IMAGE SCRAPING AND CLASSIFICATION PROJECT

Submitted by

MIENGANDHA SINHA

ACKNOWLEDGE

I would like to express my gratitude to the Company FlipRoboTechnology to give this project to me. In making of this project I hereby used to take help from the references from some websites and aslo from which is given by the some websites and also from which is given by the company as sample documentation and details related project and professional guided me a lot in the project and the other previous projects helped me and guided me with completion of the project.

INTRODUCTION

- Image scraping and classification

From the title we can understand the image scraping and its classification, scraping is the important thing because web scraping means extracting data from websites, wherein a large amount of data after extraction is stored in a local system. It can access the world wide web through https and a web browser. Scraping is a very essential skill for everyone to get data from any website. For scraping images from web, we have to keep in mind that from scraping field is making appropriate use of information that can be gathered through images by examining its features and details. Images are one of the major sources of data in the field of data science and AI. This is done to make the model more understandable.

- ## Conceptual Background of the Domain Problem

From the words Image classification, it analyzes images with their AI-based Deep Learning power, the models that can be identify and recognize a wide variety of criteria. It is a term to describe a set of algorithms and technologies that attempt to analyze images and understand the hidden representations.

- ## Review of Project

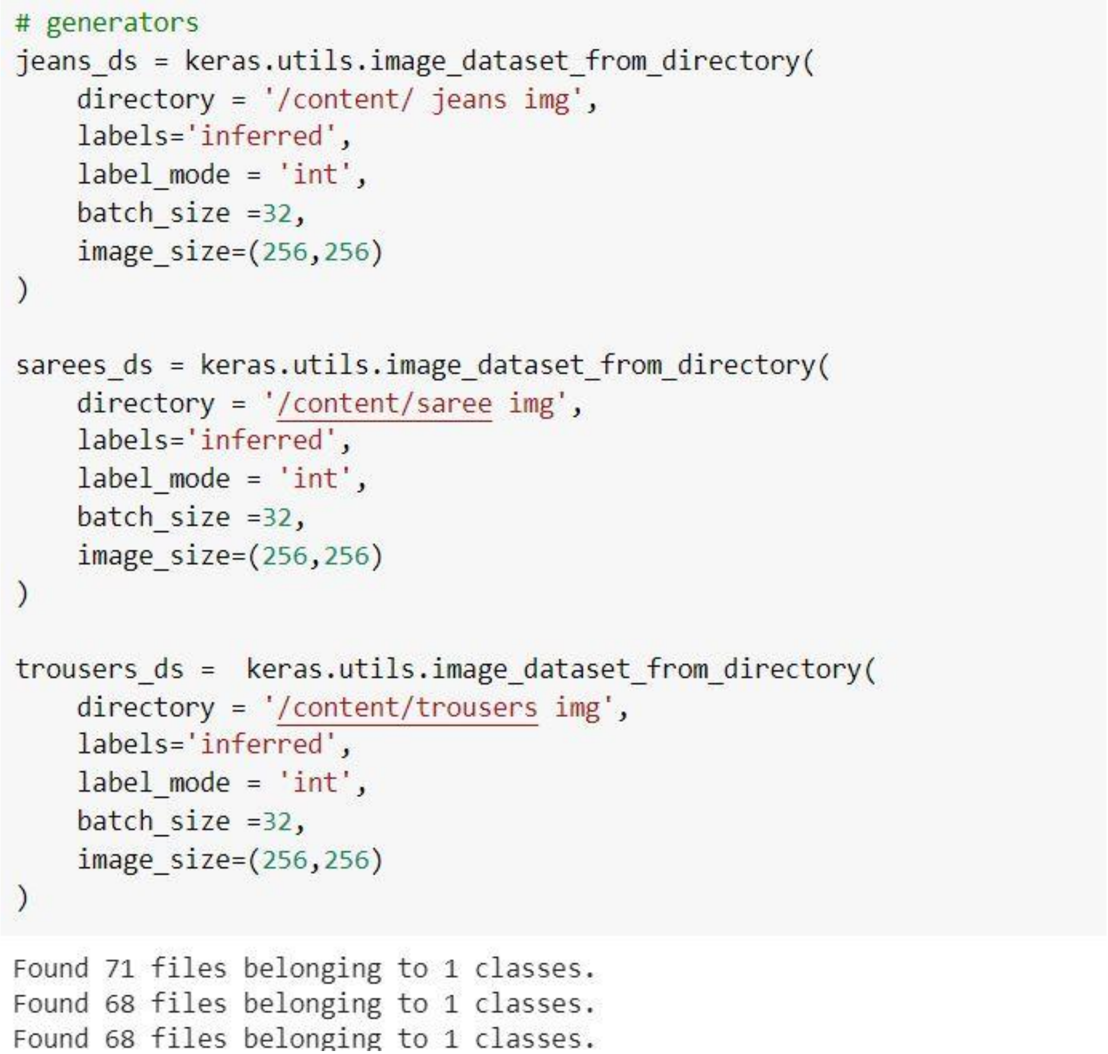
From the collecting data phase to the model building phase gives important variables in the classifying images. And then the model building do all the data visualizations, data pre-processing steps, evaluating the model, data cleaning and selecting the best model for the project.

