FLIP ROBO

Image Scraping and Classification Project

NAME OF THE PROJECT

Salman Pradhan

Submitted by:

Salman Pradhan

ACKNOWLEDGMENT

I wish to thank Flip Robo Technologies for assigning this project to me and thanks to Mr.Shubham Yadav sir for his valuable technical support on this project.

INTRODUCTION

- Business Problem Framing

First We scrape the images from Amazon.com.The clothing catagories used for scrapping will be

- Sarees (women)
- Trousers (men)
- Jeans (men)

After the data collection and preparation is done, we build an image classification model that will classify between these 3 categories mentioned above

- Conceptual Background of the Domain Problem
  Created model can be used for scrapping and classification of images
- Review of Literature
  1.First we scrap images using selenium and then save it in our local drive.After that we upload the image files in google drive according to their catagories.
  2.Then we use transfer learning and use the weights of vgg16 to classify the model.
- Motivation for the Problem Undertaken
  DataScience help us to make predictions at areas like health sectors,education, media etc. For our project we decided to implement a model which classifies the images using deep learning .

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

  The classification will based on the dataset which is scrapped from Amazon.com.

  Scrapping image files contains three catagories

  1.Sarees (women)

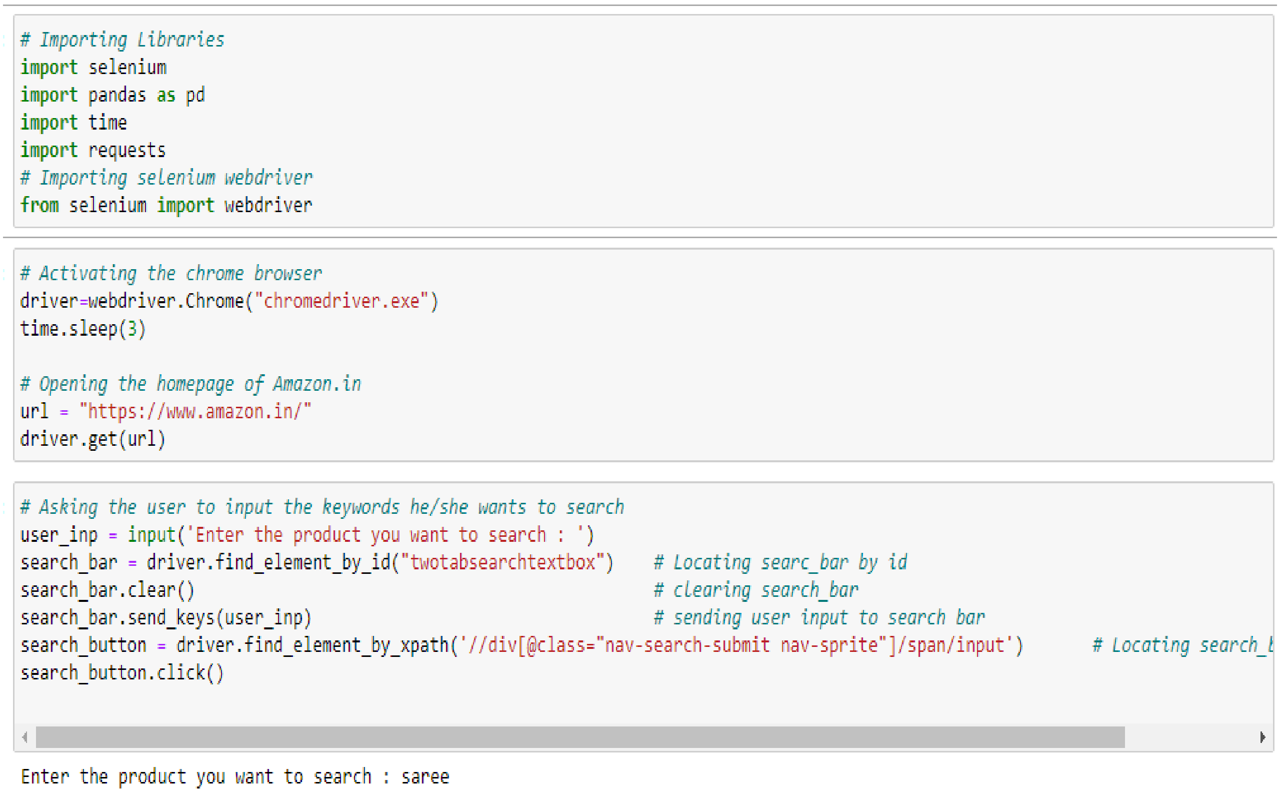
  2.Trousers (men)

