Technology Prediction Based on Entry Test Result.

Artificial Intelligence’s Project Report

# Abstract

In this report I have compared and analyzed the performance of two different machine learning algorithms: Logistics Regression and Decision Tree. And find that which one is better and why?

# Introduction

For Implementing these two algorithms I have used python’s scikit-learn and Microsoft Azure Machine Learning Studio Workbench.

Microsoft Machine Learning Studio is a platform where data scientists, and machine learning enthusiastic prepares data for training, train models, and ship them for further use. Azure presents fully drag and drop features where you pick any module form side panel and use them to Experiment the project.

Scikit-learn is python Machine learning library that comes with some well-known ML Algorithms in all sub domains of ML like Supervised, Unsupervised, Reinforced Learnings.

# Workflow:

* Data Extraction
* Data Filtration
* Feature Extraction
* Encoding categorical Features
* Selecting ML Model
* Evaluation of Results

Data Extraction

For data I download the old merit list of Mehran UET of year 2016 from the university site and run python scrip for converting raw data in to .csv format

Data Filtration

For the data filtration process all those rows which are null are removed.

Feature Extraction

I selected three features: district, area, and cpn for training the model that are strong and that can better participate in predicting the target attribute

Encoding categorical Features

For Encoding the data, I used python dictionaries to map between two values this is because to get reverse form it.

Evaluation of Results

The accuracy, performance matrix, confusion matrix of the algorithms are computed.

Basically, the Models will find the fact that students usually do while selecting the technology/ department based on their score, district, and area. For this I have chosen two algorithms Logistics Regression and Decision Tree due to the classical dataset’s attributes.

### Dataset attributes:

* Seat No.
* Name
* Father's Name
* Surname
* District [Feature Attribute]
* U/R [Feature Attribute]
* CPN [Feature Attribute]
* Discp [Target Attribute]

