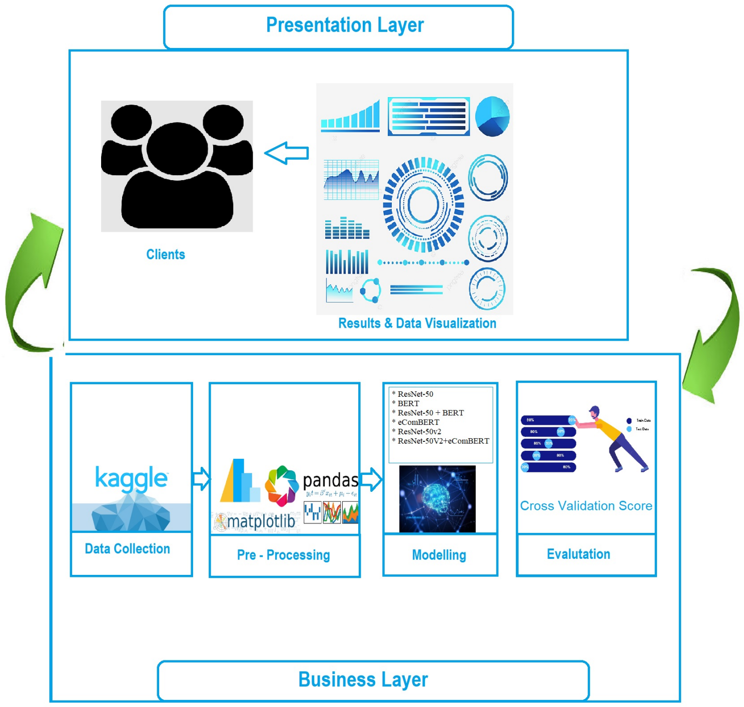


Figure 7: Proposed System Architecture

4.1 Presentation Layer

The presentation layer is where the user will be able to view the visualization as a result of the model evaluation which will be rather an outcome of the research yielded by various versions of the proposed methodologies and presented as a single entity model showing the similarity of products vs another product/offers.

4.2 Business Logic Layer

The business layer is the engine critical house of the system where core of the work is performed here , starting from inferring , interpreting and enchantment of data to be readily processed by the Resnet50v2 , BERT and TF-IDF using python as the tool kit. The cuda, cuml , torchvision , transformers , textwrap and cv2 libraries have been used for accomplishing the model generation for the product similarity match.

4.3 Data tier

Furthermore , the entire process begins at this tier and is a part of the actual business tier itself where we capture and extract the source image and text description data set from the public repository as mentioned in the section 3.1 above.

5 Implementation

5.1 Introduction

In an attempt to prove the proposed hypothesis , the considered data-sets containing image and text file are traversed and processed iteratively across multiple combinations of deep learning and specifically transfer learning methodologies . Therefore before getting into the actual implementation of these deep learning neural network based Image processing methods would like to begin by summarizing the architecture of these methodologies, the following are the methodologies in consideration ResNet50 , ResNet50 v2 in image processing and TF-IDF and e-BERT / SBERT for text processing for labels and description of the product titles .

5.2 Deep Learning for Product Matching

5.2.1 Image based Product Matching

Over the years there have been multiple developments in the product matching using image based CNN and deep learning methodologies as discussed in the section 2 . The basic conception of using the image based deep learning technique is to tweak the additional layers by altering according to the contextual necessity and therefore establish a means of reducing the vanishing gradients as discussed before . Multiple technologies such as ImageNet, MobileNet , VGG19 and ResNet have been used in the works of (Sanaullah Shariff 2022) and (Hari Krishnan 2021) , hence as a modernised approach and with an attempt to improve the previous methodologies ResNet50 v2 is considered for identifying the product similarity match in combination with text vectorization algorithm such as e-BERT . The ResNet architecture and the underlying concept of using ResNet50 v2 is described in detail in the upcoming sections below ,

ResNet - Residual Networks

In accordance with the constant enhancement in the modernised image processing using deep learning methods the advent of CNN has eventuated a revolution in this research context . As a result in the works of (Jian et al. 2016) , they have proposed a new architecture therefore labelled as ResNet. The existing issue of vanishing or exploding gradient was expected to be resolved by the ResNet - Residual Network architecture by reducing the error rate thereby increasing the number of layers. This works by subduing the connections between the links by bypassing the link layer activation towards the ensuing layers attached resulting in residual blocks stacked up to form ResNets as shown in the Figure 8 . the idea behind this concept is to allow the network accustomed to fit the residual mapping instead of learning the bottom mapping layers

Fx = Hx x which gives Hx = Fx x.

With the prolonged advancements in the modernization of ResNet architecture it has been observed that mistake percent is around 3.57 as per the study revealed in the article [2] . Traditionally all along the early 2000s VGG-16 architecture has been used for these complex image based CNN problem context.However , with the introduction of ResNet especially the enhancement of ResNet -101 has proved to be a replacement for VGG-16 which has introduced the networks of about 100 and 1000s of layers in their architecture. The layers in the CNN are stacked up over each other with an attempt to improve the performance and accuracy . Therefore this has led to the usage of Resnet50 as described in the article (Koonce and Koonce 2021) , the main idea of using the residual network is the ability of combination of another approaches for tackling the different problems arising as a result of residual stacking. The architecture of Resnet50 is illustrated in the Figure 8 as shown below ,

The RELU is the activation function used in the synthesizing of Resnet50 as shown in the figure 8 below. The fundamental agenda behind adding multiple layers is to enable the learning of more complex features . In our research context , in order to identify the e-commerce products for similarity matching the initial layers can be used to pick up the edges of the product images , the later layers are used to describe the textures , the third layer will pick up on objects and the next layer will be capturing the feature of the images against the text label description . In general , the conception is when there are more additional layers added the overall performance of the network is declined which is related to the networking initialization , the optimization function and specifically the vanishing gradient .

