Image Scraping and Classification Project

Submitted by:

Sagnik Das

ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to (Datatrained) also to my SME (Ms. Sapna Verma). Who gave me the golden opportunity to do this wonderful project on the topic (Micro Credit Defaulter Project), which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to tall of them.

INTRODUCTION

• Business Problem Framing

- 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.
  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.

• Conceptual Background of the Domain Problem

This problem has 3 steps:-

a. Data Scrapping
b. Classification of image.
c. Optimization of the dataset and model building and prediction.

• Review of Literature

Identifies different types of clothing products and differentiate between male and female wear.

• Motivation for the Problem Undertaken

As 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. The idea behind this project is to build a deep learning-based Image Classification model on images that will be scraped from e-commerce portals.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

  Numpy and MatPlotLib

- Data Sources and their formats

  Amazon.com, Flipkart.com

  .Jpeg format

- Data Preprocessing Done

  Downloaded the data. Created folder. Uploaded the data in Jupyter Notebook.

- Data Inputs- Logic- Output Relationships

  Input - Downloaded the data. Created folder. Uploaded the data in Jupyter Notebook.

  Output – Classification of images.

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

  Here, you can describe any presumptions taken by you.

- Hardware and Software Requirements and Tools Used

  Hardware – Laptop
  Software – Jupyter Notebook, Selenium, BeautifulSoup, Pandas, Time, Requests, Numpy, TensorFlow, Keras, MatPlotLib.

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  Image Scraping and Image Classification.

- Testing of Identified Approaches (Algorithms)

  Numpy, MatPlotLib, Keras, Tensorflow.

- Run and Evaluate selected models

  Numpy, MatPlotLib, Keras, Tensorflow.

- Key Metrics for success in solving problem under consideration

  Numpy array, Np argmax.

- Visualizations

MatPlotLib - plt.figure()
plt.imshow(train_images[0])
plt.colorbar()
plt.grid(False)
plt.show()

```
plt.subplot()
   plt.xticks([])
   plt.yticks([])
   plt.grid(False)
   plt.imshow()
   plt.xlabel(class_names[train_labels[1]])
```

- Interpretation of the Results

  Identify true label of the product and classify it.

CONCLUSION

- Key Findings and Conclusions of the Study

  Identify true label of the product and classify it and single prediction.

- Learning Outcomes of the Study in respect of Data Science

  Identify true label of the product and classify it and single prediction.

- Limitations of this work and Scope for Future Work

  More data would have given better results. For future work more data should be considered.