# **E-Commerce Recommendation System: Using Conventional Machine Learning Approach Based on Visual Similarity**

Deepak Halliyavar, Lohitha Vanteru, Mahe Jabeen Abdul and Pranavi Sandrugu

Department of Applied Data Science, San Jose State University

DATA 270: Data Analytics Processes

Dr. Eduardo Chan

September 21, 2022

**E-Commerce Recommendation System: Using Conventional Machine Learning Approach Based on Visual Similarity**

**Proposed Research Problem**

In the world of digital devices and pandemics, the electronic commerce industry is flourishing as the user preference for buying is transformed into online shopping extending from essential to luxury commodities. They mostly rely on search engines that accept keywords from users as input or recommend products to them depending on their purchase history and the behavior of other users who share similar interests. This kind of recommendation system can easily generate extraneous suggestions leading to inferior conditioned and unrelated products for the users due to different semantic usage for each product on different websites. So, even though the virtual stores have access to huge commodity data, without an efficient recommendation system, delivering the desired product to the consumer is a demanding task, and can hamper a potential purchase by confusing the user.

Most of the products substantially appeal to our attention based on their visual appearance. Often, we come across things like a pretty top while browsing Instagram or a cute home décor item at a random place. However, when we search them online to purchase, we would not be able to find the similar one as we are not aware of the right term to aptly describe them which leads to unnecessary scrolling through irrelevant listings. Our approach to perceiving things is a significant predictor of the type of articles we like and hence the saying “a picture is worth a thousand words” translates to an image being worth of thousand searches nowadays. A snapshot of the product contains a wealth of specifications that give a detailed view of its attributes, appearance, texture, and color. Regarding this, we are proposing an efficient recommendation system, where we can give an image as an input instead of text to display or suggest relevant products.

***Problem Statement***

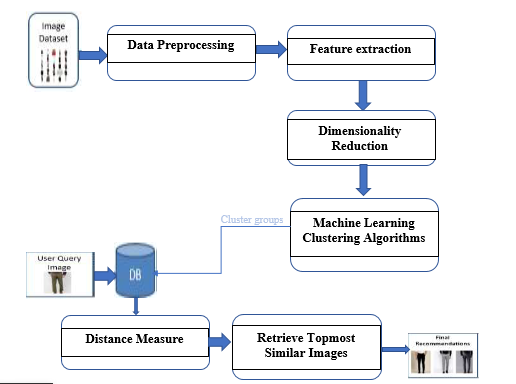
Various kinds of shopping experiences require diverse data to generate recommendations and all those systems are not designed equally nor do they need to be. The advancements in the fields of image processing and the rampant surge in the variation and volume of data available to design recommenders facilitate us to build beyond undemanding collaborative filtering and take advantage of these improvements which can consider the visual similarity of products in recommendations. It is essentially important where the commodities are assessed on their aesthetics and e-shopping is predominantly considered a visual experience. Our objective is to enhance user experience and satisfaction by not only displaying the products that look alike but also accommodating them with possibly better results by retrieving more appropriate products from the input images based on their features.

***Methodology***

We are proposing a smart search engine where users can upload screenshots or any other kind of images, and our model will extract the features and calculate the similarity score between the input and the closely grouped bunch of images in our database and retrieves the most similar items. The model will be trained with a preprocessed dataset of images for various features which accurately describe the object such as texture, shape, and color using transfer learning based on the image weights and consequently refined using dimensionality reduction techniques like Principal Component Analysis (PCA), t-distributed stochastic neighbor embedding method (t-SNE). Once the feature extraction is done, products are grouped to form as many as possible clusters using unsupervised machine learning algorithms such as K-means clustering and Gaussian mixture model algorithms. We will compute the similarity unit from the input image to the closely related cluster in terms of distance as it is the most preferred method, using either Euclidian or Cosine distance where the minimum distance is considered as the highest similarity to fetch the topmost appropriate similar products.

Figure 1

A proposed Image recommender system using unsupervised machine learning models



## *Conclusion*

The proposed model of our research can be implemented in any online commercial applications such as clothing, furniture, and fashion websites to help retailers improve their sales and enhance customer experience by providing quicker access to their desired products. One of the limitations of our model is that it currently works on images that exclusively contain the product alone and further research in areas like object detection, and human detection is needed to improve accuracy for images that contain other noises along with the product. This model can be further extended to perform recommendations using both images and metadata like size, brand, and gender to improve the suggestions to complete a look.

