Santander Customer Satisfaction **Team Uncertainty **: - Suraj Kumar Mondal, Saurabh Shetty, Vishal Kumar and Nazim Saifi 

**Motivation and About the Project**

Customer satisfaction is the key idea for an Organization to succeed.

When your customers are satisfied, they believe in the brand and become loyal. This loyalty increases sales and profitability.

The motivation comes from here that it’s a Goal for every organization and the “Santander Customer Satisfaction” is one of best challenging problem for our research.

Here we have anonymized dataset containing a large number of numeric variables. The "TARGET" column is the variable to predict. It equals one for unsatisfied customers and 0 for satisfied customers. The task is to predict the probability that each customer in the test set is an unsatisfied customer.

**Data and Labels**

1. Here the dataset has anonymous

column with numeric data

2. Data set contains many unnamed columns

3. Unnamed columns make feature

extraction tedious

4. Exploratory data analysis on the

training set can reveal latent features which contribute to the Target column

**References**

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

**Model**

We build a machine learning algorithm using the training data set and predict the total satisfied and unsatisfied customers in the Santander test data.

For initial set of experiment we have used Logistic Regression as a base model.

Here “Random Forrest” and “XG Boost” is used to build a machine learning model.

**Results**

**Conclusion and Future Work**

Accuracy score is not the best score to evaluate the performance of a model.

For imbalanced dataset ROC-AUC score and F1 score are better metrics to evaluate the model.

From these metrics, we find that XGBoost Classifier is the best model followed by Random Forest and Logistic Regression trained on the Santander Dataset.

For future work

1. Other models can be trained and ensembled in different combinations with each other to get better results.

2. We can use Bayesian optimization to find the best hyper-parameters for each of the model.

3. We can also try implementing deep learning models to the problem.