

I'd compare stock pickers to astrologers, but I don't want to bad mouth astrologers.

Eugene Fama

Eugene Fama won the Nobel Prize in 2013 for the Efficient Market Hypothesis (EMH), which states that asset prices fully reflect all available information and it is impossible to beat the market consistently.

In the era of big data, data scientists have questioned this assumption. The amount of available data and rewards for those who dare trying, make the stock market the hardest game for AI to challenge.

If an AI can beat the stock market game, it can beat any game.

Optimizing asset allocation with a multi agent approach

Lucas Oliveira Souza

Data Analyst, Machine Learning Practitioner, Certified Investment Specialist

Background

There are four main strategies used to optimize asset allocation in an investment fund:

- Transform price and volume data into technical trade indicators, such as Bollinger's Band, Japanese Candlesticks and Elliot Wave's, and use them to predict the best time to buy and sell
- Apply non linear regression algorithms on historical data, to predict stock price movement
- Use reinforcement learning to optimize investment strategies through massive backtesting
- And all of them combined.

*** A bat-neural network multi-agent system (BNNMAS) for stock price prediction**

Reza Hafezi, Jamal Shahrabi, Esmail Hadavandi

*** A Multiagent Approach to Q-Learning for Daily Stock Trading**

Jae Won Lee, Jonghun Park, Member, IEEE, Jangmin O, Jongwoo Lee, and Euyseok Hong

