Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Kollana Harish, harishkollana@gmail.com :
1. Importing Libraries
2. Loading the dataset
3. Dataset Information
4. Data Analysis on Columns
5. Feature Engineering on Columns
6. Data Cleaning
7. Feature Selection
8. Fitting the models and Hyper Parameter Tuning
    1. Linear Regressor
    2. Random Forest Regressor
    3. XGB Regressor
9. Model Comparison
10. Model Selection
11. Conclusion

Please paste the GitHub Repo link.
Github Link:- https://github.com/harishkollana/Ted-Talks-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)

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 Salman 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 and its speaker list boasts of the likes of Al Gore, Jimmy Wales, Shahrukh Khan, and Bill Gates.

After Importing and libraries and Data Set, we had went on with Data Analysis on each and every column and tried to find out the relation between independent variables and dependent variables. On the second Step we had done Feature engineering on each and every column to find out the categorical and numerical analysis. Then after we had Data cleaning i.e., Null values treatment using KNN imputer, Outliner's treatment using IQR method and Transformation using Log.

Then We had continued with Feature selection using F Regressor and saved the important features to fit into model with the help of P values

After that we fit three models for training namely Linear Regressor, Random Forest Regressor and XGBOOST Regressor. For Hyper parameter tuning we had used Randomized Search cv and for accuracy we had used Mean absolute squared error, Mean squared error and R squared error.

The Conclusion Was Random Forest Regressor is the best model with 90% accuracy Mae taken into account as it doesn't affect by outliners.