  3.Jeans (men)

- Data Sources and their formats

  This project is divided into two phases: Data Collection and Model Building.

  First we import libraries and goes to Amazon.com and scrap images .



```
# Importing Libraries
import selenium
import pandas as pd
import time
import requests
# Importing selenium webdriver
from selenium import webdriver
```

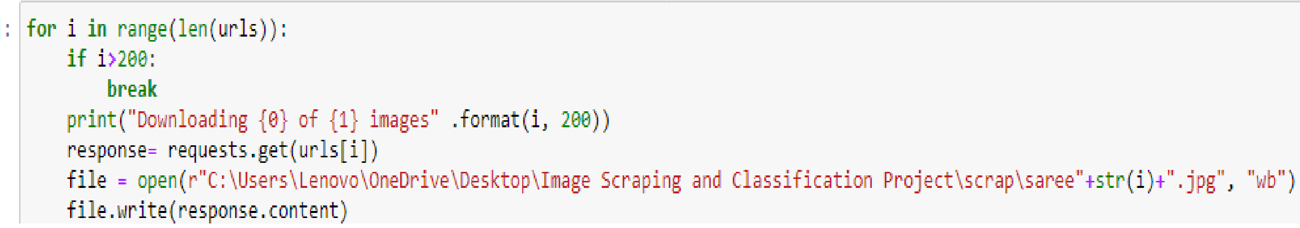
```
# Activating the chrome browser
driver=webdriver.Chrome("chromedriver.exe")
time.sleep(3)

# Opening the homepage of Amazon.in
url = "https://www.amazon.in/"
driver.get(url)
```

```
# Asking the user to input the keywords he/she wants to search
user_inp = input('Enter the product you want to search : ')
search_bar = driver.find_element_by_id("twotabsearchtextbox")    # Locating searc_bar by id
search_bar.clear()                                               # clearing search_bar
search_bar.send_keys(user_inp)                                   # sending user input to search bar
search_button = driver.find_element_by_xpath('//div[@class="nav-search-submit nav-sprite"]/span/input')       # Locating search_b
search_button.click()
```

```
Enter the product you want to search : saree
```

Then we save the images in local directory.

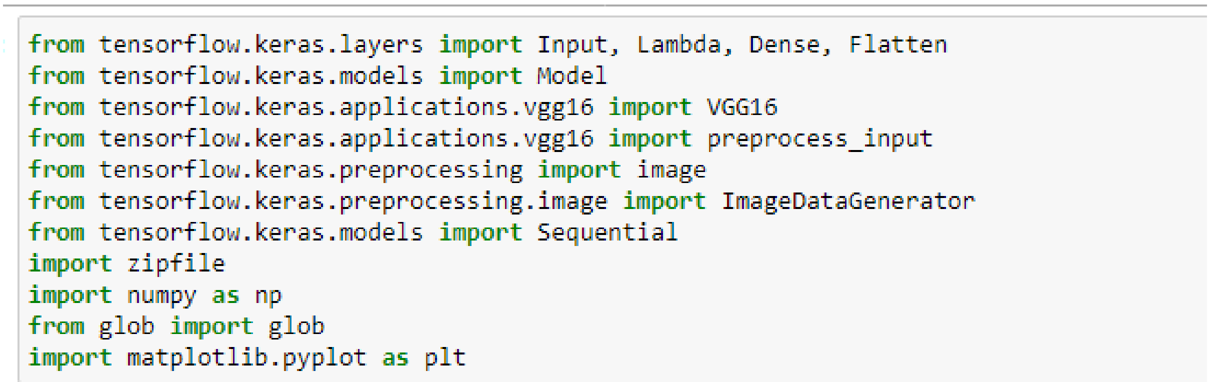


```
for i in range(len(urls)):
    if i>200:
        break
    print("Downloading {0} of {1} images" .format(i, 200))
    response= requests.get(urls[i])
    file = open(r"C:\Users\Lenovo\OneDrive\Desktop\Image Scraping and Classification Project\scrap\saree"+str(i)+".jpg", "wb")
    file.write(response.content)
```

