

Movie Rating Prediction

NAME: Khushi Rajpurohit

Data set used: the data set is of Movies rating and was retrieved from Kaggle.

Data Preprocessing:

- `head()`, `shape()`, `info()`, `describe()`, `.duplicated()` and `.isnull().sum()` methods are used to get an overview of dataset, its structure and missing values.
- Null values were dropped from the dataset.
- Some columns were converted to integer data type

Exploratory Data Analysis:

- Plotted the distribution of all the Columns
- Counted numbers of movies made per year
- Counted the top 10 genres
- Plotted bar chart to show the top ten genres with highest ratings and votes.
- Plotted the top 10 Director with highest Ratings and Votes
- Plotted the distribution of Ratings

Insights from Data Analysis:

- Highest number of movies were made in year
- Top 2 genres are: *Drama and Romance*
- Top 2 Directors are: *Mahesh Bhatt and David Dhawan*
- Top 2 Directors with Highest ratings are: *Saif Ali Sayeed, Sriram Raja*
- Top 2 Directors with highest Votes are: *Rajkumar Hirani, Ang Lee*
- Top genres with Highest ratings are: *History , Romance , Documentary and Family*
- Top genres with highest Votes are: *Drama, Adventure and Sports*

Predictive Analysis:

- First I performed Feature Engineering.
- I used Linear Regression, Polynomial Regression and Ridge Regression for predictive analysis.
- The best out of three with highest accuracy is Ridge Regression with accuracy of 70%.