International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 8 ǁ 2021 ǁ PP. 87-98

Social Media Based Recommender System for E-Commerce Platforms

Ms. Sneha P Tiwari
REVA Academy of Corporate Excellence, Reva University, *Bengaluru, India*
Mr. Ravi Shukla
Consultant, Data Science at Dell Technologies
Ms. Ramamani Venkatakrishna
REVA Academy of Corporate Excellence, Reva University, *Bengaluru, India*

Abstract *- When social media networks first began to appear just over a decade ago, no one expected these new types of forums to take off as quickly as they did. Social networking has developed into something more than merely a means for friends to communicate with one another over time. Businesses have long recognized the benefits of social media and have used it to their advantage, opening the way for others to follow. Before social media, companies had to attend live events to find a specific community of targets, making it more expensive and time-consuming. This has changed now because of social media by changing the way businesses communicate with their audience, allowing start-ups to get in front of a targeted group of people virtually, to make it easier for companies to offer value before asking for anything from a prospect, substituting cards by appearing in the stream of your viewer's feed, etc. Various business tools like Marketing analytics, networking, product promotions/discounts, informal employee learning/organizational growth, partnership building/loyalty services, and e-commerce can all benefit from social media platforms like Facebook, Instagram, Twitter, Pinterest. Despite the avalanche of possibilities that social media marketing offers, advertisers and brands will also face uphill obstacles in terms of characteristics, evolving customer desires, and other social media patterns and challenges, many brands and social media advertisers fail to stay on top. E-commerce sites employ recommendation algorithms to propose commodities for their clientele. The commodities might well be recommended based on top sellers on a site, based on customer's demographics or a study of the customer's previous buying behavior as a forecast for future purchasing behavior. This system remains inactive until the customer finds out about the brand/site and visits the site. With the increase in the businesses interact with customers on social media platforms for various activities we see an opportunity to recognize the customer's lifestyle, understanding their likes and dislikes in terms of clothing and appetite from the customer's posts, tags and captions used. Using which E-Commerce platforms can recommend the products fitting the customer's lifestyle without even waiting for the customer to visit the site*

Keywords*: Recommender Systems, social media, Customer Lifestyle, E-Commerce Platforms*

Date of Submission: 12-08-2021 Date of acceptance: 27-08-2021

I. INTRODUCTION

Within the last decade, the rapid growth of social media is an outcome of human's insatiable desires to communicate as well as the advancements made in the field of Digital Technology. Social media are interactive platforms to create, distribute, and share contents by individuals on the web. Merriam-Webster defines social media as "forms of electronic communication (such as websites for social networking and microblogging) through which users create online communities to share information, ideas, personal messages, and other content (such as videos)". According to Professors Andreas Kaplan and Michael Haenlein of the ESCP European Business School social media is "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content". Social media provides various platforms like blogs, social networking sites etc., through which individuals can communicate, share information, engage, and network. It has gotten quite possibly the most impressive hotspots for online collaboration, entertainment, news updates, networking, and viral marketing.

With the increase in human engagement with social media, various businesses have found new and innovative opportunities to evolve. E-Commerce is one such competent business that has been utilizing social media completely in all possible ways. Social media platforms are used to improve the interaction between customers and e-commerce businesses, not just to educate them on profitable transactions and new product launches. In addition to this several social groups and communities also promote the direct selling of goods.

When any E-commerce platform is integrated with social media, it helps the customers on social media to find out the new brands and discover relatable products through social content. This approach of "discovery through social content" is important for small companies who are starting up, but don't have much money to invest in ads on an e-commerce site. Even if the customer journey is long, it gives the customer adequate time to get to know the brand and provides more time to the company to develop a unique shopping experience for the customer, resulting in long-term loyalty. Social media platforms' social elements also make it simpler for companies to develop interactive, shareable campaigns that help new product launches go viral. Additionally, certain back-end technologies also make it easier to communicate information and new goods purchased by your followers and their peers. You could believe it is difficult to tell the difference between the tangible outcomes and the job put forth by the complex social media networks, but it is possible to implement and maintain parts that enhance the interaction between customers and professional organizations, such as the feeling of correspondence and service providers. Not only does it establish two-way communication in an ideal community, but it also greatly stimulates customer retention.

Fundamentally speaking social media has benefitted the e-commerce industry in many ways, be it developing new and creative business strategy, or enhancing the existing business tools for marketing, customer relationship management, analytics and tracking and providing recommendations. Among all the above-mentioned functions, recommender systems are reshaping the E-commerce landscape by evolving from niche business tools utilized by a few E-commerce sites to major business tools. Many of the major e-commerce sites currently use recommender systems to assist their consumers in finding goods to buy. A recommender learns from a consumer and suggests goods from the available options that she will find most useful.

The primary goal of a recommendation system is to improve the user experience while navigating and, therefore, produce positive commercial results. A recommendation system is a tool that makes online suggestions using a set of algorithms, data analysis, and even artificial intelligence (AI). Depending on the goal of each platform, the quantity of data gathered, and even the sort of technology utilized, these recommendations can be customized for each user by monitoring user's behavior - such as clicks, ratings, and searches. When presenting and recommending a selection of goods, the user data collected makes the decision-making process easier. This user data collected consists of impacted consumer details, other individuals who engaged with the platform previously and products to recommend. It helps businesses drive traffic to the site and increase customer engagement, on the other hand, it benefits the consumer by streamlining the purchasing process, providing a more secure purchase, encouraging profitable purchases, personalizing the shopping experience, and helping us discover new goods.

So far, we have seen many recommendation systems for various fields, using different methods. However, these systems still have few constraints that need to be addressed. This study suggests a method to improve the performance of existing recommendation systems by incorporating social media to analyze customers' behavior on it.

