iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IX **Month of publication:** September 2022

DOI: https://doi.org/10.22214/ijraset.2022.46904

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

Ecommerce Website with Recommendation System Including Chatbot and Reverse Image Search

Preeti Badave[1], Bhakti Bhomaj[2], Bhargavi Bindu[3], Riddhi Shivarkar[4], Prof Nikhil Dhavase[5]
[1, 2, 3, 4, 5]*Department of Information Technology, Marathwada Mitra Mandal's College of Engineering, Pune*

Abstract: ***In Today's generation, it has become a need to get clear vision, prioritize and effortlessly to provide appropriate information to decrease the problem of information overload, which has created a major problem for many users using internet. Recommender systems is capable to solve this major problem by using searching technique through a huge volume of dynamically generated useful data to provide users with personalized content and services depending on their interest. Recommendation system plays an important role in the internet world and is used in many applications. Recommendation systems are of three types: 1) collaborative filtering 2) content based and 3) hybrid based approach. With the advancement of machine learning for decades, there are still numerous problems insolvable, such as image recognition and location detection, image classification, image generation, speech recognition, natural language processing and so on. In this emerging field of deep learning, the research on topic like image classification has always been the topmost traditional research to be evolved. Simultaneously, Image recognition technology is also beneficial to gradually respond in better way to the development of international indicators, and betterment the development and progress in various fields. Hence, image processing technology which is based on machine learning has always been widely used in feature image segmentation, classification and recognition, and is an important topic in various fields. It allows you to find similar products in your store in a very easy way. Let the customer find the product based on the photo. What's more, you no longer have to manually match similar products, it will be done automatically by AI algorithm. The main purpose of this project is to use the concept of a machine learning algorithm to build a Recommendation and Reverse Image Search and recognition with the help of convolutional neural networks (CNN).***

Keywords: Recommendation System, Reverse Image Search, CNN, Vectorizer

I. INTRODUCTION

A recommendation system is an vast class of web applications that involves predicting the user responses to the options. Recommender systems are uncomplicated algorithms which provide the most accurate and exact relevant items to the user by filtering out needful or useful information from of a huge volume of data. Recommendation engines has ability to discover various data patterns in the data set ,which would be useful for filtering .

This can be done by learning customerchoices and interests and provides the accurate results that best co-relates to customers needs and interests.

It allows you to find similar products in your store. The customer will find the product by uploading its image. They will no longer have to manually match similar products, it will be done automatically by AI algorithm.

II. MOTIVATION

Modern e-commerce is about much more than just offering products online We aim to build a virtual shopping assistant using image recognition technologies to improve the overall online shopping experience for the users. The application has two main components image recognition service and a product recommendation system. Each component is an important part of this application.

III. AIM AND OBJECTIVES OF THE WORK

A. Project Aims

Our aim is to build a virtual shopping assistant using image recognition technologies to improve the overall online shopping experience for the users. The proposed application has two main components image recognition service and a product recommendation system. Each component is an important part of this application.

©IJRASET: All Rights are Reserved

B. Project Objectives

Recommender systems in ecommerce website has the ability to predict whether a particular user would prefer an item or not based on the user's profile history. These systems make use of information-filtering-techniques which helps to analyse given information and provide the user with more accurate/relevant products. So the main objective of recommender systems is to provide recommendations based on recorded information on the users' preferences.

It allows you to find similar products in your store in an easy way. Let the customer find the product based on the photo. Because of this, we no need to manually match similar products, it will be done automatically by AI algorithm.

IV. SYSTEM REQUIREMENT

A. Hardware Requirements

1) *Software Application Processor:* Intel ® Core TM i7- 10510U.
2) *Hard Disk:* 20 GB and above.
3) *RAM:* 4 GB.

B. Software Requirements

1) *Operating system:* Windows 7 and above
2) *Language:* PYTHON
3) *Framework:* Flask

V. SYSTEM ARCHITECTURE

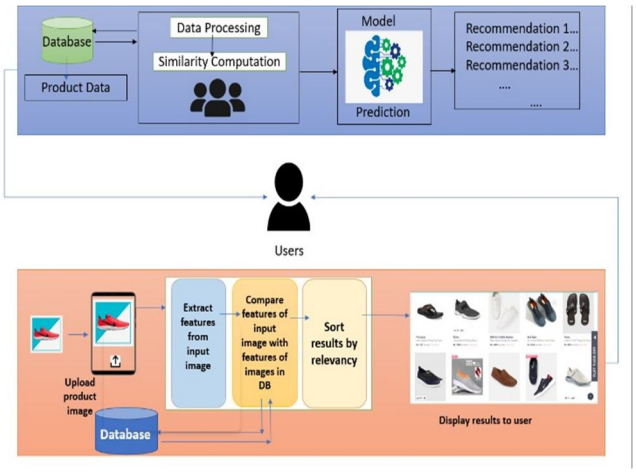


Fig1. System Architecture

VI. ARCHITECTURE SUMMARY

There are two modules in our project Recommendation System and Reverse Image Search.

A. Recommendation System

In first module Recommendation System i.e. product-details including category, name, color, type, brand, image, etc is stored in the database. Depending upon the inputs from the customer i.e. customer clicks on any product/view the product, the model will be able to compute the similarity and predict the similar products based on some parameters and then the model will display the recommended products .

B. Reverse Image Search

In this module, customer will upload product image through the system. The model will be able to extract features from input image, compare features of input image with features of images in database and then sort results by relevancy and display the results to the user.

VII. UML DIAGRAMS

A. Data Flow Diagram Level 0

A data-flow diagram (DFD) describes a flow of data in system of Login Management, User Management, Product Recommendation, Image Search. The DFD is also capable to provide outputs and inputs information of each entity and the process itself. It does not have any decision rules, control flow, loops. Specific operations based on the data can be shown by a flowchart.

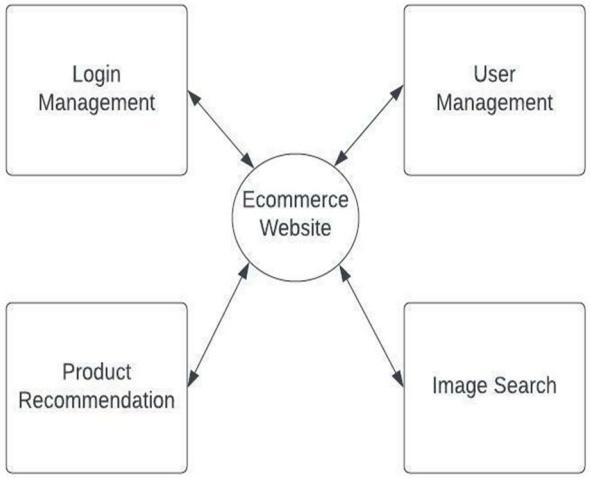


Fig2. Data Flow Diagram level 0

B. Data Flow Diagram Level 1

This is the DFD Level 1 diagram describes DFD Level 0 diagram in our system in depth. Data needs to be classified into training & testing, for training, the data should be pre-processed and after these two processes are done new data is classified by using CNN.