[2] https://medium.com/@siddheshb008/resnet-architecture-explained-47309ea9283d

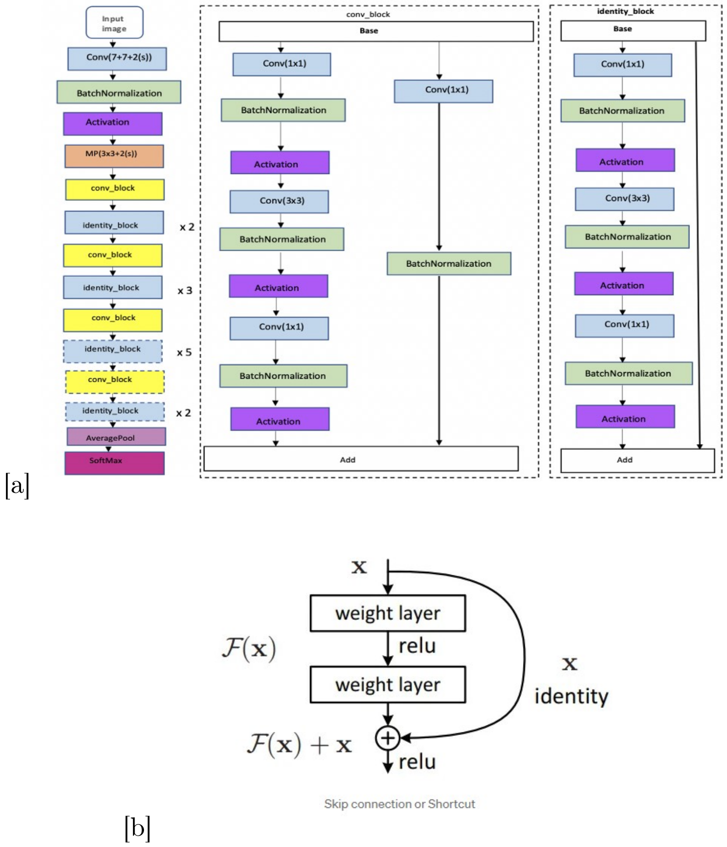
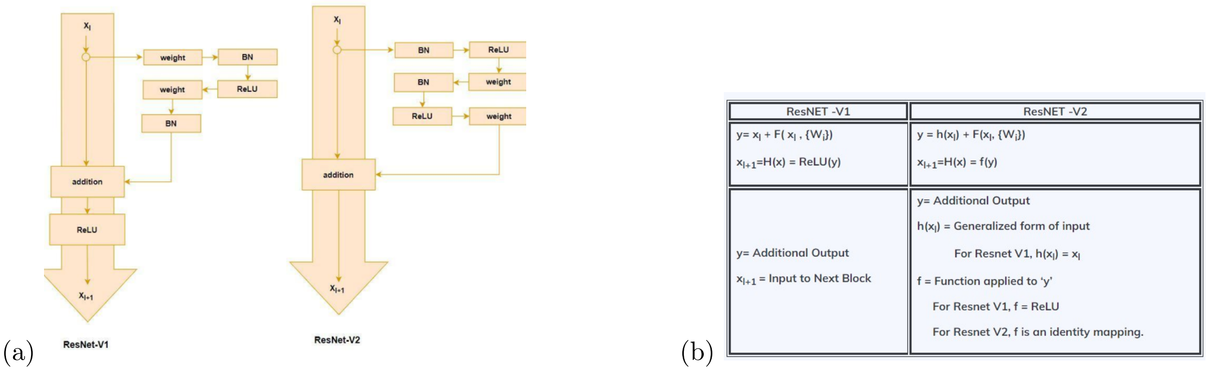


Figure 8: (a) Resnet50 Architecture / (b) RELU function Snapshot



ResNET -V1	ResNET -V2
$y = x_l + F(x_l, \{W_i\})$ $x_{l+1} = H(x) = ReLU(y)$	$y = h(x_l) + F(x_l, \{W_i\})$ $x_{l+1} = H(x) = f(y)$
y= Additional Output x_{l+1} = Input to Next Block	y= Additional Output $h(x_l)$ = Generalized form of input For Resnet V1, $h(x_l) = x_l$ f = Function applied to 'y' For Resnet V1, f = ReLU For Resnet V2, f is an identity mapping.

Figure 9: (a) Resnet50 v1 vs v2 (b) Resnet50 v1 / v2 Differences

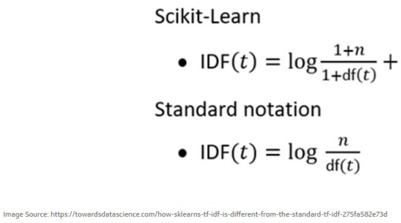
Resnet50 v2

Inspired from the works of (Rahimzadeh and Attar 2020) , we have used ResNet50v2 as an improvement of the product similarity match detection problem. This is a further enhancement of the ResNet50 Architecture . The major change in the implementation of the Resnet50 v2 is the absence of non-linearity as shown in the figure 9 which we could observe in the Resnet 50 , More specifically in Resnet50 v2 the batch normalization and the RELU activation function for the vanishing gradients is performed before musing with the weighted(W) matrix. The given figure 5.2.1 also shows the main differences between Resnet50 v1 and v2 respectively .

5.2.2 Text based Product Matching

- **TF-IDF** It is one of the text vectorization specifically used in this context as a means of concatenation with BERT as an attempt to improve the performance of the product similarity match prediction model . They have two main parts Term Frequency and Inverse Document Frequency

, while the former is computed by determining the number of repeating words , Logarithmically scalable frequency and Boolean frequency . Whereas the latter (IDF) is synthesized by how uncommon a specific word schema is embarked in the word corpus . The formula to arrive the same is mentioned below in the Figure 10 ,



Scikit-Learn

- $\mathrm{IDF}(t) = \log\frac{1+n}{1+\mathrm{df}(t)} + 1$

Standard notation

- $\mathrm{IDF}(t) = \log\frac{n}{\mathrm{df}(t)}$

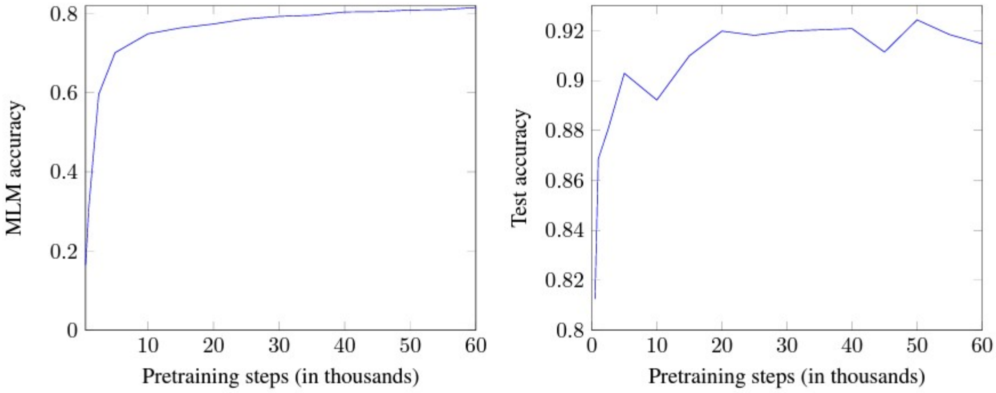
Image Source: https://towardsdatascience.com/how-sklearns-tf-idf-is-different-from-the-standard-tf-idf-275fa582e73d

Figure 10: IDF Formula

Hence the combined usage of both the common and uncommon pattern analogy is specifically useful in the product similarity context in the commerce platform . This is represented as shown below ,
$tfidf(t, d, D) = tf(t, d).idf(t, D)$

- **BERT**

Consequently the usage of BERT has proved to be effective as per the works of (Tracz et al. 2020) , despite being a complex transformer based deep learning methodology which facilitates the bidirectional connection across the elements and hence enables dynamic output enhancement . Specifically the usage of eComBERT / eBERT has proved to produced enhance results in the product matching as shown in the Figure below 11 which depicts the accuracy of both training and test data,



(a) Masked language model prediction accuracy.