II. Literature Review

Over the last few years, the use of online social media has increased rapidly. People can use Social Media Networks to interact, share, comment on, and watch many forms of multimedia material. Due to their fast change rate, vast volume, and inherent heterogeneity, this phenomenon generates a massive amount of data with Big Data characteristics [2]. Users of social networking sites actively build and join communities to share their media material and rich experiences with a wide range of individuals [10].

However, the usage of social media seems to have an influence not only on the lives of individuals in society, but also on businesses. It has fundamentally transformed the way business function, whether it is in terms of purchasing and selling or marketing. The way businesses discover and connect with their consumers has changed dramatically because of social media. Businesses benefits from social media in several ways by Providing a virtual platform for entrepreneurs to reach out to a specific audience, changing how companies connect with them. customers, by appearing up in your audience's feed, you can replace business cards (which typically wind up in the garbage), making it simple for businesses to give value before asking a prospect for anything an

Here, the term "business" majorly refers to e-commerce platforms. The transformation of businesses because of social media has not only benefited the e-commerce platforms, but it has also transformed the consumer's purchasing lifestyle [2]. Because of the rapid advancement of technology and communication channels, consumers have begun to use more online sources. Social media is the most significant of these technologies. Consumers will have far more access to the information they need regarding products and services that will be rewarded through social media. It is undeniable that social media components such as Facebook and Twitter have been increasingly popular in recent [11]. According to a research, there are eight key takeaways to understand how social media has aided in the transformation of the lives of both kinds of contributors to an e-commerce business: allows Businesses to Become Omni-Present, Increases Personalization, Generates More

Loyalty, Increases Industry Collaboration, Adds Credibility to Your Brand, Increases Referrals, Helps You Build a Personal Brand, Helps You Gauge Audience Feedback [13].

In this perspective, with the evolution of social media and social media-based applications, one cannot deny that every individual wants to utilize social media entirely [1]. A survey aimed to determine the impact of social media on our lives in recent years, as well as the impact of social networks on consumer purchasing behaviour in Pakistan. A sample of 1,000 young consumers aged 18 to 50 years old who used social media platforms and had an account on any of the social networks participated in this study. The data for the study was collected via a questionnaire and the findings from the data showed that social media has a major influence on Pakistani consumer behavior [11].

One cannot deny the fact that evolution of social media is the result of digitization, hence the combined effect of social media evolution and digitization has diversified consumer segments, and consumers now have broader choices with shorter production cycles. This has a significant impact on the fashion business. As a result, a system that effectively enables product searching and recommendation is becoming more vital. However, because of the nature of the fashion business, where design is so essential, the text-based search approach has limits. A researcher created a Sketch-Product fashion retrieval model and a vector-based user preference fashion recommendation model, both based on deep learning, for quick fashion product searches and suggestions. The vector-based fashion recommendation model performed admirably as well, and the method is anticipated to improve customer satisfaction by assisting users in more efficiently looking for fashion goods or proposing fashion products before they start searching [8].

Recently, Recommender Systems (RS) have been developed to aid users in finding "what they truly need" among vast amounts of data [2]. User textual reviews, ratings, and compared opinions are used to create RSs. The RS makes use of social network resources to propose material, articles, news, e-commerce goods, and individuals [3]. RSs are evolving from niche marketing tools employed by a few E-commerce sites to major business tools that are reshaping the E-commerce landscape. Many of the major e-commerce sites currently use recommender systems to assist their consumers in finding goods to buy [14]. A research paper explains how recommender systems helps Ecommerce sites improve sales and provide ideas for new applications of recommender systems to E-commerce. A taxonomy of RS is created by examining six sites that utilize recommender systems which includes the interfaces they provide to consumers, the technologies they employ to produce suggestions, and the inputs they require from customers [14].

Even though research on social media in RSs has grown year after year, adequate literature reviews and classification of these RS studies are still lacking and must be improved. [3]. Most on-line shopping search engines still rely heavily on their knowledge bases and employ key word matching as a search method to locate the most likely goods that customers wish to purchase, this is inefficient in the sense that product descriptions might differ greatly from vendor to the consumer A researcher attempted to solve a similar problem, by providing a smart search engine which takes photos as input and deduces information about items and classifies the input image into one of the product categories using a neural network. Then, using a different neural network, model the similarity score between pairs of images, which is to choose the nearest product [4].

The concept of recommendation isn't limited to e-commerce; it's also utilized to suggest social groups (or individuals) to other individual users on social media. It is the subject of one study where it focuses on the potential of social tagging for highlighting users' interests and defining communities, also go through several novel approaches to improve a variety of strategies that have been adopted for use in community recommendations: collaborative filtering, a random walk model, a Katz influence model, a latent semantic model, and a user-centric tag model. Each algorithm successfully incorporates social tagging information and presents an empirical evaluation based on genuine CiteULike and Last.fm datasets. The findings show that the various algorithms combined with social tagging provide considerable [10]. Another similar research paper presents and explains a unique recommender system system for big data applications that can make suggestions based on user interactions and the multimedia content created. A "user-centered" approach is used in the recommender system. An experimental campaign was conducted utilizing data from many social media networks to evaluate the suggested strategy and demonstrate how it may provide extremely promising outcomes [2].

Based on findings from the above cited research papers we perceive how a recommender system has evolved both in the field of e-commerce and social media. Further this study focus on how social media recommender systems may be used to provide recommendations from an e-commerce site.

III. Objective

As a result of the fast growth of the Internet in the previous decade, many E-commerce websites and applications have emerged. People are increasingly preferring to shop online rather than in physical establishments. Besides this, social media sites also have grown in popularity because of this evolution, accessing social media such as Twitter, Facebook, Linked In, Instagram, and YouTube has become very affordable. Since many individuals are utilizing social media sites irrespective of age groups, it has become incredibly easy for businesses to reach out to their customers via these platforms. Mentioning businesses using social media, one cannot overlook how e-commerce and social media have merged to create an exciting and lucrative shift in retail marketing. The e-commerce businesses are immensely utilizing social media by incorporating it with the major business functions including trading of products, marketing, recommendation, customer service and related analytics. Accordingly, these businesses are facing stiff competition from peer websites and applications and hence businesses are trying to make use of a variety of strategies to set themselves apart.