*** An Algorithm for Trading and Portfolio Management Using Q-Learning and Sharpe Ration Maximization**

X. Gao and L. Chan

*** Embedding technical analysis into neural network based trading systems**

T. Chenoweth, Z. Obradovic, and S. S. Lee

*** A Multi-Agent Framework for Stock Trading**

Darryl Davis, Yuan Luo, Kecheng Liu

*** Data Mining on Romanian Stock Market Using Neural Networks for Price Prediction**

Magdalena Daniela NEMEȘ, Alexandru BUTOI

*** An Automated FX Trading System Using Adaptive Reinforcement Learning**

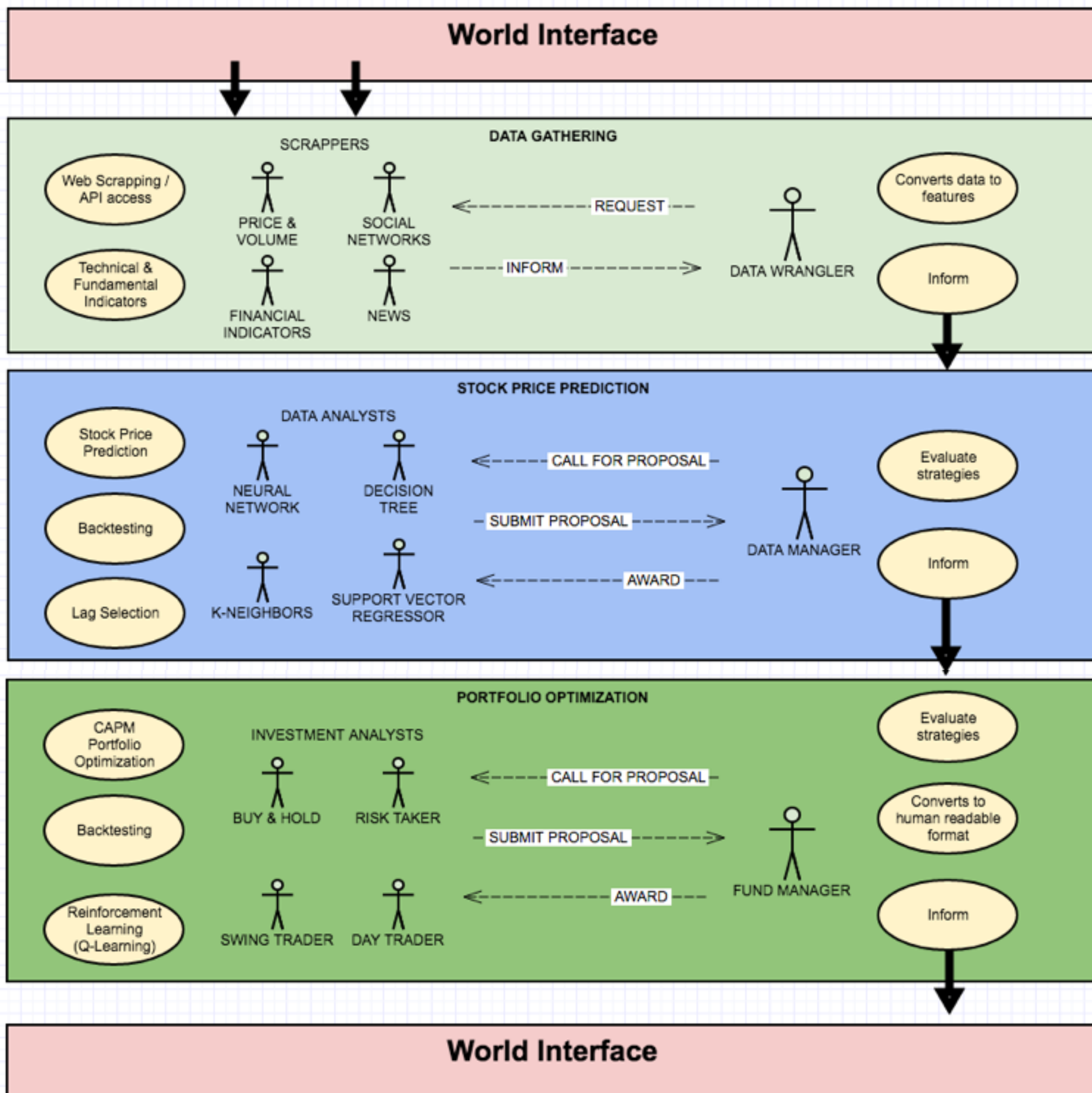
M A H Dempster, V Leemans

*** Learning to trade via direct reinforcement**

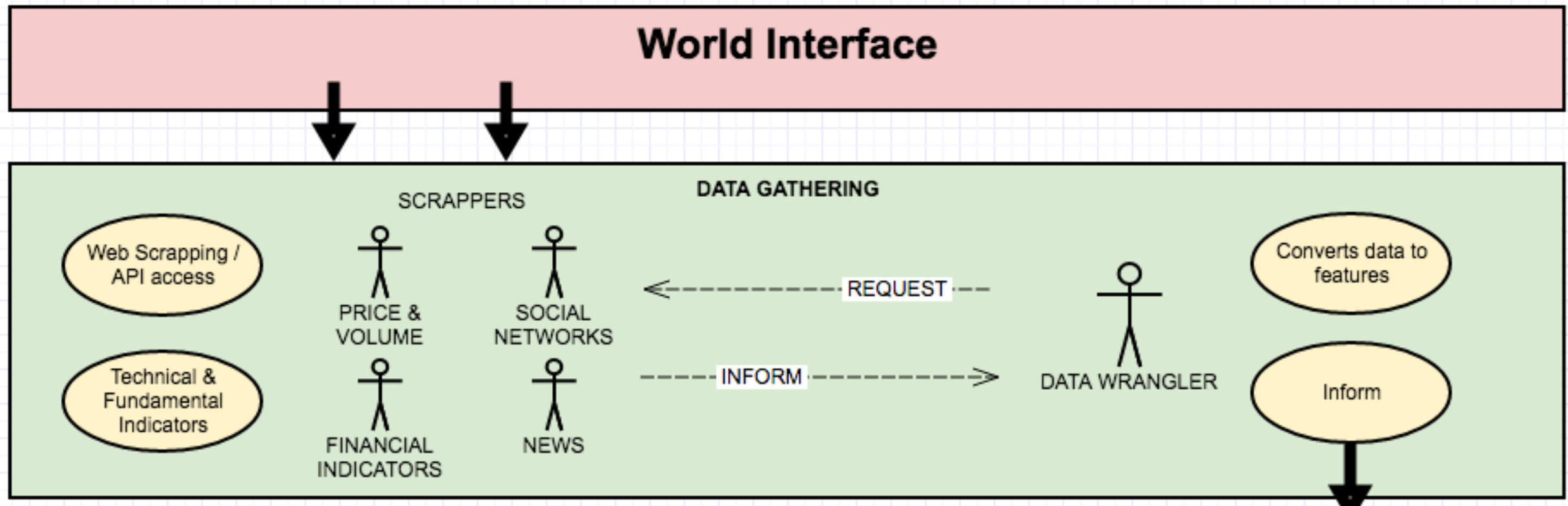
J. Moody and M. Saffell

