**DSA/ISE-5103**

**Project Report (Draft)**

*Daniel Casares and Felipe Olivera*

*Due December 4, 2017*

**Detecting fraud on credit card transactions**

The University of Oklahoma

# Executive Summary (1 page)

**–** Concise problem statement

**–** List of major concerns/assumptions (if any)

**–** Summary of findings

**–** Recommendations

# Problem description and background (including related literature for problem, solution techniques, etc.)

The use of plastic cards (i.e. credit and debit cards) as a payment method has grown signiﬁcantly over past years, unfortunately so has fraud (Bahnson 134). Plastic card fraud is defined as an unauthorized account activity committed by means of the debit and credit facilities of a legitimate account. Some successful fraud tactics observed in the industry are lost and stolen card fraud, counterfeit card fraud, card not present fraud, mail non-receipt card fraud, account takeover fraud and application fraud (Krivko 6070). Based on the latest figures gathered in 2015, card fraud accumulated $21.84 Billion worldwide in losses (The Nilson Report 6). When banks lose money due to credit card fraud, the losses partially are passed to customers through higher interest rates, higher membership fees and reduced benefits. Hence, it is both the banks’ and cardholders’ interest to reduce illegitimate use of credit cards (Maes 2).

In this work, we consider the problem of identifying whether a credit or debit card account has been subject to fraudulent activity, using real-life transaction data from a Latin American credit card processing company. The goal is to construct a supervised learning model that can detect fraud on new (previously unseen) plastic card transactions. Fraud detection is, given a set of credit card transactions, the process of identifying those transactions that are fraudulent. Thus, the transactions are classified as genuine or as fraudulent transactions (Maes 2). Different detection systems that are based on machine learning techniques have been successfully used for this problem, in particular: neural networks, bayesian learning, artiﬁcial immune systems, association rules, hybrid models, support vector machines, peer group analysis, decision tree techniques such as ID3, C4.5, and random forest, discriminant analysis, social network analysis and logistic regression (Bahnson 135, Mahmoudi 2510).

# Exploratory data analysis (the highlights; not the kitchen sink)

Undersampling (Felipe)

EDA (Dani)

Dataset description

# Analysis plan

Explanation of modeling choice – Why choose this technique? Strengths, weaknesses? Dani

We investigated the performance of three techniques in predicting fraud: Logistic Regression (LR), Support Vector Machines (SVM), and Random Forest (RF).

Feature selection, engineering, missing value, outlier plan (felipe)

Validation plan (including how do your findings compare with others?) Dani

# Results and validation of analysis

# Conclusion

# References

Bahnsen, Alejandro Correa, et al. “Feature engineering strategies for credit card fraud detection.” *Expert Systems with Applications*, vol. 51, 2016, pp. 134–142.

Krivko ,M. “A hybrid model for plastic card fraud detection systems.” *Expert Systems with Applications*, vol. 37, 2010, pp. 6070–6076.

Maes, Sam et al. “Credit Card Fraud Detection Using Bayesian and Neural Networks”. *Proceedings of NF*, 2002.

Mahmoudi, Nader, et al. “Detecting credit card fraud by Modiﬁed Fisher Discriminant Analysis.” *Expert Systems with Applications,* vol. 42, 2015, pp. 2510–2516.

“The Nilson Report.” David Robertson, 17 Oct. 2016,

www.nilsonreport.com/upload/content\_promo/The\_Nilson\_Report\_10-17-2016.pdf. Accessed 02 Dec. 2017.

# Appendix

**–** Data visualizations, tables, etc. which support the work, but are not of primary importance

**–** List of data transformations, missing value imputations, outlier treatment, etc.

**–** List of any important assumptions not otherwise included

**–** Important code excerpts or algorithms used / developed if any.