An over-view target of this project is to revise the traditional Recommender system for e-commerce businesses by designing a social media-based Recommender System that can incorporate insights from various social media contents of a given target customer such as captions, hashtags, posts, statuses, stories. These extracted contents from a social media platform for given target aid to figure out the behavior and style of life of the customer. Based on which the e-commerce sites can articulate the products for recommendations such that it fits the customer's preferences and lifestyle.

The study focuses on developing a data science-based model as well as a simple recommender engine. For a given image from the social media account of the target customer, the model attempts to identify three primary criteria for a particular image from the target customer's social media account: the apparel type of the person wearing in the image, the person's gender, and the color of the apparel. These parameters are then passed into the recommender engine, which selects similar products from the given product catalogue list of an e-commerce site to provide recommendations back to the target customer.

3.1 Relevant scopes for the study:

- Product type is scope: Fashion apparel, the fashion industry is the most fascinating and important one. Just about all the individuals are direct fashion consumers, be they shopping at stores or flea markets.
- Social media in scope: When spoken about fashion, Instagram is one such social media platform that has greatly impacted the fashion industry. Fashion bloggers, social media Influencers and other Fashions enthusiasts are seen active on Instagram
- E-commerce site in scope: Myntra.com, is major Indian fashion e-commerce. When it comes to shopping for fashion items online, it is regarded as one of the largest e-commerce platforms.

IV. Project Methodology

CRISP-DM methodology was used to execute this project structurally, it is a popular framework for data mining project planning. This is a framework that has been employed in a variety of industrial applications and has proven to be effective. This is an unrestricted process where one can move backwards and forward among different stages. The arrows indicating the necessity among the phases are equally significant; the outside circle shows the cyclical characteristics of the framework. As the outer circle figure demonstrates, CRISP-DM is not a one-time process. Every phase is a fresh learning experience from which we can learn new things, and which may lead to other business problems.

V. Business Understanding

E-commerce is any type of business or business transaction that involves the sharing of information over the Internet. It is a business model over the internet. When the purchasing and selling of any goods are made through electronic media it is often referred to as e-commerce. It has fundamentally revolutionized the way entrepreneurs do business, as a result, it is witnessing wide reach and popularity, and it has been used by everyone from tiny firms to large corporations. It has completely transformed the retail market in the last ten years. It evolved from an almost non-existent business model to a potential challenge to the recent walk-in stores and malls as we call it.

E-commerce has changed and evolved resulting in tremendous exponential growth to meet today's consumer's ever-changing preferences and expectations. It was quite limited in its early days. It was as simple as; you came upon something you liked and decided to purchase it. Customization was out of the question. Today, one may not only personalize a product online but also have it created to one's specification and purchase without ever leaving the house. Retailers and customers both stand to benefit from the e-commerce business.

The online business industry has undergone some radical reforms in recent years: Large businesses are forced to sell their products online, small businesses are now operating from social media platforms, which has increased in local enterprises, operational costs have decreased and costs of parcel delivery have increased, logistics have progressed with the development of automation techniques and artificial intelligence, Customer's purchasing patterns have shifted dramatically, social media has evolved into a tool for boosting sales and promoting brands, merchants are increasingly trading on social media platforms such as Facebook and Instagram.

In today's world, where most people prefer to purchase online, the current state of e-commerce appears to be quite favorable, with more and more people going online with their e-commerce stores, and it is projected to reach its pinnacle in the next years. For optimal growth, e-commerce will constantly require the introduction of new technologies and marketing approaches. In future, the businesses that value their e-commerce business, devote sufficient resources to develop and market online-based solutions will undoubtedly succeed.

VI. Data Understanding

After defining the problem statement and deciding the approach to be followed, the next significant step was to consult the industry experts from the field of Deep learning and the e-commerce industry to understand the details on the requirement of the dataset. As a result, this project required two different datasets, and images dataset, and a product list dataset. The image dataset is used to train the deep learning model to predict the apparel type, while the product list dataset will be utilized to pick and provide a recommendation.

6.1 Product List Dataset:

As previously mentioned, the e-commerce platform considered here is Myntra, a prominent online fashion store. A sample dataset of the product catalogue list from this e-commerce website is utilised for this project. This dataset collaboratively created by both Datastock and PromptCloud has 15,000 records for the period from 01st Jun 2019 - 31st Aug 2019. The dataset includes following listed fields:uniq_id, crawl_timestamp, uniq_id, crawl_timestamp, product_id, link, size, variant_sku, brand,care_instructions, dominant_material, title, actual_color, dominant_color, product_type, images, body, product_details, size_fit, complete_the_look, type, variant_price, variant_compare_at_price, ideal_for, is_in_stock, inventory, specifications. Among these 25 available fields, 8 relevant fields are considered for this project:

1. product_id: The data type of the column here is Integer. This is the unique id given to each product listed on Myntra's online website.
2. link: The data type of the column here is the URL. This is the hyperlink to the product on the website, that may be clicked to go straight to the product on the website.
3. size: The data type of the column here is String. This refers to the clothing size of the garment/product listed. It is the label sizes used for garments sold off-the-shelf.
4. dominant_color: The data type of the column here is String. This refers to the colour of the mentioned clothing item. In the case of multi-colour, significant colour is listed.
5. Product_type: The data type of the column here is String. This is the category type the clothing product belongs to. In the case of many sub-categories, the lowest category is listed.
6. Images: The data type of the column here is the URL. This is the hyperlink to the image of the product on the website, that may be clicked to view the image online.
7. variant_price: The data type of the column here is Integer. This is the product's maximum retail price as displayed on Myntra's online website.
8. ideal_for: The data type of the column here is String. This refers to the set of individuals for whom the listed clothing would be suited. It has 5 distinct values women, men, unisex, girls and boys.

