

Name: Bamba GUEYE

Project Guideline: **Prediction of Graduate Admission**

1. Project Idea:

Our project is called Prediction of Graduate Admission. The idea of our project is to predict the admission of graduate students. We know that most students have difficulty deciding which university to choose. Many students pursuing master's degrees struggle to select a shortlist of potential universities, due to a lack of familiarity with university rankings or having been provided inaccurate information by their peers.

Our goal is to create a form to guide students toward universities well-suited to their academic profiles through a comprehensive assessment and avoid overestimate or underestimate their profile.

2. Relevance to Sustainable Development Goals (SDGs):

This project is related to the SDGs, particularly the number 4 “Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all” which is the quality of education because ensuring that admission policies provide equal opportunities for all qualified individuals, regardless of their background, can help in promoting equitable access to education.

3. Literature Examples:

This project is inspired by the research papers following:

- “Predictive Modeling for Graduate Admissions Using Machine Learning Techniques” by I. Shanthi and Dr. K. Venkata Rao. It was published in the *Journal of Emerging Technologies and Innovative Research (JETIR)*, Volume 6, Issue 6, in June 2019.
- “A Comparison of Regression Models for Prediction of Graduate Admissions” by Mohan S Acharya, Asfia Armaan, and Aneeta S Antony. In Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019), 2019.

4. Describe Your Data:

The dataset has 500 instances and 9 attributes. It comes from the website Kaggle. The dataset consists of attributes like Serial No, GRE Score, TOEFL Score, University Rating, SOP (Statement of Purpose), LOR (Letter of Recommendation), CGPA (Undergraduate Cumulative GPA), Research, and Chance of Admit. The data format is in CSV and the data preprocessing steps required are:

- Loading CSV Dataset into Pandas Dataframe
- Quick Look into Datasets Rows

- Dimension of dataset
- Quick overview of dataframe
- Quick statistical summary
- Check and remove missing values if it exist
- replace missing value and data visualisation
- check for duplicate values and drop

If necessary:

- Data transformation/Manipulation
  - Alters axis labels
  - apply a function along in axis Apply()
  - convert the column form data type
  - convert the dataframe to list
  - query for the dataset
  - sort a dataset
  - random sample
  - filter...

#### 5. Approach (Machine Learning or Deep Learning):

I will use a machine learning approach to solve this problem because I think this task is not complex causing by the number of lines and columns one has in our datasets. We don't have a large dataset to choose a deep learning model.