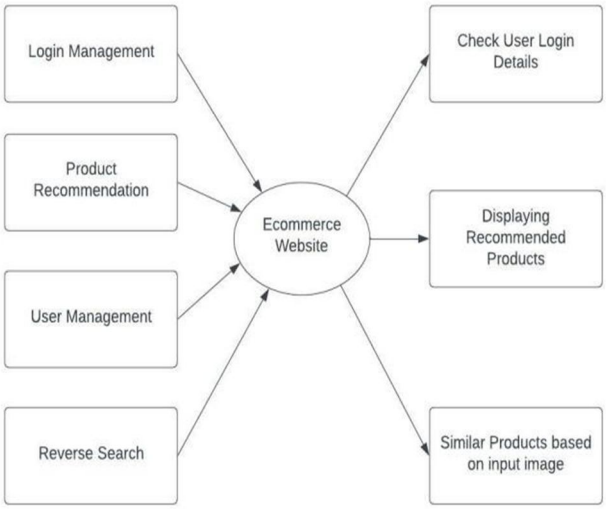


Fig3. Data Flow Diagram 1

C. Data Flow Diagram Level 2

This is the DFD Level 2 diagram describes DFD Level 1 diagram in our system in depth. First User will Login into the system with valid credentials. In first module Recommendation System i.e. details about the products including category, name, color, type, brand, image, etc is stored in the database. In this module, customer will upload product image through the system. The model will be able to extract features from input image, compare features of input image with features of images in database and then sort results by relevancy and display the results to the user.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

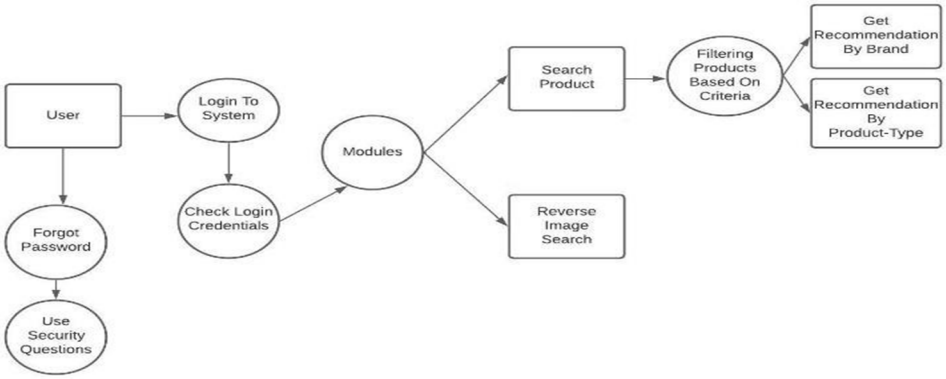


Fig4. Data Flow Diagram 2

VIII. ACTIVITY DIAGRAM

Activity diagram is another useful diagram used to describe the dynamic characteristics of the built system. Basic view of the activities can be described. Basically, it is a flowchart to represent the flow from one activity to other activities. The control flow is drawn from one operation to another to get brief idea about the flow of our system.

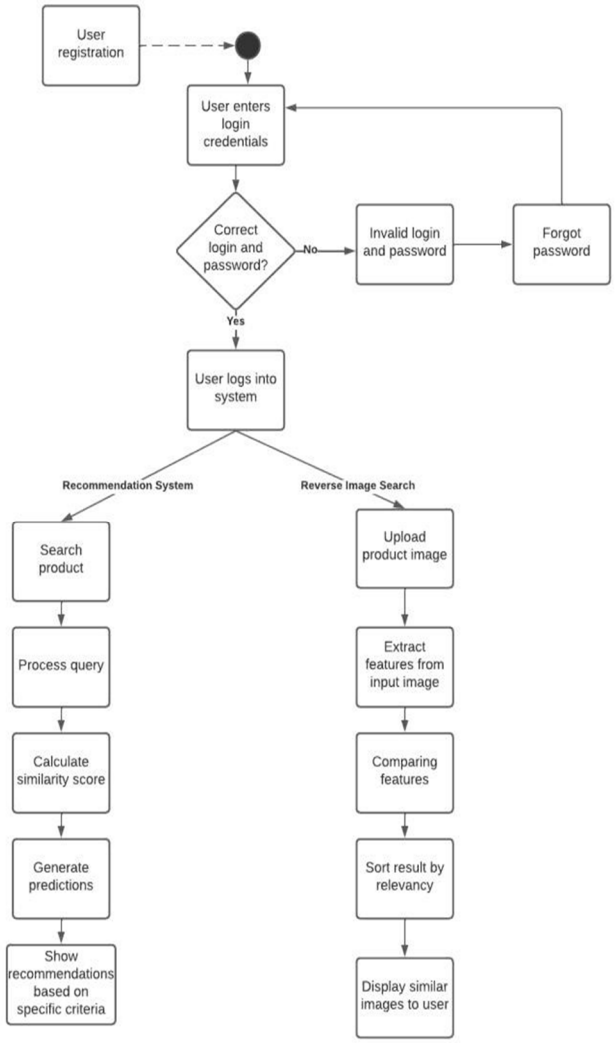


Fig 5.Activity Diagram

IX. USE CASE DIAGRAM

A use case diagram is another important diagram in UML. It basically defines the interactions between an actor and a system. The actor can be a human or any other external systems. In system engineering, use cases are considered to be important and are used at a higher level than software engineering.

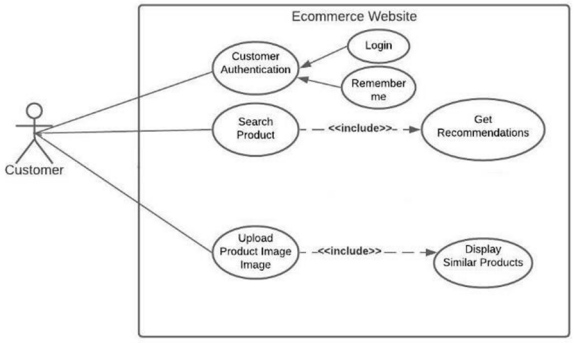


Fig6. Use Case Diagram

X. IMPLEMENTATION

A. Dataset

1) Dataset: Marketing_sample_for_myntra_ecommerce.csv

2) Rows : 15,000

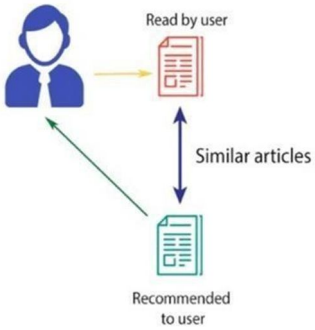
3) Columns : 24

Uniqid, 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.