6.2 Images Dataset:

This dataset created is utilised in building the apparel type prediction model, hence the dataset had to be included with all possible kinds of images of individuals wearing a variety of clothing. This was achieved by extracting images from Instagram and Google about the product_type classes listed in the Myntra product catalogue list dataset.

6.2.1 Extracting images from Instagram:

Instagram Scraper is used to scrape required images from Instagram. It is a free command-line tool that can be used on a PC to scrape public photos from the Instagram front end. It allows downloading a defined quantity of Instagram images based on given inputs like a hashtag, profile name (public account), location. As mentioned above, to include all types of apparel categories listed in the product_type column of the Myntra product list dataset, different category names are used in form of hashtags to extract the images.

6.2.2 Extracting images from Google:

To assure better training of the model with all possible images of people wearing different clothing, Google images are also included. The same is accomplished by using an online tool Image Extractor or Extract. pics, it is a simple online tool that lets one extract and download images from any publicly accessible website.

To begin, just paste the website's URL into the input area and click "Extract", the process will take few seconds. After the operation, the grid displays a list of all pictures in the input URL. Extracted Images can be downloaded in a ZIP file by clicking on “Download selected”.

Finally, a list of combined folders is created from all the photos retrieved using both techniques for all the categories. As per the product_type field, there are 261 classes of apparel types which were scaled down to 107, related hashtags and keywords were mapped and used for scraping the images. The below tables illustrate the mapping, count of the number of products and total numbers of images extracted for each class.

Description	Count
No. of classes in Product Type	261
No. of Mapped Hashtags/Keywords	107
Total no. of Images extracted	11365
Total no of Products	13580

Table No. 6.1Overall distribution of Data

VII. Data Preparation

Subsequently, post business understanding and data understanding, the CRISP-DM methodology's next major goal is to prepare the data for modelling and analysis. This entails choosing, cleansing, and converting the data to be utilized in the project. Although preparing raw data for analysis frequently necessitates a lot of work, but the phase is significant as per the old saying "garbage in, trash out. At the end of the data extraction, there were 107 classes of images. The no. of classes for prediction was very high. There was a dire need to reduce the classes further, the same was achieved by combining relevant categorized into a single class, yielding 19 final labels for data modelling. After merging the hashtags and producing a new set of Final Labels, the data must be mapped into Myntra's product catalogue list dataset to make product selection for recommendations simpler.

7.1 Splitting of the dataset into Train and Validation:

Splitting the dataset for training and validation after all the data gathering and required transformations is a crucial step of assessing data mining models. When a data set is split into a training set and a validation set, the larger section of the data is used for training and a smaller section of it is utilized for validation. It takes a random sample of the data to guarantee that the training and validation sets are comparable. You may reduce the effects of data inconsistencies and gain a better understanding of the model's features by utilizing similar data for training and validation.

7.2 Image Conversion:

Before modelling, all the images in the dataset must be normalized to one format of images. Since most of the images were in .jpg, all the images were converted to .jpg. Python Imaging Library (PIL) is used here to perform the conversion, and the same are copied into a new set of folders.

VIII. Data Modeling

After integrating and transforming the data to a required form, every data science project's main step, Data Modelling begins. The major goal of this phase is to satisfy all the business goals discussed and mentioned in the previous stages. To accomplish the main objective of developing a social media-based recommendation system, data modelling is divided into five different modules: Apparel Identification, Gender Identification, Color Identification, Instagram login and targeting, Recommendation Engine.

8.1 Module1: Apparel Identification

The apparel identification module is the most important model built in the project out of all the others, The model's purpose is to investigate the fashion clothing type that the person is wearing in the source image. With the 19 classes images dataset to be used in training the model, it becomes a multi-class classification problem. The ultimate model was achieved after developing the model in two different phases.

8.1.1 Phase 1:

Generally, when a problem statement emerges out of the Images dataset, the topic of Deep learning usually follows. Due to its high degree of performance across various types of data, deep learning is becoming a hugely appealing subset of machine learning. Building a convolutional neural network is a fantastic method to utilize deep learning to categorize pictures (CNN). The Keras Python library makes building a CNN very simple. Computers view images in terms of pixels, which are often correlated to each other. For instance, a specific set of pixels could mean an edge in the picture or maybe another pattern. Convolutional neural networks use this to recognize images. CNN multiplies the pixel matrix with a filter matrix or a core (core) and adds multiplication values. Then the convolution slides to the next pixel and repeats the same process until all the image pixels are covered.

Subsequently, an activation function for any neural network model is vital. The neurons in a neural network work following their weight, bias, and activation function. Neuronal weights and biases in a neural network are tuned based on the output error, which is known as Back-propagation. On the other hand, an activation function makes the back-propagation achievable by taking a call on whether the neuron must be activated. The goal of any activation function is to induce non-linearity into a neuron's output.

Considering CNN for this project with ReLU and SoftMax as activation functions since this is a multi-class classification problem. At Phase 1, three slightly different sets of models are built and run on the dataset.

a) **Basic CNN**: Beginning the building of a model with CNN is not a bad idea, yet one cannot expect a convincing result. The final metrics of the basic CNN were 0.4351 for training accuracy and 0.2920 for validation accuracy.

b) **Basic CNN with Augmented Images**: Based on the findings of Basic CNN, it was determined that 11364 pictures were insufficient to train 19 image classes. Following which augmenting the existing images to increase the number of images was the next best step. Image augmentation is a helpful approach for increasing the size of the training set without having to acquire additional photos while developing convolutional neural networks. The concept is straightforward: replicate photos with slight variations so that the model may learn from additional examples. Post augmentation, the same Basic CNN was executed on the updated dataset, with a training accuracy of 0.0495 and validation accuracy of 0.2040.

c) **VGG16 with Augmented Images**: With the model's accuracy not improving substantially, the next best step was to switch to a pre-trained model. For multi-class classification, VGG16 is said to astound with better results. Karen Simonyan and Andrew Zisserman of Oxford University's Visual Geometry Group Lab suggested VGG 16 in their article "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION" in 2014. With a dataset of over 14 million pictures belonging to 1000 classes, the model achieved 92.7 percent test accuracy, made it a well-known model that was submitted to the ILSVRC-2014. Here the VGG16 model was run with adding 2k images more for all 19 categories, followed by freezing few layers. A substantial shift in the metrics was found here, with training and validation accuracy of 0.9636 and 0.5356 respectively. The following snapshots show the, model params and results.