*** A multi-agent decision support system for stock trading**

Y. Luo, K. Liu, and D. N. Davis

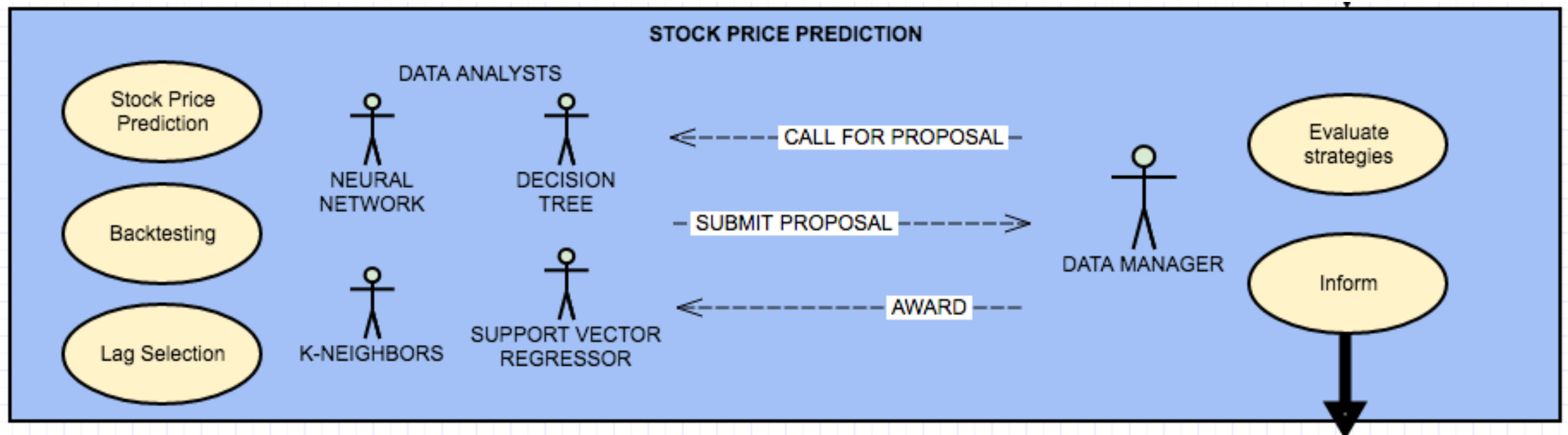


Data Gathering



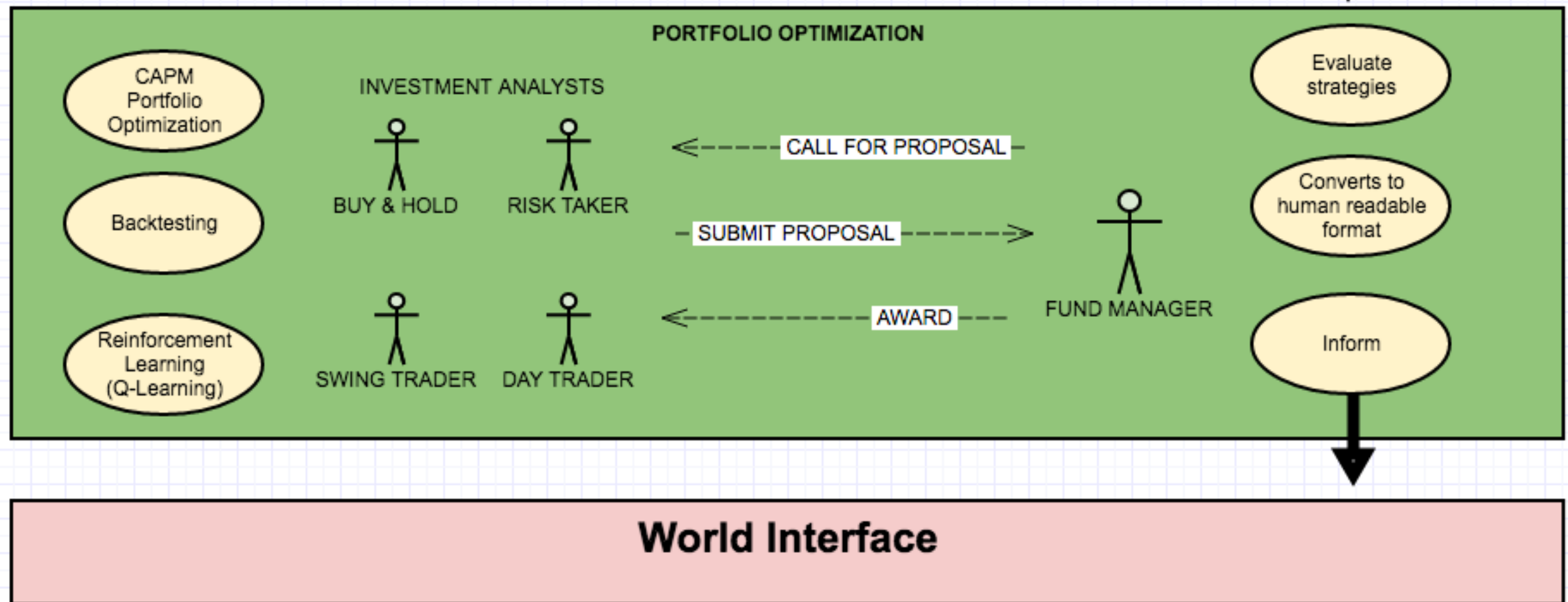
- Each SCRAPPER agent collects data from a data source, upon request of the manager (DATA WRANGLER), and sends to DATA WRANGLER.
- DATA WRANGLER consolidates the data and convert to features
- The features are passed to the next layer, to be fed into a non-linear regressor algorithm

Stock Price Prediction



- All DATA ANALYSTS receive the same information from DATA WRANGLER.
- DATA MANAGER submits a call for proposal for the agents to predict price for a given stock on a given period of time
- Each DATA ANALYST tries to predict stock price movement using a non linear regressor algorithm and the given data. Performance is measured with backtesting (crossvalidation on time-series) and submitted to DATA MANAGER.
- DATA MANAGER chooses the most performative model and inform the predictions to the next layer

Portfolio Optimization



- **FUND MANAGER** submits a call for proposal for agents to maximize return for a given amount of capital, in a given period of time and a given list of allowed assets.
- Each **INVESTMENT ANALYST** tries to maximize the portfolio return using a pre-defined policy. The policy is backtested, and it evolves through Q-Learning until convergence. The results are submitted to **FUND MANAGER**
- **FUND MANAGER** chooses the best strategy, converts the results to human readable actions and inform the human operator.

