What use of AI in finance?

ISCF Intermediate Presentation

Denis DEMKO Thomas DEROO Doris FEJZA
Elona KARAJ Hussein LEZZAIK Estia MALIQARI Yijue XIE
Supervisor: Dr. Dritan NACE

[1]University of Technology of Compiègne

November 24th 2020

Table of Contents

Table of Contents

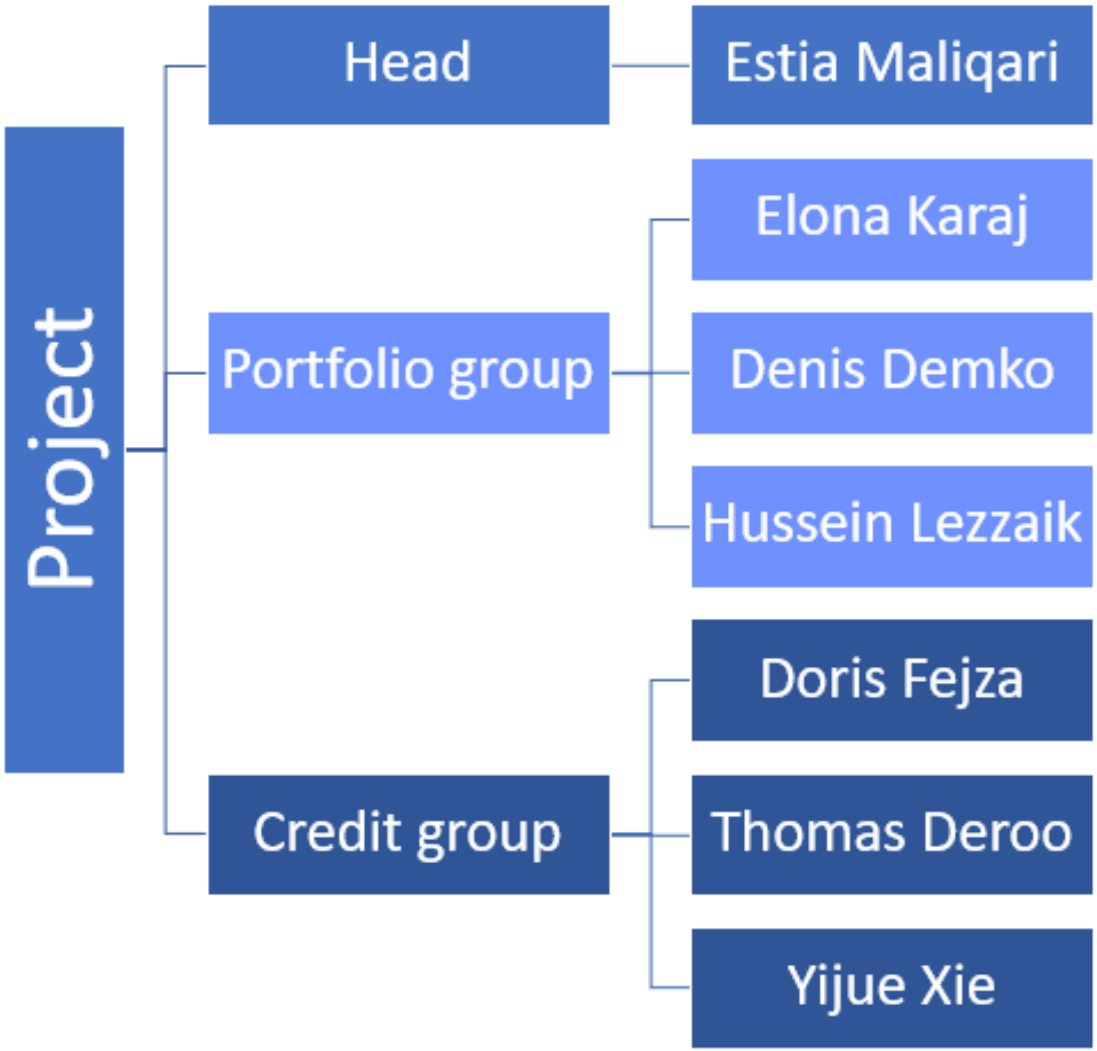
Objectives

Provide a study on the various applications of Machine Learning in finance