# Model Evaluation Microsoft Azure ML Studio

A close up of a map

Description automatically generated

Figure 1 Machine Learning Model

A screenshot of a cell phone

Description automatically generated

Figure 2 Featured Data Visualization

A screenshot of a video game

Description automatically generated

Figure 3 Featured Data Visualization 2

A screenshot of a cell phone

Description automatically generated

Figure 4 Featured Data Visualization 3

### Logistics Regression Model Evaluation

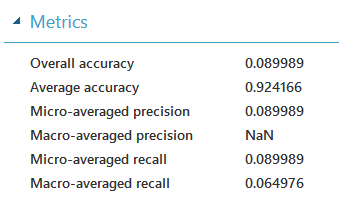


Figure 5 Logistics Regression Performance Matrics

A screenshot of a computer

Description automatically generated

Figure 6 Logistics Regression Confusion Matrix

### Decision Tree Model Evaluation

A screenshot of a social media post

Description automatically generated

Figure 7 Decision Tree Model Evaluation

# Python Sci-kit Learn Model Evaluation

## Logistics Regression Model Evaluation

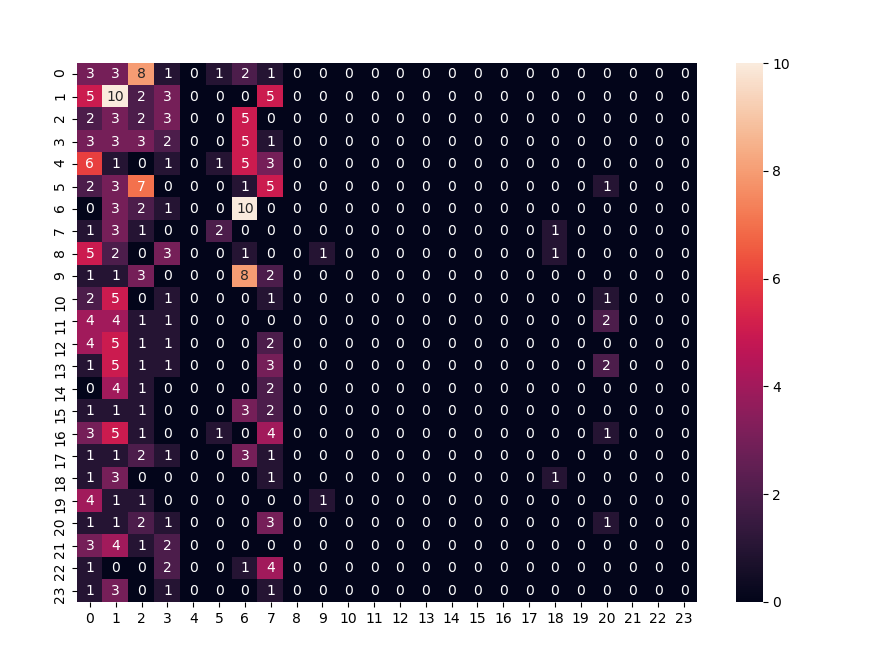


Figure 8 Logistics Regression Confusion Matrix

A screenshot of a cell phone

Description automatically generated

Figure 9 Logistics Regress accuracy measure

## Decision Tree

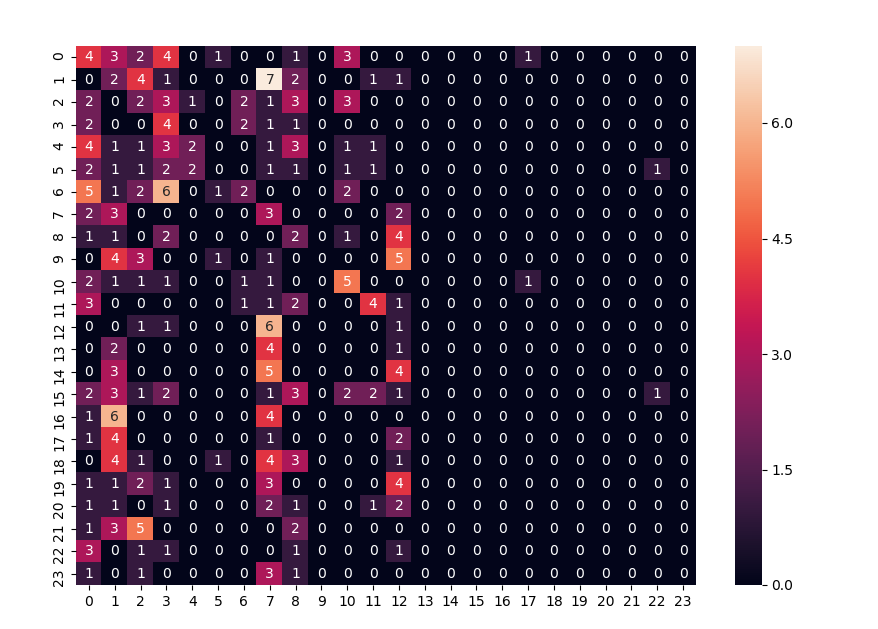


Figure 10 Decision Tree Confusion Matrix

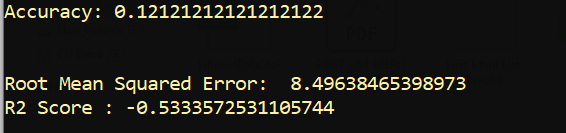


Figure 11 Decision Tree Accuracy

# Results and Findings

The Accuracy of the Azure ML algorithms and Scikit learn algorithms are quite same and the accuracy of two different algorithms are also showing same results with little variance.

By changing the number of iterations, or by changing the kernel of algorithm some effect on the accuracy of the algorithm are seen. For example, when changing the kernel function of Logistics regression model to “multiclass” it produces positive results on the accuracy of the algorithm.

The Evaluation results that I have obtained from that dataset are not up to the mark that I was expecting, but the predictor when I tested it on my raw data It predicts nicely and efficiently by comparing the results from the list.

In the Evaluation I found that sometimes the accuracy of the model can be seen low but if you have applied the right model with right technique the predicted results will be much more accurate.

# Live Data Test

It can be easily seen that the model is predicting the department with precision although its accuracy is very low.

A screen shot of a social media post

Description automatically generated

Figure 12 Live Data Test

# References

[1] Student Entry Test Result Downloadable. Link: <https://github.com/Mohammad-Bilal-Bhatti/Muet-Matit-Lists/blob/master/FIRST%20LIST%20MERIT%20F16.pdf>

[2] Different Merit List of Reference and Download. link: <https://github.com/Mohammad-Bilal-Bhatti/Muet-Matit-Lists>

[3] Medium Article on Linear Regression. link: <https://towardsdatascience.com/linear-regression-using-python-b136c91bf0a2>

[4] Medium Article on Random Forest. Link: [https://towardsdatascience.com/random-forest-in-python-24d0893d51c0](https://towardsdatascience.com/random-forest-in-python-24d0893d51c0%20)

[5] Medium Article on Logistic Regression. link: link: <https://towardsdatascience.com/building-a-logistic-regression-in-python-301d27367c24>