

# PREDICTING SANTANDER CUSTOMER SATISFACTION

*Geoffrey Loss, Serena Simkus, Srikar Varadaraj, and George Yu*

## ABSTRACT:

Customer satisfaction is an important metric for businesses to keep track of in order to retain customers. Unfortunately, it is often hard to predict until it is too late and the customer has already taken their business somewhere else. We will attempt to predict satisfaction for Santander customers from a very sparse dataset with several hundred features. A precursory glance at the dataset suggests at its sparsity, and thus we will be examining the performance of lasso regression on the dataset. Given the large amount of data features, we will also be using PCA to try and identify the most relevant features in the dataset. As a baseline, ordinary least squares and ridge regression will also be implemented to compare to our lasso regression. We will also apply fused lasso and compare our results to the original lasso technique. We believe that sparse techniques will perform well relative to other approaches due to the peculiar nature of the datasets. We will be able to verify our hypothesis using testing datasets provided online. If time permits, we would like to implement and compare the performance of several classical algorithms such as the SVM with various kernels and multilayer perceptron to our regression performance.

## REFERENCES:

Tibshirani, Robert, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. "Sparsity and smoothness via the fused lasso." *Journal of the Royal Statistical Society Series B* 67.1 (2005): 91-108. Web.