Study the portfolio management problem and the application of the SVM method

Study the tool developed for the application of credit risk assessment

Organization



Machine Learning applications in finance

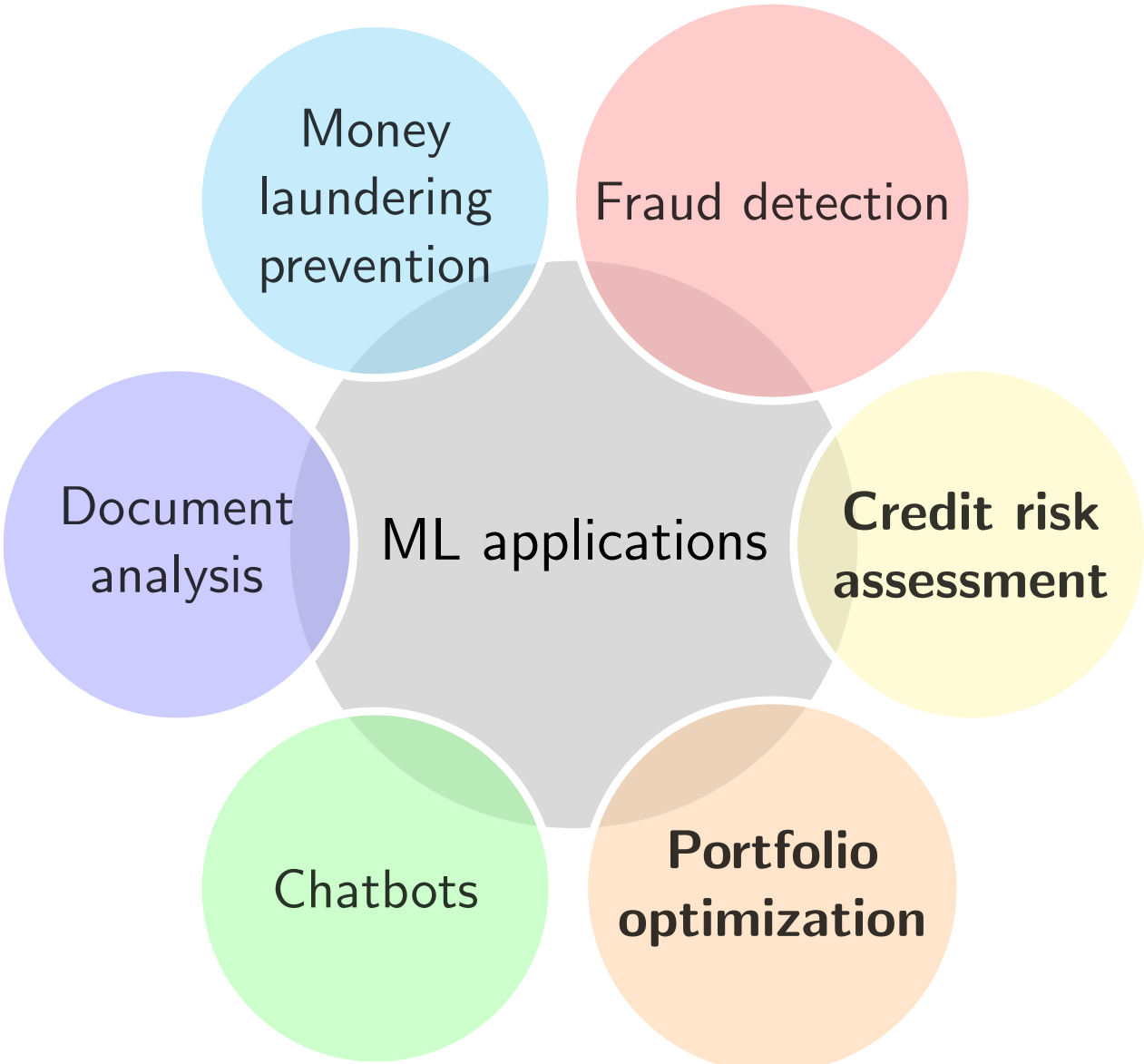


Table of Contents

Context of the project

Predicting stock market changes

2 main theories : Fundamental analysis vs Technical Analysis

Support Vector Machines(SVM)