- ## Motivation for the Problem Undertaken

Here there are files where the scraped images kept and from those they are classified and build a model. Images of jeans datasets have 71 files belonging to 1 classes. Images of sarees datasets have 68 files belonging to 1 classes. Images of trousers datasets have 68 files belonging to 1 classes. The idea behind in this project is to build a deep leaning based Image Classification model on images that will be scraped from e-commerce portal. This task is divided into two phases: Data collection and Model building.

ANALYTICAL PROBLEM FRAMING

- Mathematical/Analytical Modelling of the Problem

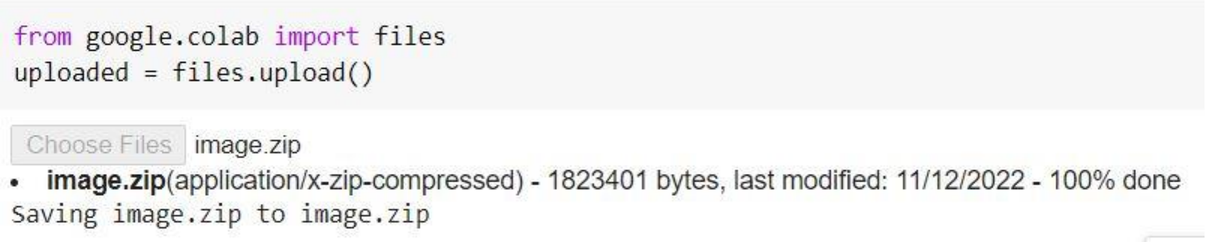


```
# generators
jeans_ds = keras.utils.image_dataset_from_directory(
    directory = '/content/ jeans img',
    labels='inferred',
    label_mode = 'int',
    batch_size =32,
    image_size=(256,256)
)

sarees_ds = keras.utils.image_dataset_from_directory(
    directory = '/content/saree img',
    labels='inferred',
    label_mode = 'int',
    batch_size =32,
    image_size=(256,256)
)

trousers_ds =  keras.utils.image_dataset_from_directory(
    directory = '/content/trousers img',
    labels='inferred',
    label_mode = 'int',
    batch_size =32,
    image_size=(256,256)
)
```

```
Found 71 files belonging to 1 classes.
Found 68 files belonging to 1 classes.
Found 68 files belonging to 1 classes.
```

- Data Sources and their formats

The data sources and their formats are from .zip file.



```
from google.colab import files
uploaded = files.upload()
```

```
• image.zip(application/x-zip-compressed) - 1823401 bytes, last modified: 11/12/2022 - 100% done
Saving image.zip to image.zip
```

- Data Inputs- Logic- Output Relationships

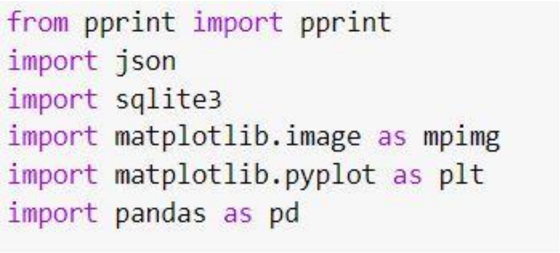
The relationships between inputs and outputs can be studied extracting weights of the trained model. Regression is that relationships between them can be blocky or highly structured based on the training data. It requires the data scientist to train the algorithm with both labeled inputs and desired outputs.

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

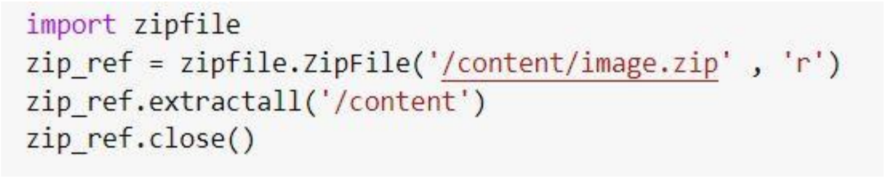
Presumptions are by using regression label encoding, classifier, selection of the best models, confusion matrix that it means the relationship between the dependent and independent variables look fairly linear. Thus, our linearity assumption is satisfied.

