Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Contributor Roles:

- **ROHIT RAJ:**
  (rajrohit2500@gmail.com)
  - **Loading Dataset**
  - **Data Analysis Using EDA**
  - **Finding and Treating Null Values**
  - **Model Development**
    - **Linear Regression**
  - **Observations**
  - **Conclusions**

- **RANJITA RAJ:**
  (ranjitaraj00@gmail.com)
  - **Importing Libraries**
  - **Loading datasets**
  - **Feature Engineering**
  - **Finding and Treatment of Outlier**
  - **Model Development**
    - **RandomForest Regression**
  - **Observations**
  - **Conclusions**

- **SANDHYARANI PATRA:**
  (Sandhyaranipatra25@gmail.com)
  - **Importing Libraries**
  - **Loading datasets**
  - **Model Development**
    - **XGB Regression**
  - **Hyperparameter Tuning**
  - **Observations**
  - **Model Selection**
  - **Conclusion**

Please paste the GitHub Repo link.

Github: https://github.com/sandhya-8368/Ted-Talk-Views-Prediction

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Project Name: TED Talks Views Prediction

Problem Statement:

A TED talk is a recorded public-speaking presentation that was originally given at the main TED (technology, entertainment and design) annual event or one of its many satellite events around the world. TED is a nonprofit devoted to spreading ideas, usually in the form of short, powerful talks, often called "TED talks." TED is dedicated to researching and sharing knowledge that matters through short talks and presentations. Their goal is to inform and educate global audiences in an accessible way.

Problem Approach :

We have Started with data loading and we have done EDA,Null value treatment,feature engineering, data cleaning,outlier treatment, target encoding feature selection, Hyperparameter Tuning then model building and finally the model selection. So we have used the below models.

- **Linear Regression Model**
- **RandomForest Regressor Model**
- **XGB Regressor Model**

Conclusion:

- **In all of the models, the errors have been in the range of 2,00,000 which is around 10% of the average views. which implies that the prediction was 90% of the time correct.**
- **RMSE is more influenced by Outliers where as MAE does not increase with Outliers. MAE is linear and RMSE is quadratically increasing.**
- **So we chosen MAE as deciding factor for our model.**