Shopee price match guarantee

*CMPE 256 Web and big data mining, Spring, 2021
Instructor: Dr Magdalini Eirinaki,
Computer Engineering Department
San Jose State University, San Jose, CA

Akash Aggarwal
SID: 013856531
akash.aggarwal@sjsu.edu
Computer Engineering Department
San Jose State University

Sowmya Mruthyunjaya
SID: 014619930
sowmya.mruthyunjaya@sjsu.edu
Computer Engineering Department
San Jose State University

Sudha Vijayakumar
SID: 014749488
sudha.vijayakumar@sjsu.edu
Software Engineering Department
San Jose State University

***Abstract*—Most shoppers dont like to overpay for the products they buy online. They shop around various retailers for the same product before making the final purchase so as to get the best deal. While e-businesses have made life easy for people where they can buy products from the comfort of their home, they are also entangled in a price war with their competitors. Every business wants to retain or add more customers and in the due process try to provide the best possible pricing. When people try to sell products online, sometimes duplicates are posted with different titles and pricing which cuts into the profits of the business. Companies use methods like marketing, advertisements, give-aways, fun competitions to attract customers. One of the ways a company can offer its products at a competitive rates to its customers is by product matching. With increase in the number of sellers and vendors, product matching has become a herculean task. In this project we are leveraging machine learning models to perform product matching so as help find a better way of providing the best price to the customers.**

***Index Terms*—product matching, price match, machine learning, competitive rate**

I. Introduction

In recent times we have seen applications that can predict the similar products based on the thumbnails provided. For example, Amazon provides recommendations based on similar titles and images of a product. The two types of recommendations are user based and item based collaborative filtering. In user based collaborative filtering, similar users are matched according to the items that they like or have previously liked. In item-item based collaborative filtering, similar items are matched using the meta data of the users that have rated an image. In this project, we are using item based collaborative filtering along with image processing.

This project primarily offers a 'Price Match Guarantee' solution to the Kaggle Challenge Shopee-Price Match Guarantee to determine if two products are the same by examining their images. Typically, we think the motive of customers when they do online shopping is to buy products from websites that provide the best price. Therefore, the aim of retail companies is to offer the best price for a particular product. Shopee is one of the leading e-commerce giants in Taiwan and Southeast Asia. Their aim is to provide an easy, fast and secure online shopping experience specific to the region. Shopee is a pioneer in providing rigid payment and logistic support. They also offer a Lower Price Guaranteed option for various products to maintain customer satisfaction and retention. The aim of this challenge is to provide efficient product matching that can support:

- More accurate product categorization.
- Relevant product recommendations.
- Uncover marketplace spam.

Also, customers will be able to make an informed decision in choosing the right product with the best price guarantee. To achieve this, we plan to perform product matching using hybrid similarity scoring techniques coupled with deep learning algorithms with the given product images and metadata. In this project, we are using image based deep learning techniques like Resnet50, Restnet152, EffnetB3, EffnetB5, EffnetB6, Eca-nFnet-10 along with Tf-Idf Vectorization. We have done Phashing to find the baseline score. Finally, ensemble learning has been performed to find the best set of working models.2

II. Approach

Figure 1 gives a high level overview of our project.The flow of the project is as follows:

