Capstone Project Submission

| **Team Member’s Name, Email and Contribution:** |
| --- |
| | **NAMES** | **E-MAIL** | **CONTRIBUTION** | | --- | --- | --- | | Sunil Kumar | 14bbt1019@gmail.com | Entire Project work. | |
| **Please paste the GitHub Repo link.** |
| **GitHub link:-**  https://github.com/Sunilkumar17-design/Regression\_Capstone.git |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |

**Summary of work**

**TED is devoted to spreading powerful ideas on just about any topic. The main objective is to build a predictive model, which could help in predicting the views of the videos uploaded on the TEDx website.**

**Started with data loading data and importing the libraries and then with exploring the data and looking into columns using visualization. I have seen a few columns for the null values and outliers as well.**

**I treated the data accordingly and since I have categorical data with me so I did target encoding and one hot ending on given columns. Since some values were varying and causing the heteroskedasticity. I have done standardization accordingly.**

**I have explored the data and looked for the trend in bi-variate analysis.**

**I have checked that views are correlated with the speakers as shown top 5 speakers with daily views exceeding 100000.**

**More talks session delivered is showing that more views and popularity also the duration is not much influenced when there are poplar speakers.**

**Finally, I started modelling and saw XGBoost, cat boost and random forest performs well. For score improvement, I have used regularization and Hyperparameter tunning.**

**I have faced many challenges during this project starting with looking for converting the categorical data into numerical and standardization also selecting the right amount of features.**

**I have made a few observations that keeping the right amount of features and right feature selection is very important and finally are responsible for your model betterment and good accuracy.**

**I have faced some challenges during converting the str type of column to the dictionary type.**