Big Data Analytics Course Project

Abstract by,
Surya Narayanan S
121015098
4ITB

Topic given: Data Science and Machine Learning

Project Task given: Apply ML Algorithm to Dataset

Abstract:

Task chosen: *Recognize Handwriting using CNN + RNN with Keras and TensorFlow package*

Dataset: https://www.kaggle.com/landlord/handwriting-recognition

Dataset size: 1.3 GB

No of Images: 3.72 Lakhs

A CRNN (Convolutional Neural Network + Recurrent Neural Network) Machine Learning Model is built to recognize and convert Handwritten Images to text using TensorFlow and Keras package in python. This model is used to predict handwritten images to its equivalent text.

Before fitting the model, pre-processing like giving high contrast to background and foreground, converting to grayscale is done.

After predicting, accuracy of characters and words are used for evaluation.

Procedure:

1. Visualize image
    a. Find anomaly in images. Remove or correct if found any.
2. Pre-process image
    a. Fix a size for all images like height and width.
    b. Make high contrast between text and background.
    c. Convert image into Gray scale, if required.
    d. Split the dataset into train and test sets.
3. Build model and specify its parameters
    a. Build Convolutional Neural Network, Convert it into RNN compatible input.
    b. Build Recurrent Neural Network.
    c. Fit the model with train dataset and Predict the data with test dataset.
4. Evaluate ML Model
    a. Find Accuracy of letters and words predicted.
    b. Improve model by varying parameters of model, if possible.
    c. Improve model by varying Test: Train ratio.

Sample Image: (to predict as Kevin)