



# Image Classification Model

Submitted by:

Aniruddha Sawant

## **ACKNOWLEDGMENT**

Aniruddha Sawant ("I") acknowledge that I have collected data from 'amazon.in' and used the data and files provided by Flip Robo ("Company") namely 'Problem Statement' to build the predictive model and have used 'sample documentation' for guidance and to write report.

# INTRODUCTION

- **Background of the Domain Problem: -**

Images are one of the major sources of data in the field of data science and AI. This field is making appropriate use of information that can be gathered through images by examining its features and details.

- **Business Problem: -**

The idea behind this project is to build a deep learning-based Image Classification model on images that will be scraped from e-commerce portal. This is done to make the model more and more robust.

- **Review of Literature: -**

There are total three categories of image data i.e Saree, Men Trouser and Men Jeans present in the data that would be used to predict the image category. But before using those images to predict the outcome we have collected the data from Amazon.in, cleaned the data, transform the data into structured format. This process ends up with a smooth flow of the data to predict the image category and build the model.

Have used Coding a Convolutional Neural Network (CNN) Using Keras Sequential API to train the data, and built convolutional, pooling and dense layers.

- **Motivation for the Problem Undertaken: -**

We are required to model Image of products with the available image data. This model will then be used to predict the image categories.

# Analytical Problem Framing

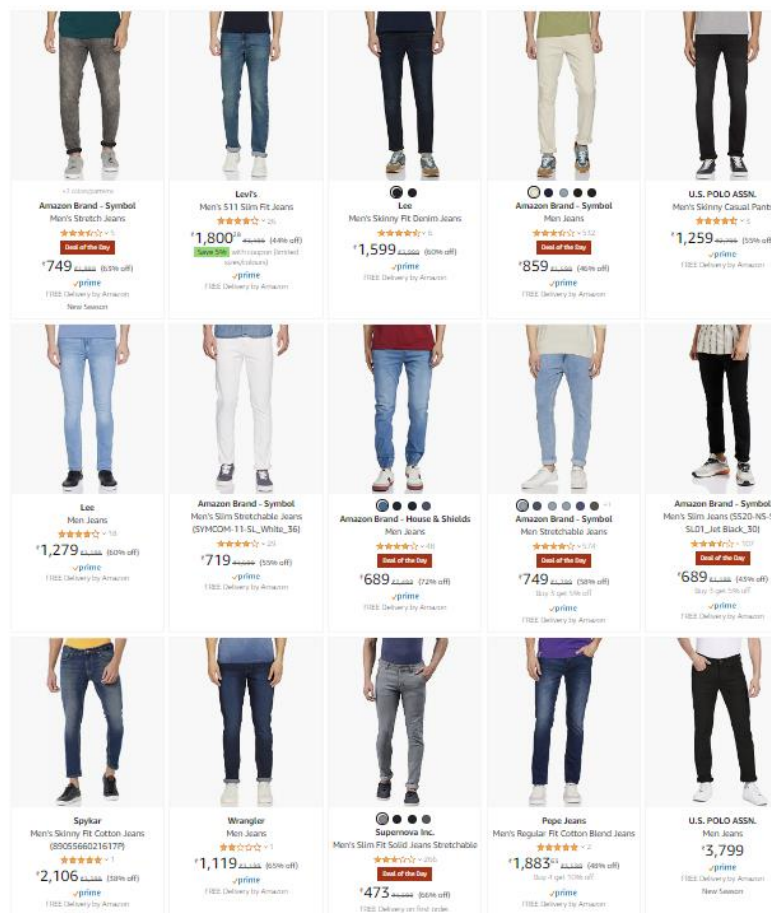
- **Mathematical/ Analytical Modelling of the Problem**

1. Identifying sources of the data
2. Data collection
3. Data structuring
4. Analysing the data
5. Cleaning and processing the data
6. Adding convolutional, pooling and dense layers
7. Evaluating model
8. Predicting outcome for test data

- **Data Sources and their formats**

I have used Amazon.in website to collect the data to build the model. I have used selenium to crawl the data. I have created three folders for each category namely Saree, Trouser and Jeans.

1. Data source sample: - ([Sample link](#))



## 2. Dataset sample: -

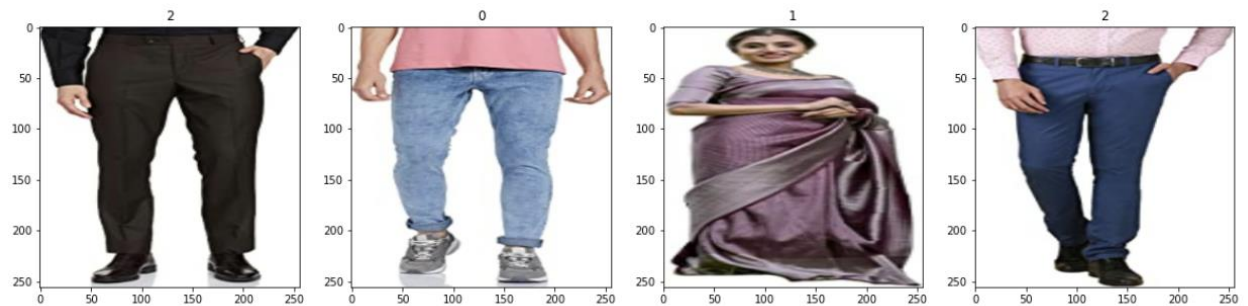
Below dataset 3 image categories and it was being cleaned using Python.