Like this we scrap images of above three catagories.
Then we upload images in google drive by divide them in train and test .We divide images by giving 4 images of each catagories in test folder and other images are in train folder.

- Data Preprocessing Done

After images uploaded to google drive,we make our model by using googlecolab.we import necessary libraries and mount the google drive.



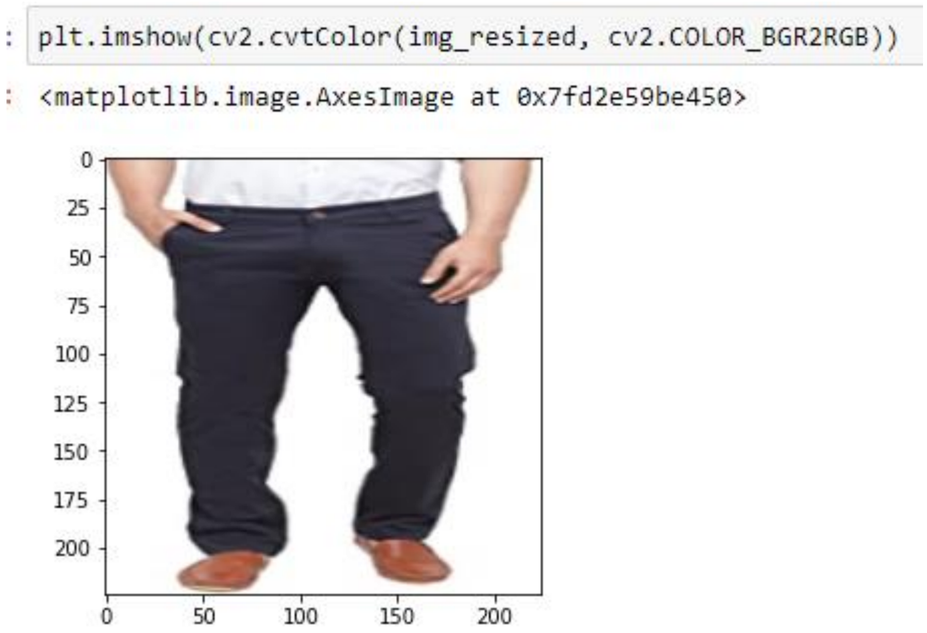
```
from tensorflow.keras.layers import Input, Lambda, Dense, Flatten
from tensorflow.keras.models import Model
from tensorflow.keras.applications.vgg16 import VGG16
from tensorflow.keras.applications.vgg16 import preprocess_input
from tensorflow.keras.preprocessing import image
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.models import Sequential
import zipfile
import numpy as np
from glob import glob
import matplotlib.pyplot as plt
```

- Data Inputs- Logic- Output Relationships

Images are in jpeg format and their sizes range between 5 kb to 10 kb.

- State the set of assumptions (if any) related to the problem under consideration

As we use transfer learning(VGG16) to solve this classificationproject,we have to make our image size (224,224,3).



```
: plt.imshow(cv2.cvtColor(img_resized, cv2.COLOR_BGR2RGB))

: <matplotlib.image.AxesImage at 0x7fd2e59be450>
```

Above picture is the sample picture of training dataset.

