AI Forecasting

BITCOIN PRICE PREDICTION

Group 4 - Team Members

- Juan Carlos Castaneda
- William Chance
- Martin Rasumoff
- Jorge Sira

UNIVERSITY OF MIAMI
MIAMI HERBERT
BUSINESS SCHOOL

Introduction

OUR PROJECT

The objective of this project is to predict the price direction of Bitcoin at the end of the upcoming 4 hours (variable time-frame).

Our aim is to use different machine learning classification techniques, and compare different model performances. We also experimented with variations on re-training size, re-training frequency, as well as the number of features (to be performed). Model outputs the predicted direction which will be interpreted as a long, short, or non-confident signal (based on certain criteria)

Then, the goal is to provide actionable information that enables the trading decision process on when, how much, and what type of trade to execute (short vs long).

Note that this project was inspired by the work performed at the Northumbria University in Newcastle* (Gerlein, Eduardo, McGinnity, Martin, Belatreche, Ammar and Coleman, Sonya, 2016 "Evaluating machine learning classification for financial trading: An empirical approach. Expert Systems with Applications")

UNIVERSITY OF MIAMI
MIAMI HERBERT
BUSINESS SCHOOL

OUR PROJECT (cont'd)

The inputs to the models include past quote data (OHLCV), which are enriched within the process with a set of technical indicators, to be used as features on the model training phase.

Part of the project consists in testing different features which includes indicators from several categories such as momentum indicators, volume indicators, volatility indicators, trend indicators, etc.

Our plan is to complete an MVP for this project consisting of creating a software solution that:

1. **Creates different sets of indicators to test a group of predictive models (dataset creation)**
2. **Models each of the datasets using processing windows and frequent re-training (sliding window durations and re-training periods to be defined) - (model creation and findings generation)**
3. **Summarizes the findings**
4. **Presents the outcome to the end-user (GUI)**

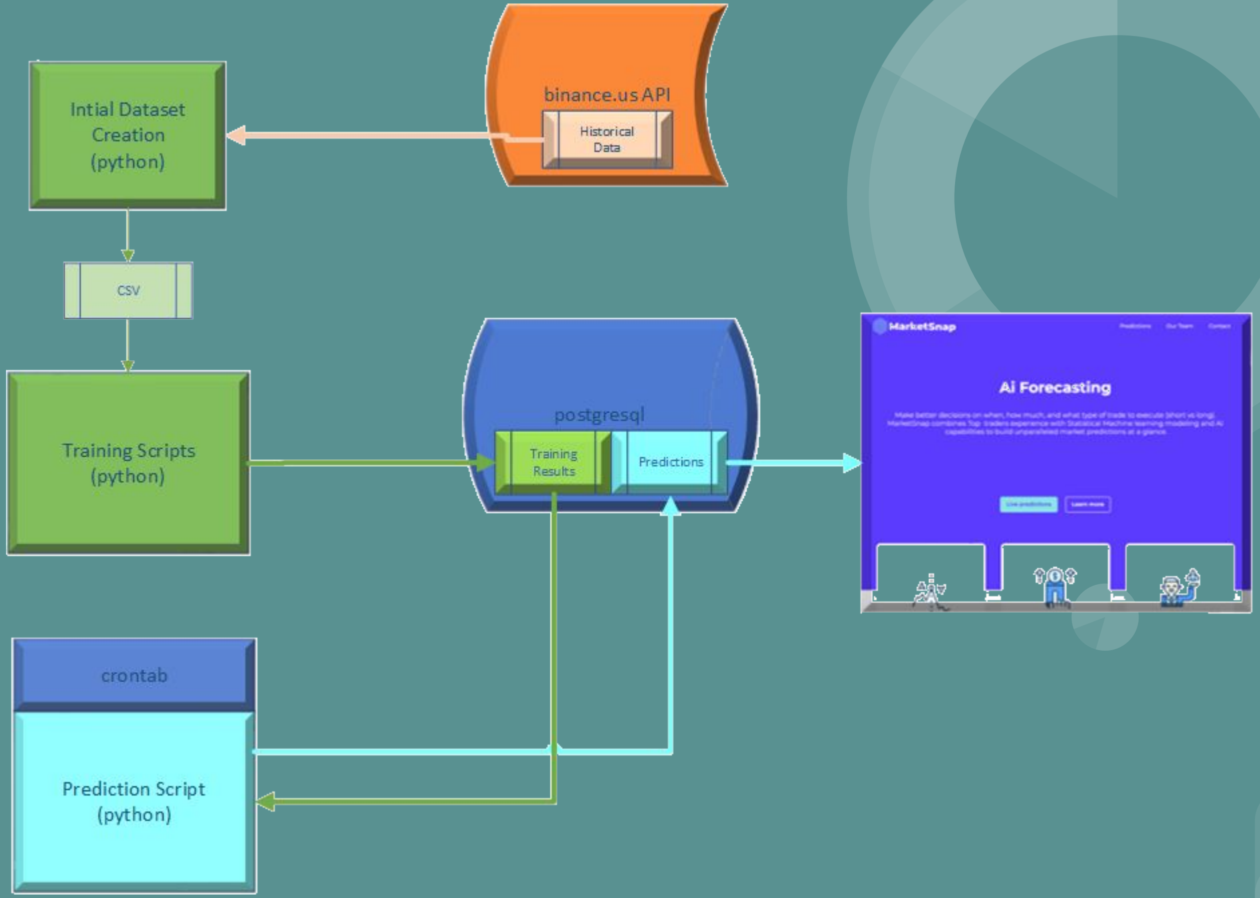
UNIVERSITY OF MIAMI
MIAMI HERBERT BUSINESS SCHOOL