8.1.2 Phase 2:

The results of Phase 1 showed that the number of pictures needed to train 19 classes was still insufficient, thus phase 2 was carried out by decreasing the classes to 5. Iterating the model's execution by increasing the number of classes one by one up to a maximum of ten. The following snapshots provide model parameter details and results.

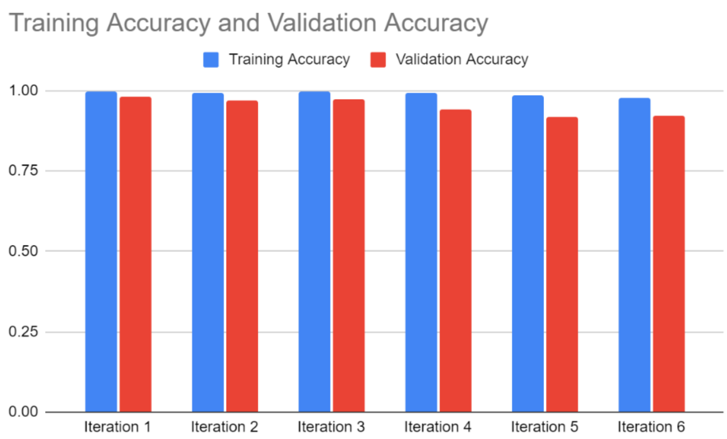

Figure No. 8.1 Phase2_Results

8.2 Module2: Gender Identification

Post apparel type prediction, identifying the gender of the person in the image becomes another significant component for the selection of products to provide recommendations. This is attained by utilizing a pre-trained model by Tal Hassner and Gil Levi in their publication “Age and Gender Classification Using Convolutional Neural Networks” [6].

This model uses a convolutional neural network to accurately identify the gender and age of a person in an image. The predicted gender may be one of ‘Male/Female’, and the predicted age may be one of the following ranges- (0 – 2), (4 – 6), (8 – 12), (15 – 20), (25 – 32), (38 – 43), (48 – 53), (60 – 100). However, only the gender predicted from the output will be made used in this study. This model is trained using the Adience dataset, built with 3 convolutional layers, it has 2 fully connected layers, each with 512 nodes, and a final output layer of softmax type.

- Convolutional layer; 96 nodes, kernel size 7
- Convolutional layer; 256 nodes, kernel size 5
- Convolutional layer; 384 nodes, kernel size 3

For this study, an instant zip file with all the required source files from data-flair.training.com is made used [6]. Here the model begins with detecting the face, classify it into male/female, classify the age as one out of the 8 age ranges mentioned above and followed by displaying the predicted gender and age. It consists of the below-listed files, where the .pb file is a protocol buffer file that has the graph definition and the trained weights of the model in binary format and the file with the .pbtxt extension hold it in text format. Concerning gender and age, the .prototxt files define the network configuration while the .caffemodel file describes the internal states of layers parameters.

- opencv_face_detector.pbtxt
- opencv_face_detector_uint8.pb
- age_deploy.prototxt
- age_net.caffemodel
- gender_deploy.prototxt
- gender_net.caffemodel
- detect.py

Upon saving the files on the PC, below are the final steps followed to witness the output: Installation of all the pre-required libraries OpenCV and argparse, Open the Command Prompt, change the working directory to the same folder where the above-listed files are saved. Run “ py gad.py –image < image_name> “on command prompt, View Output.

8.3 Module3: Color Identification:

The color identification module, one of three components for a recommendation, is primarily designed to identify and choose a smaller set of items for suggestions. This module is designed to calculate the proportion of each of the identified colors in the input image, with the dominating color shown as the third parameter for selecting recommendations.

Beginning with the previously encoded colors: 000 = BLACK, 001 = RED, 010 = BLUE, 011 = GREY, 100 = GREEN, 101 = YELLOW, 110 = ORANGE, 111 = WHITE. This pre-trained model mainly consists of four functions:

- scale_RGB() function is to calculate the primary colors red, green and blue components by passing the respective pixel values of the primary colors from the given input image.
- sigmoid() function calculates the sigmoid function for the given pixel value deduced in the recognizeColor() function and passed to the trainRGB() function.
- trainRGB() function is the primary function where instant training of the model is carried out with activation values returned from the sigmoid() function. To identify the colors in the input image.
- recognizeColor() is the final function, accepts the input image, returns height and width of the image, returns pixel values to calculate RGB components, and calculates proportions of colors identified and print the values.

8.4 Module4: Instagram login and targeting

A basic module for extracting an image of interest from the Instagram account of a specified target customer, to which the previously constructed modules for detecting three attributes may be applied. There are numerous libraries to access and extract Instagram using python, Instaloader is one such library available freely. It lets you download Instagram photos (or videos) together with their captions and other metadata. It exposes its core methods and structures as a Python module, resulting in a robust and user-friendly Python API for Instagram, allowing for more customized media and metadata extraction. The steps to get the module operating as mentioned above are as follows: Import Instaloader and required method and structure, Get an instance of Instaloader, Due to the nature of Instagram as a platform and the fact that we need to navigate to the target customer’s Instagram account, we will have to use one of our Instagram accounts to Log in with the instance created, Once after logging in, define the target customer’s user name and pass it as argument in download_profile() function, The function generates a folder containing all the photos obtained from the target account, from which a random image is chosen and sent to the predictor modules.