B. Proposed Algorithm

C. Recommender System

D. Techniques Used In Recommendation System



1) Content Based

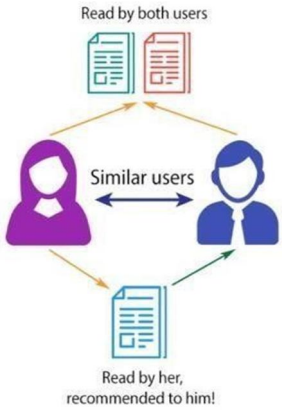
In Content based filtering, we filter data based on actual content, The system will recommend same types of articles of similar content which the user is reading. There are various fundamentals attributes that are used to calculate the similarity while checking about similar content.

a) Merits

- It can work well even with less user's data.
- It needs item data rather than user's data to start giving recommendations to users based on their interest.

b) Demerits

- Good volume of item's data is required.
- Sufficient features must be available to measure the similarity.



2) Collaborative Based

It is considered to be one of the smart recommender systems that measures the similarity between various users and items, that are widely used in an e-commerce website and also online movie websites. It has ability to analyze the taste of similar users and does recommendations based on it. Given figure shows the two different users and their interests along with the similarity between the taste of both the users. It is found that both users have similar tastes so one's interest is recommended to other and vice versa. This is the way collaborative filtering works.

E. Limitations

1) Enough users data is required to find the similarities between the users. To overcome such problems, often hybrid approaches are used.

2) Even if there are multiple items and users that are to be recommended often, there may arise problems of users and rating matrix to be sparse which becomes challenging question to filter out the users who have rated similarly to the same item.

3) The issues may occur in recommending items to the user due to sparsity issues.

F. Understanding Relationships

Product-product relationship: Product-product relationship may occur when products are similar by appearance, price, properties and description. Some examples include a product having same brand, color, type, price, structure, attributes, etc. 2. User-user relationship:

User-user relationships may occur when customers have similar liking for a specific product or service. Some examples are customers having mutual friends, similar backgrounds, similar age, etc.

G. Data & Recommender Systems

In addition to relationships, recommender systems has the following types of data:

1) *Product attribute data:* Product attribute data can be defined as the information related to the product itself such as type, color, brand in case of products, genre and cast in case of movies, cuisine in case of food.

2) *User behavior data:* Users behavior data can be defined as useful information about the liking of the user on particular product. This type of data can be gathered from, clicks, view history & ratings.

H. Recommendation Through Description Of The Content

This approach makes use of the description provided for specific item to make recommendations. As description will go deeper into the product details, like title, summary, tag lines, genre, etc., it will provide better recommendation. The format of these details are in text format (mainly string) and it's necessary to convert this text into numeric value to easily compute the similarity.

1) *Term Frequency-Inverse Document Frequency (TF-IDF):* TF-IDF is a concept used to retrieve information from data. It can be used for feature extraction purposes and it is a sub-class of natural language processing (NLP).
2) *Term Frequency:* Term Frequency is defined as the occurrence of specific word in the current document upon the total number of that word in the document. It specifies the occurrence of the term in a document and gives lesser weight when the frequency is less and vice versa, hence it is divided by the length of document to normalize.

$$\mathrm{Tf}(t) = \frac{\text{Frequency occurence of term t in document}}{\text{Total number of terms in document}}$$

3) *Inverse Document Frequency:* Inverse document frequency is defined as the total number of documents upon frequency of documents containing that word. It specifies the rarity of the word. The more the word occurs in the document, the IDF value decreases. It helps in giving a higher score to rare terms in the documents.

$$\mathrm{Idf}(t) = log_{10}(\frac{\text{Total Number of documents}}{\text{Number of documents containing term t}})$$

4) *Chatbot*

What is Natural Language Processing?

Natural language processing, or NLP is a subfield of artificial intelligence research that is focused on developing models and points of interaction between humans and computers based on natural language. This includes text and speech-based systems systems Algorithms:

a) Lemmatization and stemming
b) Word clouds
c) Keywords extraction
d) Named Entity Recognition
e) Topic Modelling
f) Knowledge graphs

NLP Algorithm – Keyword Extraction

- Keywords extraction is one of the most important tasks of the natural language processing field and is responsible with finding ways for extracting the most important words and phrases from a given text or a collection of texts. This is done in order to help us summarize, organize, store, search and retrieve content in a meaningful and efficient way. • We already have a large number of keywords extraction algorithms available and each applies a different set of principal and theoretical approaches to this problem. We have algorithms that extract only words and algorithms that extract words and phrases. We have algorithms that focus only on one text and algorithms that extract keywords based on a whole collection of texts.
- The most popular keywords extraction algorithms out there are:
- Text rank: works on the same principle behind the pagerank algorithms by which google assigns importance to different web pages on the internet
- TF-IDF: term frequency – inverse document frequency aims to better define how important a word is for a document, while also taking into account the relation to other documents from the same corpus.
- Rake: rapid automatic keywords extraction falls into the category of algorithms that can extract keywords and key phrases based only on the text of one document, without the need to consider other documents in the same collection.

I. Reverse Image Search

1) *Introduction To Convolutional Neural Network (Cnn)*

a) The convolutional neural network (CNN) is a part of deep learning neural networks.
b) CNN is an algorithm mostly used in image recognition.
c) They're usually used to examine visual imagery and are generally used in in image classification.

d) CNN are very good at extracting features from the input image, such as lines, gradients, colors, circles, along with eyes and faces.

e) This provides an unique ability to machines to view the world as humans would.It can even use the knowledge for image and video recognition, image classification and image inspection, recommendation systems, NLP, etc.

2) Layers in CNN

There are four layers in convolutional neural networks:

a) *Convolution Of An Image:* Each convolution filter defines a characteristics of image i.e. pixels and the Convolutional Neural Network algorithm studies which features includes the resulting reference i.e. input image.

We have 4 steps for convolution:

- Assigning the feature with the image.
- Multiply each image pixel to the corresponding feature pixel.
- Add those values and find the total sum of the matrix.
- Divide that total sum by the total number of pixels in that particular feature. Consider the below image

At this point, we are done with the first 2 steps i.e. assigning & multiplying. We have taken an input feature image and considered one pixel from it. We multiplied this with the corresponding pixel of existing image and this product is stored at corresponding pixel in another buffer feature image.