## Background

The research problem is about building an image-based recommender system. Earlier, most shopping sites used search engines. However, these search engines work on keyword matching, which is not efficient in getting the exact match of the searching item. L Chen, F Yang, and H Yang (2015) have built an intelligent recommender that takes objects' images instead of description text as input. They used Amazon product image data from May 1996 – July 2014, which includes 3.5 million products. They have used three models SVM, AlexNet, and VGG. SVM is a linear classification model which operates as a baseline model. The other two, AlexNet and VGG, are deep convolutional neural network classification models. They can achieve a classification accuracy of 0.5 and recommend products with a similarity score higher than 0.5.

A similar problem of image-based product recommendation systems is addressed by Alamdari et al. (2022) using Convolutional Neural Networks. They have used five convolutional neural network (CNN) models- VGG16, VGG19, ResNet50, Inception V3, and Xception to extract features of the product images. In addition, they used four versions of MovieLens Dataset - 100k, 1M, 10M, and 20M. They evaluated their experiment by comparing the results of mean absolute error (MAE), and root mean square error (RMSE) metrics between the KNN (traditional method) and the CNN (proposed methods). RMSE was lower in CNN methods than in conventional methods; based on this, CNN works better in the recommendation systems.

Furthermore, according to Tuinhof et al. (2018), they have composed a method of using a CNN classifier as an image feature extractor and k-nearest neighbors (k-NN) algorithm for the image-based product recommendation used as a ranking algorithm. They have used the publicly available fashion dataset in their research. First, they extracted the feature from an input image using a CNN classifier for the recommendation problem. Then, they submitted them to the k-NN ranking algorithm, which returns the top-k matching style recommendations.

**Related Work References**

Addagarla & Amalanathan (2021) built an efficient system called e-SimNet, which gives recommendations for visually similar products on e-commerce websites. The system is built to get the Top-N recommendations by using Deep learning (DL) algorithms and approximation nearest neighbours oh yeah (ANNOY). Extraction of image features and embeddings was generated using DL algorithms while ANNOY was used for Index tree construction. The SqeezeNet Architecture of the convolutional neural network (CNN) approach did exceedingly well with an error rate of 0.0378% and 96.2% accuracy compared to other architectures such as VCC, DenseNet, VGG and ResNet. Out of the approximation approaches used FLAN and LSH outperformed ANNOY in accuracy, while the system for visual recommendation ANNOY was the best choice.

Pang-Ming Chu & Shie-Jue Lee (2017) proposed a recommended system to address problems caused due to high dimensionality and sparse data, which might result in a bad and inefficient recommendation system. Their novel system uses the comments of users who brought items reflecting their opinions. Word2vec is used to analyze the semantics of text words in order to evaluate consumer reviews, and each word is symbolized by a unique vector. Using the dimension reduction technique, the item vectors are brought down to lower dimension space. At last, the iterative cluster method for calculating robust rating scores (ICRRS) is used to create small clusters of items involved and recommend to users the results generated. To measure the performance of the proposed method and for comparing it with ICRRS, maximum absolute error (MAE) and root mean square error (RMSE) is used. MAE for the proposed method is 1.13 while for the ICRRS it is 1.498 and RMSE is 2.316 for ICRRS and 1.288 for the novel approach presented. The novel recommendation system performs well on real-world data sets, according to experimental results.

Gharaei et al. (2021) designed a clothing recommendation system to address the increase in online shopping due to the covid-19 pandemic situation. The designed method considers the new items’ features and solves the cold start issue. A Deep Neural Network(DNN) is used in this proposed design. DNN is used for feature extraction from item images and feedback from users is also incorporated for system evaluation. Experiments conducted on this system prove the system to be efficient and accurate with 73.7%. The advantages of this system are high speed and serendipitous recommendations of items. We plan to compare our system to other content-based clothing recommendation algorithms that do not use deep learning methods in the future, as well as measure and improve the performance of our proposed method.

