أكاديمية سدايا
SDAIA Academy

Vote

Predict of movie

Using Web Scraping and Regression

By :
Rana Alturki
Reem Binzeraiban

Objective

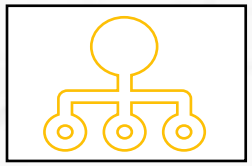
Web scraping provides an effective way to extract large amounts of data quickly for further analysis.

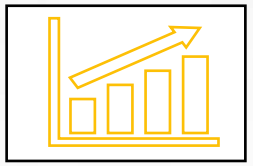
Data set 1000 Record and 10 columns.

The goal of this project is to takes data from IMDb website:

- Uses regression models to predict **vote** of movies.

Methodology





Gathering Data

Using Web Scraping

Exploratory Data Analysis

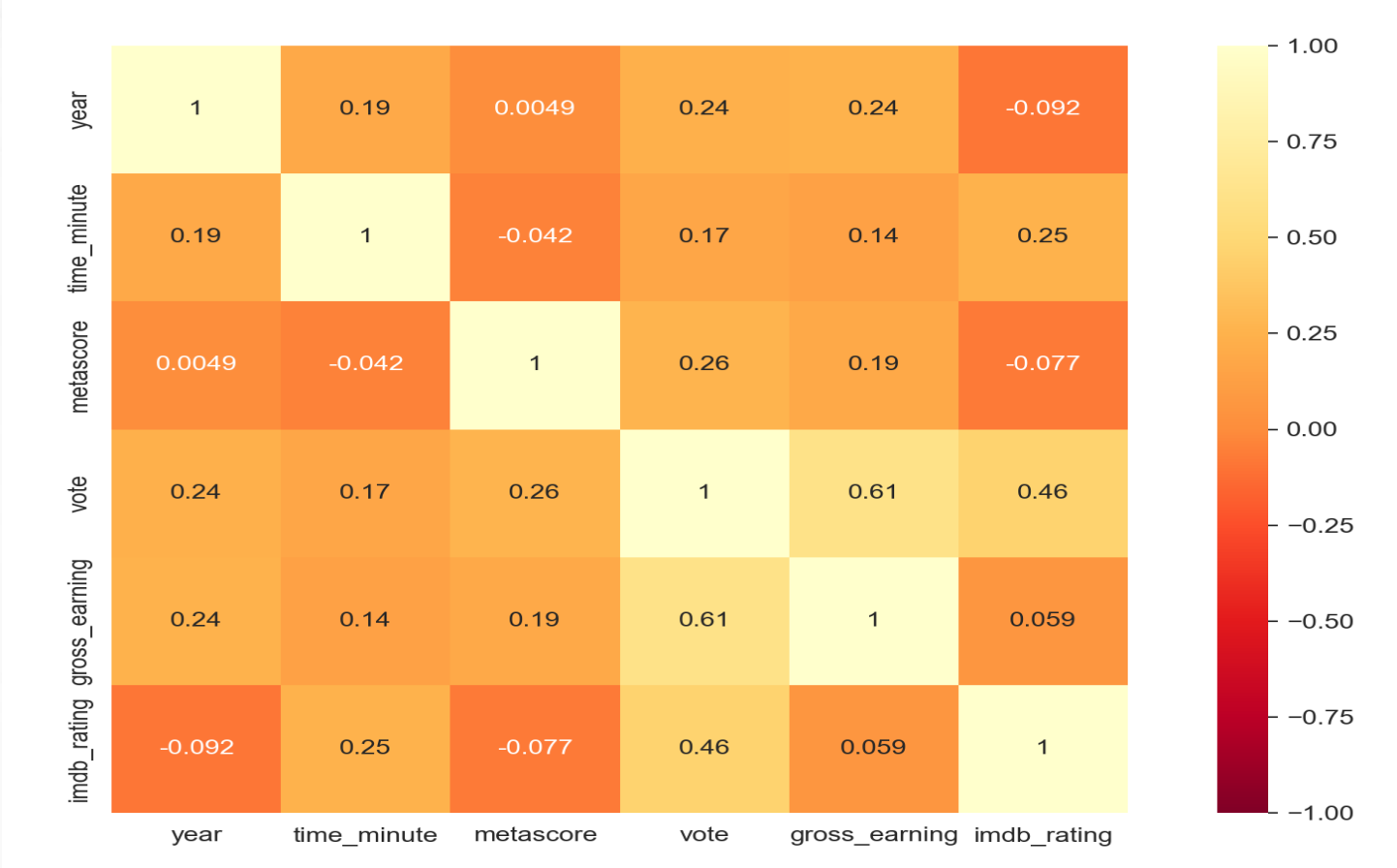
- Cleaning
- Analysis and visualizing

Data Preparation and Regression

Correlation

We noticed through the map that there is a strong relationship between the target(vote) and feature such as

- gross_earning (0.61)
- imdb_rating (0.46)