(b) Performance of BERT-based fine-tuned model (eComBERT) trained with CR batch construction strategy.

Figure 11: eBERT Inclusion - Proposed system

5.2.3 HyperTuning

The combination of various methodologies in consideration are employed , we begin by generating a model for prediction of product match using Resnet 50 , then perform the same using eBERT and then TF-IDF , these are then combined to evaluate the improvements, the same is applied using ResNet50 V2 on the given image and text(csv) data set which will be discussed int he upcoming section 6 below .

6 Evaluation

In this section, evaluation ,analysis of the finalized model provided for the identification of product similarity match for shopee.com is elaborated . Further more pyplot library from ”matplotlib” and "gc" libraries are used for visual representation of the models generated results as graph or some other visualizations. The effective performance of the models developed are measured by using some of the evaluation metrics as described in the below sections ,

- **F1 Score**

F1 score is a machine learning metrics which are used in the classification models which is a measure to scale the improvement of the models using the combination of precision and recall scores of a given model . This uses the components of the confusion matrix using the various metrics for evaluating the classifiers such as accuracy , recall, precision and f1score . The formula of arriving at the f1 score is as shown in the Figure 12 below ,

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Figure 12: F1 Score

- **Cross Validation Score**

The cross validation score is a measure of evaluation of a model which helps in assessing the overall performance of the dataset instead of evaluating the test / train split . The number of folds is defined (default is 5) and the data in the dataset are split according to the defined number of folds . This has been considered as the main evaluation measure following the works of (Hari Krishnan 2021) and (Sanaullah Shariff 2022) . This can be used for both regression and classification models which is characterised by the following salient features as described below ,

- **estimator** - The model object to use to fit the data
- **X** - The data to fit the model on
- **y** - The target of the model
- **scoring** - The error metric to use
- **cv** - The number of splits to use

- **Cosine Similarity**

Cosine similarity is a means of understanding the semantic similarity between two different sets in consideration according to the article published in (Han, Kamber and Pei 2012) . This is mainly used in this context since we aspire to extract the similarity between product titles across various products in both image data sets and csv files respectively , this is synthesized by the formula as shown in the Figure 13.

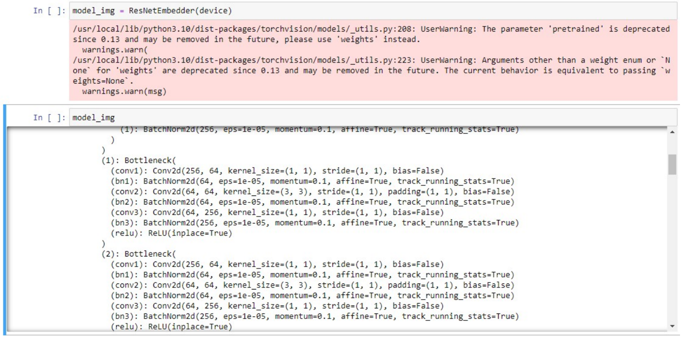
$$\text{cosine similarity} = S_C(A,B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2 \cdot \sum\limits_{i=1}^{n} B_i^2}},$$

Figure 13: Cosine Similarity

6.1 EXPERIMENT 1 : Image Based Product Similarity - ResNet50

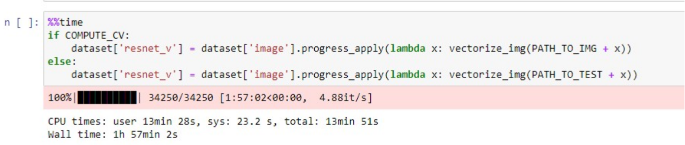
We begin our analysis using the given image dataset containing about 34250 images where we consider 34200 for training and the remaining for testing . They are pre-processed and normalized before subjecting to the model , using the ResNet50 the product similarity match is arrived as shown in the figure

14 below. The response time taken by using ResNet50 for the dataset in consideration is about 117 minutes as shown in the figure 15 Further once the above model is generated they are then calibrated for identifying the product similarity , which further gives the prediction in the tabular format as shown in the Figure 16 .



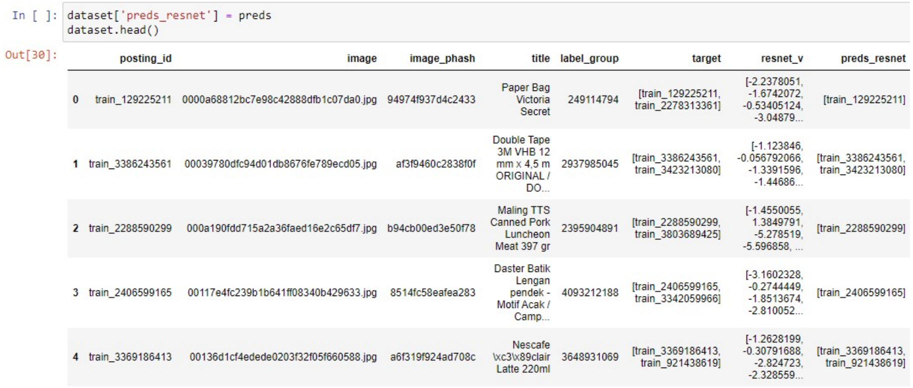
```
In [ ]: model_img = ResNetEmbedder(device)

/usr/local/lib/python3.10/dist-packages/torchvision/models/_utils.py:208: UserWarning: The parameter 'pretrained' is deprecated
since 0.13 and may be removed in the future, please use 'weights' instead.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/torchvision/models/_utils.py:223: UserWarning: Arguments other than a weight enum or `N
one` for 'weights' are deprecated since 0.13 and may be removed in the future. The current behavior is equivalent to passing `w
eights=None`.
  warnings.warn(msg)

In [ ]: model_img

          (1): BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
        )
      )
      (1): Bottleneck(
        (conv1): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (bn1): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
        (conv2): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False)
        (bn2): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
        (conv3): Conv2d(64, 256, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (bn3): BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
        (relu): ReLU(inplace=True)
      )
      (2): Bottleneck(
        (conv1): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (bn1): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
        (conv2): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False)
        (bn2): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
        (conv3): Conv2d(64, 256, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (bn3): BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
        (relu): ReLU(inplace=True)
```

