**GRADUATE SCHOOL RECOMMENDATION SYSTEM**

Individual Project - CMPE 256

Instructor: Shih Yu Chang

Semester: Summer, 2019

Name: Nithya Kuchadi

Student ID: 013769665

**Introduction:**

As competition for admission into higher education increases, it becomes even more important for applicants to find graduate schools that fit their requirements and expectation. Selecting appropriate schools to apply, however, is a time-consuming process, especially when looking for schools at graduate level due to the various factors in decision making imposed by the schools and applicants. This problem can be addressed by developing a recommender system for graduate admission seekers which can help them to choose graduate school matching their academic profile.

**Existing Work:**

Ramkishore Swaminathan, Joe Manley Gnanasekaran and Swetha Krishnakumar in the paper “University Recommender System for Graduate Studies in USA” have implemented a recommendation system that recommends a single university based on undergrad university, gre, gpa and research experience features and have achieved an accuracy of 53.4%. In this paper, they have built a recommendation system based on K-NN, SVM and Random Forest classification models.

**Limitations of the Existing Work:**

1.The models they have used are recommending only a single university.

2.The accuracy for these models is around 53% which is less.

**Motivation & Approach:**

To recommend more than one university to users and also to improve the accuracy of above mentioned classification algorithms for university prediction.

Implemented 2 models

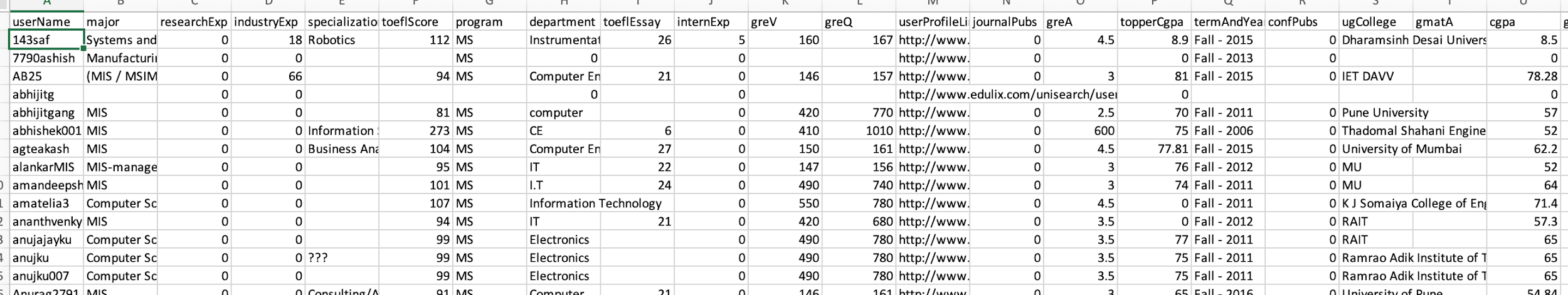
K Nearest Neighbors- To recommend n universities to the user

CNN- To improve the University prediction accuracy

**DATA**

Data is scraped from edulix.com. The required data is extracted from the HTML by using the python library ‘BeautifulSoup’.

WebScraped Data:



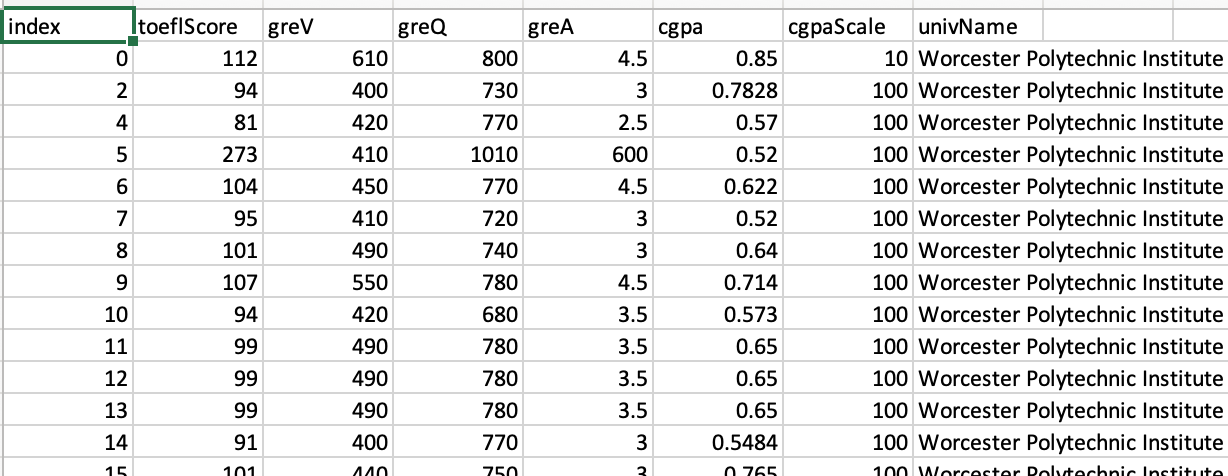
**Data Pre-processing Steps:**

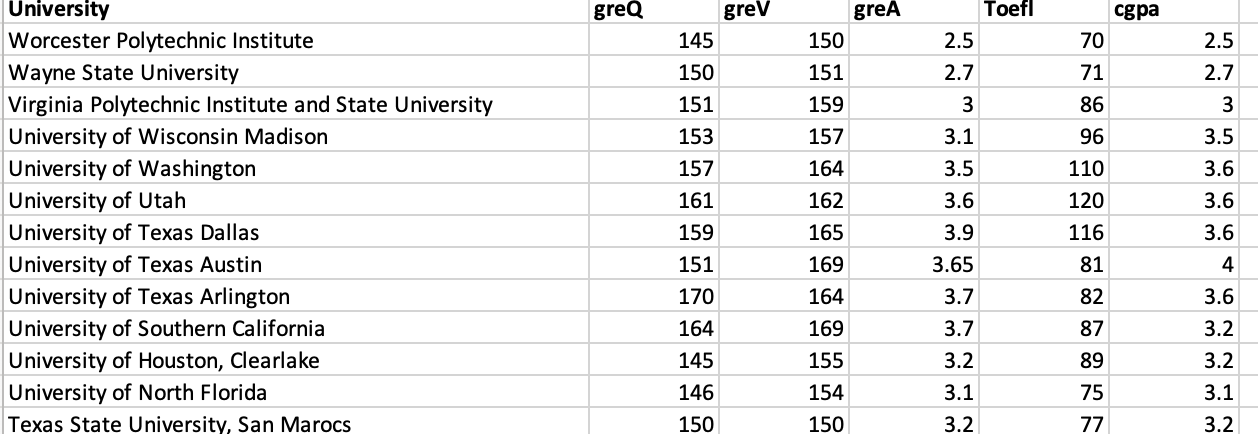
1. GRE Quant and Verbal scores are cleansed since they contained the scores of both old and new versions of the examination. The old scores are converted to new scores.

2. All the GPA scores were uniformly scaled to 4 point scale.

3.The empty fields are replaced with the average value.

