**Capstone Project Submission**

| **Team Member’s Name, Email and Contribution:** |
| --- |
| 1. Anjali Tidke ( [anjalitidke123@gmail.com](mailto:anjalitidke123@gmail.com) )  * Importing Libraries * Loading Dataset * Data Analysis using EDA * Model Development * Extra Tree Regression * Random Forest Regression * XGBoost Regression * Gradient Boosting regressor * Model Selection * Conclusions  1. Shubham Dukare ([shubhamalmabetter@gmail.com](mailto:shubhamalmabetter@gmail.com))  * Importing Libraries * Loading datasets * Feature Engineering * Model Development * Lasso Regression * Ridge Regression * KNearestNeighbors Regressor * Gradient Boosting regressor * Model Selection * Observations * Conclusions |
| **Please paste the GitHub Repo link.** |
| Github Link(Shubham Dukare):-  <https://github.com/shubham4955/TED-Talks-VIews-Prediction>  Shubham Drive Link:-  <https://drive.google.com/drive/folders/11KIgA_l7Vt9GcgoMeYrOc1oN0N3yt-ax?usp=sharing>  Drive Link(Common):-  <https://drive.google.com/drive/folders/1VPc6PO2rxOJiPlNgPCzYoP_FUlHmn3C7?usp=sharing>  Github link(Anjali Tidke):-  <https://github.com/AnjaliTidke/TED-talk-Capstone-project>  Anjali Tidke Drive Link:-  <https://drive.google.com/drive/folders/1nVU78G12eCUq2zCClbIQxWO5Yy4DnMvv?usp=sharing> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**   | 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.  We have Started with data loading and we have done EDA, feature engineering, data cleaning, target encoding feature selection and then model building.  So, we have used this models:   * Ridge Regressor * Lasso Regressor * KNearestNeighbors Regressor * Gradient Boosting Regressor * Random Forest Regressor * Extra Tree Regressor * XGB Regressor   As we know, RMSE is more influenced by outliers MAE doesn't increase with outliers.  MAE is linear and RMSE is quadratically increasing.  So, We chosen MAE as a deciding factor for our model.  On the basis of MAE, the best performing regression model is Random Forest Regressor. | | --- | |