Ankit Grover
189303158
CCE
C

Refer this link for more implementation details: REgreSHUN | Kaggle

Salary Prediction

In the following problem statement we have a dataset of Jobs and their various associated Levels, for example Level-1 for a Junior engineer,etc.
We can leverage the level position to predict the salary.
A sample from the dataset has been shown below.

| Business Analyst	1	45000
Junior Consultant	2	50000
Senior Consultant	3	60000
Manager	4	80000
Country Manager	5	110000

We use Linear, Logistic and Polynomial Regression. We want to reduce the RMSE, to have high accuracy predictions but making sure we do not overfit.

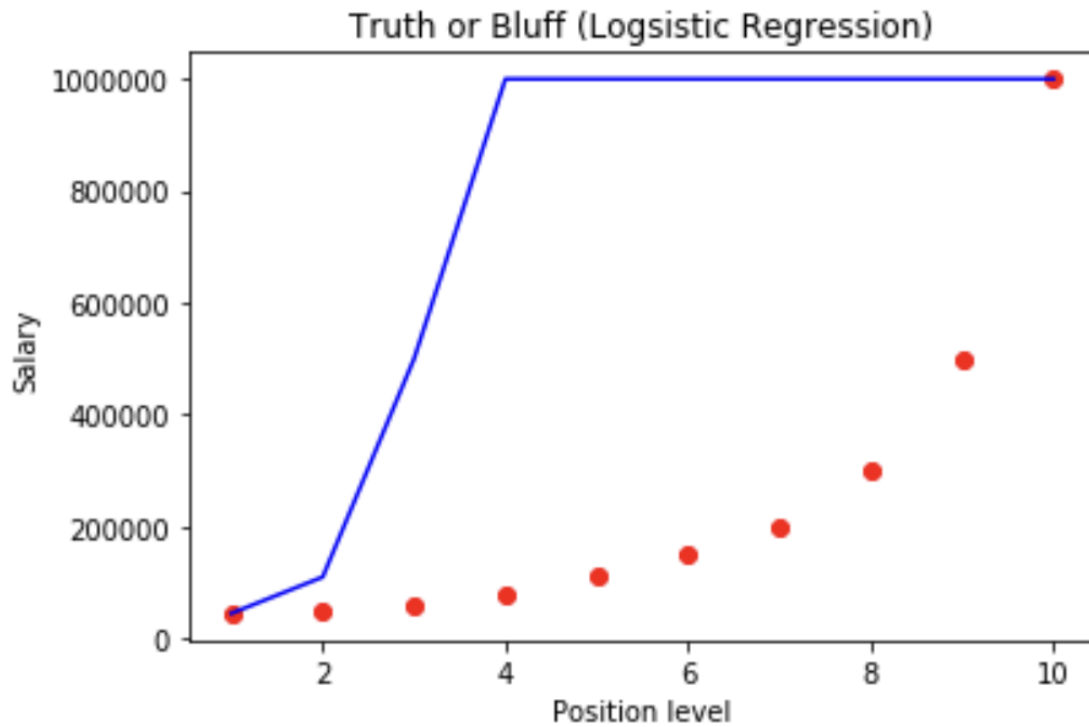Polynomial Regression of Degree 4 overfits.

