Ahsanullah University of Science and Technology
Department of Computer Science and Engineering

CSE 4108: Artificial Intelligence Lab

Fall 2020

Project 01 Report

Codeforces Rating Prediction

Lab Section: B1

Submitted by

Alam Khan 170204084

Submitted To

Md. Siam Ansary	Ashek Seum
Department of CSE, AUST	Department of CSE, AUST

Date of Submission: 09 September, 2021

Contents

1 Problem Description

Codeforces is the best online judge for competitive programming. Competitive programmers do practice in this OJ. If a programmer does better in a programming contest then he will get more rating points. So it will be better if we can predict the rating of a contestant by using some characteristics of his/her.In my project, I used some features of a contestant to predict his/her rating. I used two models in my project to predict the rating points and showed necessary comparisons between them to know which one performed better.

2 Dataset Description

The dataset contains the information of the users of famous online judge Codeforces. It has 150 rows and 5 columns in total. There are 5 features in each row of the dataset. They are: Problems Solved, Followers, Contribution, Registered Years, Ratings. Here the feature Ratings is the targeted feature in our model. The other four features are independent. Generally, if a user has more ratings then he/she has a higher number of problems solved count, more followers, more contribution points. Besides, it is said that if a user practiced for more years then he/she can have higher ratings. I used 2/3 portion of the dataset to train my models and 1/3 portion of the dataset to test my models.

3 Description of the Models

In my project, I used to two models to predict the result of my project. They are:

3.1 Model 1: Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

3.2 Model 2: Decision Tree Regression

Decision Tree is one of the most commonly used, practical approaches for supervised learning. It can be used to solve both Regression and Classification tasks with the latter being put more into practical application. It is a tree-structured classifier with three types of nodes. The Root Node is the initial node which represents the entire sample and may get split further into further nodes. The Interior Nodes represent the features of a data set and the branches represent the decision rules. Finally, the Leaf Nodes represent the outcome. This algorithm is very useful for solving decision-related problems.

4 Result Table

Linear Regression and Decision Tree Regression, these two models are used to measure the results. To compare the results five performance metric scores are executed. They are Mean Absolute Error, Mean Absolute Percentage Error, Mean Squared Error, Root Mean Squared Error, R2 score.

Performance Metric	Linear Regression	Decision Tree Regression
Mean Absolute Error	200.49	319.18
Mean Absolute Percentage Error	0.12	0.17
Mean Squared Error	66167.05	235293.94
Root Mean Squared Error	257.23	485.07
R2 Score	0.83	0.41

From the result table, it is seen that Linear Regression Model has less error in Mean Absolute Error, Mean Absolute Percentage Error, Mean Squared Error, Root Mean Squared Error and more score in R2 Score..

5 Discussion

Now let's see a graph of Actual vs Predict of our Model 1, Linear Regression Model:

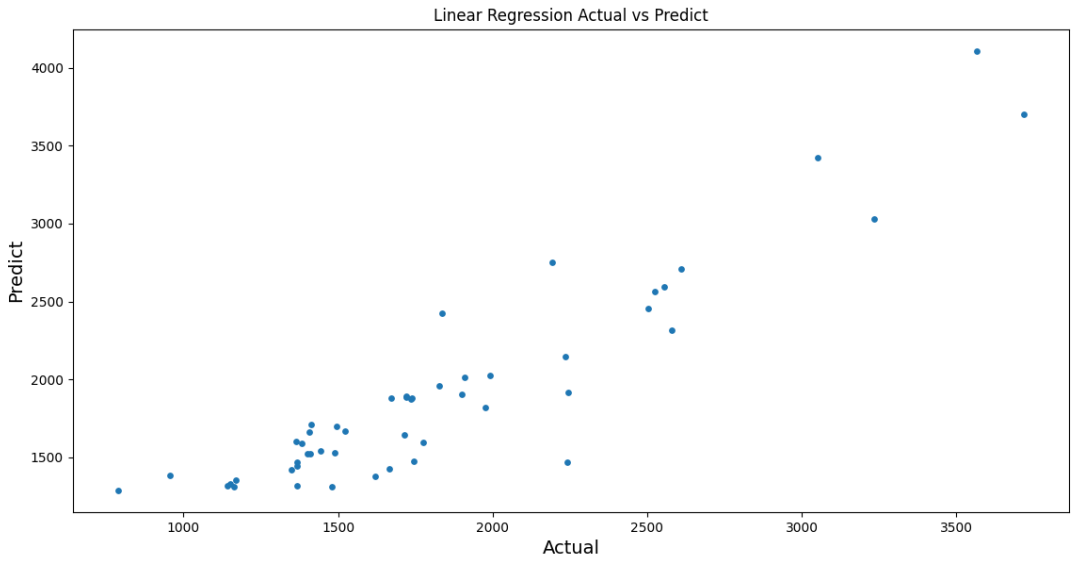

Figure 1: Actual vs Predict Graph of Model 1

A histogram of Actual vs Predict of our Linear Regression Model:

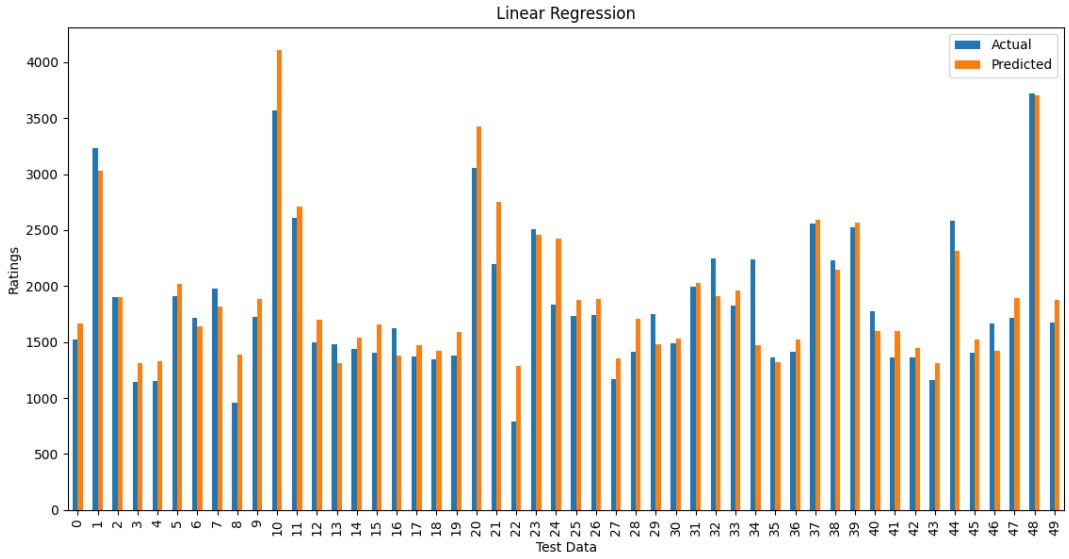

Figure 2: Histogram of Actual vs Predict

A graph of Actual vs Predict of our Model 2, Decision Tree Regression Model:

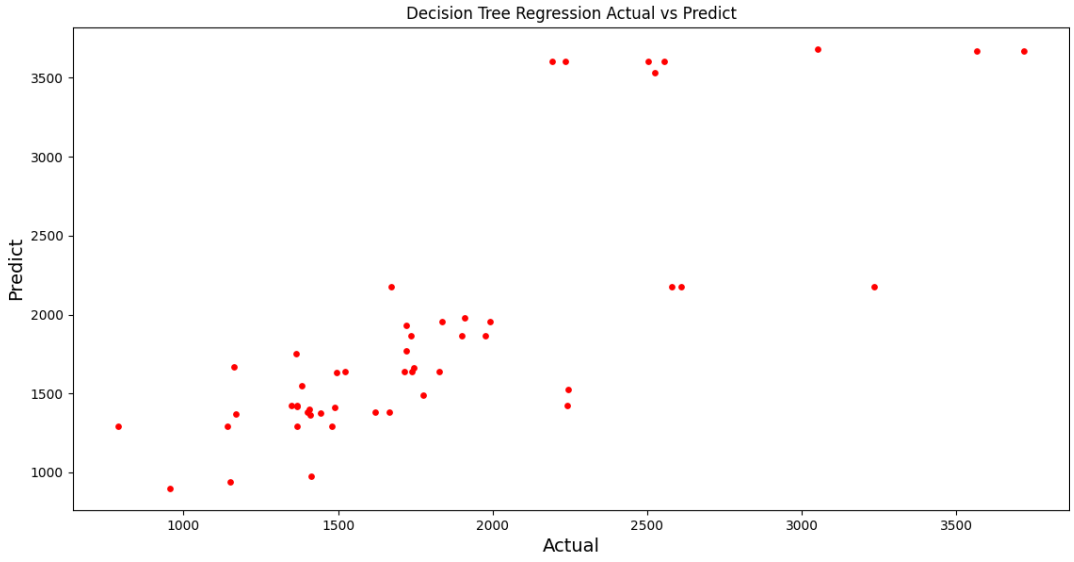

Figure 3: Actual vs Predict Graph of Model 2

A histogram of Actual vs Predict of our Decision Tree Regression Model:

Figure 4: Histogram of Actual vs Predict

Again, now let's see a comparison of our two models according to Actual vs Predict graph

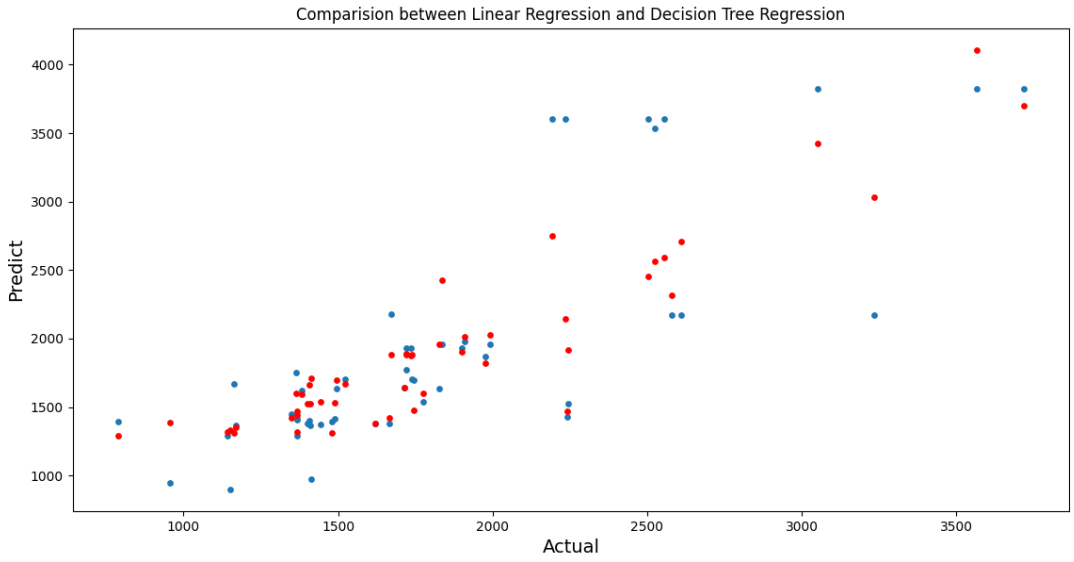

Figure 5: Comparision between Linear Regression Model and Decision Tree Regression Model

From the graphs, we can see that our model 1, Linear Regression Model performed better than our model 2, Decision Tree Regression Model.

From our result table, we got that Mean Absolute Error of our model 1 is 200.49 and model 2 is 319.18. So our model 1, Linear Regression Model performed better in this performance metric. Mean Absolute Percentage Error of our model 1 is 0.12 and model 2 is 0.17. Here is also our model 1 performed better. Mean Squared Error of our model 1 is 66167.05 and model 2 is 235293.94. Again, our model 1 performed better. Root Mean Squared Error of our model

1 is 257.23 and model 2 is 485.07. Our model 1 performed better in this performance metric too.

If we analyze the R2 score of our two models, we can see that our model 1 got 0.83 and model 2 got 0.41. R2 coefficient of determination is a statistical measure of how well the regression predictions approximate the real data points. So, our model 1 performed better in this performance metric too.

Finally, after analyzing all the graphs and result table values we can say that our model 1, Linear Regression Model performed better than our model 2, Decision Tree Regression Model.