Start where you get and that

I understand that the client would want some of the predictors no t to be droped and it would worked around that

Why using f1 score instead of auc\_roc:

Base reason, because your data is inbalance, auc score takes averages for all tresholds and that wont work in inbalance data

Read <https://stackoverflow.com/questions/44172162/f1-score-vs-roc-auc>

Model building:

First with default parameters – see notes in pycharm

Next trying max debth, goal sacrifice training score for improvement in test:

[3, 5, 10]

Started with xgb as it it is faster and slightly better test core, in case I am not sucesfull in increasing, ill try tuning regulaariztion parmaeters for rf

Great, we have increased training socre to 96 and test score to 97

In rf, tried to increase speed and improve test score on the cost of training score,

In xgb tried improving overall score

Try using standard scaler, see if improves – no use for xgb and rf

For EDA, put that you droped all columns which information would not be available to loan officer (such as…s)

Say that that this project is for increase knowledge

SVC, knn taes a lot of time

The most important part from the modeling perspective is eda, in my opinion

We ended with 5 old var and one newly created, so in total we have 5 predictors and one depended binary variable. In analysis, I always took into account the variable definition and they probable impact oon the prediction

I approached project: fictional client is a bank and the model will be used by loan officer in decideding whether to grant loan to a fairly new business. Mentioned approach played a significant role in feature selection presented in the EDA.

I would break this project into 3 parts: EDA, training models, and deployment of the created model

EDA is the part where I’ve spent the most and the part which can always be improved. I will not go into details here, comments.

For the modeling part, I have used 2 algotithms. Decided on the f1 score due to skewness of the predictor. After I was satisfied with the prediction, it was time to deploy the mdel. I have decided for the Heroku (full name). Created flusk application and moved it to the platform.

I hope you will enjoy the material and for any sugestions or comment, please feel free to contact me on:

Cheers!

Vanja