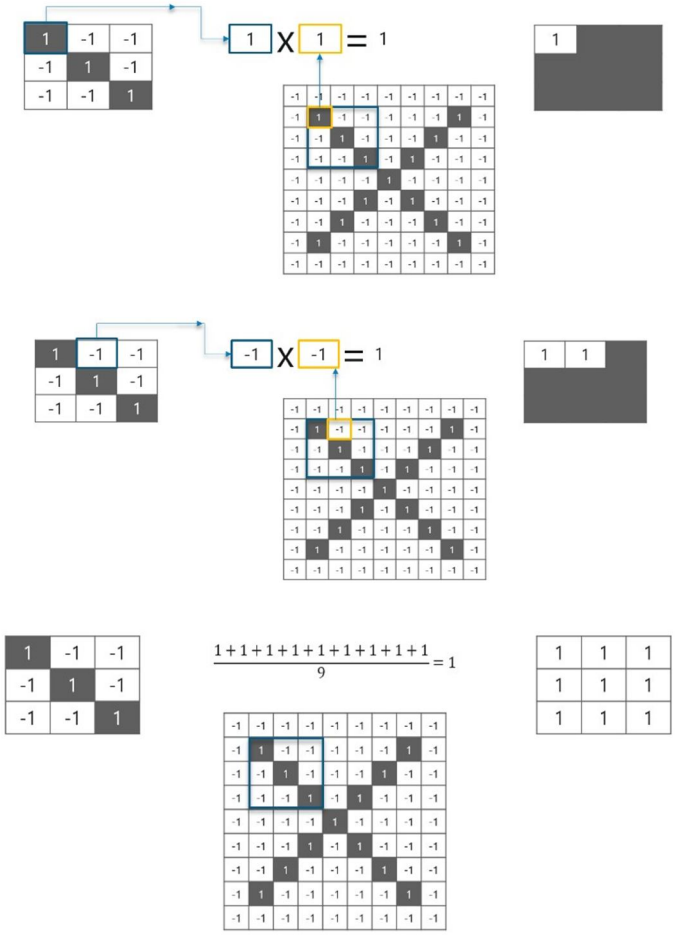


Fig.10 Convolution

With this image, we completed the last 2 steps i.e added the values which gave the sum and divide this number by the total number of pixels in the given input feature image. After this, the final value which is obtained is placed at the middle pixel of the filtered image matrix as shown below:

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

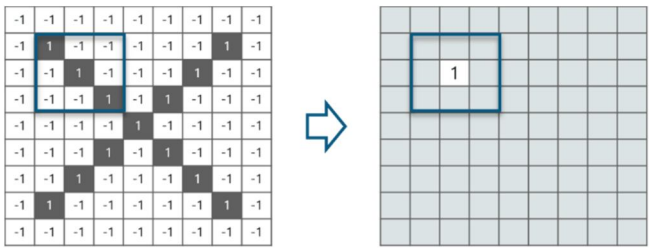


Fig 10.1

Now, we keep on considering different patches of feature image and move this filter around and follow the same steps at any pixel in the image. For brief understanding, let's consider an example:

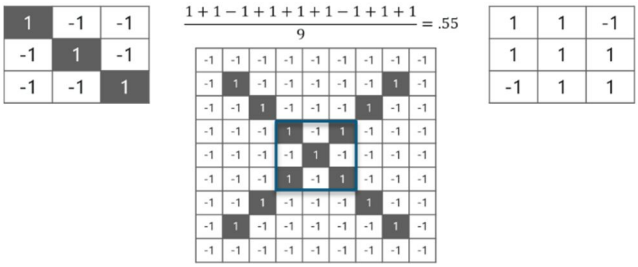


Fig 10.2

After performing the first 4 steps specified in the algorithm, we have the value as 0.55. We place this value in the image as explained before. This is done in the following image:

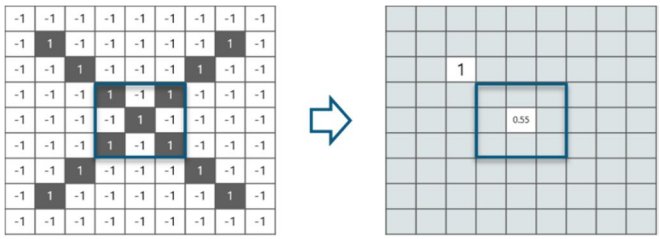


Fig 10.3

Similarly, we go through all patches and move the feature to every other position in the image and see how the feature is similar to that area. So after doing this, we will get the output as:

0.77	-0.11	0.11	0.33	0.55	-0.11	0.33
-0.11	1.0	-0.11	0.33	-0.11	0.11	-0.11
0.11	-0.11	1.0	-0.33	0.11	-0.11	0.55
0.33	0.33	-0.33	0.55	-0.33	0.33	0.33
0.55	-0.11	0.11	-0.33	1.00	-0.11	0.11
-0.11	0.11	-0.11	0.33	-0.11	1.00	-0.11
0.33	-0.11	0.55	0.33	0.11	-0.11	0.77

Fig 10.4

©IJRASET: All Rights are Reserved

Here we considered only one filter. Similarly, we will do the same convolution with each and every filter to get the convolution of all the filters present in the input feature image.

b) *ReLU Layer:* ReLU is an activation function which is used as transformation function. Rectified Linear Unit (ReLU) activates (assigns a value) a node if the input is greater than a specific value, while the input is below zero, the output is zero, but when the input rises above a specific value, it will have a linear relationship with the dependent variable and a certain value is assigned to it.

Consider the below example:

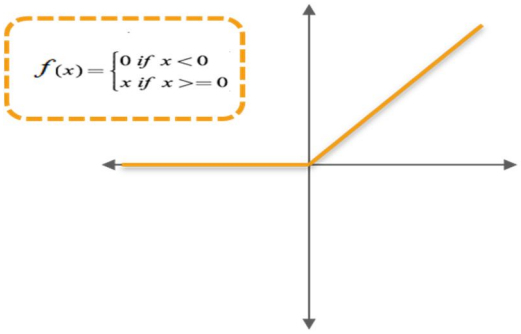


Fig 11.1 ReLU

We have taken a simple function with the values as mentioned above. So this activation function only computes certain operation if that value is obtained by the dependent variable i.e it has certain value greater than zero. For this given example, these values are obtained:

x	f(x)=x	F(x)
-3	f(-3) = 0	0
-5	f(-5) = 0	0
3	f(3) = 3	3
5	f(5) = 5	5

Why do we require ReLU here?

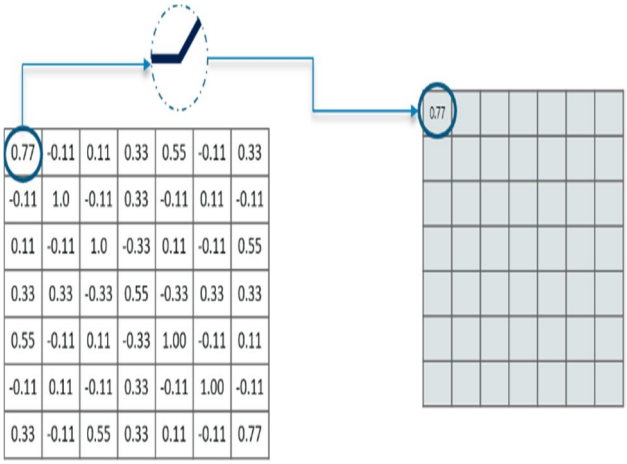
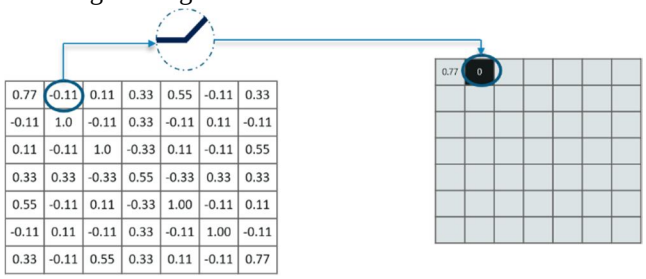