**Data after Pre-processing:**

****

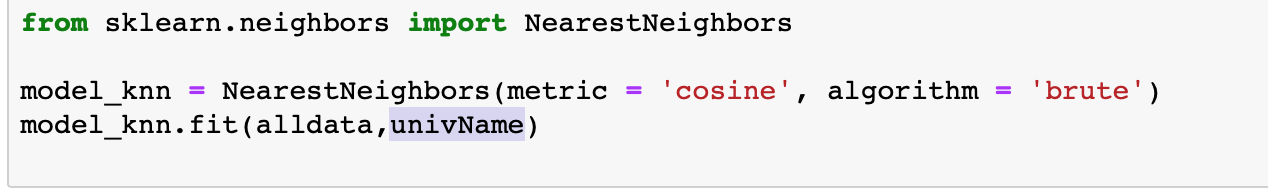
****

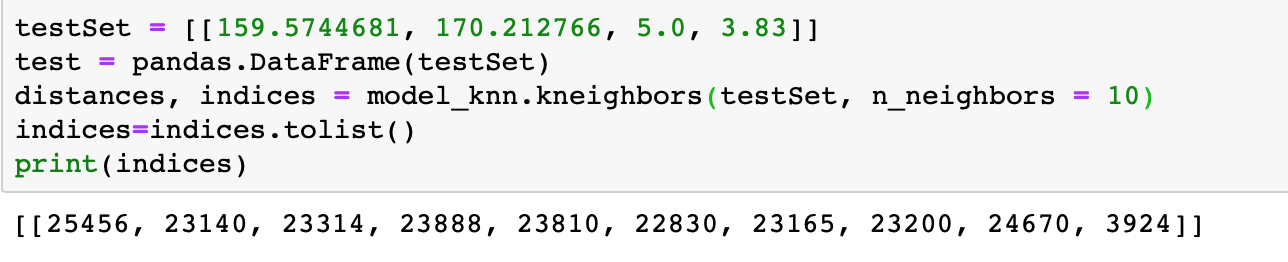
**Models:**

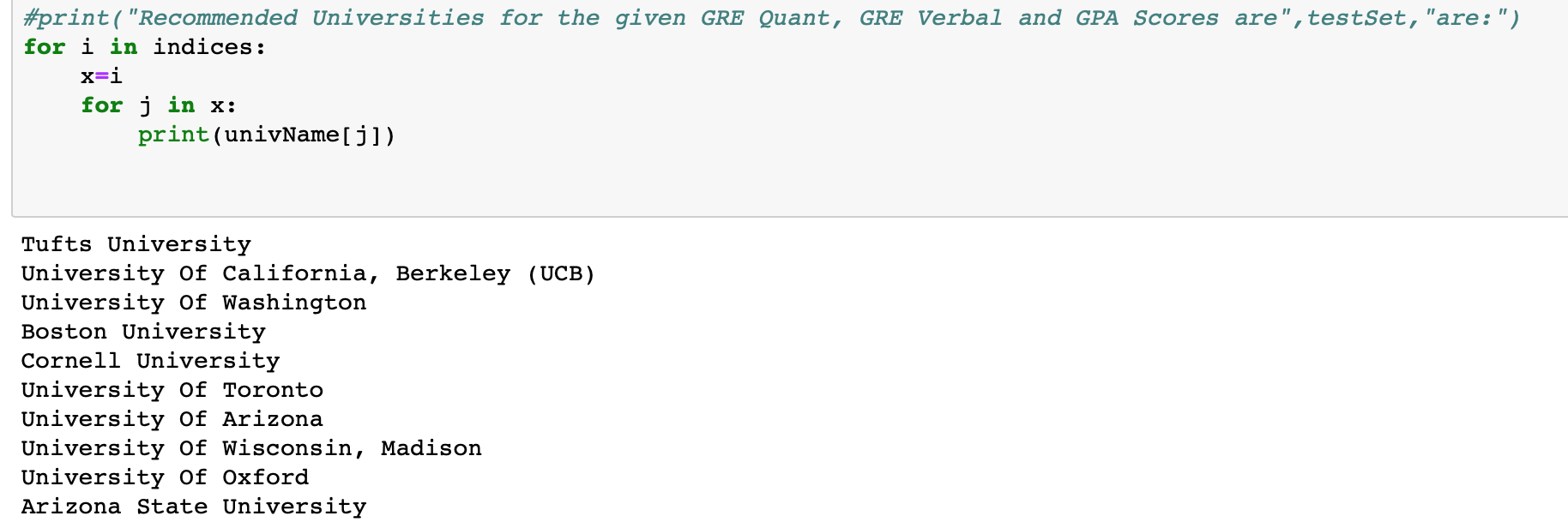
**K Nearest Neighbors:**

K-Nearest Neighbor is a commonly used recommendation algorithm in Information Retrieval. The core idea of the KNN algorithm is: if the majority of the k most similar neighbors of sample in the feature space belongs to a certain category, then the sample is considered to belong to this category. Traditionally it finds a set of users similar to a query user. From these neighbors a set of recommended items it constructed.