Technology

INFRASTRUCTURE

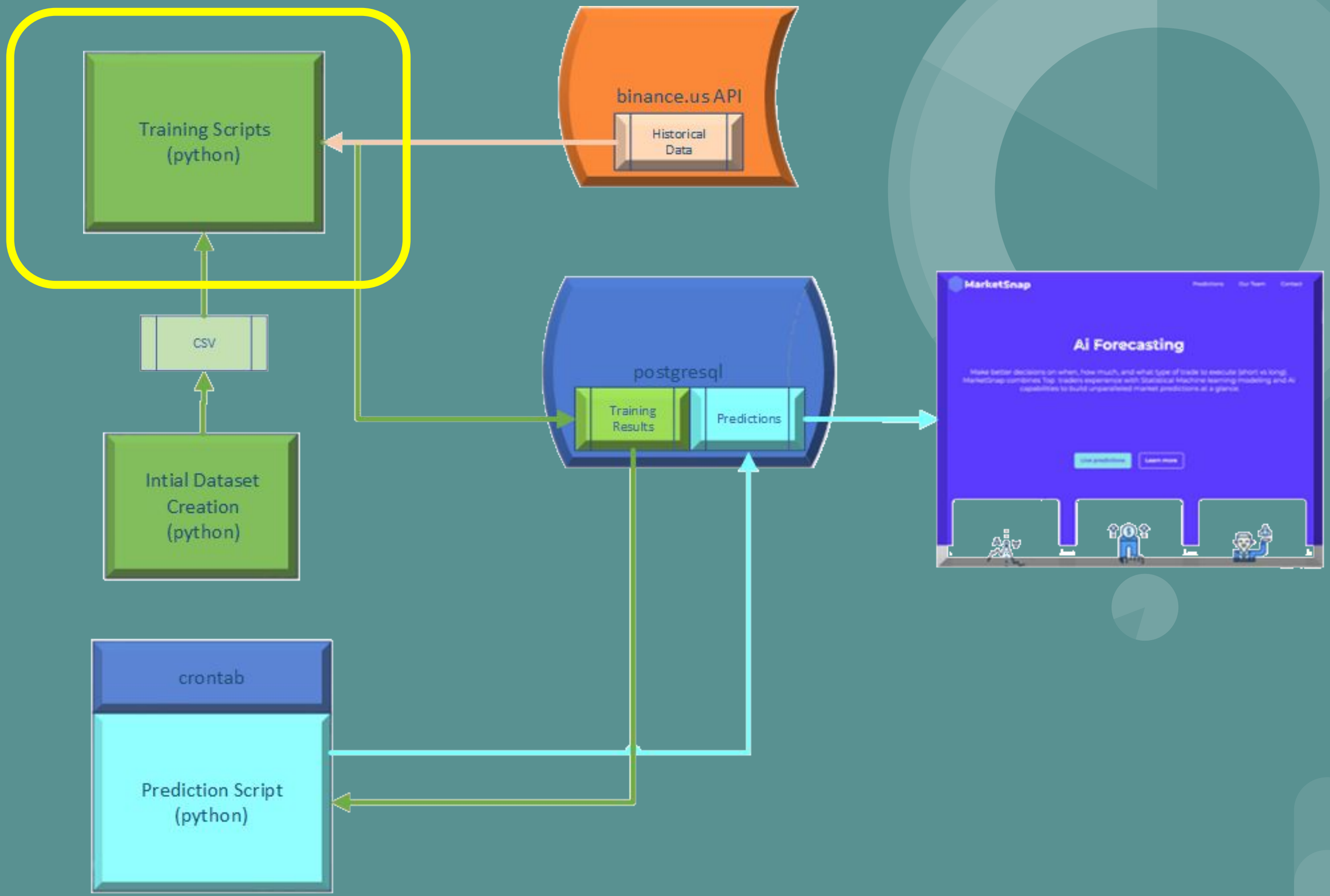
- 3 Virtual Machines running - Debian Linux OS
- Backend - Python, Crontab
- Frontend - Docker, Nginx, Bootstrap + Flask
- Database - Postgresql