SVM is the machine learning algorithm used to classify predictions

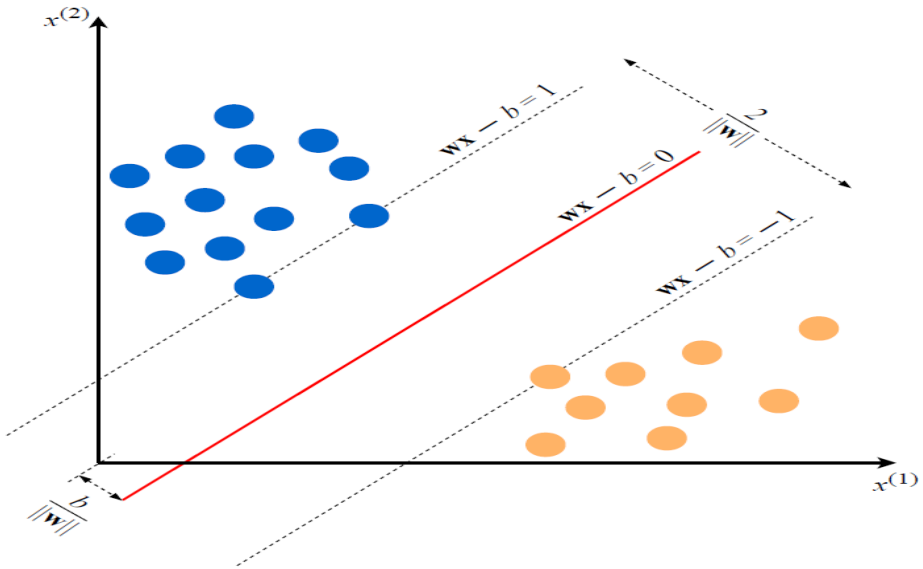


Figure: SVM in 2 dimensions

Financial Indicators

Key Features for Market Prediction

- Price
- Moving Average (MA)
- Adaptive Moving Average (AMA)
- Bollinger Band
- Average True Range (ATR)
- Moving Average Convergence Divergence (MACD)
- Relative Strength Indicator (RSI)
- Stochastic Indicator

