

**Student:** Xiaomin Wu

**Week project:** week 35, Predict the profit via machine learning

**Data preparation for the training:**

Dataset creation: import data to pandas dataframe, then take everything except last one as x\_values, and the last one as y\_values, set 80% of data as training set, rest of those as test set.

Scaling: I didn't realize if I scaling data or not in this case. I think I didn't.

Dummy variables: in this case, I didn't use dummy variables. But I learnt what's and how to use dummy variables. It's like when there is an independent variable is in string form to describe a feature, we transfer it into number 1 or 0, that means true or false. And we only take m-1 independent variables.

**Relevant metrics for the case:**

Mean Absolute Error is : 7320.441614848141

Mean Squarred Error is : 77506468.16885436

Root Mean Squarred Error is : 8803.775790469357

R2 score of model is : 0.9393955917820569

**Your conclusions of the results:**

The R2 score of model is 0.9394..., it means the accuracy of my model is 93.94%. In my opinion it's an acceptable margin of error. And there is a close variable in data source, I can get almost same value of the profit.

So it's good enough and the result with new data is usable.