8.5 Module5: Recommendation Engine

The major goal of any recommender engine is to select and provide suggestions of various available products based on the input. This module is designed to accept all the attributes predicted by the respective models from the image extracted from the target customer's account on Instagram. These predicted attributes: Apparel type, Gender and Color of the apparel are further utilized to pick products for recommendations from Myntra's product list dataset, which results in list of products similar to the apparel wore by the target customer in the extracted image. This simple module begins with accessing Myntra's dataset using pandas as a data frame. Assign the three attributes resulting from the prediction models to variables, that are passed as condition checks to select the item's hyperlinks from the data frame. The selected items are ultimately displayed as recommended products. Further, these hyperlinks can be clicked to navigate directly to the product on Myntra's online website.

IX. Model Evaluation

In the initial four stages of the CRISP-DM process, after exploring data and discovering the relevant patterns solving the business objective, it's time to ask the question: Are the results good? Not just evaluating models, but also the technique utilized to make them, as well as their potential for practical application. Generally, data evaluation comprises three major tasks evaluating tasks, reviewing the process, and determining the next steps. In Phase 1, a simple CNN model did not produce any compelling results. After running a pre-trained model VGG16 with an updated dataset that included augmented images and 2k additional images after Phase 1, we saw a substantial boost in training and validation accuracy. The same may be seen in the table below.

Phase1	Training Accuracy	Validation Accuracy
Basic CNN	0.4351	0.292
Basic CNN + Augmented Images	0.0495	0.206
VGG16 + Augmented Images + 2K images	0.9636	0.5356

Table No. 9.1 Phase1_Results Comparison

Based on the results of the last model in phase 1, which had a training accuracy of 0.9636 and a validation accuracy of 0.5356, we believe that 19 classes are a big number for a model with a dataset of 88967 pictures. Given the time constraints, it was necessary to limit the amount of data collected, thus label reduction was the next best option. The below table demonstrates 6 iterations performed starting from 5 labels up to 10 labels. Here the previous VGG16 + Augmented Images + 2K images are executed.

	Classes	Train Images	Val Images	Training Accuracy	Validation Accuracy
Iteration 1	5	19024	4368	0.9965	0.9812
Iteration 2	6	22926	5223	0.9935	0.9712
Iteration 3	7	26804	6024	0.9964	0.9737
Iteration 4	8	30691	6931	0.9923	0.9406
Iteration 5	9	34481	7765	0.9866	0.9169
Iteration 6	10	38424	8708	0.9775	0.9206

Table No. 9.2 Phase2_Datasize, Labels and Results

Considering the findings from phase 2, the model executed at Iteration 3 with 7 classes and validation accuracy of 0.9737 is the best model that can be considered for the deployment. Below snapshots are from the output of the Iteration 3 model.

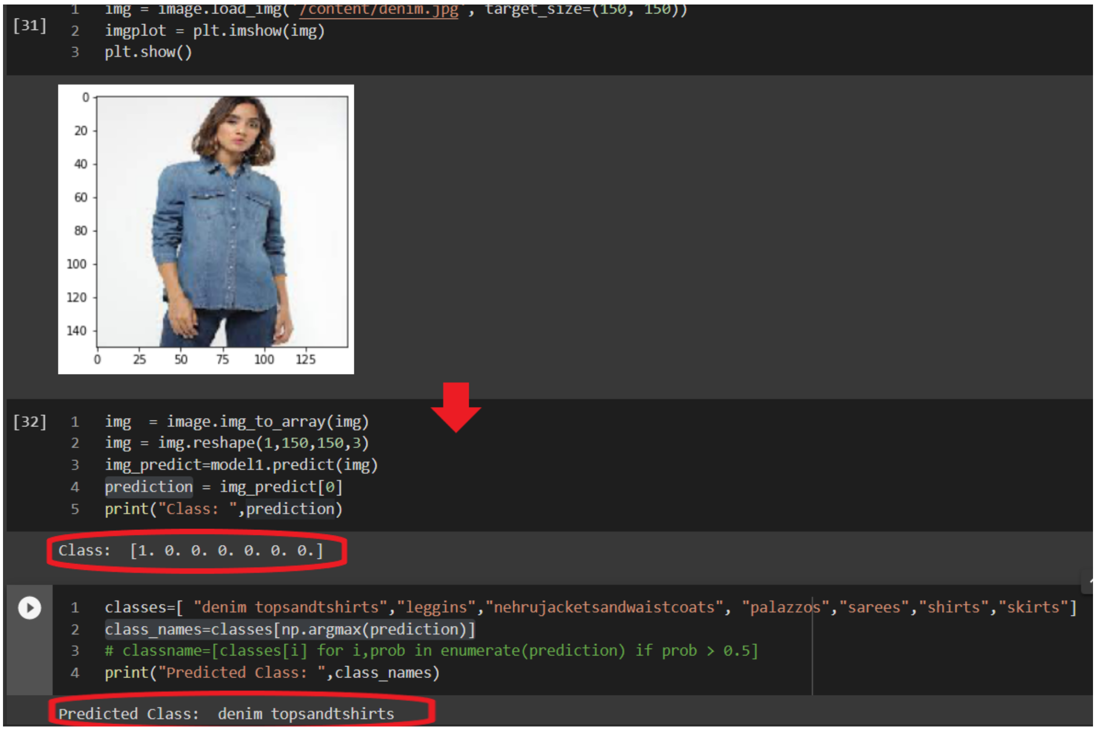

Figure No. 9.1 BestModel_Output

X. DEPLOYMENT

All the efforts in the last five phases are rewarded in this stage. It doesn't matter how brilliant your findings are, or how precisely your models match the data, if you don't utilize them to enhance the current business model, it continues to be just an untested hypothesis. A model is not useful unless the results can be accessed by the stakeholders. In the case of this project, the primary stakeholders are customers active on Instagram and Myntra, an e-commerce platform. The overall scope of deployment for this project is to demonstrate a complete end-to-end flow of the social media-based recommender system built for key stakeholders by creating a web page that can be hosted locally. The integration of all the modules developed as part of Data modelling is made by utilizing high-level Python web frameworks Django and Bootstrap.