Dataset

Figure: CAC40 Dataset

Implementation

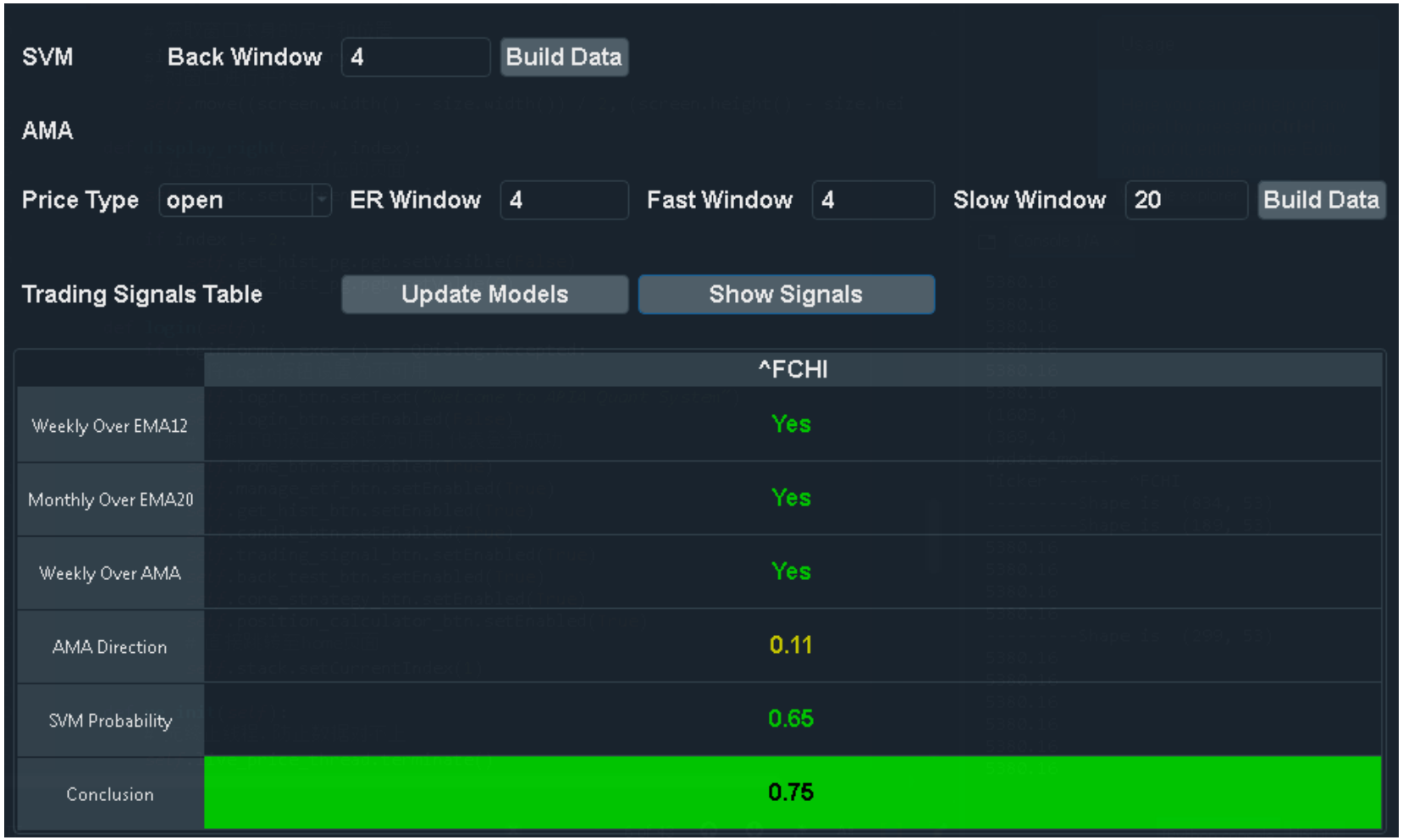


Figure: Trading Signal

Future Work

Add Deep Learning to the state of the art methods

- Develop Deep Neural Architecture for market prediction
- Train on same dataset CAC40
- Compare performance/accuracy with previous model

Table of Contents

Context of the project

- Problem studied
  Automation of the process of approving an overdraft
  - What is an overdraft?
    Overdraft is a type of revolving credit, granted from the bank to a client, typically when its account reaches zero
  - Difference between overdraft and loan
    Overdraft has a fixed duration of one or two years and a fixed fee
- Objective
  Find the most adequate method for the overdraft data based on the research work conducted on the loan data

Methodology

Data Analysis

Heavily imbalanced data, 725 bad clients out of 11397 samples in total

Data Cleaning

Remove inaccurate samples such as duplicated rows

Data split

Separate training, validation and testing data set prevent the result from over-fitting.

Feature engineering

Add multiply features to make the data set richer.

Feature engineering

- Time Characteristics
- Numerical Variables
  - Magnitude of the data is too large
  - Distribution does not conform to a normal distribution
- Combining Features
  - Perform linear operations on features
- Feature bucketing
  - Discretize continuous variables and merge discrete variables with multiple categorical values
- Feature selection

Model exploration

- Balanced Random Forest (BRF)
- Improved BRF
- Balanced AdaBoost
- Multi-model training with voting system
- Improved voting
- Improved BRF with splited training

$$\text{specificity} = \frac{\text{number of both actual bad and predicted as bad}}{\text{number of actual bad}}$$

$$\text{sensitivity} = \frac{\text{number of both actual good and predicted as good}}{\text{number of actual good}}$$

$$\text{Total accuracy} = \frac{\text{number of correct classification}}{\text{the number of evaluation sample}}$$

- By using 20 features selected, with the balanced random forest model, we've got a total AUC of 0.75.