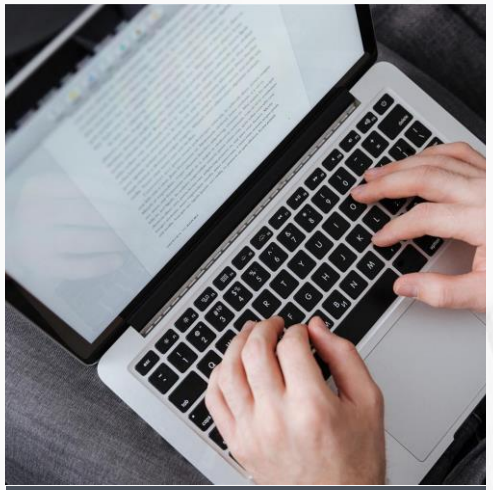


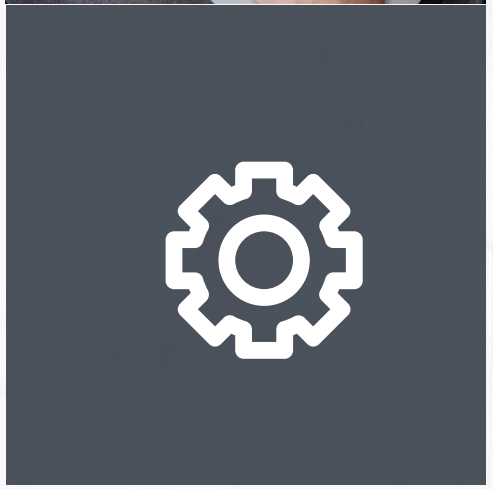
Data Preparation

- Feature Selection
- Splitting Data

60% Train	20% Validation	20% Test

- Feature Engineering
    - Dummy Variables
    - Add new columns
    - Impute zero value with mean
- Regression models

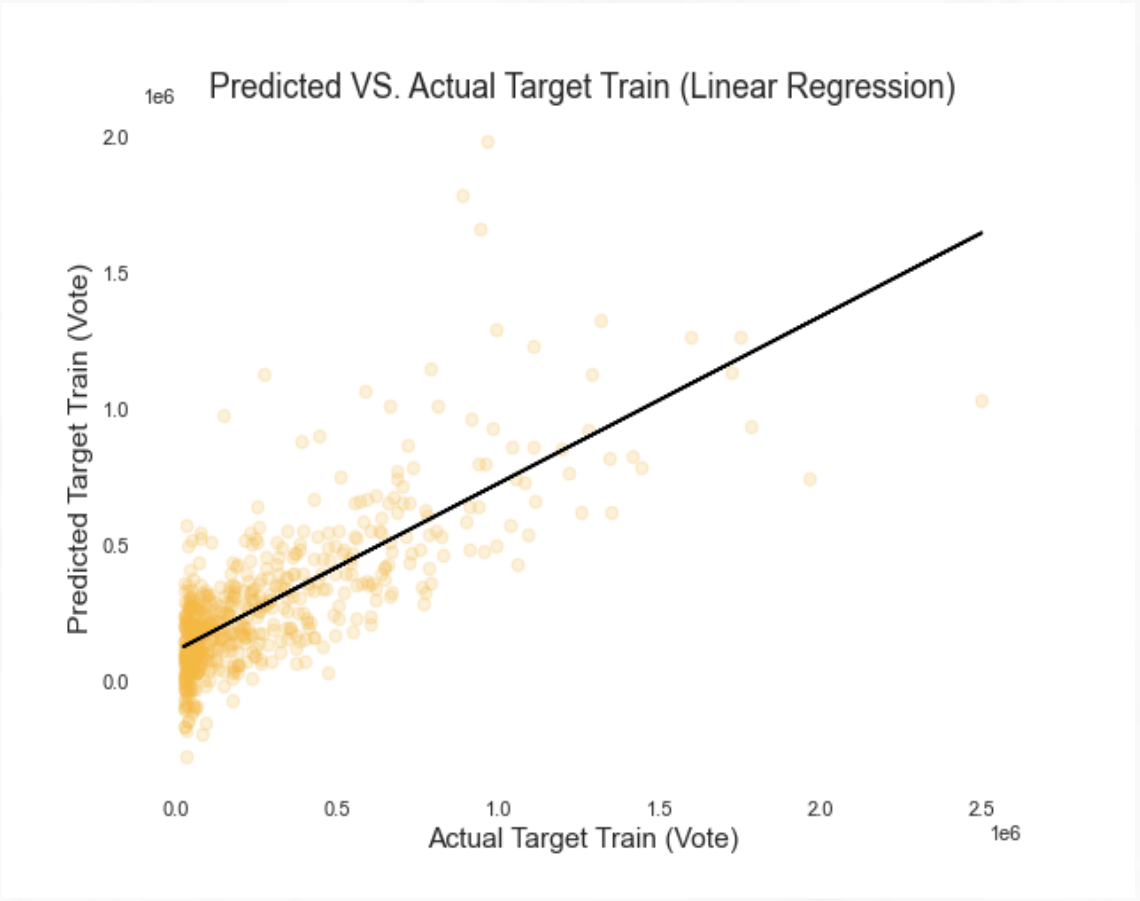




Analysis and Results

Algorithm	Train	Validation
Liner Regression	0.61	0.60
K-fold Liner Regression	0.61	0.59
Polynomial	**0.77**	**0.63**
Ridge Regression (alpha = 0.2)	0.61	0.60
Tuned Ridge Regression (alpha =1)	0.61	0.59
Lasso Regression Cross (alpha = 1)	0.61	0.59

Regression ML Models



Test score :0.69

Conclusion

Polynomial Regression shows the best prediction of “Vote”

Future Work：

- More data for Imdb
- Group Categorical more features.

Thank you!

We hope you enjoy it ! 😊