IMAGE SCRAPING AND CLASSIFICATION

Submitted By,
Tinu Shiby
Data Science Intern
Flip Robo Technologies

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to my mentor Ms. Sapna Verma Robo Technologies for his inspiration and guidance at every stage of the project. The blessing, help and guidance given by him will certainly help me in achieving better things in future.

I wish to express my profound gratitude to the DataTrained family for providing an opportunity to undertake this internship.

Lastly, I thank the almighty and my parents for their every small support and encouragement which helped me to complete this project successfully.

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.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.I have scraped images of jeans,sarees and tops and build data from it. That data will be provided as an input to your deep learning problem.

Data Description

The data in this project is collected from amazon.in. Scrapped almost 928 images from the category of jeans, sarees and tops and saved the images in a folder.

IMAGE SCRAPING

The images are collected from the website amazon.in and scraped almost 928 images from the category jeans, sarees and tops.

TRAINING MODEL FOR CLASSIFICATION

The images are downloaded and are saved as a folder with three subfolders for jeans, sarees and tops.
Then labelled the images as follows:
jeans were replaced by 0, sarees were replaced by 1 and tops by 2. Then resized the image to (180,180,3).

Splitted the data as X and y where x is the resized images and y is the cloth label.Then converted X and y to arrays and scaled the data. By using the train-test split method the dataset is splitted as training and testing data.

The algorithms used for the training and testing is convolutional neural network.For our model development I have used the optimizer adam and for loss I used SparseCtaegoricalCrossentropy.

On the basis of accuracy and F1 score we choose convolutional neural network after data augmentation with dropout layer being applied.

CONCLUSION

In conclusion, this research is about image classification by using deep learning via framework TensorFlow. The roles of epochs in DNN was able to control accuracy and also prevent any problems such as overfitting. Thus we were able to classify the three categories of images with high testing accuracy of 100%.