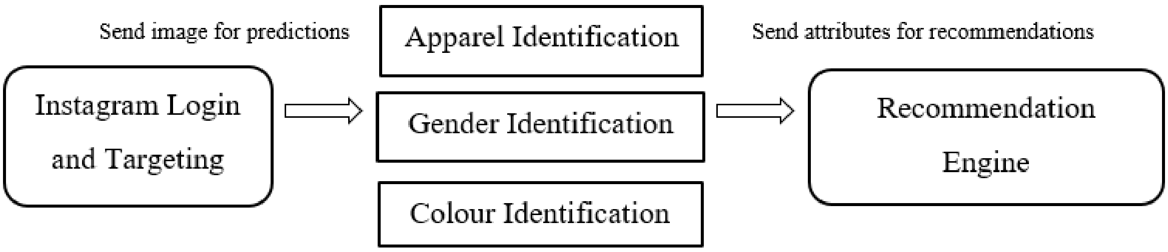

Figure No. 11.1 Deployment Block Diagram

The purpose of this webpage is to show the complete flow of the modules produced in earlier phases, as well as their integration and corresponding functions. The webpage provides a text box “Enter Target Username” to enter the target customer’s Instagram profile name, accepting which the application downloads the images posts from the account and selects one image at random for further processing. Upon clicking on the “Get Recommendations” button, the user is navigated to the result page. Here the page is populated with the three predicted attributes and the list of recommendations. These recommendations are in the form of product images along with the product description and a “Find This on Myntra” button, by clicking it one may navigate to the respective products on Myntra’s online site.

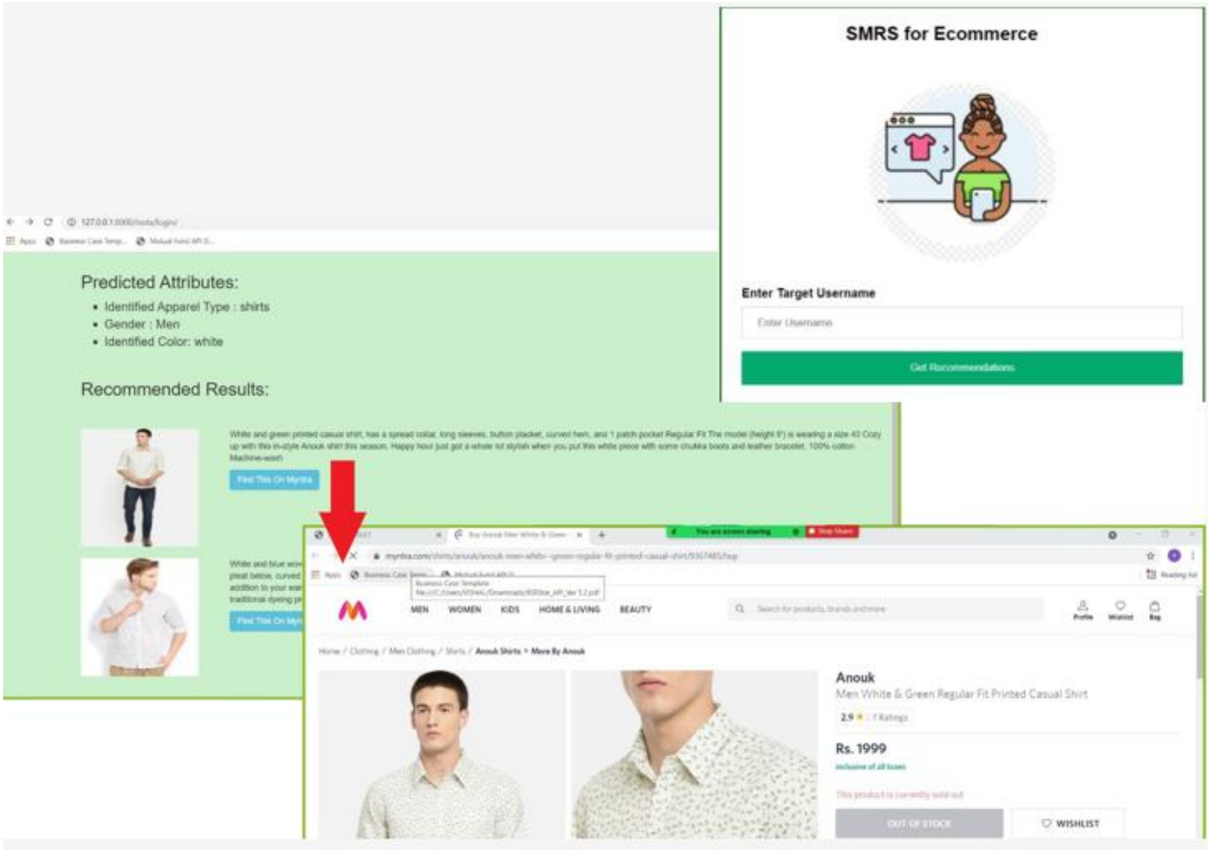

Figure No. 10.1 Deployment_webpage

XI. CONCLUSION

As discussed, the overall goal of this project of creating a social media-based Recommender System for an E-commerce platform is demonstrated by creating a locally hosted web page. The Instagram username of the target customer is keyed into the web page, upon clicking on Submit, the image picked by the model for prediction is displayed along with the predicted criteria considered by the recommender engine to provide recommendations. These recommendations are paginated on the webpage together with an image of the product and hyperlink that navigates to a particular product on Myntra's website.

