Data mining with python:
Automated FOREX trading

Kevin Voss Sjøbeck(s103451)
Benjamin Maksuti(s103449)
Anders Wessberg(s103477)

Abstract—**This project will utilize the oandapy API to get realtime data from the currency market to analyze, as well as enter and exit trades on based on the analytics. In the analytics we are going to use a short and a long moving averages, furthermore we are going to use MCAD as an extra precaution before we enter trades and use recovery zones to prevent losses on the account**

I. INTRODUCTION

Here will be written an introduction

II. ANALYSIS

A. Getting data

We obtain all the financial data of the financial instrument EUR vs USD on a 5 minute timeframe from oanda using their REST API for python. From oanda API we get the financial charts of the last 7 years from 2007-10-24 to 2014-10-24. After obtaining the data, it is stored in json fileformat in a file called "fxdata.txt", this allows us to use this data will then be used both to form our hypothosis of forex trading on the EUR vs USD currency pair, as well as performing a trading simulation on this set of data. However in order for us to use the naive Bayes classifier, we need to have two dataset's and not just a single one. In our domain the one dataset needs to be the trades which gained profit, and the other set needs consist of the trades which lost. When we then have the possibility to analyze the data with different sets of features.

The pseudo code of the random trading algorithm, where we will keep the orders for no more than 20 minutes:

while $streamFinancialData$ **do**
$\quad i = random(0, 10)$
$\quad$ **if** $i = 0$ **then**
$\quad\quad openOrder(short)$
$\quad$ **else if** $i = 1$ **then**
$\quad\quad openOrder(long)$
$\quad$ **end if**
$\quad$ **for** $order\ in\ orders$ **do**
$\quad\quad$ **if** $order.takeProfit <= order.value$ **then**
$\quad\quad\quad profitList.append(order)$
$\quad\quad\quad order.close()$
$\quad\quad$ **else if** $order.stopLoss >= order.value$ **then**
$\quad\quad\quad lossList.append(order)$
$\quad\quad\quad order.close()$
$\quad\quad$ **else if** $order.duration >= 4$ **then**
$\quad\quad\quad order.close()$
$\quad\quad$ **else**
$\quad\quad\quad order.duration+ = 1$
$\quad\quad$ **end if**
$\quad$ **end for**
end while
$saveListToFile(profitList, "profit.txt")$
$saveListToFile(lossList, "loss.txt")$

This algorithm should theoretically give us equally profitable trades in both the long and the short direction.

B. Specific features

To analyze the data we use a supervised classification learning algorithm, or to be more specific we use a naive Bayes classifier, the idea is to use a naive Bayes classifier on our financial data we gathered using our random algorithm. This isn't a new concept, in fact there are a couple of scientific articles on this already, as can be seen in the article by KAWABATA and TAKATA. We will train the classifier on some specific features, however we will only use 2-3 features per training round. The training will be made in two rounds and the features are the following

1)
  - SMA(period=10)
  - SMA(period=40)
  - RSI(period=100)
2)
  - openBid(period=1)
  - closeBid(period=1)
  - volume(period=1)
  - spread(period=1)
3)
  - Parabolic SAR default settings (0.02, 0.2)
  - MACD default settings (fast=12, slow = 15, period=9)