**KNN Implementation Screenshots:**







**Convolutional Neural Network:**

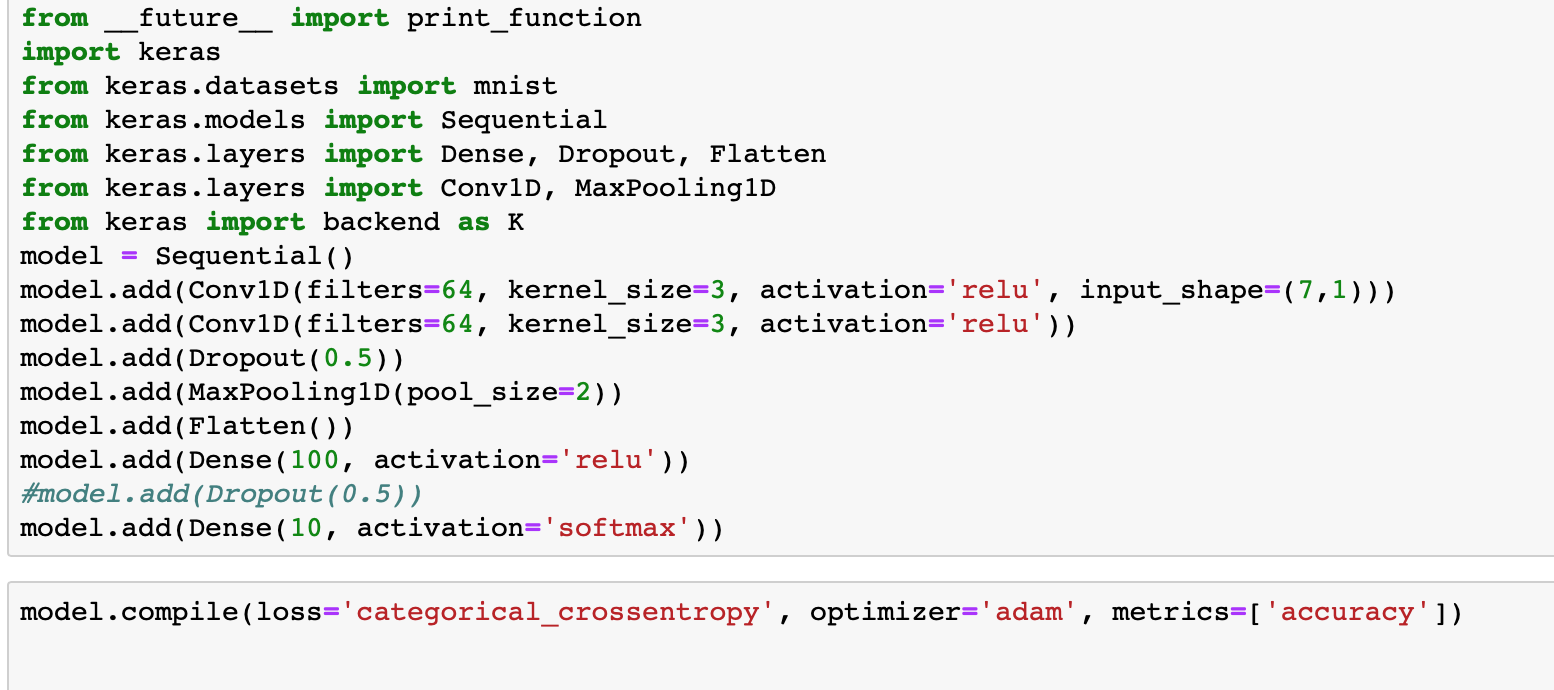
Convolutional Neural Network is a class of deep, forward-feed artificial neural networks, most commonly applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. It takes the input image and assigns learnable weights and biases to the various features in the image. A convolutional network is able to successfully capture the Spatial andTemporaldependenciesinanimagethroughthe application of relevant filters. The filter is used to detect patterns and slides over the entire image to capture relevant data.