Figure 14: ResNet50 Model for Product Matching



```
n [ ]: %%time
       if COMPUTE_CV:
           dataset['resnet_v'] = dataset['image'].progress_apply(lambda x: vectorize_img(PATH_TO_IMG + x))
       else:
           dataset['resnet_v'] = dataset['image'].progress_apply(lambda x: vectorize_img(PATH_TO_TEST + x))

       100%|██████████| 34250/34250 [1:57:02<00:00,  4.88it/s]

       CPU times: user 13min 28s, sys: 23.2 s, total: 13min 51s
       Wall time: 1h 57min 2s
```

Figure 15: ResNet50 Model Processing Time



```
In [ ]: dataset['preds_resnet'] = preds
        dataset.head()
Out[30]:
```

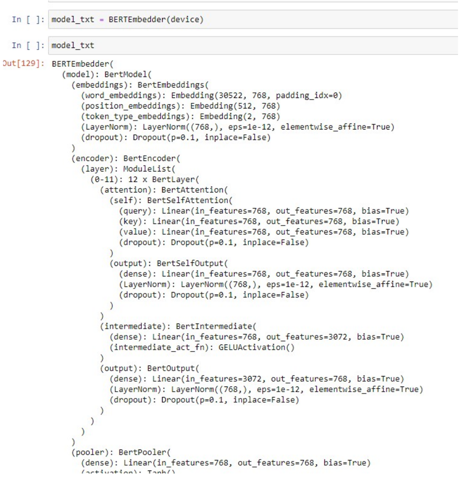
	posting_id	image	image_phash	title	label_group	target	resnet_v	preds_resnet
0	train_129225211	0000a68812bc7e98c42888dfb1c07da0.jpg	94974f937d4c2433	Paper Bag Victoria Secret	249114794	[train_129225211, train_2278313361]	[-2.2378051, -1.6742072, -0.53405124, -3.04879...	[train_129225211]
1	train_3386243561	00039780dfc94d01db8676fe789ecd05.jpg	af3f9460c2838f0f	Double Tape 3M VHB 12 mm x 4,5 m ORIGINAL / DO...	2937985045	[train_3386243561, train_3423213080]	[-1.123846, -0.056792066, -1.3391596, -1.44686...	[train_3386243561, train_3423213080]
2	train_2288590299	000a190fdd715a2a36faed16e2c65df7.jpg	b94cb00ed3e50f78	Maling TTS Canned Pork Luncheon Meat 397 gr	2395904891	[train_2288590299, train_3803689425]	[-1.4550055, 1.3849791, -5.278519, -5.596858, ...	[train_2288590299]
3	train_2406599165	00117e4fc239b1b641ff08340b429633.jpg	8514fc58eafea283	Daster Batik Lengan pendek - Motif Acak / Camp...	4093212188	[train_2406599165, train_3342059966]	[-3.1602328, -0.2744449, -1.8513674, -2.810052...	[train_2406599165]
4	train_3369186413	00136d1cf4edede0203f32f05f660588.jpg	a6f319f924ad708c	Nescafe \xc3\x89clair Latte 220ml	3648931069	[train_3369186413, train_921438619]	[-1.2628199, -0.30791688, -2.824723, -2.328559...	[train_3369186413, train_921438619]

Figure 16: ResNet50 Prediction table

Therefore the accuracy of the product match using ResNet50 is measured using f1Score and the cosine similarity is used in the identification of similar titles and the cv score obtained as a result of ResNet 50 Model is 62.5 % .

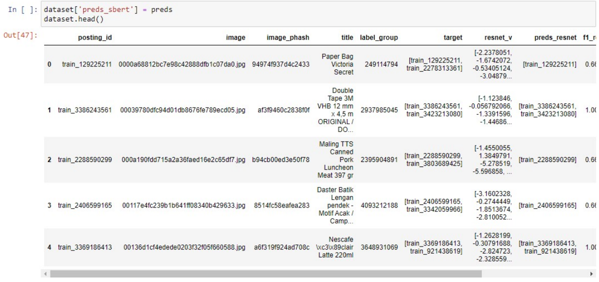
6.2 EXPERIMENT 2 : Text Based Product Similarity - eBERT

The second attempt in the product matching journey of the proposed system is by using e-Commerce based BERT algorithm for identifying the common product titles .Here the csv dataset with 21290 records containing multiple product titles across multiple categories and groups identified using the group-id . The model arrived is as shown below in the Figure 17.The BERT Model took about 39 minutes for processing the label titles and the predictions made by the BERT model is as shown in the Figure 18 below , Further evaluation of the arrived model proves that the accuracy of the product matching model is about 54.5% which is only less compared to the previous image based prediction. Therefore in the next experiment it has been attempted to improvise by adding up further text vectorization methods for enriching the similarity detection,



```
In [ ]: model_txt = BERTEmbedder(device)

In [ ]: model_txt

Out[129]: BERTEmbedder(
  (model): BertModel(
    (embeddings): BertEmbeddings(
      (word_embeddings): Embedding(30522, 768, padding_idx=0)
      (position_embeddings): Embedding(512, 768)
      (token_type_embeddings): Embedding(2, 768)
      (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
      (dropout): Dropout(p=0.1, inplace=False)
    )
    (encoder): BertEncoder(
      (layer): ModuleList(
        (0-11): 12 x BertLayer(
          (attention): BertAttention(
            (self): BertSelfAttention(
              (query): Linear(in_features=768, out_features=768, bias=True)
              (key): Linear(in_features=768, out_features=768, bias=True)
              (value): Linear(in_features=768, out_features=768, bias=True)
              (dropout): Dropout(p=0.1, inplace=False)
            )
            (output): BertSelfOutput(
              (dense): Linear(in_features=768, out_features=768, bias=True)
              (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
              (dropout): Dropout(p=0.1, inplace=False)
            )
          )
          (intermediate): BertIntermediate(
            (dense): Linear(in_features=768, out_features=3072, bias=True)
            (intermediate_act_fn): GELUActivation()
          )
          (output): BertOutput(
            (dense): Linear(in_features=3072, out_features=768, bias=True)
            (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
            (dropout): Dropout(p=0.1, inplace=False)
          )
        )
      )
    )
    (pooler): BertPooler(
      (dense): Linear(in_features=768, out_features=768, bias=True)
      (activation): Tanh()
```

Figure 17: BERT Model



```
In [ ]: dataset['preds_sbert'] = preds
        dataset.head()
```

