Eva Lipska

C5-T4

**Customer Default Identification Report**

**Problem**

Credit One is at the risk of losing business caused by an increase in the number of customers who have defaulted on loans.

The use of predictive analytics models may reduce the risk of loan default and credit fraud.

**The goal of the analysis**

Developing a model that can be used to predict the risk of customer default, in particular:

* How does the probability of default payment vary by categories of different demographic variables?
* Which variables are the strongest predictors of default payment?

**Analysis**

The analysis revealed patterns based on demographical data: gender, education, marital status, and age.

24 % of men and 20 % of women defaults. Women are the majority – 60 % – of the customers. 26 % of divorced customers default, however, they are the minority (1%) of all the customers. Married customers default slightly more (23%) than singles (21%). Customers ages 30 to 40 are much less likely to default payment (20%) than older groups and customers ages 20 to 30 (22 % of the defaults).

Unfortunately, the judgment about rules based on these patterns is moot, which was confirmed by the predictive modeling. Four predictive models were created: Random Forest, K-Nearest Neighbors, Support Vector Machine and Linear Regression, and all of them met the same problem, which should be a subject to further analysis.

Even though initial model evaluation in all the cases indicates quite good results (accuracy of 77-80 %), further analysis of the evaluation metrics (Kappa, Sensitivity, Specificity, Precision) revealed important drawbacks. All of the models, even though correctly predicted non-default status, failed to predict default status. Their prediction is not better than the one that could be achieved by a dumb model always predicting the most frequent class, in this case – 0. The dumb model would be right 78 % of the time, which means that our models don’t differ from a dumb model.

High accuracy of default status prediction is a business objective, and cannot be compromised.

**Summary**

Even though the analysis provides useful insights, that should be included in the business strategy, at this point prediction of the default status of customers given their demographical and payment history measurements cannot be based on the existing models.