Change of Money by using different model to bet

**Regression(normalization)**

svr income: 51.300000000000004

gbrt income: 28.700000000000003

**Regression(Without normalization)**

svr income: 0

gbrt income: -5.7

**Classification(Without normalization)**

lr Income: 3.3

nb Income: 5.300000000000001

svm Income: 0.5999999999999996

rf Income: -8.399999999999999

**Classification(Normalization)**

lr Income: 3.1999999999999993

nb Income: 5.300000000000001

svm Income: -34.6

rf Income: -8.8

Total, using all model :+ $40.19999999999999

Best: SVR with Regression(normalization), earning $51.3

Betting strategy:

The betting strategy in regression and classification is different, but they have something in common, I am trying to avoid risky and low win odds race. On the other hand, I look at the historical data of the jockeys and trainers, which is to avoid betting on the jockeys and trainers with bad average ranking.

For the regression part, I find out that the prediction of finishing time of the two model is quite difference. SVR has predict time between each horse with a larger gap, therefore this affect the strategy of betting using the SVR and GBRT model.

Moreover, the model without using the data without normalization is quite bad,

The difference of time between each prediction become very small, they are very close to each other, therefore it is hard to use those data to find out a more grantee move.

The normalization reduces the difference between features, which suit them into

[-1,1], This is very important to regression, as it need a similar parameter standard of the d dimension space, as to reduce the value of least square error.

For the classification part, normalization become less important, the accuracy between those models are quite similar, except for the SVM. The model without normalization is surprising winning more money. But the running time of that is like forever.

May be the normalization process reduce the importance of some important features. Because classification is mainly about the likelihood, just to find out a line to separate the class, therefore the difference between each feature are less important.

About the strategy, for the SVR-regression, I bet on the case which predict with the shortest finishing time, and the 4<=win odds <= 30, time between the first 2 horse must be >=40ms, the average ranking of the jockey and the trainer must be <=8.

For the GRNT-regression, it is similar but just with a smaller time gap between the top 3 horses, because the prediction data is more concrete in GRBT.

For the classification models, it is quite similar, except there is no condition about the time difference among the top horses.

The win odds grantee a high expected value of the bet, because I find out that is no wisely to bet on some horse with low win odds, the expected value of those is smaller than $1. Moreover, win odds is a quite important feature of the bet, almost all high win odds game loses. Which make it is not worthy to bet on those case, and so we avoid betting on them. Thirdly, the time between the first two horse, as the time between the first 2 horse increase, it means higher chance of our desire horse win, and so it rises the probability of winning. Finally, the average ranking of the trainers represents the skill of the trainers, and that of jockeys represent the quality of the living standard of horse, which reflect the condition of the horse.

Combining those factor, I choose to avoid the risky and low trade off game. Only buying game with high expected value, as to avoid losing.

All in all, my strategy is to avoid losing and chose the game with high trade back.