

# **ML PROJECT: SANTANDER VALUE PREDICTION**

**Yilin Liu N18292931**

**Ning Liu N15207037**

## **The problem we are trying to solve.**

We choose this problem from Kaggle, the name of which is Santander Value Prediction Challenge. In their website, they said Santander Group aims to go a step beyond recognizing that there is a need to provide a customer a financial service and intends to determine the amount or value of the customer's transaction. This means anticipating customer needs in a more concrete, but also simple and personal way. So our goal is to help them identify the value of transactions for each potential customer.

## **Our dataset**

We download data from kaggle's website. There are train data and test data. Since our dataset is too big to upload, we didn't upload them to the github. If you want to check our raw data, you can download them from this link: <https://www.kaggle.com/c/santander-value-prediction-challenge/data>.

## **The knowledge that we used.**

Actually, we used so many machine learning knowledges that we learned in this semester, including K-means, Principle Components Analysis(PCA), etc.. Besides, we also use a new algorithm which is LightGBM to train our model and predict the target data.

## **What you will see in our project**

There will be training result, Graphic Analysis and the plot of PCA and variance. For the details, you can check our code in jupyter notebook.