PROCESS FLOW



Experimentation and Deployment

Experimentation Methodology (Training Script)



High-level Experimentation Process

Example using a Training-Window of 500 and a re-training period of 5

How often to re-train

Test Size Size

Table 5. Consolidated results for the experiments with variations in retraining set size, retrain period, and number of attributes.

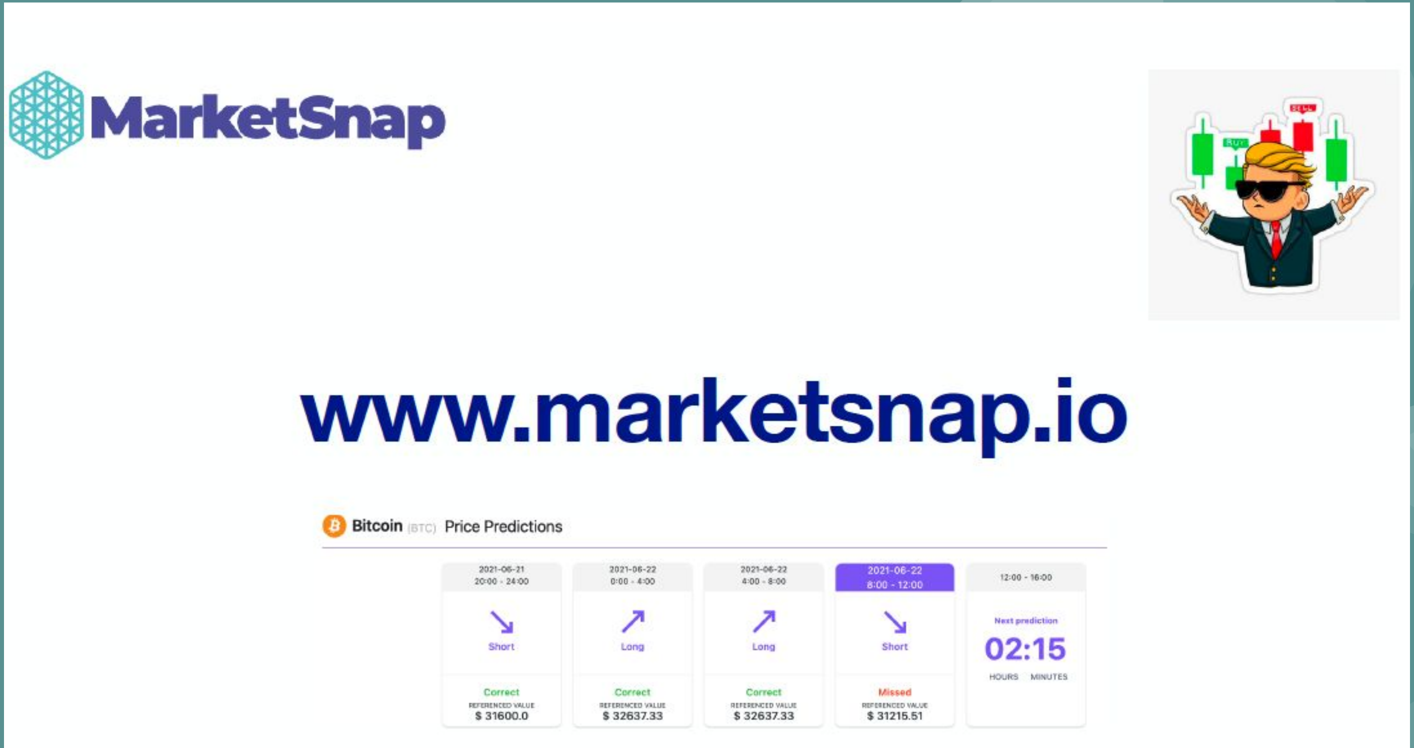
Currency Pair:				USDJPY						
Experiment Setup	*Retrain Set Size*	*Retrain Periods*	*# of Attributes*	**Metrics**	**Machine Learning Models**					
					OneR	**C4.5**	**Jrip**	**LMT**	**Kstar**	**NaiveBayes**
Setup 1		5	*5*	**Accuracy**	49.29	53.38	51.79	53.17	51.79	52.99
				DOWN Accuracy	48.35	52.82	51.01	52.86	51.19	52.25
				UP Accuracy	50.16	53.82	52.43	53.39	52.81	53.65
				10-Fold Cross-Val.	53.71	51.90	53.51	53.31	48.90	50.90
				Cumulative Return	**-55.79**	**116.88**	**18.25**	**106.17**	**69.94**	**142.89**
Setup 2	*500*	10	*5*	**Accuracy**	49.56	52.97	51.41	53.00	51.48	52.74
				DOWN Accuracy	48.61	52.35	50.59	52.64	51.16	51.98
				UP Accuracy	50.41	53.48	52.05	53.25	52.77	53.42
				10-Fold Cross-Val.	51.70	52.10	52.71	52.30	52.71	51.30
				Cumulative Return	**-53.02**	**116.04**	**6.97**	**107.96**	**64.02**	**123.55**
Setup 3		15	*9*	**Accuracy**	50.49	52.00	51.92	52.76	49.77	52.20
				DOWN Accuracy	49.61	51.29	51.29	52.33	50.44	51.50
				UP Accuracy	51.36	52.53	52.34	53.04	52.06	52.75
				10-Fold Cross-Val.	51.30	49.10	51.90	51.10	51.90	53.31
				Cumulative Return	**-14.69**	**18.61**	**41.86**	**41.48**	**62.47**	**54.76**
Setup 4		5	*5*	**Accuracy**	50.01	53.89	52.37	53.72	51.01	53.58
				DOWN Accuracy	49.11	53.18	51.53	53.57	51.29	52.86
				UP Accuracy	50.87	54.53	53.16	53.82	52.94	54.23
				10-Fold Cross-Val.	50.15	50.85	51.25	50.05	50.85	52.95
				Cumulative Return	**-15.35**	**146.06**	**84.50**	**145.19**	**70.22**	**136.94**
				Accuracy	49.98	53.69	52.96	53.70	51.17	53.56
				DOWN Accuracy	49.09	52.96	52.25	53.59	51.37	52.84

