Ruochen Qiu
10455875

MEMORANDUM

Highlights

- The clients who fall into the age bucket between has the highest risk to experience financial distress in the next two years.
- The clients who gain around $3000 per month has the highest risk to experience financial distress in the next two years.
- The clients who have over 5 dependents are more likely to experience financial distress in the next two years, while those families have 3-4 dependents.

Comparison

[auc-roc] In this case, Random Forest works best (0.70), while there is no much difference between Ada Boost, Gradient Boosting, Logistic Regression and Gaussian NB (all around 0.65).

[precision at 5] In this case, Extra Trees works best (0.99), while Random Forest also works fine (0.75).

[recall at 5] In this case, Decision Tree (0.74) and Extra Trees (0.73) works best.

[accuracy at 5] In this case, Extra Trees (0.98) and Random Forest (0.96) are the top two best classifier, while KNN(0.92), Ada Boost(0.93) and Gradient Boosting(0.93) also works well.

[F1 at 5] In this case, similar as under recall metric, Decision Tree (0.85) and Extra Trees (0.84) works best with relatively higher precision.

[precision at 10] In this case, Ada Boost (0.67) and Gradient Boosting(0.67) both works well but not in a relatively low precision.

[recall at 10] In this case, Extra Trees (1), Ada Boost (0.99) and Gradient Boosting (0.99) works best with strongly high precision.

[accuracy at 10] Similarly, in this case, Extra Trees (0.97), Ada Boost (0.97) and Gradient Boosting (0.97) works best with pretty high precision.

[F1 at 10] Still, Extra Trees (0.81), Ada Boost (0.80) and Gradient Boosting (0.80) works well with relatively low precision compare with ***accuracy*** and ***recall*** in this case.

[precision at 20] Extra Trees (0.34), Ada Boost (0.34) and Gradient Boosting (0.34) are the top three classifiers but the precision all pretty low in this case.

[recall at 20] In this case, Extra Trees (1), Ada Boost (1) and Gradient Boosting (1) works

extremely well.

[accuracy at 20] Still, in this case, Extra Trees (0.87), Ada Boost (0.87) and Gradient Boosting (0.87) are the top three classifiers but not works as well as under ***recall*** metric.

[F1 at 20] In this case, Extra Trees (0.51), Ada Boost (0.51) and Gradient Boosting (0.51) are the top three classifiers but not works as well as under other two metrics above.

Timing

Classifier	Run Time
RandomFores	0.140374184
DecisionTree	0.030362129
KNeighbors	1.997873068
ExtraTrees	0.176604033
AdaBoost	0.084233046
GradientBoosting	0.108603001
LogisticRegression	0.349344969
GaussianNB	0.033869028

From the result above, we can see that Decision Tree, GaussianNB and AdaBoost are the top three fast classifiers.

Recommendation

In sum up, by considering the time and precision under different parameters and metrics, AdaBoost should be the best option for the credit model. Gradient Boosting and Extra Trees can also be considered if the timing is not priority concern.