Samples: -

“Men Jeans”: 0

“Saree”: 1

“Trouser”: 2



## • Data Pre-processing

1. Clean and organized the data.
2. Bifurcate images with filter of 'jpeg', 'jpg', 'bmp', 'png'
3. Classified the data into several batches
4. Divided Numpy array with 255 to make the array range between 0 to 1
5. Divided data into three parts i.e training, validation and testing data

- **Data Inputs**

1. Saree – Images of Sarees downloaded from amazon.in
2. Trouser – Images of Trouser downloaded from amazon.in
3. Jeans – Images of Jeans downloaded from amazon.in

- **Hardware and Software Requirements and Tools Used**

1. Libraries and packages used to Scrap the data
  - import selenium
  - from selenium import webdriver
  - from selenium.webdriver.common.by import By
  - import requests
  - import time
  - import warnings, warnings.filterwarnings('ignore') – To ignore unwanted Warnings
2. Libraries and packages used in model building file
  - import numpy as np
  - import pandas as pd
  - import seaborn as sns
  - from matplotlib import pyplot as plt
  - import os
  - from IPython.display import Image
  - import tensorflow as tf
  - import cv2
  - import imghdr
  - from tensorflow.keras.metrics import Precision, Recall
  - from tensorflow.keras.models import load\_model
3. Models used
  - from tensorflow.keras.models import Sequential
  - from tensorflow.keras.layers import Conv2D, MaxPooling2D, Dense, Flatten
4. Hardware used – 11th Gen Intel(R) Core (TM) i3-1115G4 @ 3.00GHz 3.00 GHz with 8.00 GB RAM and Windows 11
5. Software used – Anaconda and Jupyter Notebook to build the model, clean and structured the data.

# Model/s Development and Evaluation

- **Identification of possible problem**

1. The data was not structured and organized hence cleaned the data using various data cleaning and pre-processing techniques.
2. There are so many images which are wrongly downloaded which are required to remove from database.
3. Accuracy was little low since the trouser and jeans seems same to the model.

- **Testing of Identified Approaches**

CNN with Sequential Model the algorithms which have been used to train and test data.

- **Run and evaluate selected models**

1. **Sequential: -**

```
1 for batch in test.as_numpy_iterator():
2     X, y = batch
3     yhat = model.predict(X)
4     pre.update_state(y, yhat)
5     re.update_state(y, yhat)
```

```
1/1 [=====] - 4s 4s/step
1/1 [=====] - 1s 640ms/step
1/1 [=====] - 0s 471ms/step
```

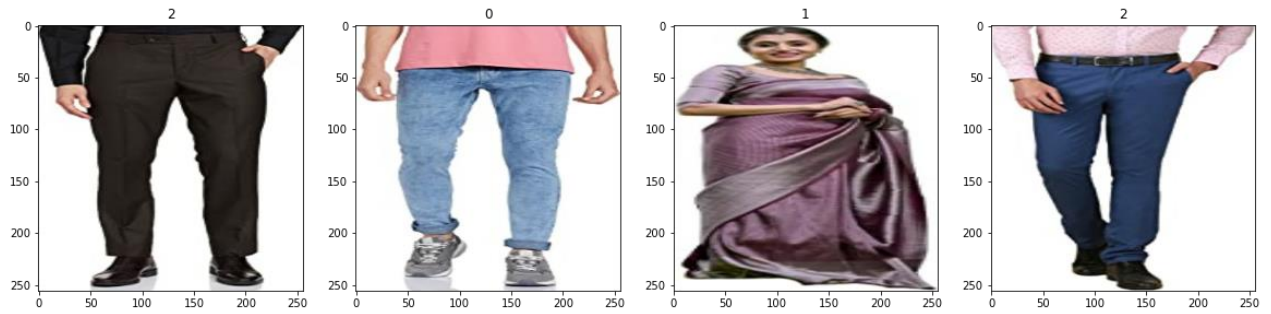
```
1 print(pre.result(), re.result())
```

```
tf.Tensor(0.6455696, shape=(), dtype=float32) tf.Tensor(1.0, shape=(), dtype=float32)
```