High-level Experimentation Process

Example using a Training-Window of 500 and a re-training period of 5

Front-end

Front End - https://www.marketsnap.io/



Questions and Answers

Thank you

Source with API to access historical Cryptocurrencies data, in particular **BITCOIN**

- **BINANCE US**
  - (**https://www.binance.us/**)
  - **We are generating a Python script that allows us to obtain the required data on demand**
  - **This script also allows to obtain historical data for different assets (pairs) and timeframes (e.g. 1-min vs. 4hrs candles)**

UNIVERSITY OF MIAMI
MIAMI HERBERT BUSINESS SCHOOL

Resources

- Mediant — Mediant FinTech Trends Report
- I know First - https://iknowfirst.com/stock-forecast-algorithm
- *Evaluating machine learning classification for financial trading: an empirical approach. https://nrl.northumbria.ac.uk/id/eprint/34544/1/Evaluating%20machine%20learning.pdf

CONCEPTS

Machine learning algorithms in FinTech are potentially better fortune tellers than any human being could be. The vast volumes of trading operations result in tons of historical data, hence an unlimited potential for learning. At the same time, ML algorithms have the capability of monitoring and feeding from a very big number of data sources which most of them available in real time. News and trade results, to name some adjacent aspects, have an impact on stock market dynamics.

Our task as algorithmic traders is to create the machine that determines which ML algorithms to include in our strategies and ensure that the chosen algorithms and their parameters continue to be "healthy"

CONCEPTS

Machine learning algorithms work best for pattern identification. They detect correlations among tons of sequences and events, extracting valuable information that's camouflaged among vast data sets. Such patterns are often missed or simply can't physically be detected by humans. The ability of ML to learn and predict enables FinTech providers to recognize new business opportunities and work out coherent strategies.