Fig 11.2

The main function is to remove all the negative values from the convolution. All the positive values remain the same as they are greater than zero but all the negative values get changed as shown below:



After this, we process this specific feature and get the following output:



Fig 11.3

Now, similarly we follow the same process to all the other feature images in the input image as well:

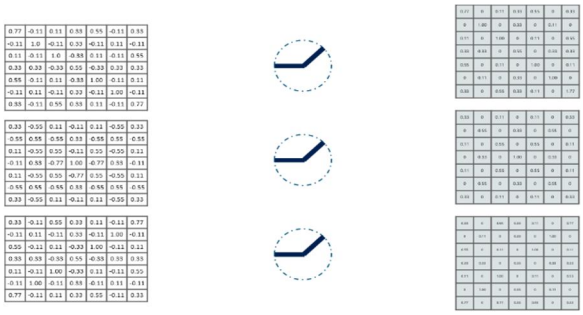


Fig 11.4

c) Pooling Layer

- Function of this layer is to shrink the image stack into smaller size than first. Pooling is performed after complete passing through the activation or transformation layer. We can do this operation by implementing the given following four steps:
- Pick the window size (usually 2 or 3)
- Pick the stride (usually 2)
- Walk your input window through the filtered image patches that you have captured.
- From each filtered window, consider the as maximum value from it.
- Consider this example. In this example we are performing pooling with a window size of 2 and stride of 2.

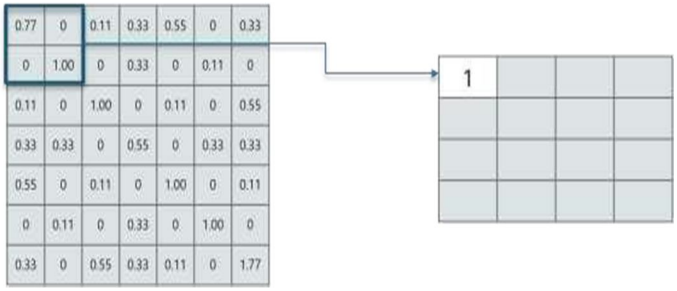


Fig 12.1 Pooling

So in this given example, we considered window size to be 2 and we had options as 4 values to choose from. From those 4 values, the maximum value in this example is 1 so we selected 1. We performed out with a 7×7 matrix but now the same matrix after pooling operation reduced to 4×4 matrix.

But we need to move the small patches window across the complete image. The similar procedure is carried out for remaining windows until the entire image is covered.

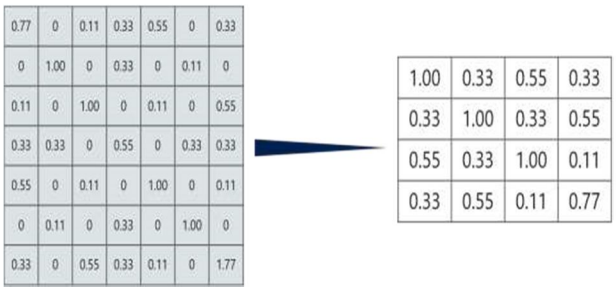


Fig 12.2

Fig 12.5

The last layer in the network is fully Till now we have carried out the operations for connected layer, which means that only one filter. We have to do the same for 2 other neurons of previous layers are connected filters as well. After doing this we get the to every neuron in consecutive layers. following result: This tells high level reasoning where all possible ways from the input to output are considered.

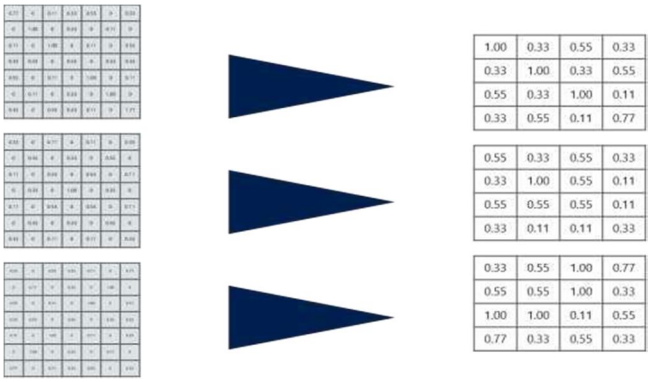


Fig 12.3

Next step is to stack up all these layers.

d) Stacking Up The Layers

Here we have 4×4 matrix from a 7×7 matrix after passing the input through 3 layers i.e. Convolution, ReLU and Pooling as shown below:

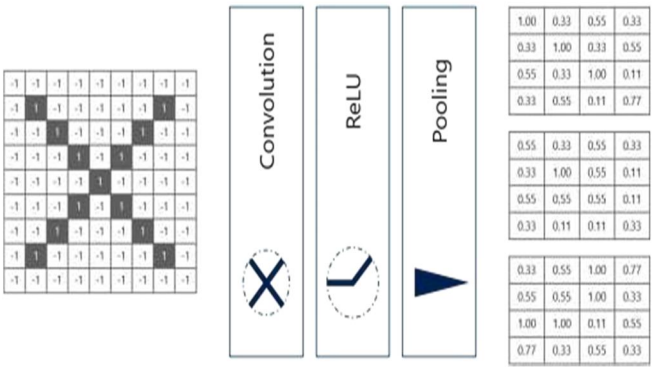


Fig 12.4

Also, fully connected layer is the last layer where the actual classification is performed.
Here we consider our filtered and compressed images into one single list. For example, list is shown below:

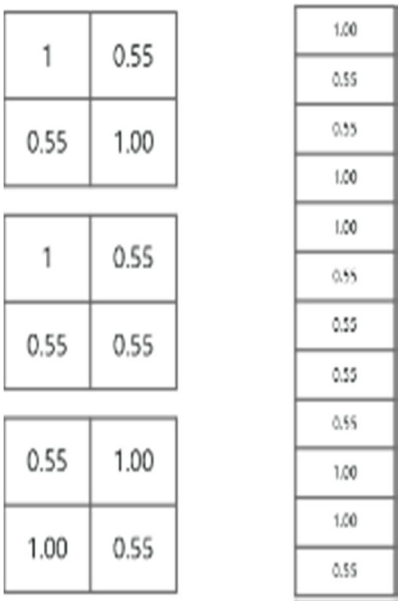


Fig 12.5

So next, when we give input in, 'X' or 'O' there will be an element in the vector that will be high. In the image below, as you can see for 'X' there are different elements that are high and similarly, for 'O' we have different elements that are high:
We have to perform these 3 operations in an iteration after the first pass. So after the second pass we get at a 2×2 matrix as shown below in the figure:

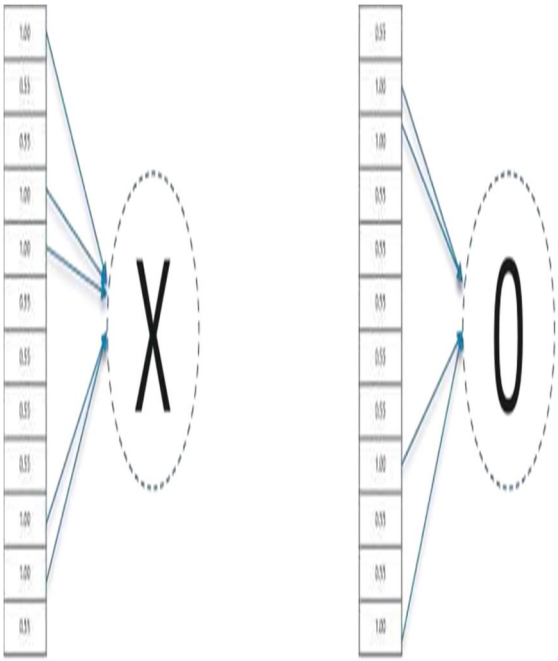


Fig 12.6

From this we can conclude that, when the first, fourth, fifth, tenth and eleventh values are high or more, we can say that the input image is 'x'. The operation is similar for the other alphabets as well. When specific values are specified the way they are, they can be assigned to an actual or original letter or a numeric value which we expect as an output.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

e) Prediction Of Image Using Convolutional Neural Networks – Fully Connected Layer

- At this stage, we have trained the network and we can start to predict and examine the working of the classifier. Let's take an example:

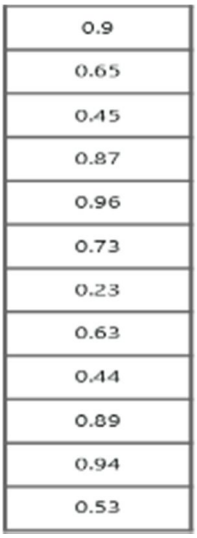


Fig 13.1

- In the above image, we have a vector with 12 values which is obtained after passing the input image of a random letter through all the layers of our network which was specified earlier while understanding with the operations.
- We will get predictions based on the output data by comparing the obtained values with list of 'x' and 'o' list which is computed earlier.

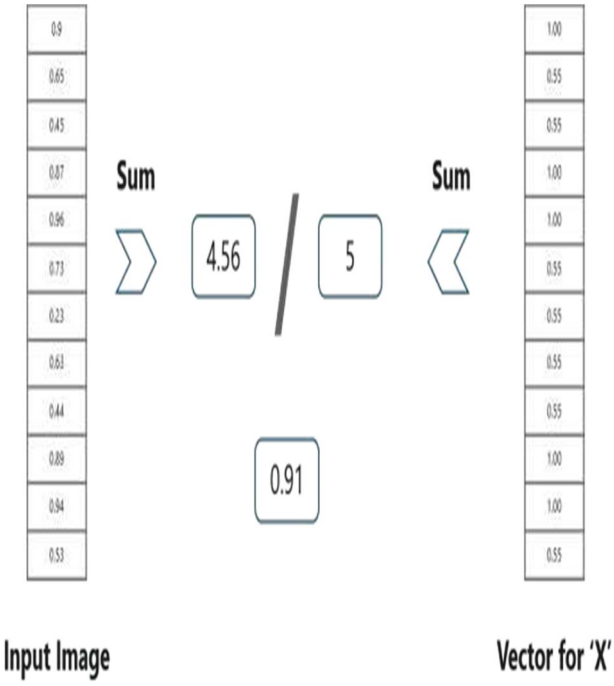


Fig 13.2

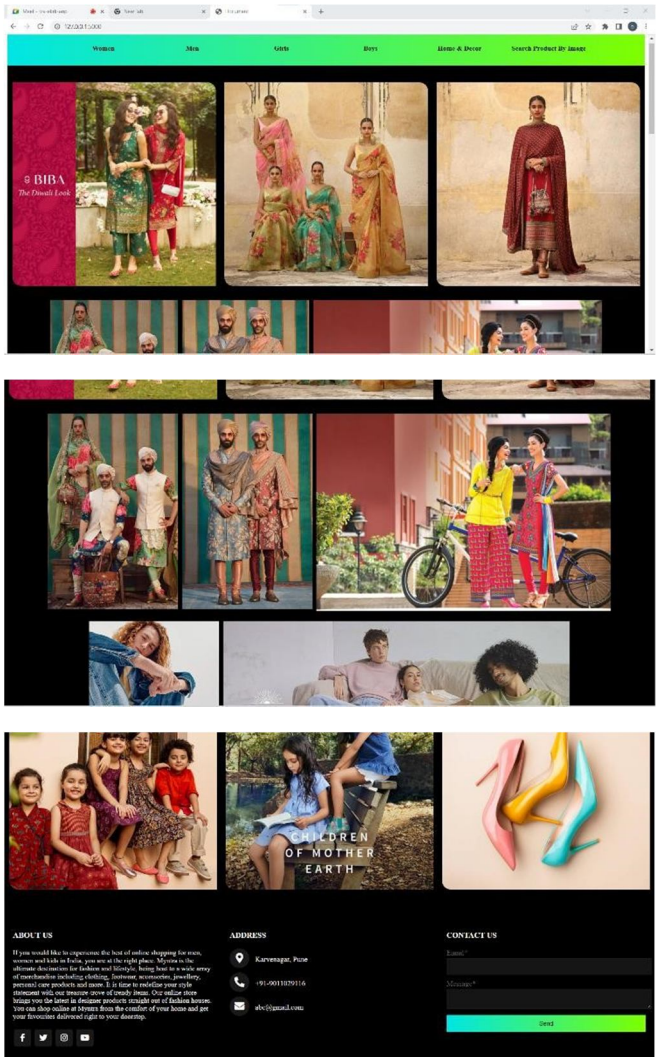
- We added the values which was found out as high (first, fourth, fifth, tenth and eleventh) from the vector table of X and we got the sum to be 5. We did the exact same process with the input image and got a value of 4.56.
- When we divide this value we will get a probability match as 0.91 Now same is done with vector table 'o'.