***Precision = 64.56 and Recall = 100***

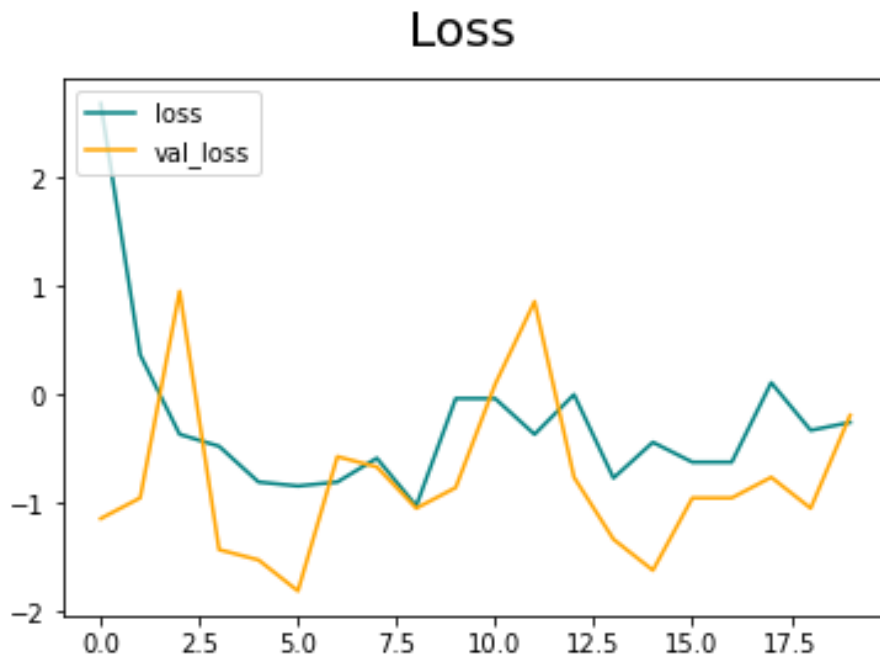
- **Visualizations**

**1. Plotting Numpy arrays as images: -**



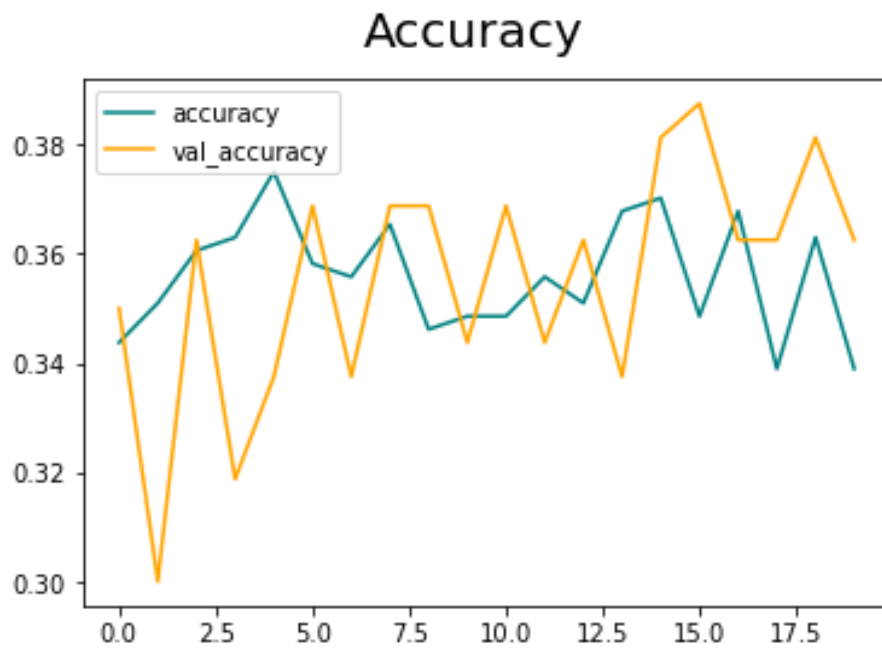
Observations: - Saree Images are represented by number 1, trouser by number 2 and jeans by number 0.

**2. Loss History: -**



Observations: - loss is decreasing at a moderate rate however Val\_loss i.e Validation loss is not decreasing.

### 3. Accuracy History: -



Observations: - Validation accuracy is increasing at moderate rate however accuracy is not increasing.



# CONCLUSION

- **Key Findings and Conclusions of the Study**

1. Have used Sequential Model to predict the image categories.
2. Precision is 65% and Recall is 100% for this model

- **Learning Outcomes of the Study in respect of Data Science**

1. Data processing helps to bifurcate the categories and clean the data which will be used to make findings.
2. Data visualization helps understand and analyse the data.
3. Model building helps to predict outcomes, in this case Sequential model fits perfect for this dataset.

- **Limitations of this work and Scope for Future Work**

1. It is necessary to keep an eye on new and updated data to further train the model and make decision as per new data.
2. To build the deep learning model it is important to use a big dataset however due to hardware limitations it is difficult to handle a big data.