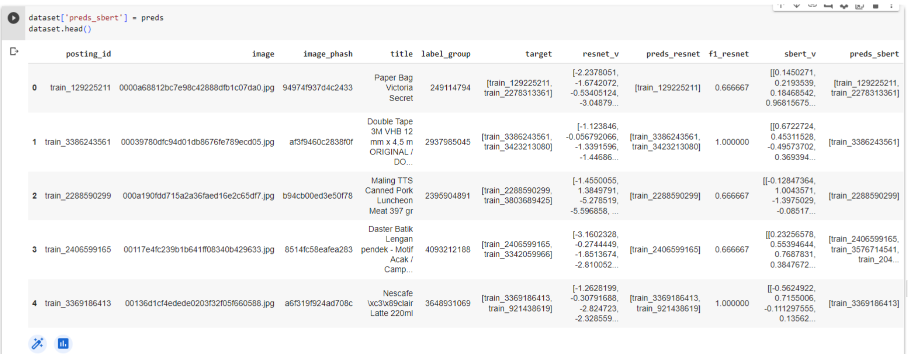
Out[47]:

	posting_id	image	image_phash	title	label_group	target	resnet_v	preds_resnet	f1_r
0	train_129225211	0000a68812bc7e98c42888dfb1c07da0.jpg	94974f937d4c2433	Paper Bag Victoria Secret	249114794	[train_129225211, train_2278313361]	[-2.2378051, -1.6742072, -0.53405124, -3.04879...	[train_129225211]	0.6
1	train_3386243561	00039780dfc94d01db8676fe789ecd05.jpg	af3f9460c2838f0f	Double Tape 3M VHB 12 mm x 4,5 m ORIGINAL / DO...	2937985045	[train_3386243561, train_3423213080]	[-1.123846, -0.056792066, -1.3391596, -1.44686...	[train_3386243561, train_3423213080]	1.0
2	train_2288590299	000a190fdd715a2a36faed16e2c65df7.jpg	b94cb00ed3e50f78	Maling TTS Canned Pork Luncheon Meat 397 gr	2395904891	[train_2288590299, train_3803689425]	[-1.4550055, 1.3849791, -5.278519, -5.596858, ...	[train_2288590299]	0.6
3	train_2406599165	00117e4fc239b1b641ff08340b429633.jpg	8514fc58eafea283	Daster Batik Lengan pendek - Motif Acak / Camp...	4093212188	[train_2406599165, train_3342059966]	[-3.1602328, -0.2744449, -1.8513674, -2.810052...	[train_2406599165]	0.6
4	train_3369186413	00136d1cf4edede0203f32f05f660588.jpg	a6f319f924ad708c	Nescafe \xc3\x89clair Latte 220ml	3648931069	[train_3369186413, train_921438619]	[-1.2628199, -0.30791688, -2.824723, -2.328559...	[train_3369186413, train_921438619]	1.0

Figure 18: BERT Model Prediction Table

6.3 EXPERIMENT 3 : Concatenating Resnet50 + e-BERT

Both the individual models have not performed considerably well and could not improve on the accuracy and response time , assuming that it could be possible to improve the same by concatenating the models from 14 and 6.2 respectively . The combined model has considerably generated predictive model at a lesser response time about a minute since we are working on the already pre-trained data-sets and models. The predictions for product similarity match by this concatenated model (ResNet50 + eBERT) is as shown in the figure 19 . Therefore after further evaluation of the model using the F1 Score and CV score the accuracy obtained is about 64.86 % respectively . Hence this is not a potential improvement the hyper-tuning of combining multiple text vectorization model is performed which is described in the upcoming section .



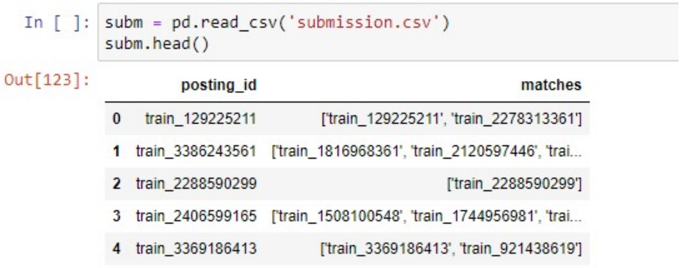
```
dataset['preds_sbert'] = preds
dataset.head()
```

	posting_id	image	image_phash	title	label_group	target	resnet_v	preds_resnet	f1_resnet	sbert_v	preds_sbert
0	train_129225211	0000a68812bc7e98c42888dfb1c07da0.jpg	94974f937d4c2433	Paper Bag Victoria Secret	249114794	[train_129225211, train_2278313361]	[-2.2378051, -1.6742072, -0.53405124, -3.04879...	[train_129225211]	0.666667	[[0.1450271, 0.2193539, 0.18468542, 0.96815675...	[train_129225211, train_2278313361]
1	train_3386243561	00039780dfc94d01db8676fe789ecd05.jpg	af3f9460c2838f0f	Double Tape 3M VHB 12 mm x 4,5 m ORIGINAL / DO...	2937985045	[train_3386243561, train_3423213080]	[-1.123846, -0.056792066, -1.3391596, -1.44686...	[train_3386243561, train_3423213080]	1.000000	[[0.6722724, 0.45311528, -0.49573702, 0.369394...	[train_3386243561]
2	train_2288590299	000a190fdd715a2a36faed16e2c65df7.jpg	b94cb00ed3e50f78	Maling TTS Canned Pork Luncheon Meat 397 gr	2395904891	[train_2288590299, train_3803689425]	[-1.4550055, 1.3849791, -5.278519, -5.596858, ...	[train_2288590299]	0.666667	[[-0.12847364, 1.0043571, -1.3975029, -0.08517...	[train_2288590299]
3	train_2406599165	00117e4fc239b1b641ff08340b429633.jpg	8514fc58eafea283	Daster Batik Lengan pendek - Motif Acak / Camp...	4093212188	[train_2406599165, train_3342059966]	[-3.1602328, -0.2744449, -1.8513674, -2.810052...	[train_2406599165]	0.666667	[[0.23256578, 0.55394644, 0.7687831, 0.3847672...	[train_2406599165, train_3576714541, train_204...
4	train_3369186413	00136d1cf4edede0203f32f05f660588.jpg	a6f319f924ad708c	Nescafe \xc3\x89clair Latte 220ml	3648931069	[train_3369186413, train_921438619]	[-1.2628199, -0.30791688, -2.824723, -2.328559...	[train_3369186413, train_921438619]	1.000000	[[-0.5624922, 0.7155006, -0.111297555, 0.13562...	[train_3369186413]

Figure 19: Resnet50 + BERT Concatenated Model Prediction Table

6.4 EXPERIMENT 4 : ResNEt50 + BERT + TF-IDF

Since , the previous experiments have used BERT which purely relies on the matching pattern of the label description of the product titles a novel approach of concatenating the TF-IDF model along with BERT is generated which are then combined with the previous model in the 6.3 . The combination of TF-IDF and BERT model , the text vectorizaiton concatenation has yield in a model with a prediction accuracy arrived using the CV score is about 61.39% . This is then used as the baseline for validation against the test data which has resulted in the refined product matching as shown in the Figure 20 below. The above model generated is the baseline for our proposed system which has yielded an accuracy of about 71.78% respectively .