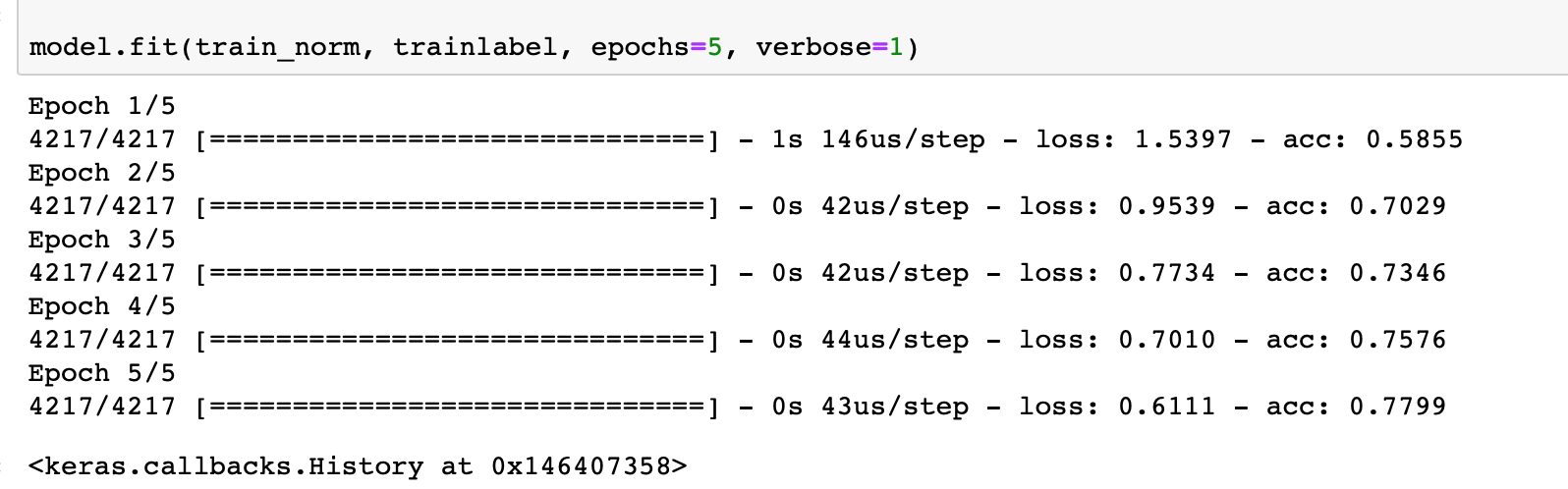
In CNN, input from one layer is passed to the next, wherein each layer is treated as an object that feeds data to the next layer. Below diagram depicts the architecture of a CNN based classification model.

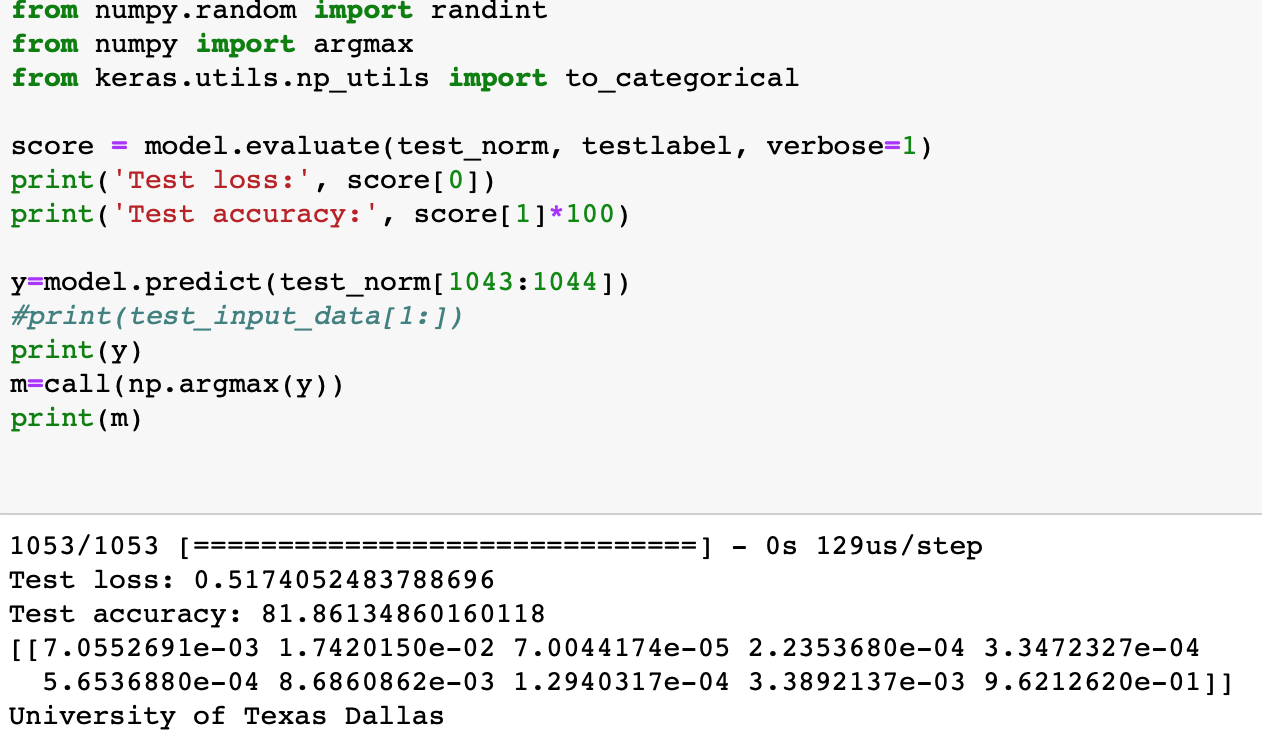
First layer is always convolutional layer, which, based on defined filter detects low-level features such as edges, color, gradient orientation, etc. Subsequent convolutional layers provides and wholesome understanding of images. The convolutional layer output is passed through the activation function ReLU (rectified linear unit) which replaces negative pixels with 0.