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

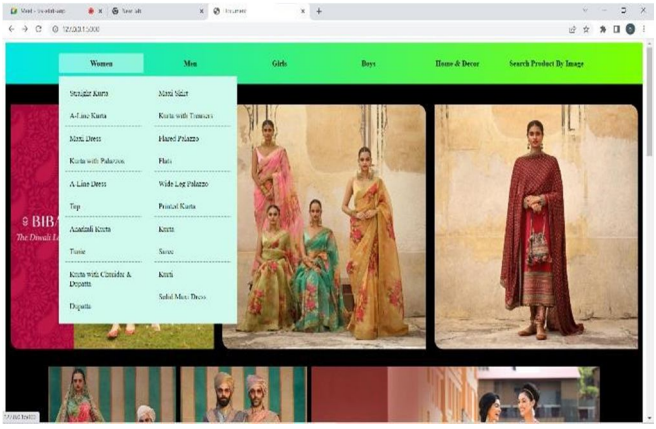
XI. RESULT AND EVALUATION

A. *Home page*



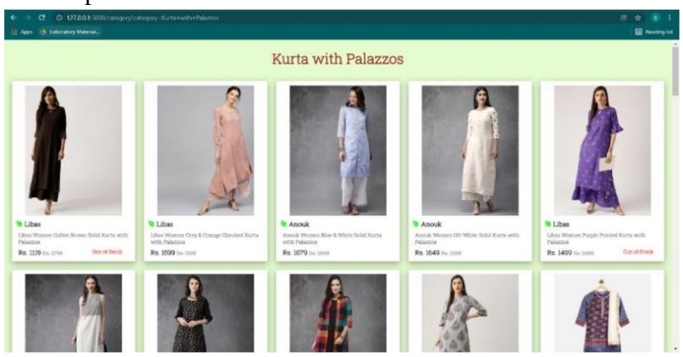
B. *Recommendation System*

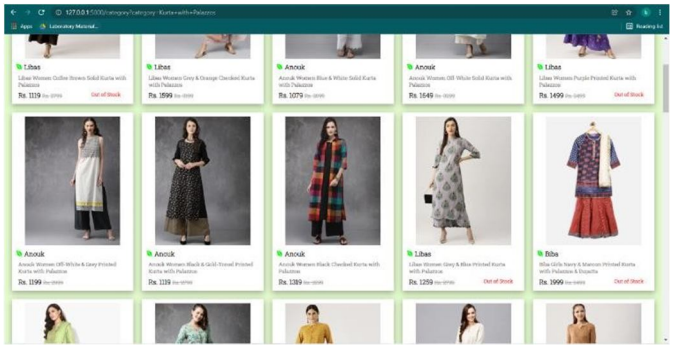
1) *Step 1:* Click on dropdown to view more sub categories



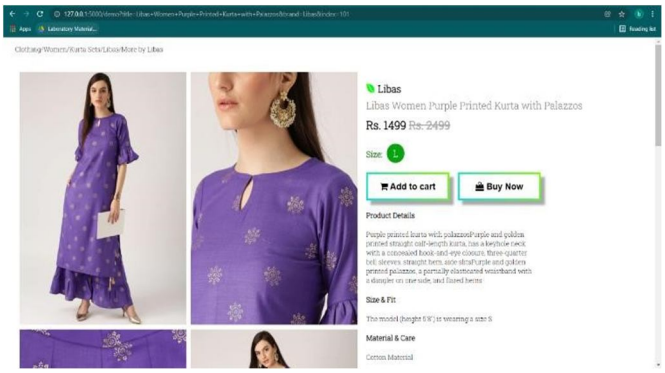
©IJRASET: All Rights are Reserved

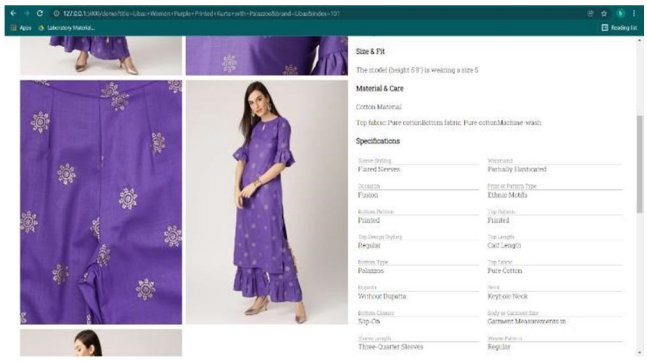
2) Step 2 : Clicked on Libas Women Purple Printed Kurta with Palazzos