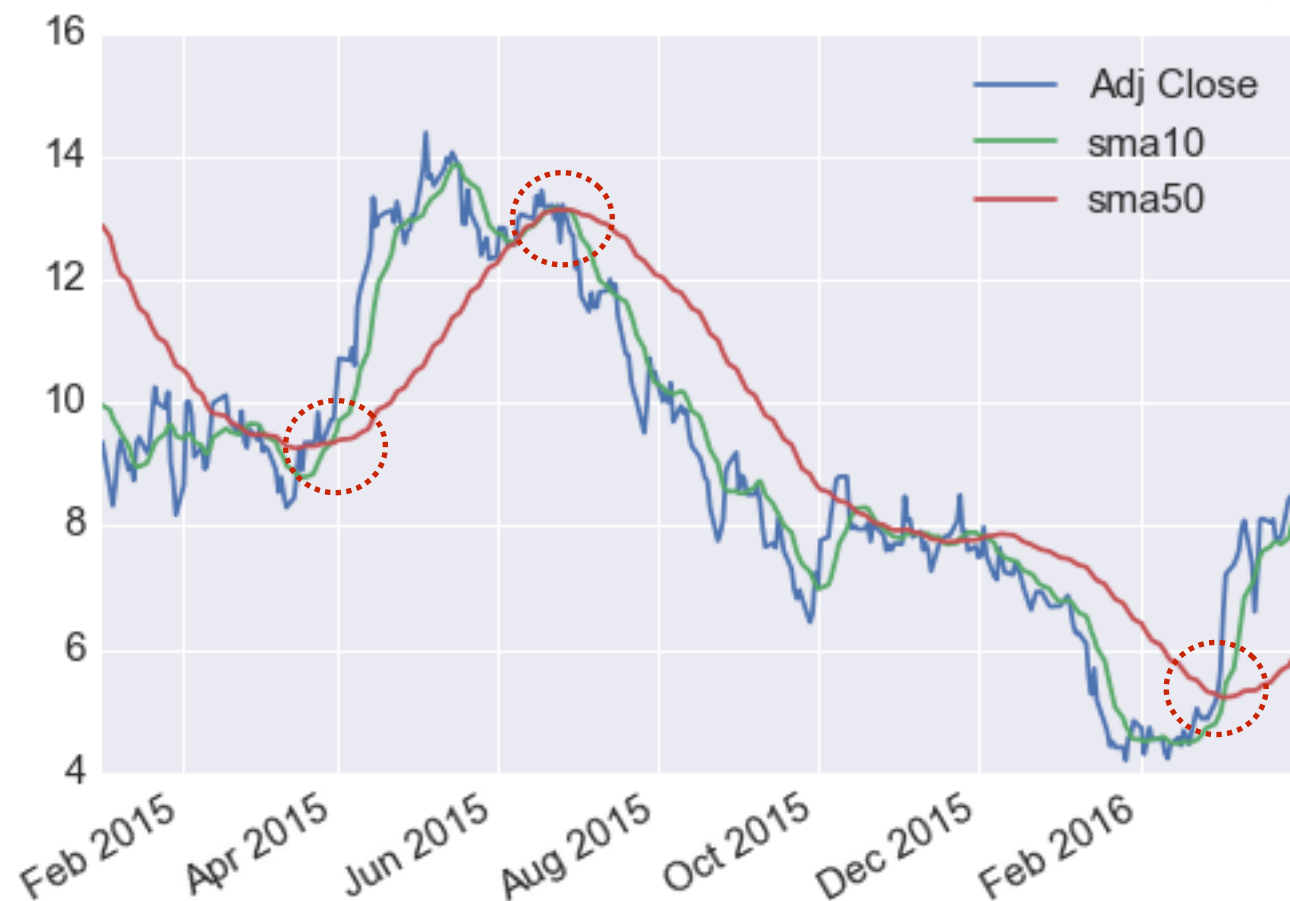