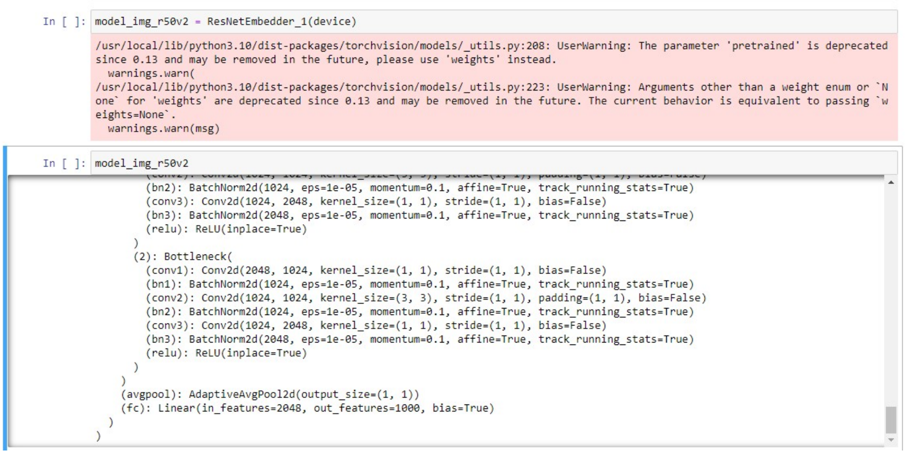
```
In [ ]: subm = pd.read_csv('submission.csv')
        subm.head()

Out[123]:
```

	posting_id	matches
0	train_129225211	['train_129225211', 'train_2278313361']
1	train_3386243561	['train_1816968361', 'train_2120597446', 'trai...
2	train_2288590299	['train_2288590299']
3	train_2406599165	['train_1508100548', 'train_1744956981', 'trai...
4	train_3369186413	['train_3369186413', 'train_921438619']

Figure 20: ResNet50+BERT+TF-IDF Preditction table

6.5 EXPERIMENT 5 : Image Based Product Similarity - ResNet50 v2

After evaluating the existing system and improvising by adding the modality in product matching by cross evaluating with the combination of multiple text vectorization models , this section describes the generation of a image based product similarity match model using ResNet50 v2. Similar to the 14 we are using the same training data set containing 34200 images,the model arrived is as shown in the figure 21 , The above model had a response time of only 12 minutes which is vast improvement considering the previous model using ResNet50 in the experiment 1 in the above section 14 . Evaluating the model further has resulted in the accuracy of about 60.89% . The prediction table obtained using the above model is as shown in the Figure 22 respectively .

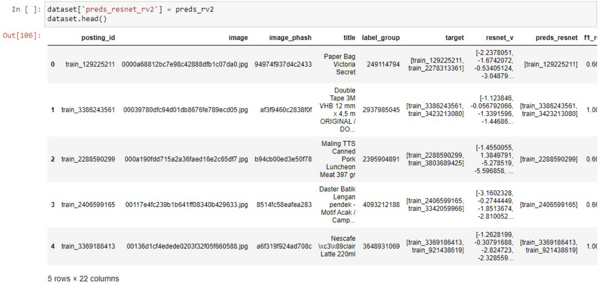


```
In [ ]: model_img_r50v2 = ResNetEmbedder_1(device)

/usr/local/lib/python3.10/dist-packages/torchvision/models/_utils.py:208: UserWarning: The parameter 'pretrained' is deprecated
since 0.13 and may be removed in the future, please use 'weights' instead.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/torchvision/models/_utils.py:223: UserWarning: Arguments other than a weight enum or `N
one` for 'weights' are deprecated since 0.13 and may be removed in the future. The current behavior is equivalent to passing `w
eights=None`.
  warnings.warn(msg)

In [ ]: model_img_r50v2

          (conv2): Conv2d(1024, 1024, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False)
          (bn2): BatchNorm2d(1024, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
          (conv3): Conv2d(1024, 2048, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (bn3): BatchNorm2d(2048, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
          (relu): ReLU(inplace=True)
        )
        (2): Bottleneck(
          (conv1): Conv2d(2048, 1024, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (bn1): BatchNorm2d(1024, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
          (conv2): Conv2d(1024, 1024, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False)
          (bn2): BatchNorm2d(1024, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
          (conv3): Conv2d(1024, 2048, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (bn3): BatchNorm2d(2048, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
          (relu): ReLU(inplace=True)
        )
      )
      (avgpool): AdaptiveAvgPool2d(output_size=(1, 1))
      (fc): Linear(in_features=2048, out_features=1000, bias=True)
    )
  )
```

Figure 21: ResNet50v2 Model

6.6 EXPERIMENT 6 : ResNet50 v2 + BERT

In line with the previous experiment for ResNet50 in the section 6.3 we have concatenated the previous eBERT model generated in the section 17 along with the Resnet50v2 . The prediction table yielded using the concatenated model is as shown in the figure 23 The model generated using the experiment 6 is then evaluated and the accuracy obtained was about 60.40% , similar to the previous encounter with



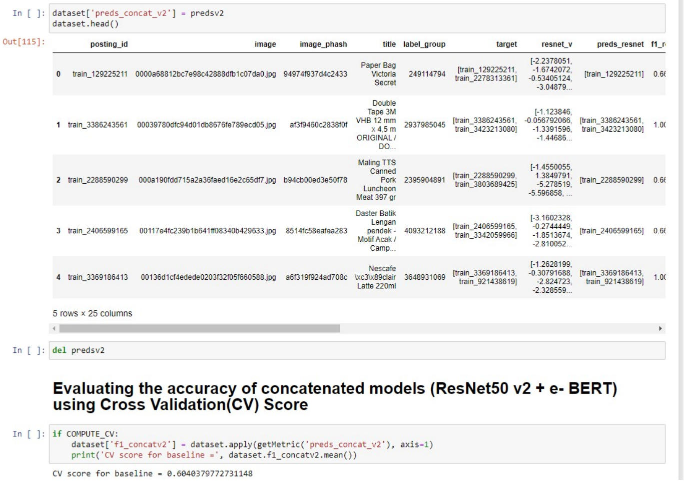
```
In [ ]: dataset['preds_resnet_rv2'] = preds_rv2
        dataset.head()
```

