AI Assignment-4

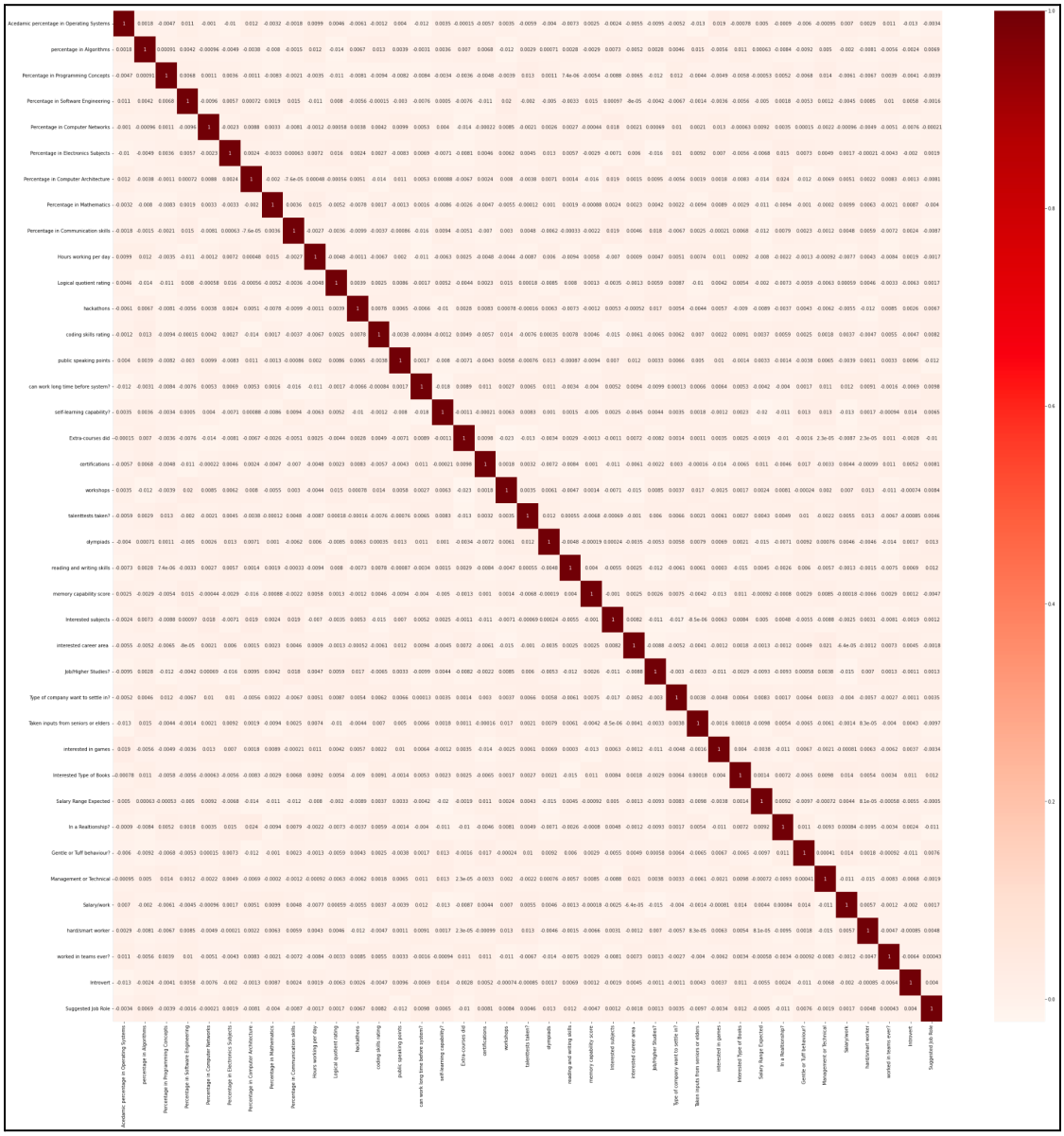
Pankaj Kumar
2019262

Dataset

- There are 20000 rows and 39 columns in the dataset.
- That implies there are 38 characteristics and 1 variable target.
- And the dataset has 20000 instances.

Feature Analysis

- Upon further examination of the dataset, we discovered a very low degree of correlation between the features.
- Some graphs are shown below. Here is the Heatmap.



Reclassifying Target variables and other modifications

- I categorized the target variable into 7-8 groups to enhance the accuracy of prediction.
- I have used Admin, Analyst, Developer, Manager, Architect, Support, Engineer, and Associate Roles courses.
- Following Reclassification, the accuracy rating increased from 5% to 20%.
- This categorization served the model well.
- Additional Bucketization of Grades didn't enhance the model's accuracy Therefore I removed it.
- There were no NULL values.
- We have also bucketize the numerical values into low,medium and high.

Label Encoding

- I performed label encoding to make the data more compatible with machine learning algorithms.
- I applied label encoder fit_transform to each column to convert them into numbers and classes.
- It can also be used to convert non-numerical labels to numerical labels, provided they are hashable and comparable.

ANN Model Predictions Accuracy

I have constructed a 5-Layer neural network using MLP Classifier Here are some of the Parameters used for the Training Purpose (128,64,32 & Relu)

Split	Accuracy Achieved
90-10	0.182
80-20	0.18775
70-30	0.18216666666666667
60-40	0.183875

Accuracies achieved using Other Models

Logistic Regression

Split	Accuracy Achieved
90-10	0.1915
80-20	0.1865
70-30	0.1865
60-40	0.18225

Random Forest Classifier

Split	Accuracy Achieved
90-10	0.172
80-20	0.17025
70-30	0.17233333333333334
60-40	0.1745

XGB Classifier

Split	Accuracy Achieved
90-10	0.183
80-20	0.179
70-30	0.18933333333333333
60-40	0.175625

SVM Classifier

Split	Accuracy Achieved
90-10	0.188
80-20	0.18075
70-30	0.182
60-40	0.1805