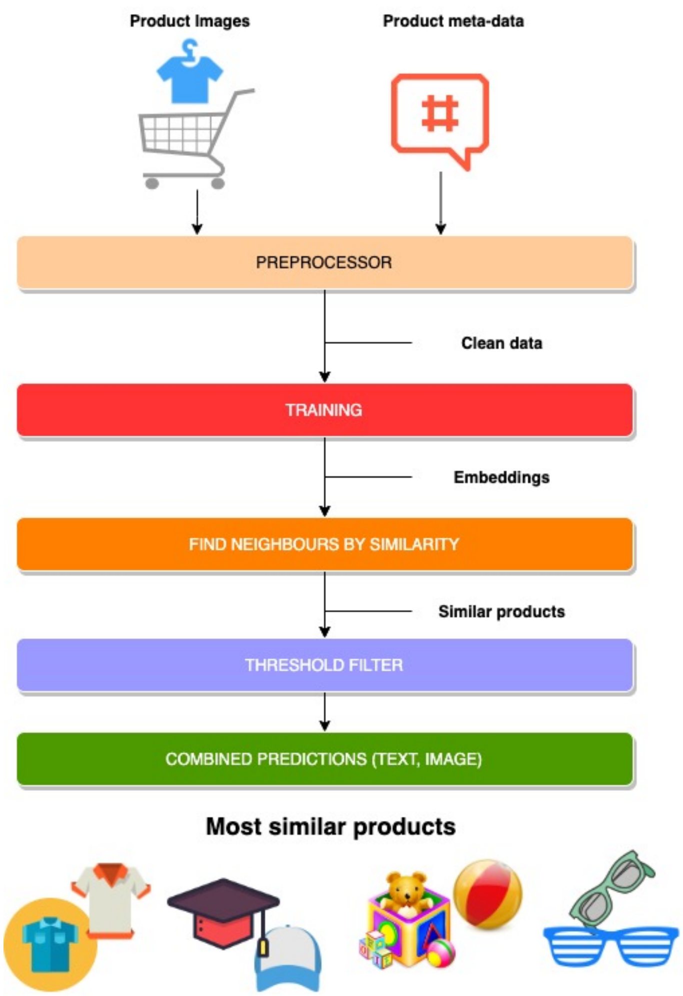

Fig. 1. High level project overview

- Exploratory Data Analysis and preprocessing - This section will discuss the basic EDA performed on the training data set and extract more useful information about the attributes provided in the training data.
- Model selection - This section describes the various models we have used in our project. It also shows the hyper parameters that are tuned to beat the baseline and the performance of the chosen models.
- Experimental study -In this section, we will discuss the steps of implementations required in our project.
- Result analysis - This section will analyze the results from our implementation of the deep learning models. We will evaluate the performance based on the F1 score of all the models.
- Conclusion - This section concludes the model that worked best for the project, results obtained. It also includes a brief summary of the future works to improve the performance of our work.

III. Exploratory Data Analysis

In any machine learning problem, it is always good to first understand the data and get as much insights into as possible before preprocessing it or feeding it to any model for training. Since our team entered the kaggle challenge, the dataset was available as part of it. We performed our initial investigation on the data to see if we can uncover any anomalies, find any correlation between features, generate summary statistics etc. The dataset contained two components - Images and metadata in csv file for both train and test sets.

```
In [11]: train.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34250 entries, 0 to 34249
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   posting_id   34250 non-null  object
 1   image        34250 non-null  object
 2   image_phash  34250 non-null  object
 3   title        34250 non-null  object
 4   label_group  34250 non-null  int64
dtypes: int64(1), object(4)
memory usage: 1.3+ MB
```

Fig. 2. Summary of train set

The csv file contained info such as posting id, image, image phash, posting title and label group. A peek at the data revealed that there were a total of 34250 train images with 5 characteristics mentioned above and 3 test images with 4 characteristics. Fig 2 provides a summary of train data.

