IFT6390
Final Project
14th of November 2018

Team: Cisse Cheikh Tidiane 20092990
Mohamed Chibane 925760
Fall Cheikh Ahmed Tidiane 20088392
Skirej Selma 20113718
Hack Yacoubou Soka 20125793

Abstract: We would like to see how classifiers behave on a set of 2 very different financial problems. The first one applies to trading stocks daily using technical and economic indicators as tools for determining order placement. We would like to see how classifiers fare on predicting daily price direction on a ETF tracking the S&P 500 index (NYSE: SPY). The second financial problem we seek answers to is, from a perspective of achieving an efficient business expansion strategy, considering the characteristics (relevant features) of the client base and records of previous marketing campaigns, will a new client be likely to subscribe to a term deposit at this banking institution. To some extent this might be useful to implement new products and assess the receptiveness of the client base and better partition it.

Chosen classifiers:

1. **Logistic regression/Multinomial Logistic Regression:** The outputs will have a probabilistic interpretation.
2. **Decision tree/Random Forest:** It implicitly determines which features are most important, will be easy to interpret and any nonlinear relationship between the features will not affect the performance of the tree.
3. **Multi Layer Neural Network:** A highly performing algorithm which can possibly be our most performant.
4. **SVM soft margin:** Will allow the modeling of non-linear decision boundaries.

Chosen datasets:

1. S&P 500 daily price action from January 1, 2000 until the November 9, 2018 (day open, day high, day low, day close, day volume, %up/down). From those 6 price and volume states and a few economic indicators we will relevant features such as a 15-day simple moving average, on-balance volume, current interest rate for example to predict whether we will place a buy order, sell order or no trade.

2. Banking term deposit subscription dataset. It holds features arranged in categories such as bank client data, last contact answers for the current campaign, social and economic context attributes and responses of clients concerning submission to a term deposit (output y = yes or no). Based on the records we gathered about the client base, we want to detect likelihood of a positive response of a new prospect client or a new product.