Out[106]:

	posting_id	image	image_phash	title	label_group	target	resnet_v	preds_resnet	f1_r
0	train_129225211	0000a68812bc7e98c42888dfb1c07da0.jpg	94974f937d4c2433	Paper Bag Victoria Secret	249114794	[train_129225211, train_2278313361]	[-2.2378051, -1.6742072, -0.53405124, -3.04879...	[train_129225211]	0.6
1	train_3386243561	00039780dfc94d01db8676fe789ecd05.jpg	af3f9460c2838f0f	Double Tape 3M VHB 12 mm x 4,5 m ORIGINAL / DO...	2937985045	[train_3386243561, train_3423213080]	[-1.123846, -0.056792066, -1.3391596, -1.44686...	[train_3386243561, train_3423213080]	1.0
2	train_2288590299	000a190fdd715a2a36faed16e2c65df7.jpg	b94cb00ed3e50f78	Maling TTS Canned Pork Luncheon Meat 397 gr	2395904891	[train_2288590299, train_3803689425]	[-1.4550055, 1.3849791, -5.278519, -5.596858, ...	[train_2288590299]	0.6
3	train_2406599165	00117e4fc239b1b641ff08340b429633.jpg	8514fc58eafea283	Daster Batik Lengan pendek - Motif Acak / Camp...	4093212188	[train_2406599165, train_3342059966]	[-3.1602328, -0.2744449, -1.8513674, -2.810052...	[train_2406599165]	0.6
4	train_3369186413	00136d1cf4edede0203f32f05f660588.jpg	a6f319f924ad708c	Nescafe \xc3\x89clair Latte 220ml	3648931069	[train_3369186413, train_921438619]	[-1.2628199, -0.30791688, -2.824723, -2.328559...	[train_3369186413, train_921438619]	1.0

5 rows × 22 columns

Figure 22: ResNet50v2 Prediction Table



```
In [ ]: dataset['preds_concat_v2'] = predsv2
        dataset.head()
```

Out[115]:

	posting_id	image	image_phash	title	label_group	target	resnet_v	preds_resnet	f1_r
0	train_129225211	0000a68812bc7e98c42888dfb1c07da0.jpg	94974f937d4c2433	Paper Bag Victoria Secret	249114794	[train_129225211, train_2278313361]	[-2.2378051, -1.6742072, -0.53405124, -3.04879...	[train_129225211]	0.6
1	train_3386243561	00039780dfc94d01db8676fe789ecd05.jpg	af3f9460c2838f0f	Double Tape 3M VHB 12 mm x 4,5 m ORIGINAL / DO...	2937985045	[train_3386243561, train_3423213080]	[-1.123846, -0.056792066, -1.3391596, -1.44686...	[train_3386243561, train_3423213080]	1.0
2	train_2288590299	000a190fdd715a2a36faed16e2c65df7.jpg	b94cb00ed3e50f78	Maling TTS Canned Pork Luncheon Meat 397 gr	2395904891	[train_2288590299, train_3803689425]	[-1.4550055, 1.3849791, -5.278519, -5.596858, ...	[train_2288590299]	0.6
3	train_2406599165	00117e4fc239b1b641ff08340b429633.jpg	8514fc58eafea283	Daster Batik Lengan pendek - Motif Acak / Camp...	4093212188	[train_2406599165, train_3342059966]	[-3.1602328, -0.2744449, -1.8513674, -2.810052...	[train_2406599165]	0.6
4	train_3369186413	00136d1cf4edede0203f32f05f660588.jpg	a6f319f924ad708c	Nescafe \xc3\x89clair Latte 220ml	3648931069	[train_3369186413, train_921438619]	[-1.2628199, -0.30791688, -2.824723, -2.328559...	[train_3369186413, train_921438619]	1.0

5 rows × 25 columns

```
In [ ]: del predsv2
```

Evaluating the accuracy of concatenated models (ResNet50 v2 + e- BERT) using Cross Validation(CV) Score

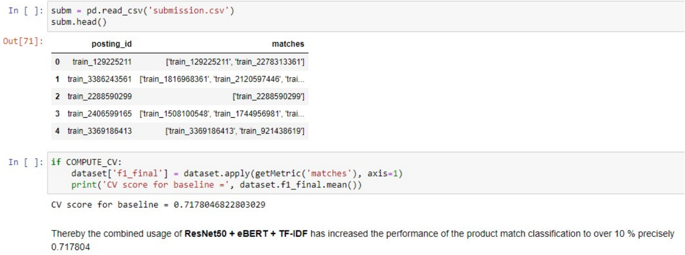
```
In [ ]: if COMPUTE_CV:
            dataset['f1_concatv2'] = dataset.apply(getMetric('preds_concat_v2'), axis=1)
            print('CV score for baseline =', dataset.f1_concatv2.mean())

        CV score for baseline = 0.6040379772731148
```

Figure 23: ResNet50v2+BERT Prediction Table

the ResNet50 even here we could observe the drop in the accuracy of product similarity match this is because of the fact that complexity of the text vectorization method is not enhanced containing only limited features and semantics to be compared against ,also the usage of KNN and cosine similarity was not ideal however this can be improved in the future . Therefore , there was a necessity of finalized model similar to the Experiment 4 in the section 6.4 .

6.7 EXPERIMENT 7: ResNEt50 v2 + BERT + TF-IDF

The finalized model combining the salient features of BERT and TF-IDF along with the ResNet50 v2 has effected in a refined model which has been validated against the test data and the prediction table obtained from the finalized model is as shown in the Figure 24 . The final model has produced a considerable improvement in the response time of only 20 minutes and the evaluation of the model resulted in the accuracy prediction of about 71.79% .



```
In [ ]: subm = pd.read_csv('submission.csv')
        subm.head()
```

Out[71]:

	posting_id	matches
0	train_129225211	['train_129225211', 'train_2278313361']
1	train_3386243561	['train_1816968361', 'train_2120597446', 'trai...
2	train_2288590299	['train_2288590299']
3	train_2406599165	['train_1508100548', 'train_1744956981', 'trai...
4	train_3369186413	['train_3369186413', 'train_921438619']

```
In [ ]: if COMPUTE_CV:
            dataset['f1_final'] = dataset.apply(getMetric('matches'), axis=1)
            print('CV score for baseline =', dataset.f1_final.mean())

        CV score for baseline = 0.7178046822803029
```

Thereby the combined usage of **ResNet50 + eBERT + TF-IDF** has increased the performance of the product match classification to over 10 % precisely 0.717804

Figure 24: ResNet50v2+BERT$_{T}F - IDFPredictionTable$

7 Comparison of various Models

The given table below illustrates the comparison various models generated as a part of the proposed system and the same has been described in the table 2 below. Although the improvement between using ResNet50 and ResNet50 has effected in improvement of 1% accuracy this can be considered as a future improvement which will be discussed in the next section.

