Sherlyne Ndiwa 12331442
Visual Data Science
10th January 24

Assignment: Model Report

1. Describe your modeling approach

I decided to use two modelling algorithms: linear regression and random forest to compare which model would perform better on my dataset.
Linear regression is a simple and widely used regression model that assumes a linear relationship between the independent variables and the dependent variable. It is appropriate for predicting happiness scores as it considers the relationship between different factors such as social support, GDP per capita, and healthy life expectancy.

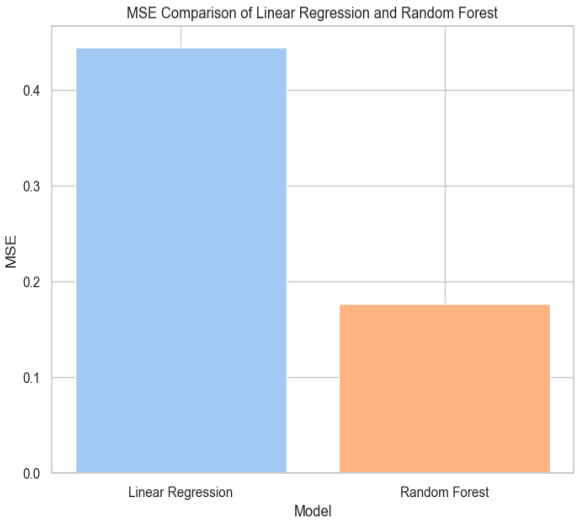
Random forest is an ensemble learning algorithm that combines multiple decision trees to make predictions. It is more robust to noise and outliers than linear regression, making it suitable for predicting happiness scores.

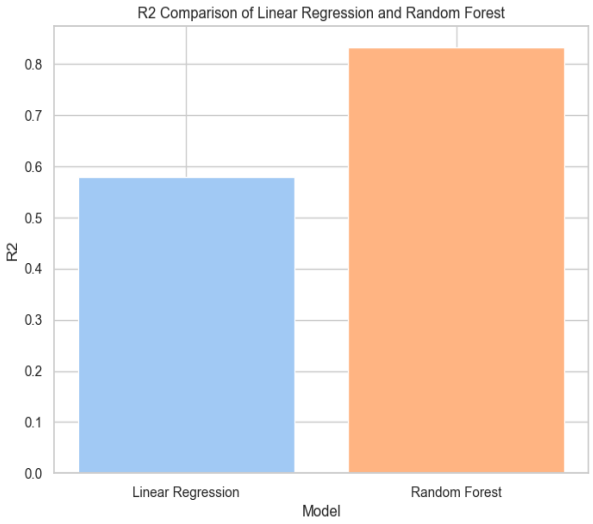
For assessing the performance of the chosen modeling algorithms, I selected Mean squared error and R2 score, and further used cross validation on the models to ensure that the results of the modeling are not biased by a split of the data. A lower MSE indicates better performance and a higher R2 score indicates better performance of the model.