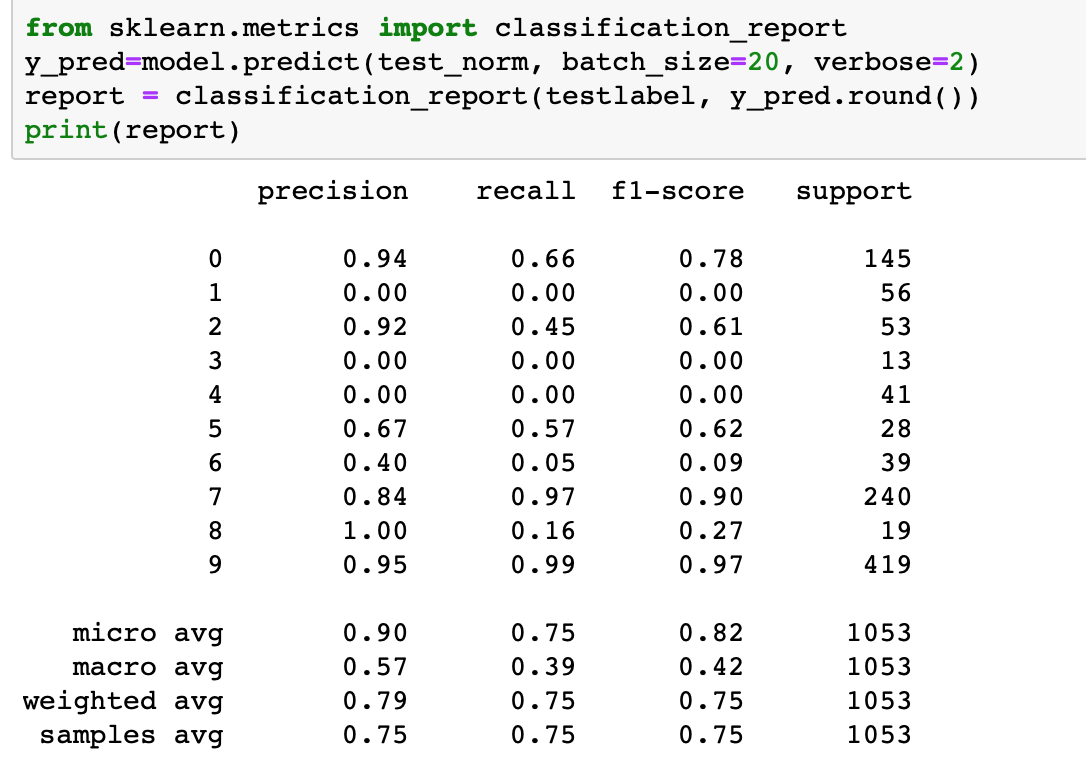
Similar to the Convolutional Layer, the Pooling layer reduces dimensionality of each feature map and retains the most important features. A dropout layer may be introduced in between which prevents overfitting. Before sending the output from dropout layer, to dense layer, the multi-dimensional array needs to be flattened using flatten layer. The final layer is the dense layer i.e. fully connected layer where every input is connected to every output by weight. Softmax activation function provides probabilistic distribution in different classes.