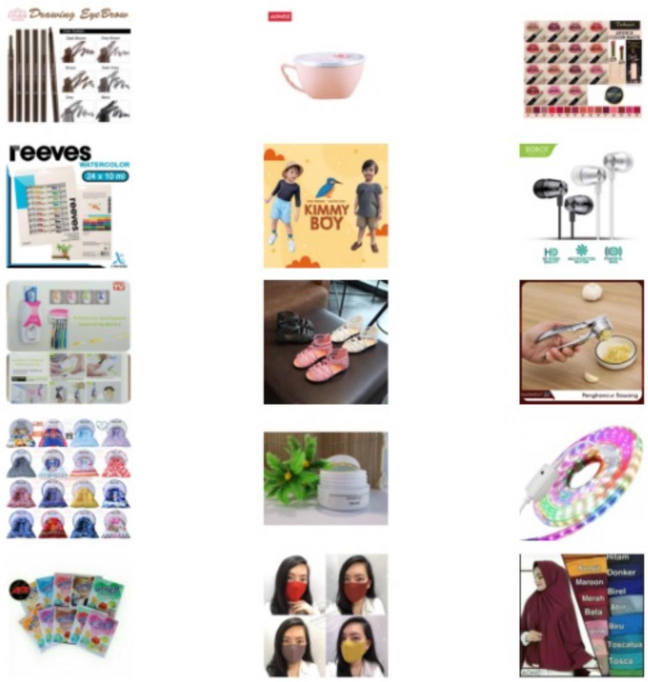

Fig. 3. Sample train images

Fig. 4. Sample of test images

Fig 3 and Fig 4 shows sample training and testing images. Fig 5 provides insight into the metadata of the project. It shows the top 50 duplicate items present in the training set.

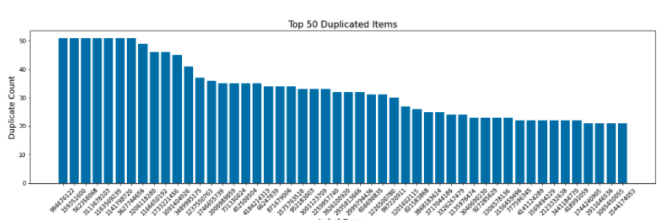

Fig. 5. Top 50 Duplicated Items according to label

Fig 6 provides the relationship between duplicate and unique item count. It shows that as the index of unique item increase, the duplicate count decreases.

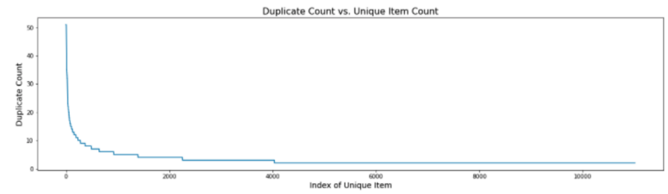

Fig. 6. Duplicate Count vs Unique Item

After performing EDA, we found that there are multiple ways to find similar products:

- We can use title find similar product titles.
- we can use product images find similar product images.
- We can use image hash find similar product images.
- We can use image OCR Text matching find similar product images.

IV. Model Selection

Since we are dealing primarily with images, we decided to choose some CNN backbones that have provided exemplary results on various well known images datasets.

A. Resnet

Resnet is the short form for residual network which serves as the backbone for many computer vision tasks. Deep neural networks suffer from overfitting and vanishing gradient problems. Resnet addresses those problems through skip connections where the gradients from one layer are passed to another layer deep in the network through identity blocks.

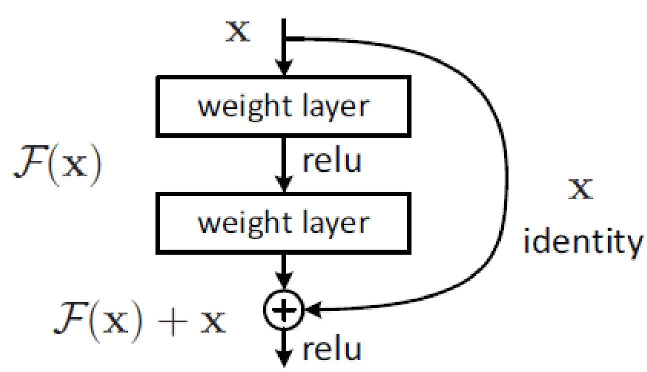

Fig. 7. Resnet

In this project we have used resnet152 and resnet50. The numbers 152 and 50 correspond to the number of layers in the network.