- Hardware and Software Requirements and Tools Used

By importing many libraries are



```
from pprint import pprint
import json
import sqlite3
import matplotlib.image as mpimg
import matplotlib.pyplot as plt
import pandas as pd
```



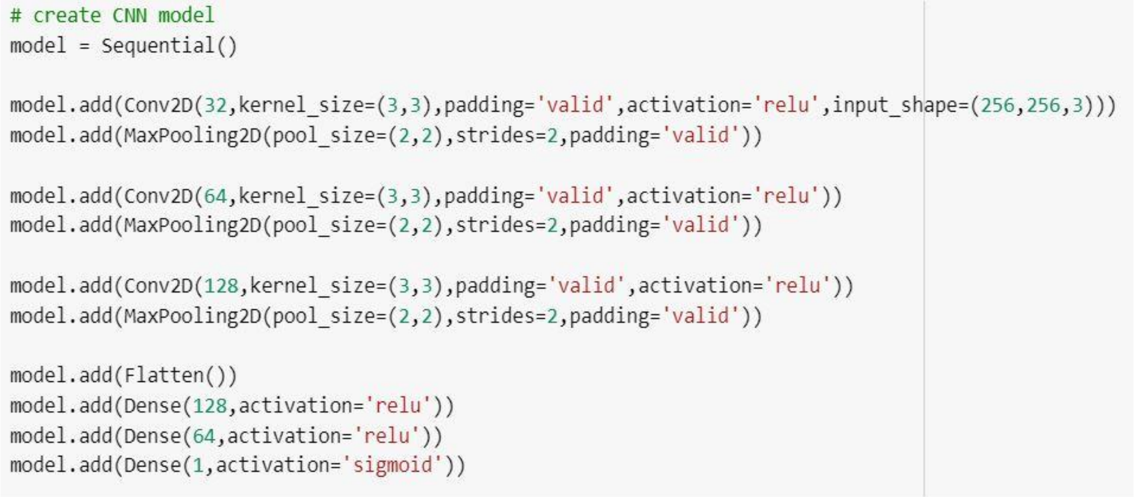
```
import zipfile
zip_ref = zipfile.ZipFile('/content/image.zip' , 'r')
zip_ref.extractall('/content')
zip_ref.close()
```

MODEL/s DEVELOPMENT AND EVALUATION

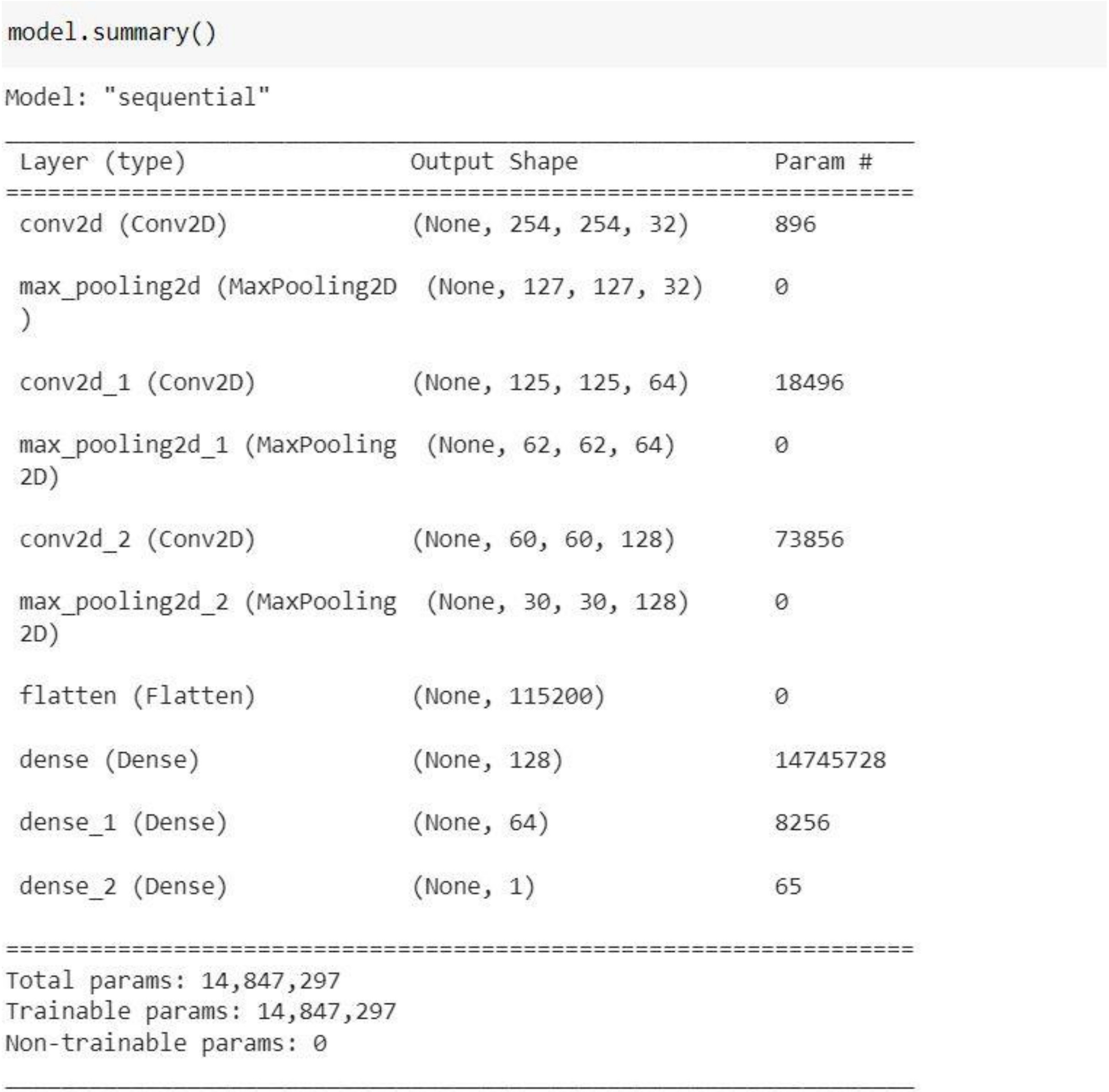
- Identification of possible problem-solving approaches(methods)

