**Movie project introduction:**

Here we can predict how far the film would be succeeded with certain film features, so we chose the most important features, the most powerful pre-processing techniques and model to get a realistic accuracy.

**Pre-processing Techniques:**

* Dropping any missing data



* Choosing the columns of features that affect the output

Text

Description automatically generated

* Transfer string data values into numerical values

Text

Description automatically generated

A picture containing text

Description automatically generated

* Function feature encoding to return list of numeric values

Text

Description automatically generated

**Analyzing performance:**

Here we use the correlation to analyze the movie data set,

Correlation steps had done to study the affection of feature data on the output

**Graphical user interface

Description automatically generated with medium confidence**

**Used features:**

The effected features we chose “run time” and there are the effected features like (genres, rotten tomatoes, Netflix, year and age, Directors, Country, Language) we did not use name as it is not affected feature.

**Models:**

* Polynomial Regression

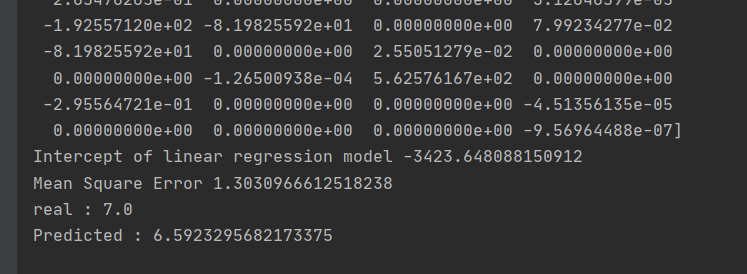
To convert the original features into their higher order terms

By using Polynomial Regression model with degree = 2

Graphical user interface, text

Description automatically generated

By increasing Polynomial Regression model to degree = 3



* Multi linear regression

considers the influence of multiple independent variables on a dependent variable, variance of MSA

Text

Description automatically generated

* Linear Regression

By using Linear Regression model

Text

Description automatically generated

**Date usage:**

We assign 30% of data for testing data,

And 70% of data for training data.

**Further techniques:**

we can improve the result by using the SVM technique or using the KNN technique.

A simple implementation of KNN regression is to calculate the average of the numerical target of the K nearest neighbors, so it is leading to best results.

**Conclusions:**

Here we need to normalize the output so it will make the MSE more normalized, to deal with data we face missing data so to rid of it we used function to dropping missing data and it has given better results, Taking features with correlation has given less MSE error, models are we use Linear Regression, Polynomial Regression and Multi linear regression.