Truth or Bluff (Polynomial Regression)
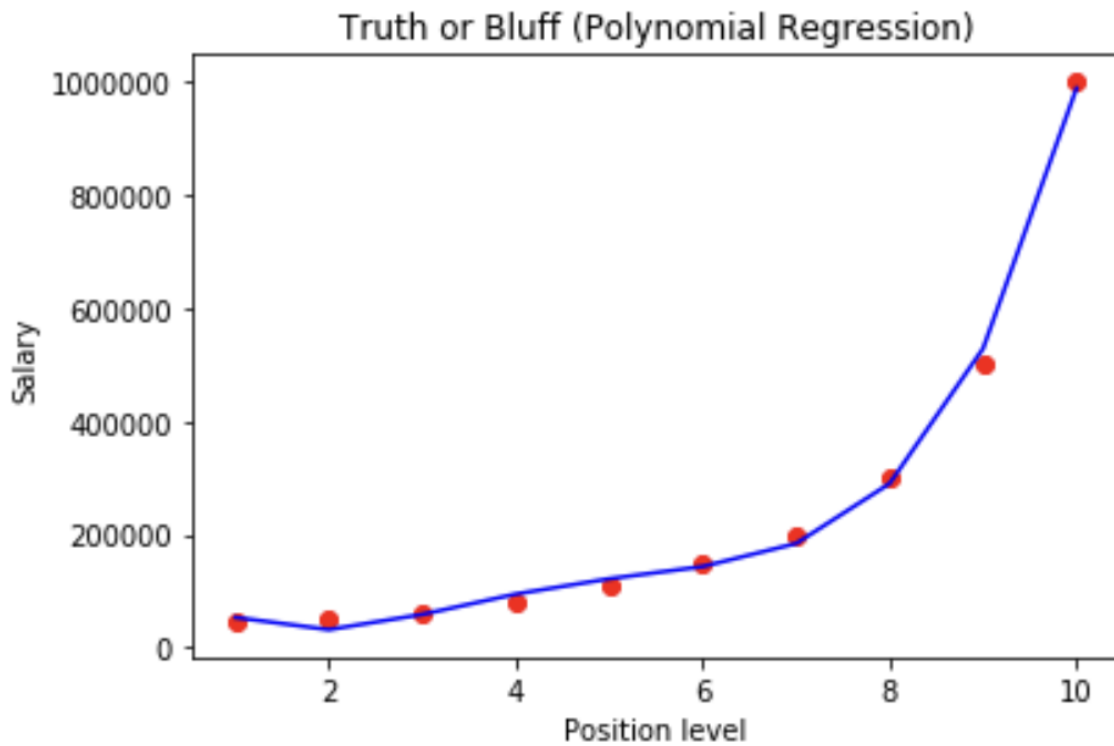
Polynomial Regression of Degree 3 doesn't overfit to such a great extent.

A Linear Regression model doesn't overfit as well but leads to huge RMSE.

Truth or Bluff (Polynomial Regression)