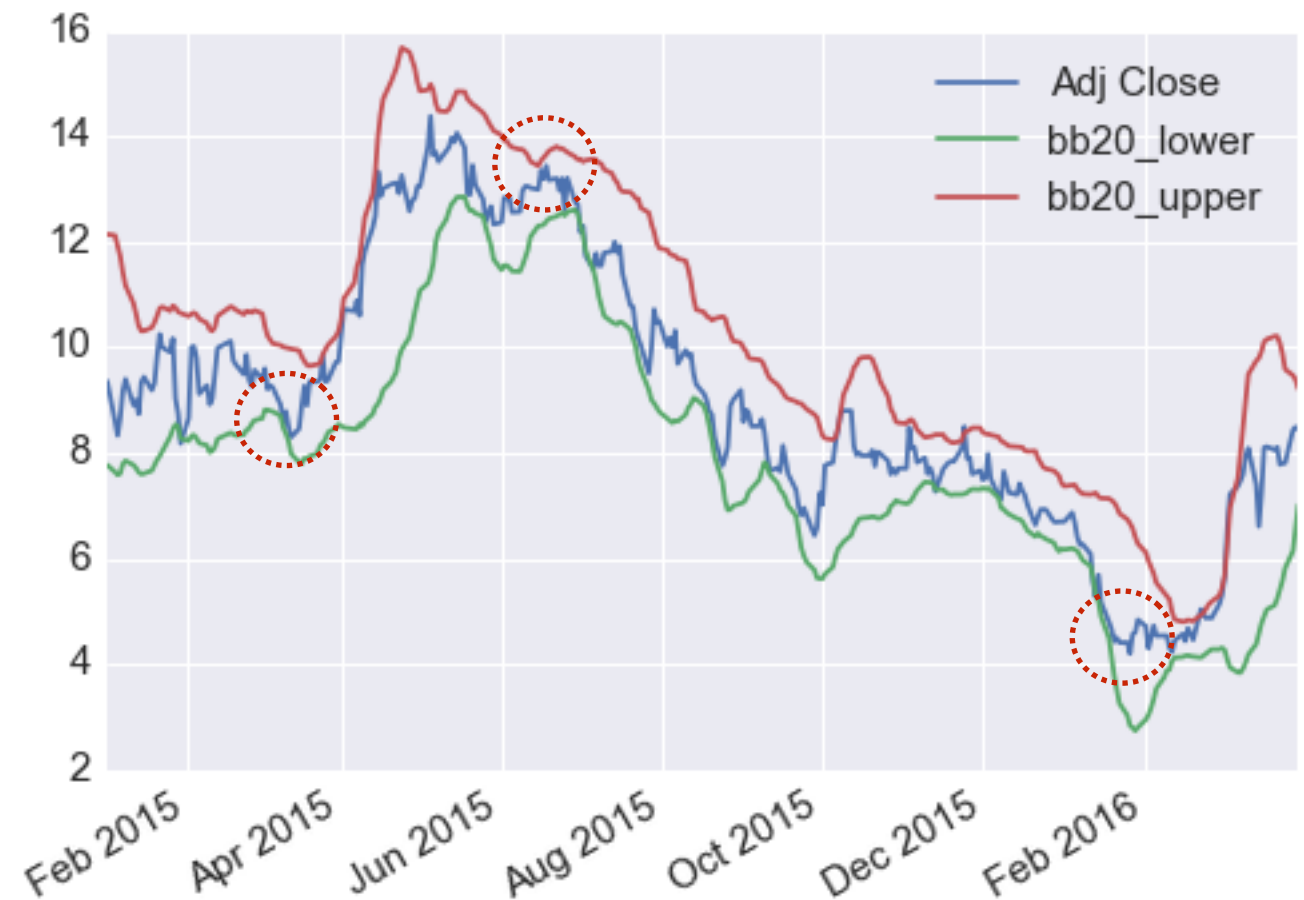
Challenges