B. Efficient net

EfficientNet is a convolutional neural network that scales the depth (which is the number of layers in the network), width (number of filters in a convolutional layer) and resolution (height and width of the input image) of the network uniformly using a fixed compound scaling coefficient. The way compound scaling works is if the input image is bigger, more layers are needed in the network to increase the receptive field and more channels to capture more fine-grained patterns on the bigger image.Efficinet provides very high accuracy while also improving efficiency by reducing parameters and

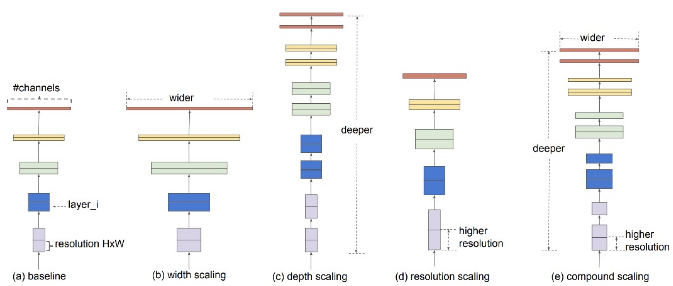

Fig. 8. Resnet

C. Efficient Channel Attention net (eca-nfet-10)

Eca-nfnet is a pretrained model on ImageNet and eca-nfnet-10 is a variant of the NFNet (Normalization Free) model family. Eca net uses very few parameters and less computation to increase the network performance.It avoids dimensionality reduction and reduces cross channel interaction in an efficient way. Eca-nfnet model variant results in an improved runtime characteristics, throughput on a GPU accelerator. It utilizes Efficient Channel Attention that is ECA instead of Squeeze-Excitation. It also features SiLU activations instead of the GELU.

Like other models in the NF family, this model contains no normalization layers (batch, group, etc). The models make use of Weight Standardized convolutions with additional scaling values in lieu of normalization layers.

V. Experimental study

Figures 9 and 10 shows the flowchart of finding similar images by product title and product image. For finding similar images using product image, after the initial EDA and preprocessing the data (resizing the images), we applied transfer learning by using pre-trained models like resnet152, ecanfnet-10, efficient net [3,5,6]. We then generated image emdeddings to bring down the dimensionality of these images. The key to good embedding is to train the model so that similar images are converted to similar vectors. K nearest neighbor algorithm and a similarity measure like cosine similarity is then used to predict similar products. This list is then passed through a suppression filter to generate the most similar products. For finding similar images using product title, the feed in clean title to the pretrained models. Here we generate title embeddings so that similar titles have same vectors. The rest of the steps are same as image based filtering.