**CNN Implementation Screenshots:**

****

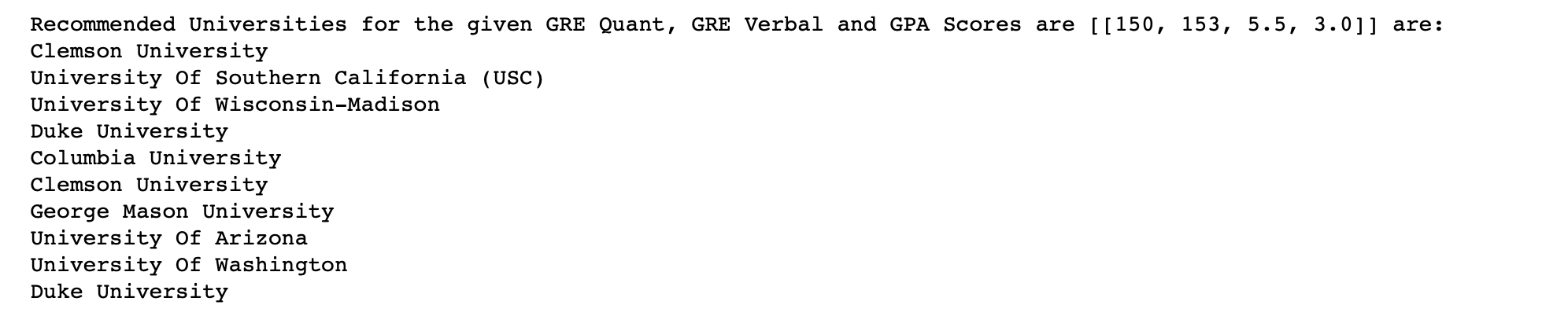
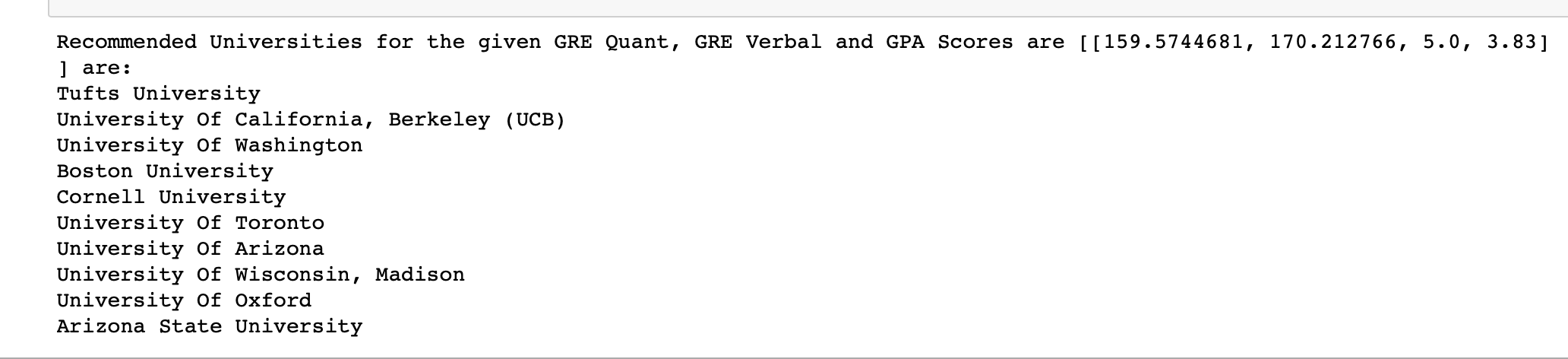
****

****

****

**Results:**

**Result of KNN:**

****

**Result of CNN for [159,170,5.0,3.83]:**

****

**References:**

1. Ramkishore Swaminathan, Joe Manley Gnanasekaran and Swetha Krishnakumar. University Recommender System for Graduate Studies in USA.