

**CSE303: Statistics for Data Science**

**[Summer 2021]**

**Project Report**

**Submitted by:**

|  |  |  |
| --- | --- | --- |
| **Student ID** | **Student Name** | **Contribution Percentage** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Introduction

Provide an overview of your project work including the performance of different models. Must not be more than 250 words.

# Dataset Characteristics and Exploratory Data Analysis

In this section, introduce your dataset. Mention number of rows, columns and other characteristics. Provide the histograms of data distribution and correlations among the variable with a suitable discussion. Try to stand out and creative in your presentation!

**Show the correlation of the columns here and Answer Task 5 in this section.**

# Machine Learning Models

Provide a brief description of the machine learning models you used such as Linear Regression, variant of Linear Regression, Logistic Regression and Support Vector Machine for regression and classification tasks.

Don’t copy-paste directly from the Internet! Write in your own words. Maximum 2 Pages.

# Data Preprocessing

Explain your data preprocessing steps including filling null values, dropping duplicates, encoding, dimensionality reduction etc. whichever is applicable. Appropriate data preprocessing can hugely impact your model’s performance.

# Different Models

Discuss the different models that you implemented in this project with their parameters. Provide a detailed description of their parameters. Use tables as necessary.

### 5.1 Regression Models

Put the description of Regression Models here. **(Task 1)**

### 5.2 Classification Models

Put the description of Classification Models here. **(Task 2)**

# Performance Evaluation

Use charts, figures appropriately to visualize and compare the performance of different models. Prove me that you are a master of visualization!

### 6.1 Regression Models

Write the performance evaluation of Regression Models here. **(Task 1)**

### 6.2 Classification Models

Write the performance evaluation of Classification Models here. **(Task 2)**

# Discussion

Most important section!

Analyze the performance of the models and provide your hypothesis behind their performance, e.g. Why some models are performing better than others? Provide appropriate reasonings of your hypothesis.

Show the ROC curve of the best performing classification model in this section. **[Task 4]**

Answer the question – “how can the university increase satisfaction level” in this section too. **[Task 6]**

# Conclusion

What’s on your mind after finishing this project?

Put remarks, comments, challenges, opportunities – anything in this section.

Remember, this is the last paragraph you are writing for this course. Make it convincing enough that you really worked hard for this course (if appropriate)!