Image Scraping and Classification Project

Submitted

by: Vijay

Venkatesh

ACKNOWLEDGMENT

The data was collected from amazon website to work on this project.

INTRODUCTION

- Business Problem Framing

  To build a deep learning-based Image Classification model on images that will be scraped from the e-commerce portal.
- Conceptual 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.
- Review of Literature

  There is not much research performed as the Data and related information was provided by the source itself, which was been taken into consideration based on the information given by Flip Robo.
- Motivation for the Problem Undertaken

  The Project was assigned by flip Robo as part of the internship phase for better understanding the concept and getting the idea of the industry.

Analytical Problem Framing

- Data Sources and their formats

  900 images from each category from e-commerce portal, Amazon.com. The clothing categories used for scraping are:
  - Sarees (women)
  - Trousers (men)
  - Jeans (men)

  .
- Data Pre-processing Done

  The data is in folder format which has images in jpg fromat saved in form of category to perform ML
- Hardware and Software Requirements and Tools Used

  The system with a 16 core processor was been used, The operating system was Windows 10, Anaconda 3 was been used for performing ML Libraries:

  import pandas as

  pd import numpy as

  np import seaborn

  as sns

  from keras.preprocessing.image import load_img,

  img_to_array import matplotlib.pyplot as plt

  import os

from keras.layers import Dense, Input, Dropout, GlobalAveragePooling2D, Flatten, Conv2D, BatchNormalization,
Activation, MaxPooling2D

from keras.models import Model,

Sequential from

tensorflow.keras.optimizers import Adam

from sklearn.metrics import

accuracy_score

Model/s Development and Evaluation

.

- Testing of Identified Approaches (Algorithms)