- Hardware and Software Requirements and Tools Used

NumPy: Base n-dimensional array package
Matplotlib: Comprehensive 2D/3D plotting
Seaborn: For plotting graph
Pandas: Data structures and analysis
Scikit-learrn: provides a range of algorithm
Selenium: For scrpping data from websites
Keras:For deep learning algorithms
Cv2: Read images

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

1.Understand business problem
2.Scrapping images and save it in google drive
3.Data analysis
4.Visualization with cv2
5.Data preprocessing
6.Feature scaling

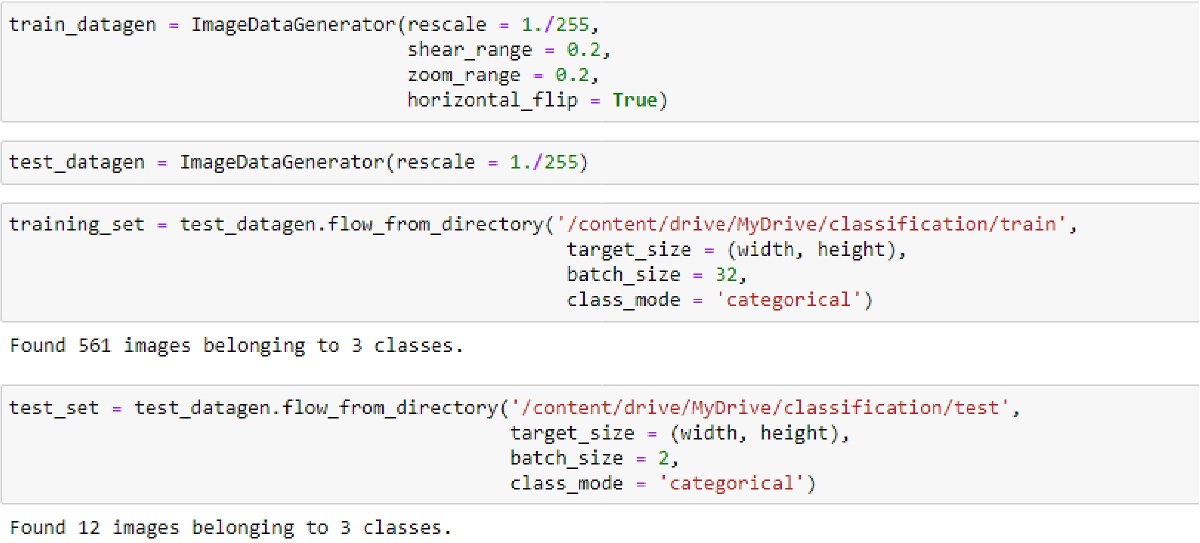
7.Model bulding with transfer learning technique(VGG16)

- ## Testing of Identified Approaches (Algorithms)

  We use weights of Vgg16 model for prediction purpose.

- ## Run and Evaluate selected models

  We divide the dataset in train and test folder like below and scaling down its value betwwn 0 to 1 because our convergence will be faster.



```
train_datagen = ImageDataGenerator(rescale = 1./255,
                                   shear_range = 0.2,
                                   zoom_range = 0.2,
                                   horizontal_flip = True)
```

```
test_datagen = ImageDataGenerator(rescale = 1./255)
```

```
training_set = test_datagen.flow_from_directory('/content/drive/MyDrive/classification/train',
                                                 target_size = (width, height),
                                                 batch_size = 32,
                                                 class_mode = 'categorical')
```