Method	specificity	sensitivity	accuracy	AUC
BRF	0.785	0.707	0.712	0.746

Results

Model explainability

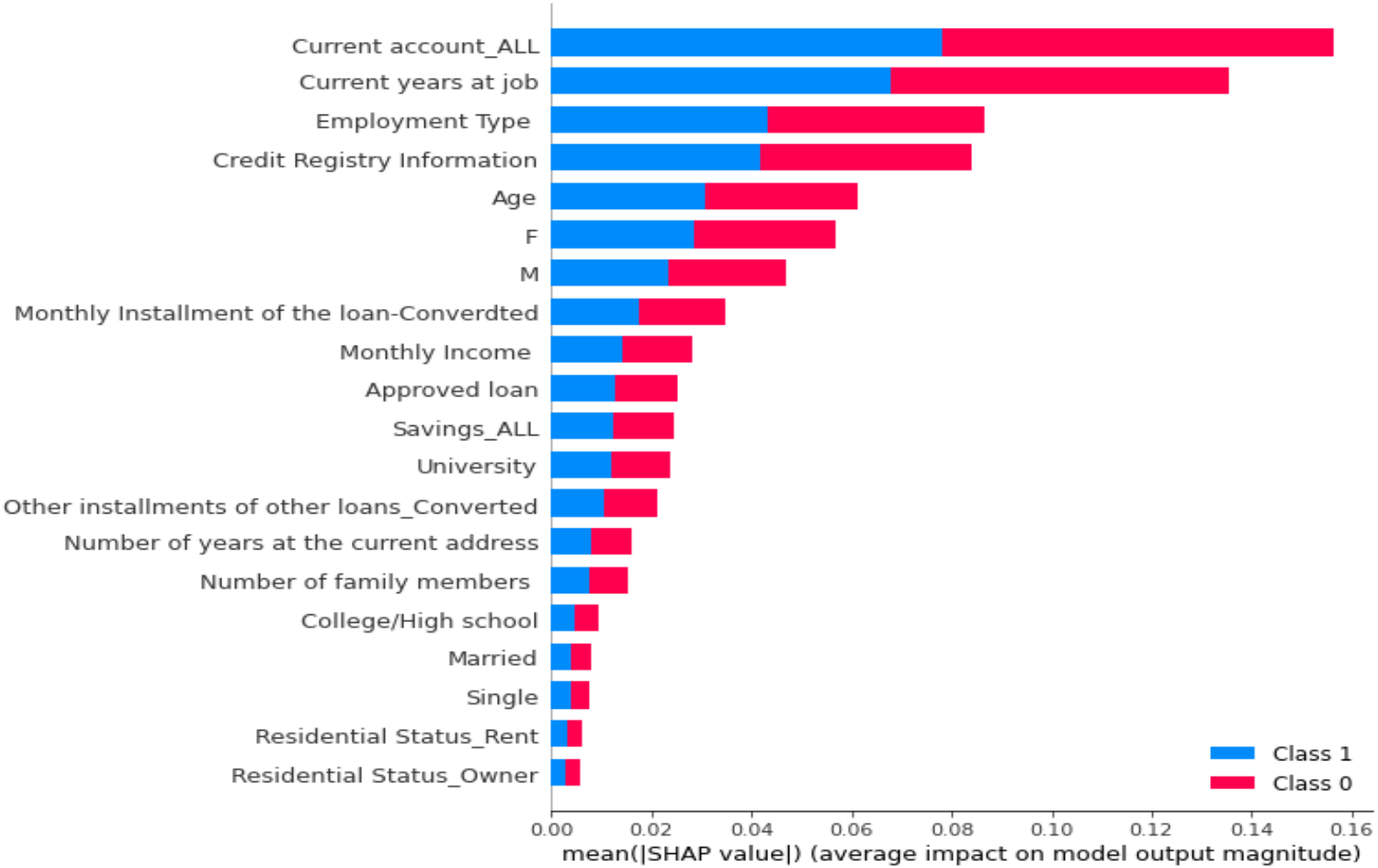


Figure: Global feature importance

Future works

Generalization of the tool

Generalize the tool developed for the application of credit risk assessment, so it can be used not only with the studied datasets of loan and overdraft, but with any type of standard customer data.

Table of Contents

Conclusions

- The study of ML applications helped to create an idea of how the ML and finance world are connected as well as how our two applications are placed.
- For the portfolio management project we studied the previous works and the necessary algorithms and concluded the implementations for the french stock market dataset.
- For the credit risk assessment project we analysed the loan data and after implementing various methods we concluded that the best method to use is Balanced Random Forest.