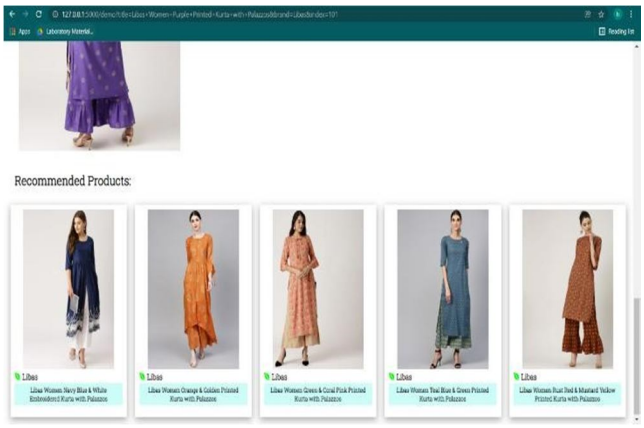


3) Step *3:* View product in detail

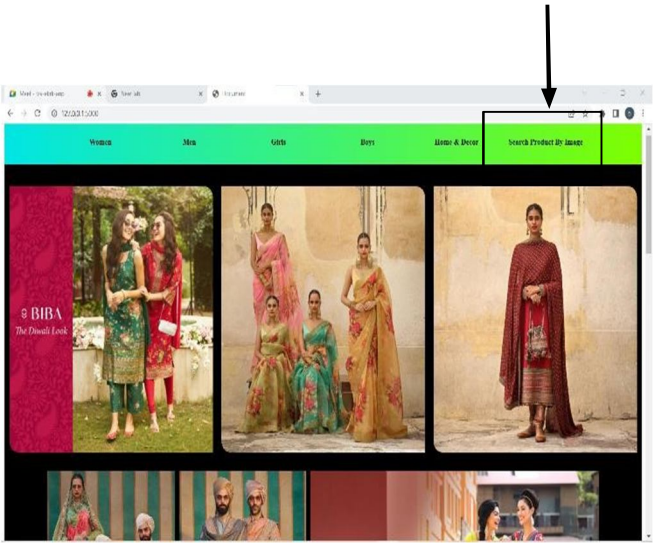




4) *Step 4:* recommended products

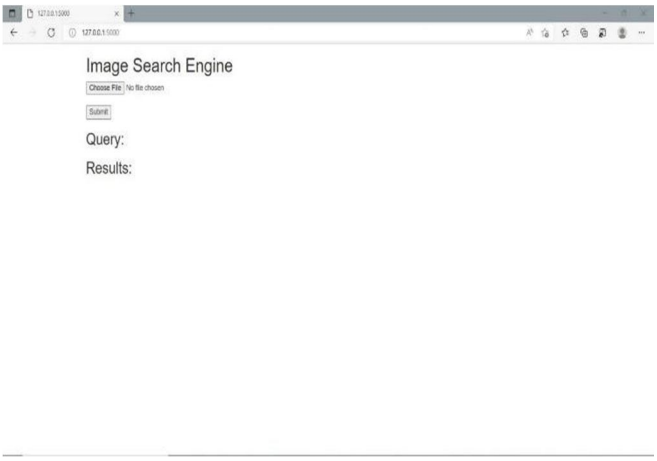


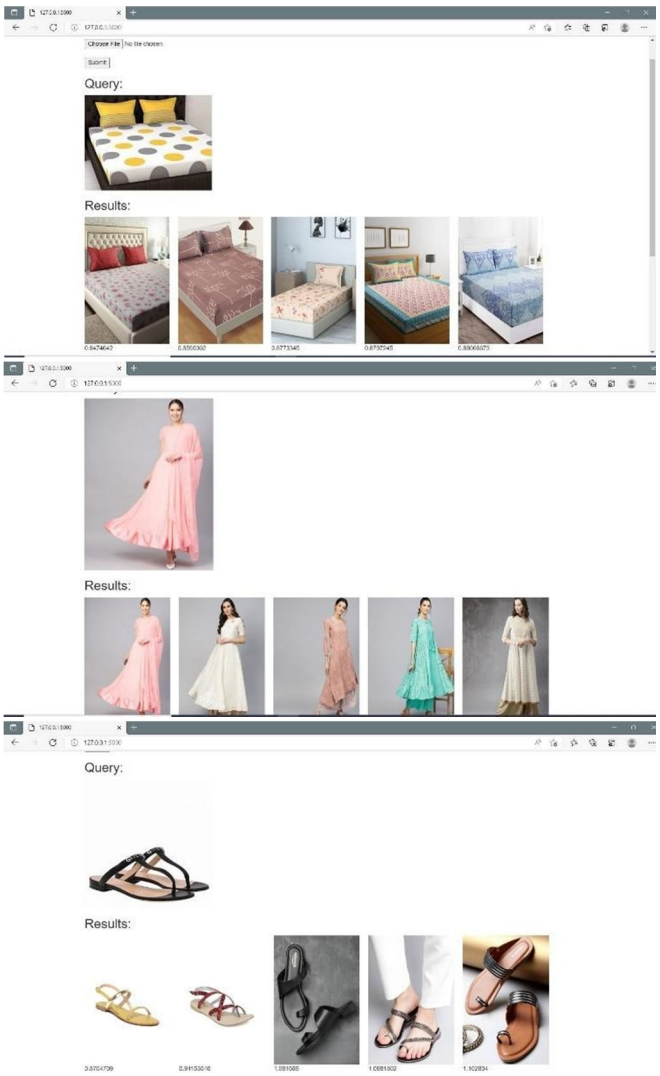
C. Reverse Image Search



D. Home Page

1) *Step 1:* Select a photo from your device.
2) *Step 2:* Click on submit button.





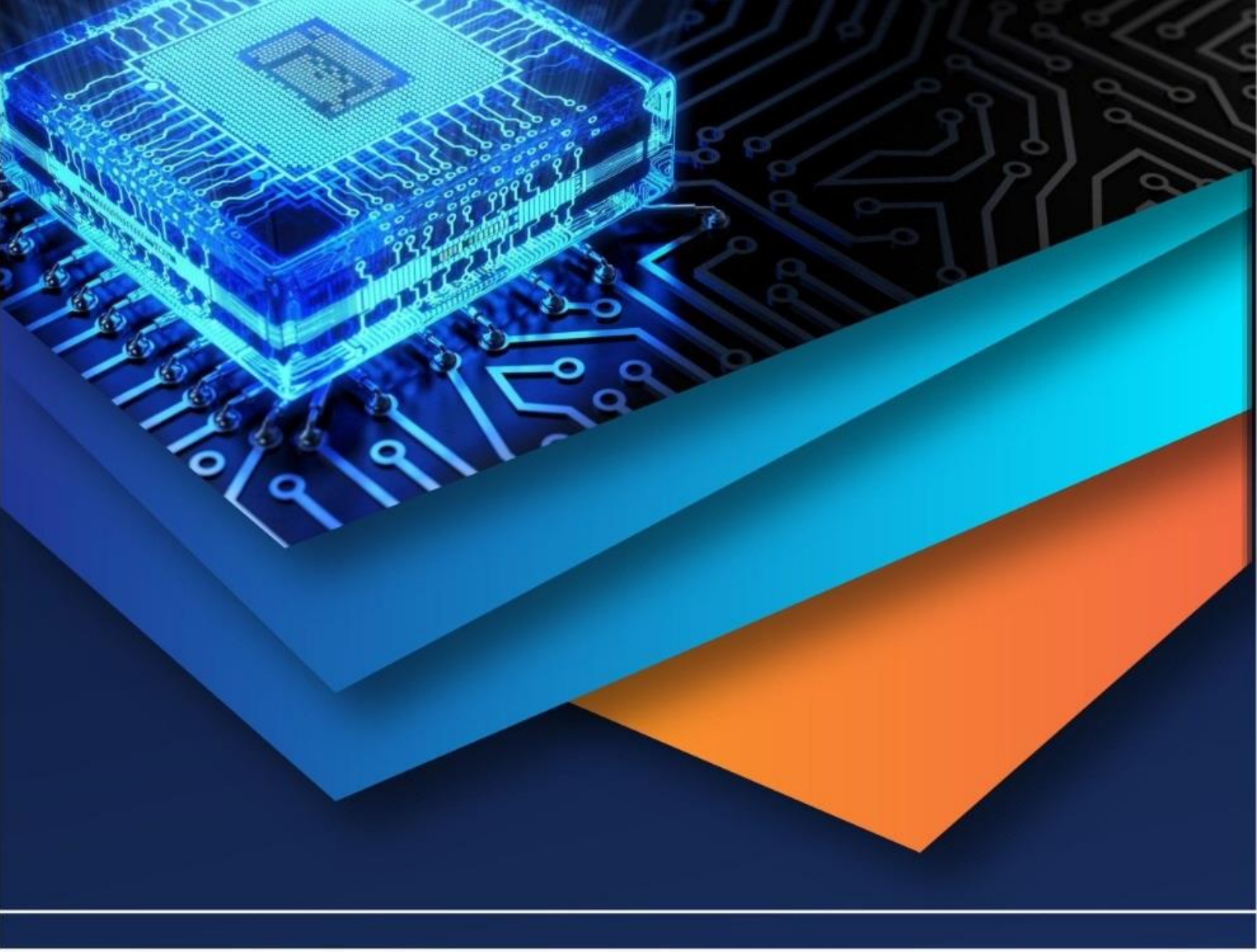
XII. CONCLUSION AND FUTURE WORK

The proposed application will have two main components, i.e Recommendation System and Reverse Image Search service. Each of these components is an important part of this application. However, image recognition service is the core functionality of the application which enables users to extract meaningful information from the images which are hard to describe in words, for example, categories, brand name, colour features of a product.

In future we will try to enhance the accuracy of all two functionalities to make the user interface more reliable.

REFERENCES

[1] Shakila Shaikh; Sheetal Rathi; Prachi Janrao , "Recommendation system in E-commerce websites: A Graph Based Approach" IEEE - 2017 IEEE 7th International Advance Computing Conference
[2] Mohsen Jamali, Martin Ester , "A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks" ACM
[3] Neha Sharma, Vibhor Jain, Anju Mishra "An Analysis Of Convolutional Neural Networks For Image Classification" IEEE
[4] Deepika Jaiswal, Soman Kp, Sowmya Vishvanathan "Image Classification Using Convolutional Neural Networks" ACM.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)