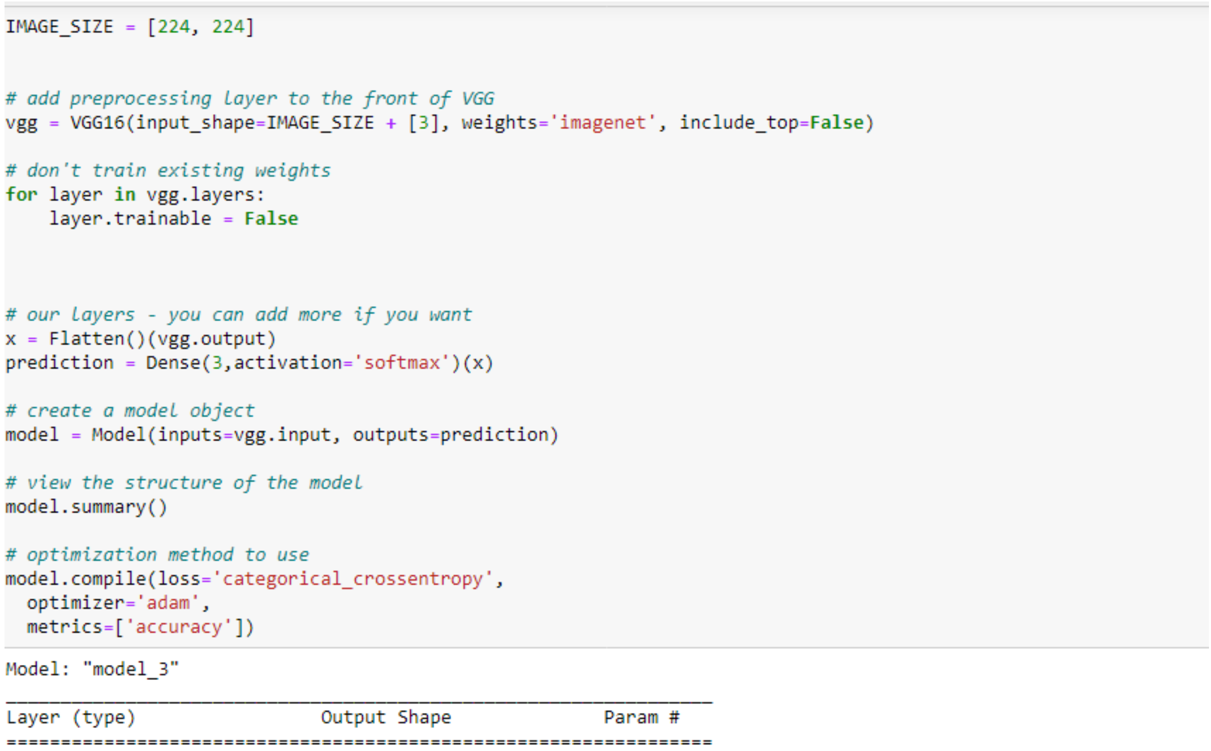
```
Found 561 images belonging to 3 classes.
```

```
test_set = test_datagen.flow_from_directory('/content/drive/MyDrive/classification/test',
                                            target_size = (width, height),
                                            batch_size = 2,
                                            class_mode = 'categorical')
```

```
Found 12 images belonging to 3 classes.
```

- ## Key Metrics for success in solving problem under consideration

  1.we have to solve this classification problem statement by VGG16.
  2.So we have to make our image size (224,224,3).
  3.We use Softmax activation function in the last layer and we use Adam optimizer for better accuracy.



```
IMAGE_SIZE = [224, 224]


# add preprocessing layer to the front of VGG
vgg = VGG16(input_shape=IMAGE_SIZE + [3], weights='imagenet', include_top=False)

# don't train existing weights
for layer in vgg.layers:
    layer.trainable = False



# our layers - you can add more if you want
x = Flatten()(vgg.output)
prediction = Dense(3,activation='softmax')(x)

# create a model object
model = Model(inputs=vgg.input, outputs=prediction)

# view the structure of the model
model.summary()

# optimization method to use
model.compile(loss='categorical_crossentropy',
  optimizer='adam',
  metrics=['accuracy'])
```

```
Model: "model_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
```

- ## Interpretation of the Results

We got an accuracy score of 100% in 10 epochs .

CONCLUSION

- ## Key Findings and Conclusions of the Study

By doing small small modification in above code, we can scrap images of any commercial websites and classify images of different catagories such as family photos classifier,vehicle classifier,animal classifier etc.

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

We get good accuracy by using only 10epochs and Vgg16 algorithms

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

By doing small small modification in above code, we can classify images of different catagories such as family photos classifier,vehicle classifier,animal classifier etc.