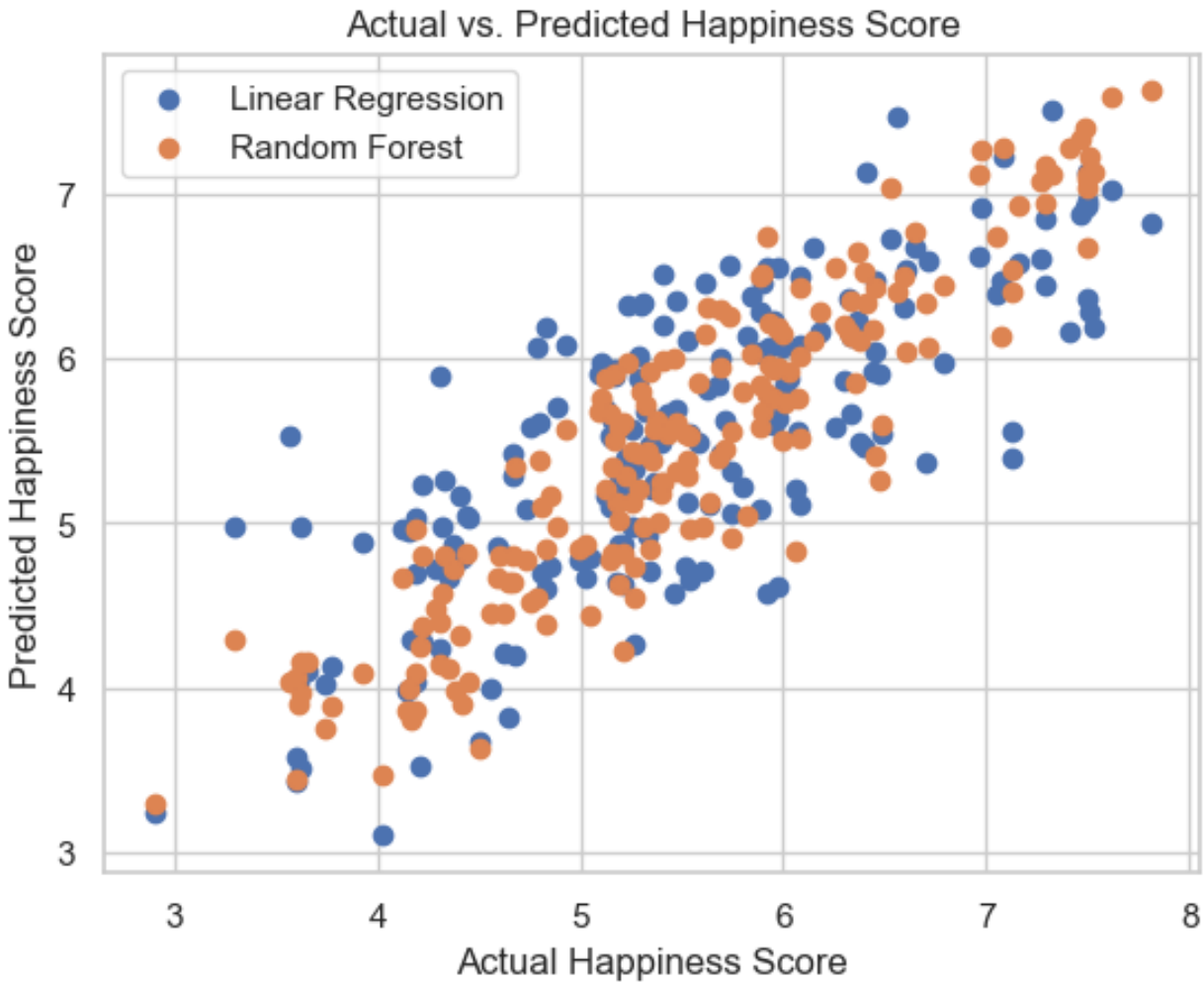
2. Describe the results of the modeling approach with a visualization.

```
Linear Regression CV Mean MSE: 0.7360955892187585
Random Forest CV Mean MSE: 0.45342896945302424
Linear Regression Test MSE: 0.4444752042726613
Linear Regression Test R2: 0.5800569191064995
Random Forest Test MSE: 0.17732486832827685
Random Forest Test R2: 0.8324623042883403
```

The results show that the random forest model has a lower MSE and a higher R2 score compared to the linear regression model, indicating that the random forest is a better model for this dataset in terms of generalizing and predicting performance.


MSE Comparison of Linear Regression and Random Forest
MSE
0.0
0.1
0.2
0.3
0.4
Linear Regression
Random Forest
Model


R2 Comparison of Linear Regression and Random Forest
R2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Linear Regression
Random Forest
Model


Actual vs. Predicted Happiness Score
Linear Regression
Random Forest
Predicted Happiness Score
3
4
5
6
7
3
4
5
6
7
8
Actual Happiness Score