Comparison of the Models			
Experiment	**Techniques Used**	**Results**	**Run Time**
Experiment 1	ResNet-50	0.62515	01:57:02
Experiment 2	eBERT	0.5458	00:38:25
Experiment 3	ResNet-50+ eBERT	0.6486	00:02:00
Experiment 4	ResNet-50 + eBERT +TF-IDF	0.7178	00:24:00
Experiment 5	ResNet-50v2	0.6089	00:14:59
Experiment 6	ResNet-50v2 + eBERT	0.6040	00:02:00
Experiment 7	ResNet-50v2 + eBERT +TF-IDF	0.7179	00:21:18

Table 2: Comparative highlights of models used

7.1 Conclusion and Future Work

With a critical consideration with respect to the overall enhancement of the customer experience in the e-commerce platform , the proposed system can be used as the baseline for refining the product matching in the e-commerce platform . This product matching can therefore be enhanced further to targeted promotions and offer campaigns for enriching the user experience . Though there is a continuous void in the area of product similarity match the need keeps evolving as the e-commerce continues to grow exponentially . Despite that fact that there are multiple strategies available in identify the similarity since the usage of deep learning methodology is considered it is a simple and effective way to determine the similarity using the k- Nearest neighbours and cosine similarity for identifying the relevant product titles . This can also be innovate and improved by considering feature engineering practices and involving the usage of other measures such as jaccard similarity especially in the e-commerce context as discussed in the works of (Hari Krishnan 2021) .

Though the result of the research have been successful in the aspect of generating a refined model for product similarity match using ResNet50v2 , BERT and TF-IDF , there have been difficulties when training and processing these image data which took the predominant time and can be handled effectively by big data practices managing them in a hadoop framework enabling the parallel processing therefore could produce a significant improvement in the response time along with the addition of feature attributes in the text file to enhance the semantics of product label descriptions . The same experimentation could also be subject to multiple other e-commerce data sets as a means of evaluating the hypothesis further .

7.2 Acknowledgement

I would sincerely love to appreciate and convey my special thanks to my mentor Mrs. Qurrat Ul Ain for her constant effort in attributing her valuable table towards the guidance of this research work which has enabled me to accomplish the completion of the project well within the stipulated deadline proposed . Further more it has been a humongous pleasure to a part of National college of Ireland and therefore extend my humble thanks to my family and friends for their constant support and motivation .

References

Abluton, Alessandro (2022). 'Visual Recommendation and Visual Search for Fashion E-Commerce'. In: *Similarity Search and Applications: 15th International Conference, SISAP 2022, Bologna, Italy, October 5–7, 2022, Proceedings.* Springer, pp. 299–304.

Abolghasemi, Amin, Suzan Verberne and Leif Azzopardi (2022). 'Improving BERT-based query-by-document retrieval with multi-task optimization'. In: *Advances in Information Retrieval: 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10–14, 2022, Proceedings, Part II.* Springer, pp. 3–12.

Bianchi, Federico, Bingqing Yu and Jacopo Tagliabue (2020). 'BERT goes shopping: Comparing distributional models for product representations'. In: *arXiv preprint arXiv:2012.09807.*

Han, Jiawei, Micheline Kamber and Jian Pei (2012). 'Data mining concepts and techniques third edition'. In: *University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University.*

Hanif, Muhammad Shehzad and Muhammad Bilal (2020). 'Competitive residual neural network for image classification'. In: *ICT Express* 6.1, pp. 28–37.

Hari Krishnan, Kannan (2021). 'E-commerce Product Similarity Match Detection using Product Text and Images'. PhD thesis. Dublin, National College of Ireland.

Hendriksen, Mariya (2022). 'Multimodal Retrieval in E-Commerce: From Categories to Images, Text, and Back'. In: *Advances in Information Retrieval: 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10–14, 2022, Proceedings, Part II.* Springer, pp. 505–512.

Jain, Sourabh and Prakash Hegade (2021). 'E-commerce Product Recommendation Based on Product Specification and Similarity'. In: *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT).* IEEE, pp. 620–625.

Jian, S et al. (2016). 'Deep residual learning for image recognition'. In: *IEEE Conference on Computer Vision & Pattern Recognition*, pp. 770–778.

Kejriwal, Mayank et al. (2021). 'An evaluation and annotation methodology for product category matching in e-commerce'. In: *Computers in Industry* 131, p. 103497.

Koonce, Brett and Brett Koonce (2021). 'ResNet 50'. In: *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, pp. 63–72.

Le, My Hong and Alexander Hinneburg (2022). 'An Application of Learned Multi-modal Product Similarity to E-Commerce'. In: *Similarity Search and Applications: 15th International Conference, SISAP 2022, Bologna, Italy, October 5–7, 2022, Proceedings.* Springer, pp. 25–39.

Li, Juan et al. (2020). 'Deep cross-platform product matching in e-commerce'. In: *Information Retrieval Journal* 23, pp. 136–158.

Mehta, Karan et al. (2022). 'Multimodal Classification in E-Commerce: A Systematic'. In.

Peeters, Ralph and Christian Bizer (2022). 'Supervised contrastive learning for product matching'. In: *arXiv preprint arXiv:2202.02098.*

Peeters, Ralph, Reng Chiz Der and Christian Bizer (2023). 'WDC Products: A Multi-Dimensional Entity Matching Benchmark'. In: *arXiv preprint arXiv:2301.09521.*

Rahimzadeh, Mohammad and Abolfazl Attar (2020). 'A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2'. In: *Informatics in medicine unlocked* 19, p. 100360.

Rajpoot, Chour Singh and Santosh Kumar Vishwakarma (2023). 'Comparative Analysis of Recommendation System Using Similarity Techniques'. In: *International Conference on Data Management, Analytics & Innovation.* Springer, pp. 119–127.

Sanaullah Shariff, Zahra Fathima (2022). 'Product Matching for E-commerce Platform based on Text and Image Similarity using Deep Neural Network Architecture'. PhD thesis. Dublin, National College of Ireland.

Tracz, Janusz et al. (2020). 'BERT-based similarity learning for product matching'. In: *Proceedings of Workshop on Natural Language Processing in E-Commerce*, pp. 66–75.

Zhang, Denghui et al. (2020). 'E-BERT: Adapting BERT to E-commerce with Adaptive Hybrid Masking and Neighbor Product Reconstruction'. In.

Zheng, Mingkai et al. (2022). 'Simmatch: Semi-supervised learning with similarity matching'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14471–14481.