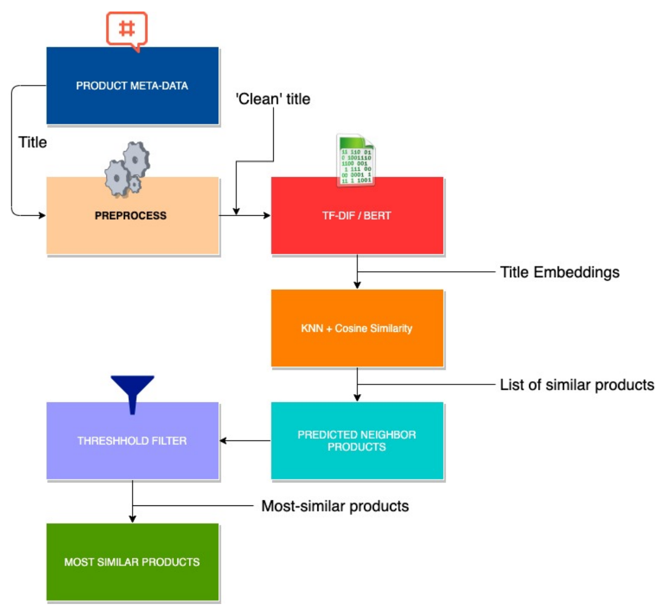

Fig. 9. Flowchart of finding similarity using product image

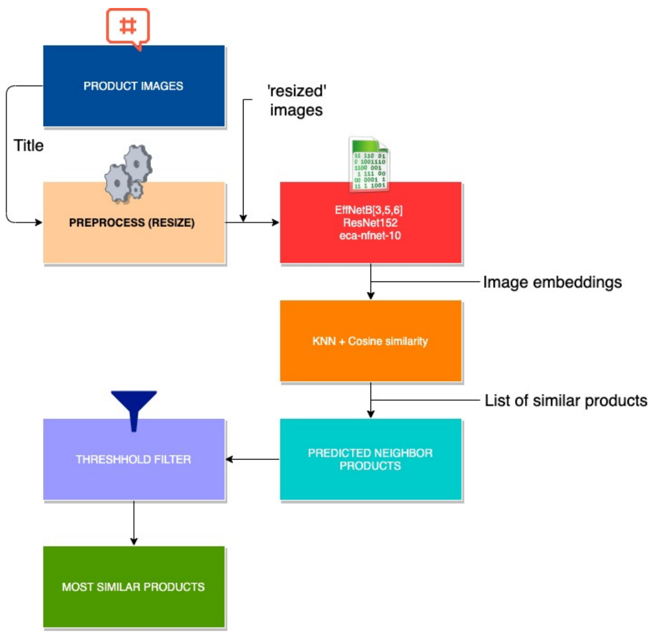

Fig. 10. Flowchart of finding similarity using image

Steps involved in project implementation:

- *Step 1:* Train Resnet152 and compute the F1 score.
- *Step 2:* To replace [RESNET152] model with the EfficientNet [3,5,6] models and train them on the SHOPEE Dataset, to see if the current [F1] Score can be improved.
- *Step 3:* To replace [EfficientNet] model with the [eca-nfnet-10] model and make inferences on the SHOPEE Dataset, to see if the current [F1] Score can be improved.
- *Step 4:* Ensemble [EfficientNetB6] + [eca-nfnet-10] to see if the [F1]-Score improves.

VI. Result Analysis

In our project, we employed various deep learning models and found the following F1 score respectively: Table 1 shows the model vs F1 score:

Model Name	**F1 Score**
Simple PHash	0.585
ResNet152	0.69
EffnetB-3	0.712
EffnetB-5	0.723
Eca-nfnet-10	0.725
EffnetB-6	0.729
EffnetB-6 + Eca-nfnet-10	0.733

TABLE I
F1 Score for various models

The following figure shows a bar graph comparing the performance of all the models based on their F1 score.

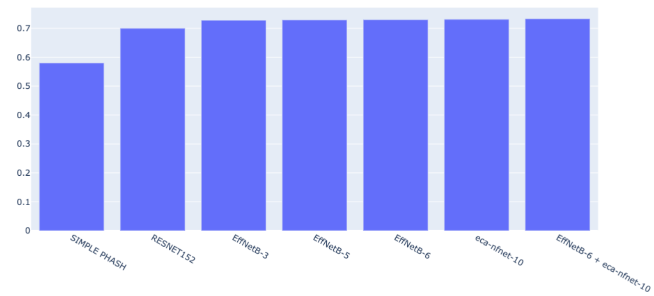

Fig. 11. Flowchart of finding similarity using image

VII. Conclusion

After submission our project to Kaggle, we achieved the best test F1 score of 0.733, ranking us in the top 90 percentile of all submissions in the competition so far. The model that performed the best was ensemble model of EffecientB-6 and Eca-nfnet-10. Generally, ensemble models perform better than the standalone models and it was apparent in our result analysis as well. While the results are satisfactory given it is the first time we performed deep learning models for image processing, we intend to continue to improve the results.

VIII. Contribution

Individually all three of us spent some time looking for projects that we can work on, one that aligns with the course work of CMPE 256. After two rounds of brainstorming, we as a team decided to enter the Kaggle competition for Shopee Price Match Guarantee. After we got the nod from our course instructor, we officially entered the competition. We spent some time understanding the problem statement. Then we held regular project checkpoint meetings to address any concerns, share progress, exchange notes. As a team we wrote the project proposal, explored the dataset, performed exploratory data preprocessing, drafted the project report and prepared presentation slides. Individually we each tried to work on different models to see if we could beat the baseline of the challenge. Akash worked on Eca-nfnet-10, Sudha worked on Efficient-Net and Sowmya worked on Resnet-152 and finally everybody worked on the ensemble modeling to achieve best reults.