1. Multiagent system, distributed programming
2. Stock price prediction with non linear regressors, such as neural networks
3. Portfolio management using Q-learning, online learning
4. Extraction of technical and fundamental analysis indicators
5. Lag selection
6. Data wrangling of massive time series dataset

Feature Selection

Examples of technical analysis features

Red circles shows inflection points

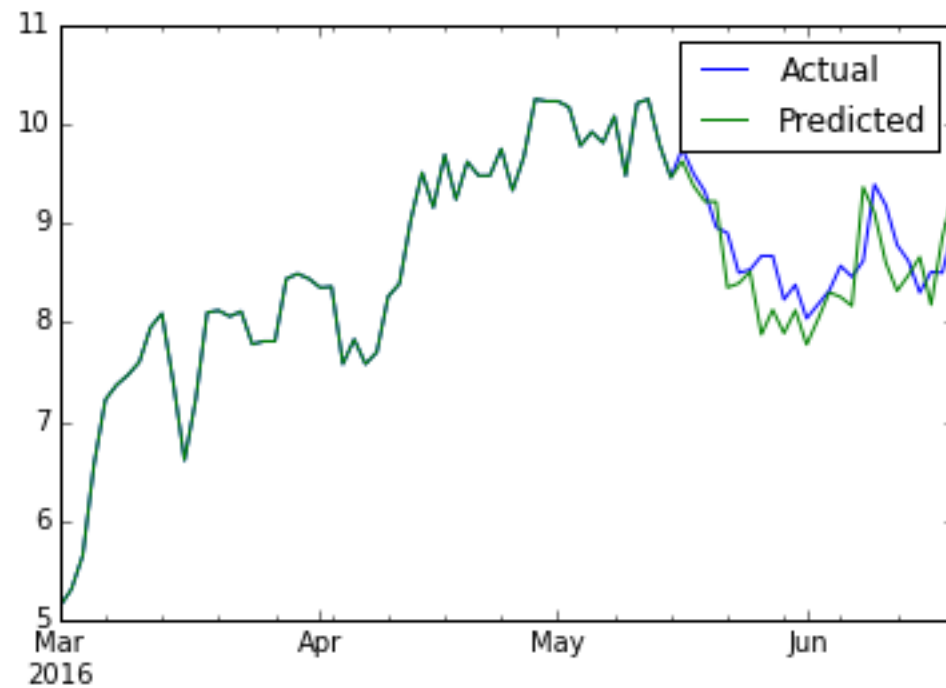


20 days Bollinger bands

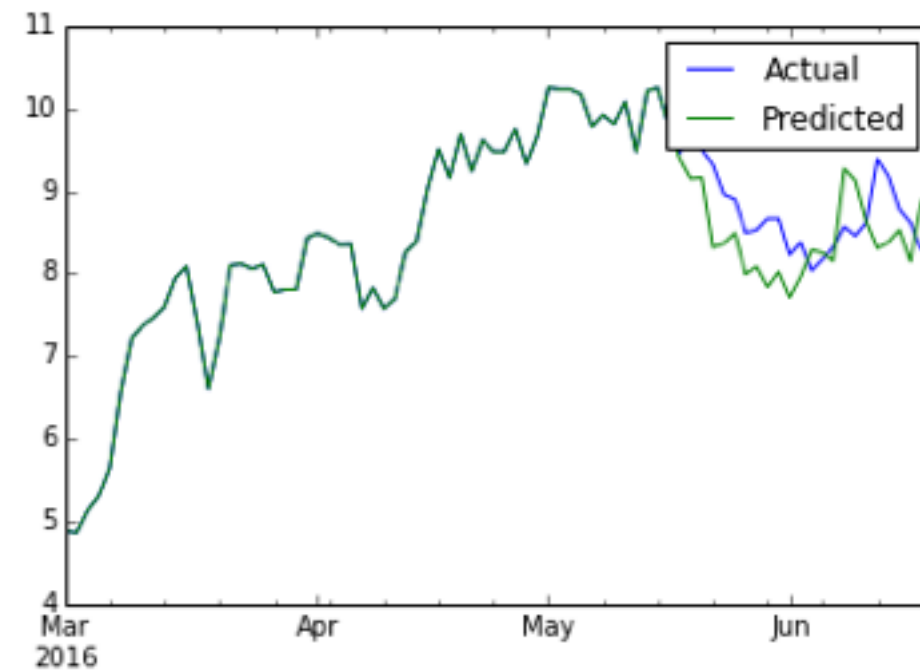
10 and 50 days moving averages

Lag Selector

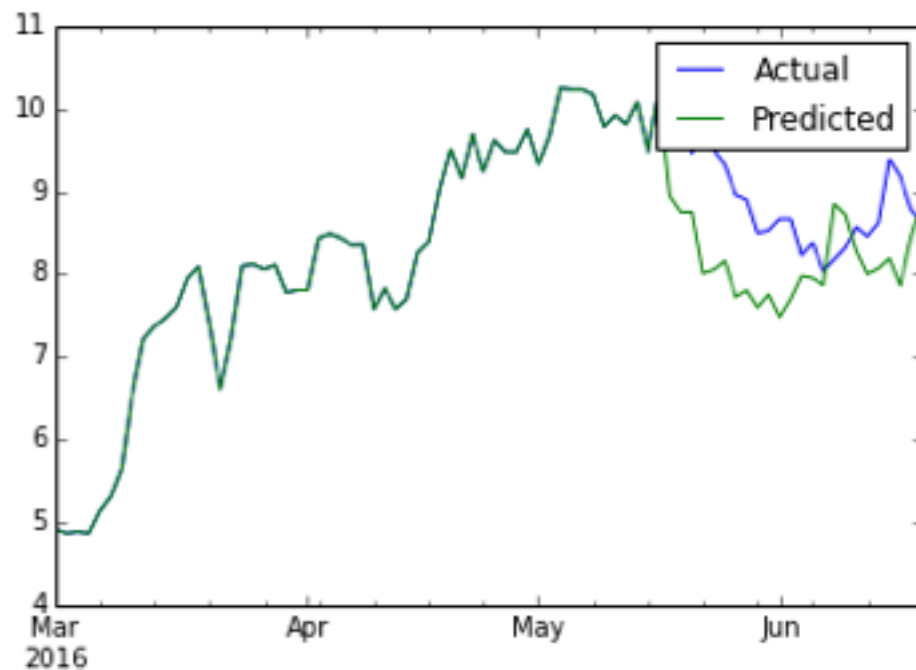
*Multilayer Perceptron with 10 hidden layers.
PETR4, Mar-Jun 2016 (80 days)*



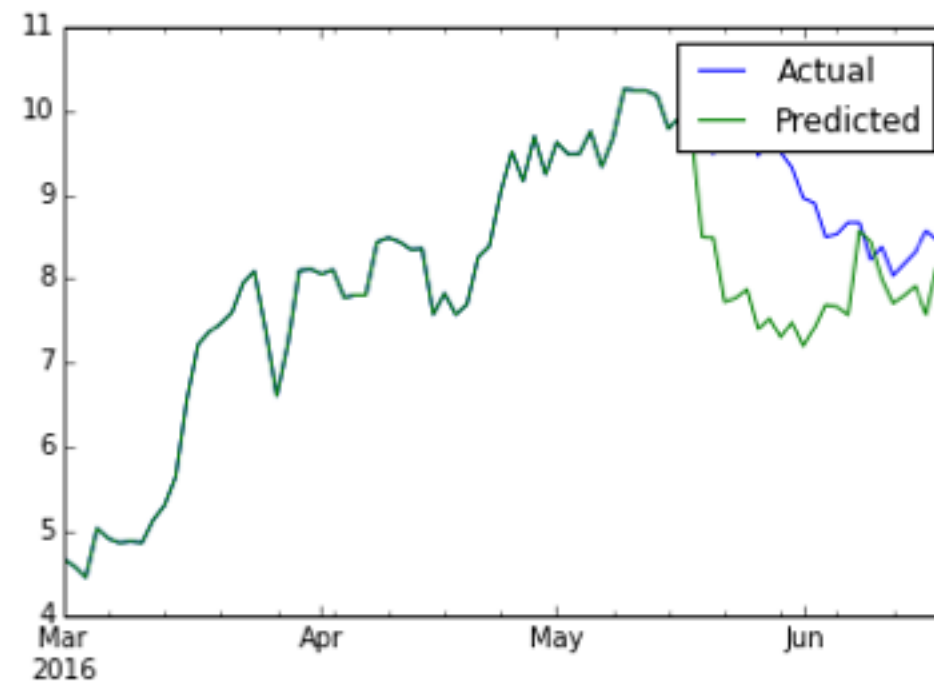
Lag: 1 day



Lag: 3 days



Lag: 5 days



Lag: 9 days

Implementation

- Main stock data collected from Bovespa and Yahoo Finance. Stocks filtered by total trading volume higher than R\$ 1 million at 10-06-2016, total 132 stocks.
- To include data features extracted from sentiment analysis of news, twitters and facebook posts.
- Implementation in Python 2.7.10. Machine learning algorithms adapted from 'scikit-learn', data wrangling with 'numpy' and 'pandas', plotting with 'matplotlib' and 'ggplot2'.
- FIPA compliant multi agents environment based on SPADE, Smart Python Agents Development Environment.
- Rodeo and Jupyter Notebook as IDEs
- Local infrastructure, 8 cores Mac laptop with 16gb RAM. To be defined larger provider.

Thank you!

Questions?