This was achieved by building five different modules, out of which three modules are deep learning-based models developed each for identifications of the fashion apparel type, gender identification and apparel's color identification using the scraped images from Instagram and Google across various apparel types mentioned in the Myntra's product catalogue list.

The apparel identification module based on VGG16, a pretrained model with few variations on the layers showed impressive results on classification of different apparel type up to 10 labels, out of which model with 8 labels and test accuracy of 0.9737 was chosen for deployment. However, classification of apparel types based on top and bottom wear separately is still lacking, which can one of the recommendations for future work. The gender identification module built by Tal Hassner and Gil Levi is made used as a filter to specifically select the products for recommendations. As claimed by the author the model performed brilliantly well across all the images of individuals of different age group (Eran Eidinger, Roee Enbar and Hassner, 2014). Further the age predicting ability of this module can be utilized as another feature to pick out the recommendations. The color identification module is another model built to filter recommendations. This model is designed to identify eight different colors and return the proportions of each of these colors from the given input image. It is inferred that the model can group all the various colors from the input image to either of these eight colors. The fourth module is developed for log in to Instagram, navigate to the target customer's Instagram account, download all the image posts, and pick an image for parameter predictions. The last module is built to accept three predicted parameters and suggest products passing all the three conditional parameters against Myntra's product catalogue list and the results are made as recommendations back to the customer. Overall, the built system is apt to identify eight different apparel types, eight different colors and the gender aligning to which the products are fetched from an ecommerce product list to provide recommendations.

As E-Commerce websites have evolved, customized recommendations to the customers have gained interest recently. The approach described in this paper is intended to overcome some of the major limitations of the current recommender system like a cold-start problem for e-commerce, the problem of outdated and updating of customer data used to provide recommendations and problem limited resources, which, in the end, can provide helpful recommendations to customers. Customer satisfaction can be improvised further by incorporating the solutions provided into the many existing e-commerce websites. So far, we have tried to recognize the limitations of existing systems and attempted to find solutions to them. Future work could include combining several areas to troubleshoot the cold start and incorporating the mentioned approach with many other existing recommender systems to provide more effective recommendations.

Considering the above-designed social media-based Recommender System and outputs, below are some of the recommendations for future work:

- Tuning the existing algorithm to include more categories of apparel types for predictions
- Developing more advanced models to include other fashion accessories such as footwear, bags, goggles, caps etc. for predictions.
- Providing outfit ideas from the products being recommended.
- Utilizing other social media content like captions, hashtags, stories, and status to provide more effective and real-time recommendations.
- Integrating the approach mentioned here along with multiple social media platforms for a given e-commerce site.
- Considering user feedback for the to make recommendations.

REFERENCES

[1]. Alexandre Wrg. (2019). Image recommendation engine — leverage transfer learning. https://towardsdatascience.com/image-recommendation-engine-leverage-transfert-learning-ec9af32f5239

[2]. Amato, F., Moscato, V., Picariello, A., & Piccialli, F. (2019). SOS: A multimedia recommender System for Online Social networks. Future Generation Computer Systems. https://doi.org/10.1016/j.future.2017.04.028

[3]. Anandhan, A., Shuib, L., Ismail, M. A., & Mujtaba, G. (2018). Social Media Recommender Systems: Review and Open Research Issues. IEEE Access. https://doi.org/10.1109/ACCESS.2018.2810062

[4]. Chen, L., Yang, F., & Yang, H. (2015). Image-based Product Recommendation System with CNN. Http://Cs231N.Stanford.Edu/Reports/2017/Pdfs/105.Pdf.

[5]. Dataflair. (2019). Interesting Python Project of Gender and Age Detection with OpenCV. https://data-flair.training/blogs/python-project-gender-age-detection/

[6]. Eran Eidinger, Roee Enbar, and T. H. (2014). Age and Gender Estimation of Unfiltered Faces. https://talhassner.github.io/home/projects/Adience/Adience-data.html

[7]. Forbes, L. P., & Forbes, L. P. (2013). Does Social Media Influence Consumer Buying Behavior? An Investigation of Recommendations And Purchases. Journal of Business & Economics Research (JBER). https://doi.org/10.19030/jber.v11i2.7623

[8]. Jo, J., Lee, S., Lee, C., Lee, D., & Lim, H. (2020). Development of fashion product retrieval and recommendations model based on deep learning. Electronics (Switzerland). https://doi.org/10.3390/electronics9030508

[9]. Kilgo, D. K., Harlow, S., García-Perdomo, V., & Salaverría, R. (2018). A new sensation? An international exploration of sensationalism and social media recommendations in online news publications. Journalism. https://doi.org/10.1177/1464884916683549

[10]. Kim, H. N., & El Saddik, A. (2013). Exploring social tagging for personalized community recommendations. User Modeling and User-Adapted Interaction. https://doi.org/10.1007/s11257-012-9130-3

[11]. Madni, G. R. (2014). Consumer's Behavior and Effectiveness of social media. 7. file:///C:/Users/sneha/Downloads/1582-1-1561-1-10-20150216.pdf

[12]. McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. SIGIR 2015 - Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. https://doi.org/10.1145/2766462.2767755

[13]. MEGAN MAHONEY. (2021). The Impact of social media on Business in 2021. https://www.singlegrain.com/blog-posts/impact-of-social-media-in-todays-business-world/

[14]. Schafer, J. Ben, Konstan, J., & Riedl, J. (1999). Recommender systems in e-commerce. ACM International Conference Proceeding Series. https://doi.org/10.1145/336992.337035

[15]. Tuinhof, H., Pirker, C., & Haltmeier, M. (2019). Image-based fashion product recommendation with deep learning. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-13709-0_40

[16]. Ullah, F., Zhang, B., & Khan, R. U. (2020). Image-Based Service Recommendation System: A JPEG-Coefficient RFs Approach. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2962315

[17]. Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning-based recommender system: A survey and new perspectives. In ACM Computing Surveys. https://doi.org/10.1145/3285029