The first round of features, is using some of the most common technical indicators on the FX market, this should give a good indication if those are just famous for being simple or if you actually are able to earn profit using these indicators. The second round another of the more simple techniques that some of the professional FX traders should use according to chat on the different trading forums, this isn't really a technical indicator, it is much more a way to look at the market and only make a trade if the current price reaches a specific value. Our hypothosis is that the traders who use price actions looks only at the price, so that is exactly what we are going to do as well, and by giving the classifier the features of openBid, closeBid, volume, and spread we hope to achive the behaviour of a price action trader. On the third and last round of features, we
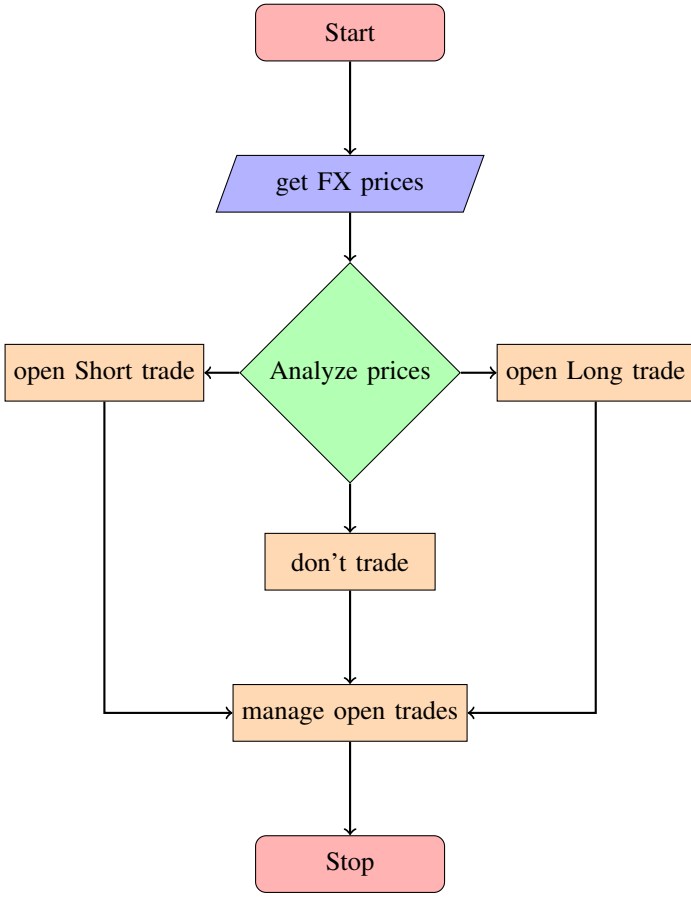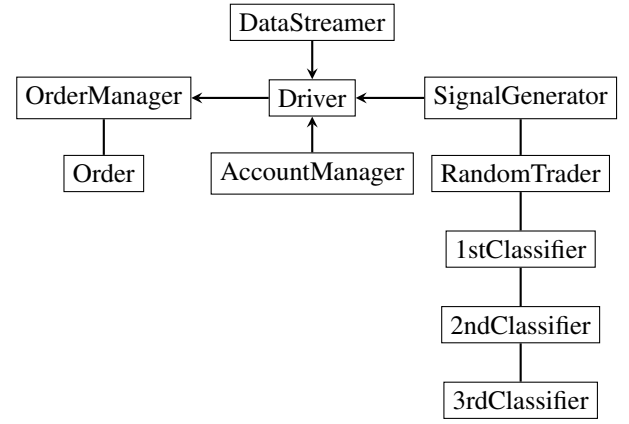
Fig. 1. Flowchart of decision making

Fig. 2. Overview of Classes

24th of October 2014 corresponding to about 500.000 times 5 minutes. All of the classifiers have been tested using a strategy with take profit, and stop losses of a risk of 1:2 meaning that for every dollar you risk, you can only earn 2 dollars, also the order is automatically stopped if a stop criteria hasn't been meet in 20 minutes.

B. Results

VI. DISCUSSION

VII. CONCLUSION

VIII. TERMS & ABBRIVATIONS

Domain	Term or Abbreviation	Meaning
Trading	FX	Forex
Trading	Bid	An offer made by an in
Trading	Ask	The price a seller is wil
Trading	SMA	Simple Moving Average
Trading	MACD	Moving Average Conve
Trading	SAR	Parabolic SAR(Paraboli
Machine Learning	NBC	Naive Bayes Classifiers

REFERENCES

Kimihisa KAWABATA and Hitoshi TAKATA. Fx trading using logistic regression analysis and naive bayes model. Technical report, Kyushu Sangyo Universit JAPEAN and Kagoshima University JAPAN. URL http://www.jsst.jp/e/JSST2012/extended_abstract/pdf/11.pdf. Accessed November 2014.

searched the internet for a forex trading strategy on the EUR vs USD pair, to see how our strategies of the first two rounds of features compare to a random strategy you can find on the internet. The strategy we found consist of using SAR indicator as well as three different MACD indicators.

III. DESIGN

Before we start trading, we need to start to train our classifier based on the feature set we would like to test. After the training is completed we can then start to trade based on our historical data. When the historical back testing is completed, the result of the trade is then shown on a graph, showing the balance of the account of a 5 minute interval. Here is a flow chart, showing how the flow of the forex data, and the decisions are being made.

IV. IMPLEMENTATION

Here we explain about the program itself

V. TEST & RESULTS

A. Test

All of the classifiers will be tested using both a sort term test, and a long term test, the short term test is from September 2014 to October 2014, stretching over 5000 times 5 minutes. The long term test is stretching from 24th of October 2007 to