Truth or Bluff (Linear Regression)
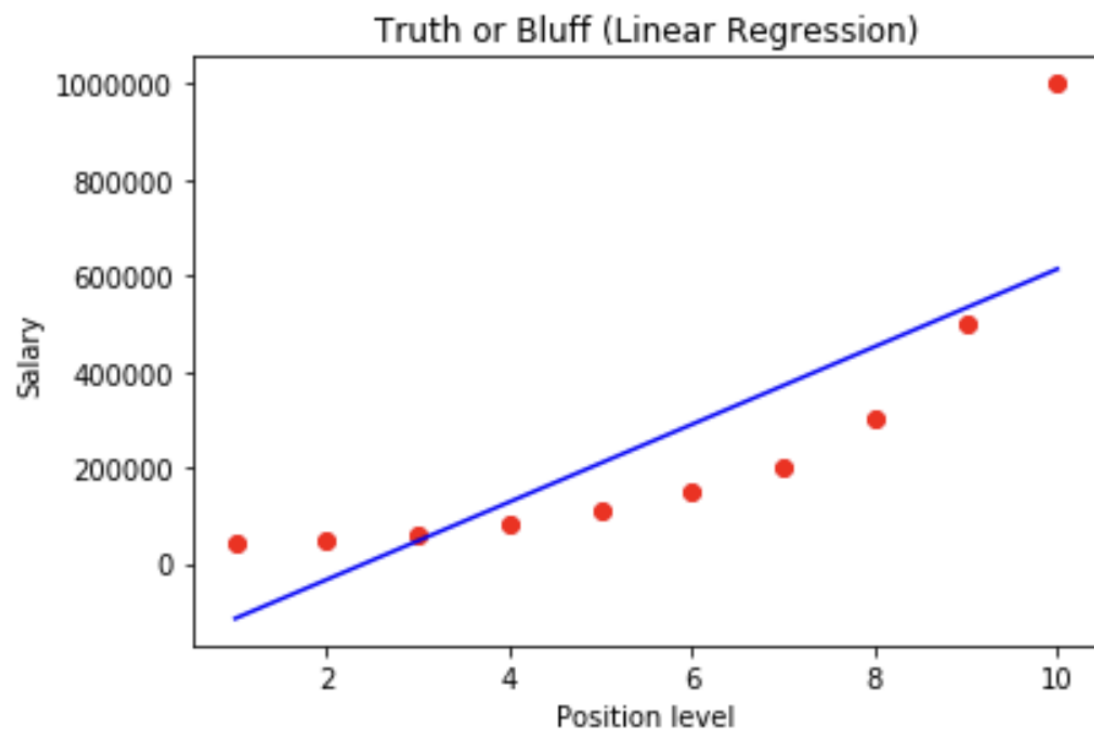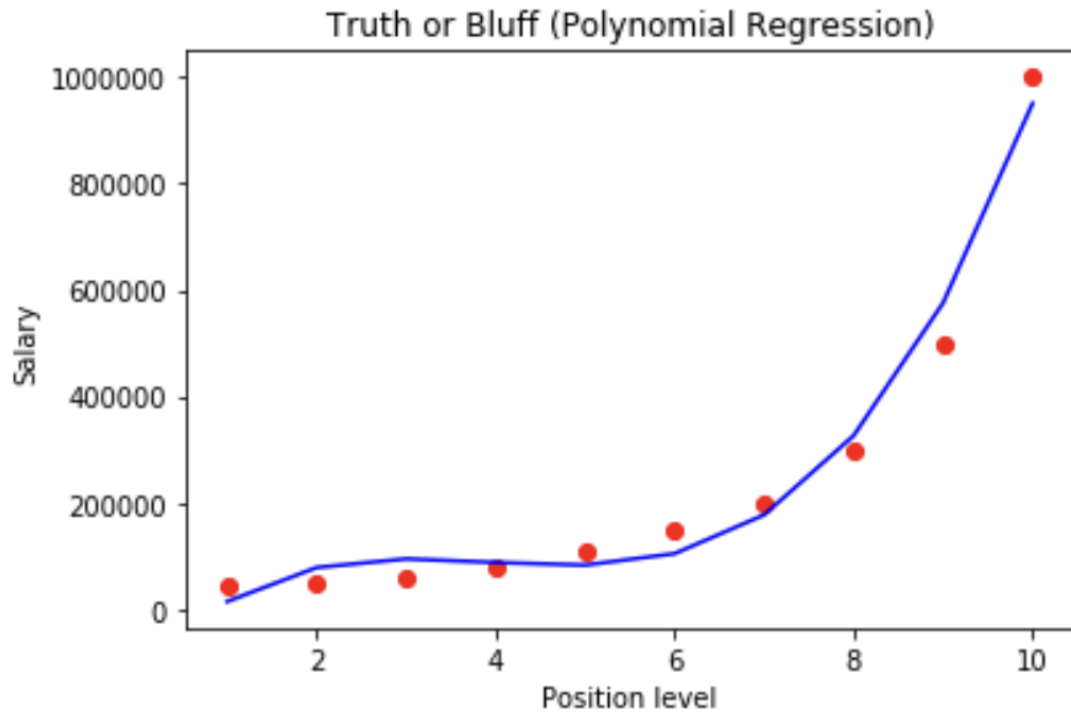
Logistic Regression again fails as we can see from the plot which resembles a logistic curve and not the behaviour of our given dataset which is polynomial in nature to an extent.

Truth or Bluff (Logsistic Regression)

Despite all of the above models used **we need more data to provide a conclusive answer as to which model to use in production, since the data accounted for is very less However, with the current data we can use Polynomial Regression of Degree 2.**