CNN model has been used which gave 81%

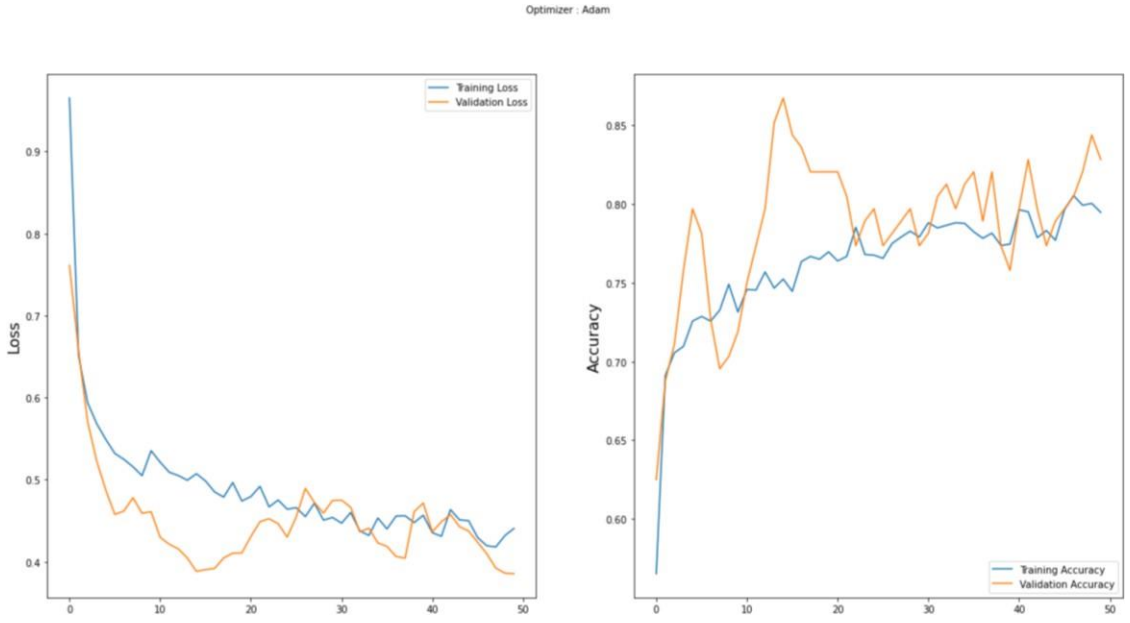
accuracy 4 convolutional layers

2 fully connected layers

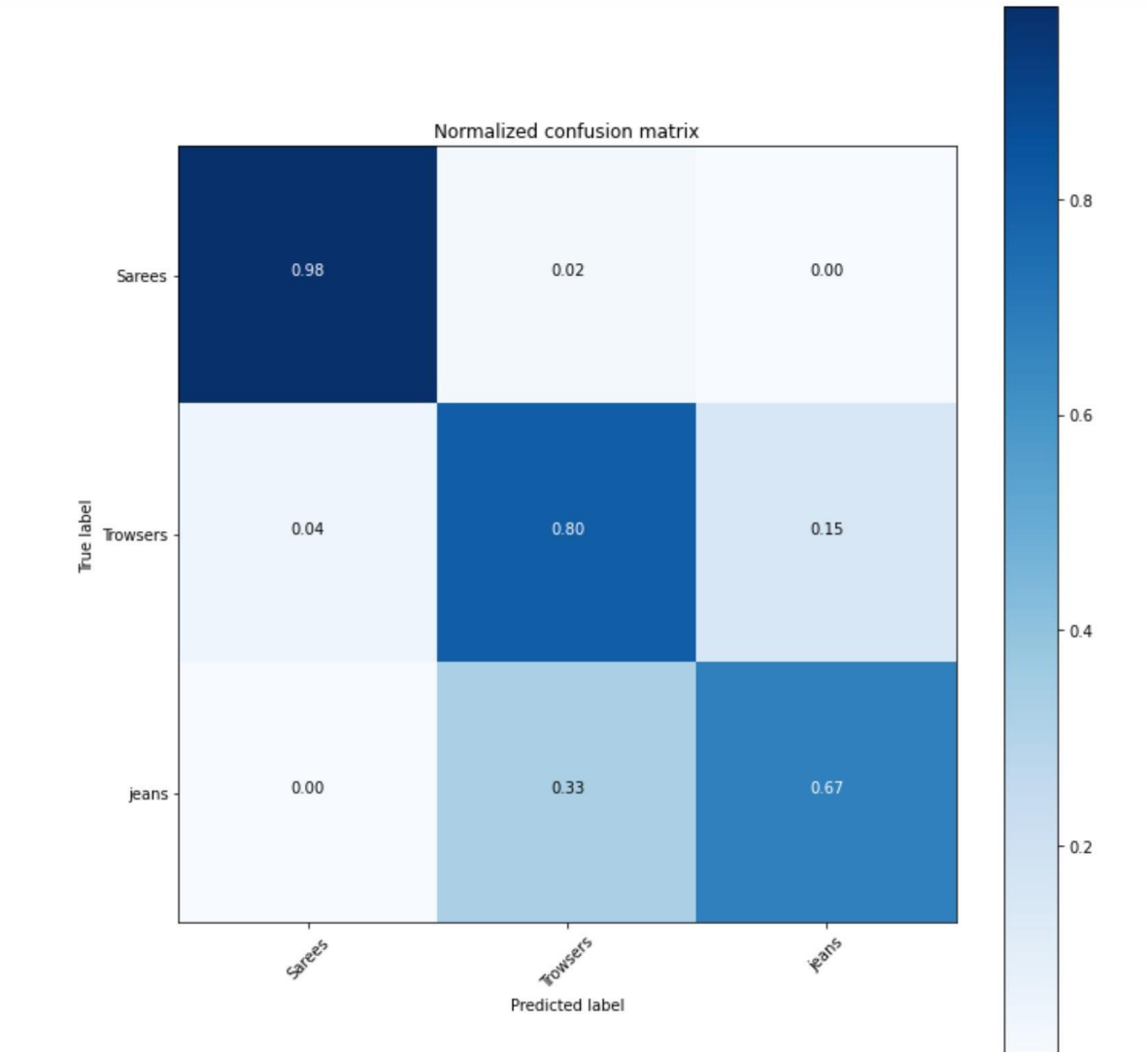
The convolutional layers will extract relevant features from the images and the fully connected layers will focus on using these features to classify well our images.

- Visualizations
  Random images from dataset

Loss and Accuracy on the train and validation sets



Confusion matrix



- Interpretation of the Results
  CNN model was been used which gave 81% accuracy , accuracy score was used measuring accuracy between prediction and test lables

CONCLUSION

- Key Findings and Conclusions of the Study

The model has a great performance for 81% accuracy. As the data wasn’t simulated and was been extracted from real-world applications. The performance was quite reasonable

- Learning Outcomes of the Study in respect of Data Science

  Adding more data and minor tweaks can improve the accuracy can help to increase the accuracy.

- Limitations of this work and Scope for Future Work

  As per observation under random images, there was a point where the model confuses between Jeans and Trousers as the input images had some similar features which were considerably difficult to distinguish between to classes.