**ABSTRACT**

**COLLEGE ADMISSION PREDICTION FOR MASTERS**

**Collecting the Data:**

In order to get the admission in the universities, we will predict the chances based on certain factors they are as follows:

* GRE Score
* IELTS/TOEFL/DUOLINGO score
* University Rating
* SOP
* LOR
* CGPA
* Research
* Chance of Admit

**Exploring the dataset:**

* The data has a total of 500 rows and 9 columns.
* Calculating the basic statistics on numeric columns.
* Returns different datatypes for each columns.
* Returns true for column having null values, else false.
* Rename the columns with appropriate names.

**Calculations:**

* Based on different factors such as the GRE, Toefl /IELTS/University ratings, SOP, LOR, CGPA we can calculate and predict the chances of getting into the required university.

**Identifying the basic computing needs:**

* The data we have taken for this project is the college statistical visual database from different entrance exams.
* If the data is un-structured, we will cleaning and remove the missing values and format the data in a structured data format.
* **Importing** - We are importing the data from the csv (comma separated value) file where we can find all the data related to our project.
* **Store** - we will store the data in the database and read the data whenever required based on the requirements.
* **Data file conversion Format**: We can open the given csv data and convert the file into txt format based on the external required format i.e text reader etc using the simple python functionality.
* **Upload and interpret Data:** We will upload and interpret the data based on certain factors and pre - process and clean the data.
* **Work performing on this Data:** The work we will we performing on this data is the data preprocess, clean, calculate the statistics, data visualizations and check if there are any computing requirements.

**Data Visualization:**

* Visualizing the GRE features.
* Visualizing the TOEFL features.
* Visualizing the University Rating features.
* Visualizing the Distribution of SOP features.
* Visualizing the LOR Rating.
* Visualizing the Distribution of CGPA features.
* Visualizing the Distribution pf Research papers.

**Data Cleaning:**

* Checking the redundant and inconsistent values in the given data.
* Remove the serial numbers and the column.
* Removing the null values from the columns like GRE, TOEFL, University Rating, SOP, LOR, CGPA, by Nan and calculating the missing values.

**Building the Model:**

* Splitting the dataset into features and label
* Calculating the function to calculate the best model for this problem.
* Check which algorithm has the highest accuracy, the selected model for this problem is the best model.
* Now, Based on our predictions, we can predict the best university based on the values using the trained model.