Meena et al. (2018) proposed a recommendation model for users which takes input in the form of images and images’ features. To differentiate images into different types, features such as color and shape in images are considered. This study uses mean and standard deviation matrices of the images for the classification and to differentiate from other images, a distance matrix is calculated. Accuracy is measured by calculating various measures such as precision and recall and observing satisfactory results. This study is restricted to the use of small data sets. Using Big datasets and applying Deep Learning technologies to these datasets or more specific efficient models is the future scope of this work.

Batuhan et al. (2019) provide a very unique approach for product recommendation without requiring users’ shopping historical data. The proposed solution is low cost and consumes less power by using a scalable embedded system. Using just a single picture this system recommends more specifically a cloth. Convolutional neural networks for the prediction part and feed-forward neural networks for the recommendation are used. The results show 86%, 98% and 75% accuracies for cloth pattern and gender, colour prediction and on clothing recommendation respectively.

Sejal et al. (2016) proposed retrieving relevant images becomes necessary when we want to meet the requirement of the users. An incredible number of images can be found online, and suggestions based on images can be given by using the search engine. It presented an algorithm that uses the image recommendation with absorbing Markov chain (IRAbMC) to find images that match a customer's input query in the form of an image. Ranking of Images by finding the keyword relevance probability among keywords from the users’ input and annotated keywords is done. The absorbing Markov chain is used to compute keyword relevance. Image visual features are used to rerank images. IRAbMC method performs much better compared to the Markovian Semantic Index (MSI) as seen from the experimental results. The proposed method boosted the relevance score of retrieved ranked images.

Angadi et al. (2021) developed a system to address the trend in online shopping wherein the brands depend on keywords used by customers in search engines to return relevant items. Similar to keywords the search engines can also take images as input and return similar images. A basic Neural network was used initially for input image classification into item categories. Next, the identification of closer images was done using a CNN model. Finally, to enhance the speed and accuracy a transfer learning model was employed. A similarity score of greater than 0.5 for recommending items was achieved. Building a more accurate design is the future aim of this work.

Shrivastava et al. (2019) proposed a textual similarity learning-based recommendation system for women’s apparel. The vectorization techniques used for recommendation are Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words(BOW). Converting the product text description into an n-dimensional vector is the first step and measuring the Euclidean similarity between the searched product and the vector is done at a later stage. The experimental results showed that the system gives a good recommendation of items that are very close in similarity with the queried item. The authors plan to use word2vec and also Deep learning models to improve the system in the future.

**Data Source and Datasets**

We will collect image data related to various categories like clothing, accessories, home, and kitchen appliances from multiple e-commerce websites like Amazon, Myntra, and Walmart by scraping their web pages using Selenium which is a python library that can scrape data dynamically. The dataset will have images grouped into categories and sub-categories which can be used to train our model.

Examples of some of the datasets that can be used are given below:

[**https://vision.cs.utexas.edu/projects/finegrained/utzap50k/**](https://vision.cs.utexas.edu/projects/finegrained/utzap50k/)

UT Zappos50K (UT-Zap50K) is a large shoe dataset of 50,025 catalog images collected from [Zappos.com](http://www.zappos.com/). It divides the images into four categories: shoes, sandals, slippers, and boots. Examples of the dataset are Ordered pairs, Fine-grained Pairs of open, pointy, sporty, and comfortable models such as men's running shoes, a woman's high heels, and a man's slipper. Furthermore, GIST and LAB color features. In addition, each image has eight associated meta-data (gender, materials) labels used to filter the shoes on Zappos.com.

<https://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html>

[The DeepFashion-MultiModal dataset](https://github.com/yumingj/DeepFashion-MultiModal) has a large-scale clothes database. It contains over 800,000 diverse fashion images ranging from well-posed shop images to unconstrained consumer photos. Each image in this dataset has labels with 50 categories, 1,000 descriptive attributes, a bounding box, and clothing landmarks. It also contains over 300,000 cross-pose/cross-domain image pairs.

<https://fashionpedia.github.io/home/Fashionpedia_download.html>

This dataset contains images of people wearing a variety of clothing types in a variety of poses. It contains 48826 files of data and includes images with both segmented apparel categories and fine-grained attributes and images with segmented apparel categories only. It consists of two parts: (1) an ontology built by fashion experts containing 27 main apparel categories, 19 apparel parts, 294 fine-grained attributes, and their relationships; (2) a dataset with everyday and celebrity event fashion images annotated with segmentation masks and their associated per-mask fine-grained attributes, built upon the Fashionpedia ontology.