

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email, and Contribution:

Team Name: Self-Project

Member No.	Name	Email I'd	Contribution
1.	Surabhi Mali	surabhimali1299@gmail.com	Machine learning Model

Please paste the GitHub Repo and Google Drive link.

Github Link:-

<https://github.com/surabhimali/Ted-Talk-Views-ML-prediction>

Google Drive Link:-

https://drive.google.com/drive/folders/1u768-u5Wo4YiaaoZNwnrJ_4Ds_OXVvLL

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

TEDtalk Views Prediction

Summary:

TED is devoted to spreading powerful ideas on just about any topic. These datasets contain over 4,000 TED talks including transcripts in many languages. Founded in 1984 by Richard Salmen as a nonprofit organization that aimed at bringing experts from the fields of Technology, Entertainment, and Design together, TED Conferences have gone on to become the Mecca of ideas from virtually all walks of life. As of 2015, TED and its sister TEDx chapters have published more than 2000 talks for free consumption by the masses. The main objective is to build a predictive model, which could help in predicting the views of the videos uploaded on the TEDx website. In Machine Learning, we use various kinds of algorithms to allow machines to learn the relationships within the data provided and make predictions based on patterns or rules identified from the dataset. So, regression analysis is a machine learning technique where the model predicts the output as a continuous numerical value. Many models were trained, from simple parametric models like Linear Regression to tree based models. Then we evaluated these models using some parameters called R square (R^2) and RMSE. R^2 is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. Root Mean Squared Error (RMSE) is a common way of measuring the quality of the fit of the model.

Conclusion:

If we try comparing the prediction accuracy among different linear regression (LR) models then RMSE is a better option as it is simple to calculate and differentiable. And the number of predictor variables in a linear regression model is determined by adjusted R^2 . As we are more concerned about evaluating prediction accuracy among different LR models we can choose RMSE over adjusted R^2 . If we compare RMSE, Optimal Random Forest and as well as Extra Tree is performing well. But if we consider RMSE along with the adjusted R^2 , Optimal Random Forest is best performer.