The collection and interpretation of data in order to uncover patterns and trends. It is a component of data analytics. Statistical analysis can be used in situations like gathering research interpretations, statistical modelling or designing surveys and studies. The approaches/methods of identification are descriptive and inferential statistics which are describes as the properties of sample and population data, and inferential statistics which uses those properties to test hypotheses and draw efficient conclusions in terms of outputs.

- Model Building



```
# create CNN model
model = Sequential()

model.add(Conv2D(32,kernel_size=(3,3),padding='valid',activation='relu',input_shape=(256,256,3)))
model.add(MaxPooling2D(pool_size=(2,2),strides=2,padding='valid'))

model.add(Conv2D(64,kernel_size=(3,3),padding='valid',activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2),strides=2,padding='valid'))

model.add(Conv2D(128,kernel_size=(3,3),padding='valid',activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2),strides=2,padding='valid'))

model.add(Flatten())
model.add(Dense(128,activation='relu'))
model.add(Dense(64,activation='relu'))
model.add(Dense(1,activation='sigmoid'))
```



```
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 254, 254, 32)      896

 max_pooling2d (MaxPooling2D  (None, 127, 127, 32)     0
 )

 conv2d_1 (Conv2D)           (None, 125, 125, 64)      18496

 max_pooling2d_1 (MaxPooling  (None, 62, 62, 64)       0
 2D)

 conv2d_2 (Conv2D)           (None, 60, 60, 128)       73856

 max_pooling2d_2 (MaxPooling  (None, 30, 30, 128)      0
 2D)

 flatten (Flatten)           (None, 115200)            0

 dense (Dense)               (None, 128)               14745728

 dense_1 (Dense)             (None, 64)                8256

 dense_2 (Dense)             (None, 1)                 65

=================================================================
Total params: 14,847,297
Trainable params: 14,847,297
Non-trainable params: 0
_________________________________________________________________
```

- Interpretation of the Results
  - A human analyst attempting to classify features in an image uses the elements of visual interpretation.
  - Deep Learning using Convolutional Neural Network (CNN) which is reputable for working with images.
  - Image classification is a subset of image recognition which has widespread use in the security industry (facial recognition), virtual search engine (object finder in stores), healthcare (emotion detection in patients) and gaming and augmented reality.
  - The purpose of this project to make an exposure of how an end to end project is developed in this field.
  - The task is divided into two phases: Data Collection and Model Building.

CONCLUSION

In conclusion, combining the formerly know the datasets analyze and to improve the model building. This assignment is all about the subset of image recognition which has widespread use in the security industry (facial recognition), virtual search engine (object finder in stores), healthcare (emotion detection in patients) and gaming and augmented reality. In a way, smartphone cameras have made all these advances possible with multitudes of pictures that can be easily created.

THANK YOU