IX. Future work

- To try different ensemble with different neighbour thresh hold to see if the score improves.
- To try [Fastai + Annoy] (Match by image) along with BART tokenizer(match by title) to find similar products.
- To try domain specific proven model Large Scale Multimodal Classification Using an Ensemble of Transformer Models and Co-Attention

X. Acknowledgement

We would like to express our sincere gratitude to Dr. Magdalini Eirinaki for her guidance, knowledge and patience throughout the course. She has answered all our queries with utmost patience and has guided us during times of need.

XI. Github Link To Project

Project Repository

References

[1] P. Ristoski, P. Petrovski, P. Mika, and H. Paulheim, "A machine learning approach for product matching and categorization," *Semantic Web*, vol. 9, no. 5, p. 707–728, 2018.

[2] S. Bell and K. Bala, "Learning visual similarity for product design with convolutional neural networks," *ACM Transactions on Graphics*, vol. 34, no. 4, p. 1–10, 2015.

[3] L. Yang and K. Easwar, "Matching the words and the features: The effect of matching advertisement language and product attributes on attitude," 2010.

[4] L. Ma, *Computationally efficient models for high-dimensional and large-scale classification problems*. PhD thesis.

[5] V. Chordia and V. K. BG, "Large scale multimodal classification using an ensemble of transformer models and co-attention," *arXiv preprint arXiv:2011.11735*, 2020.

[6] PyTorch, "Resnet-152."

[7] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, "Classification of remote sensing images using efficientnet-b3 cnn model with attention," *IEEE Access*, vol. 9, pp. 14078–14094, 2021.

[8] Y. Yousfi, J. Butora, E. Khvedchenya, and J. Fridrich, "Imagenet pre-trained cnns for jpeg steganalysis," in *Proceedings of the IEEE International Workshop on Information Forensics and Security, WIFS*, 2020.

[9] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10781–10790, 2020.

[10] P. Sun, X. Jin, W. Su, Y. He, H. Xue, and Q. Lu, "A visual inductive priors framework for data-efficient image classification," in *European Conference on Computer Vision*, pp. 511–520, Springer, 2020.

[11] I. Bello, "Lambdanetworks: Modeling long-range interactions without attention," *arXiv preprint arXiv:2102.08602*, 2021.

[12] S. Ray, "Disease classification within dermascopic images using features extracted by resnet50 and classification through deep forest," *arXiv preprint arXiv:1807.05711*, 2018.

[13] Y. Chu, X. Yue, L. Yu, M. Sergei, and Z. Wang, "Automatic image captioning based on resnet50 and lstm with soft attention," *Wireless Communications and Mobile Computing*, vol. 2020, 2020.

[14] X. Xu, W. Li, and Q. Duan, "Transfer learning and se-resnet152 networks-based for small-scale unbalanced fish species identification," *Computers and Electronics in Agriculture*, vol. 180, p. 105878, 2021.

[15] L. D. Nguyen, D. Lin, Z. Lin, and J. Cao, "Deep cnns for microscopic image classification by exploiting transfer learning and feature concatenation," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, IEEE, 2018.

[16] C. Bircanoglu and N. Arıca, "Effects of network depths on semantic image segmentation by weakly supervised learning," in *2020 28th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, 2020.

[17] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Supplementary material for "eca-net: Efficient channel attention for deep convolutional neural networks","

[18] R. Wightman, *pytorch-image-models*.

[19] A. Brock, S. De, S. L. Smith, and K. Simonyan, "High-performance large-scale image recognition without normalization," *arXiv preprint arXiv:2102.06171*, 2021.

[20] R. Wightman, "Pytorch image models." https://github.com/rwightman/pytorch-image-models, 2019.

[21] S.-H. Tsang, "Review: Resnet — winner of ilsvrc 2015 (image classification, localization, detection)," Sep 2018.

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21]