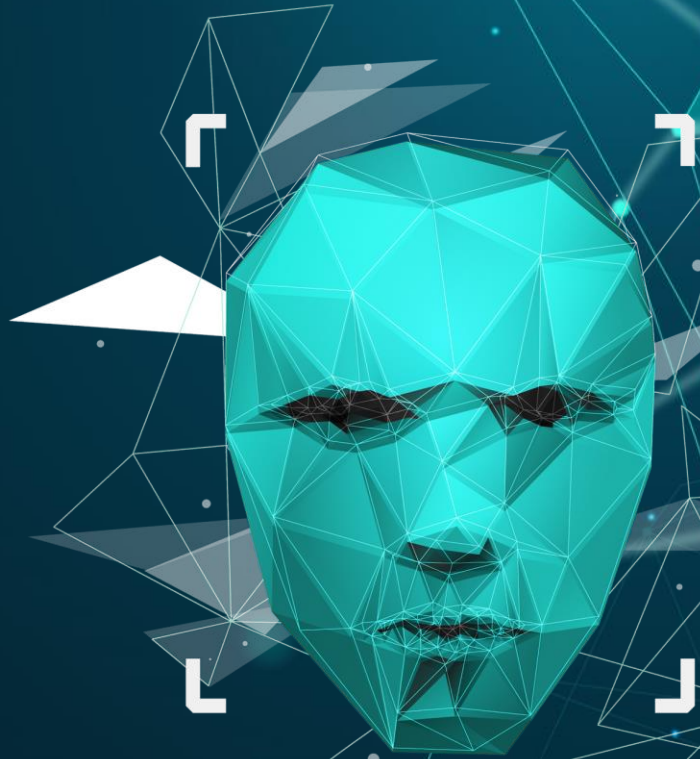
RUBIKS project

Diego Bettega



AGENDA

01 PERSONAL INTRODUCTION

02 OBJECTIVE

03 FLOWCHART

04 RUBIKS PROJECT STEPS

- Historical Data
- Target
- Feature Engineering
- Model
- Analysis
- Backtesting
- Optimizer
- Deployment

05 APPENDIX

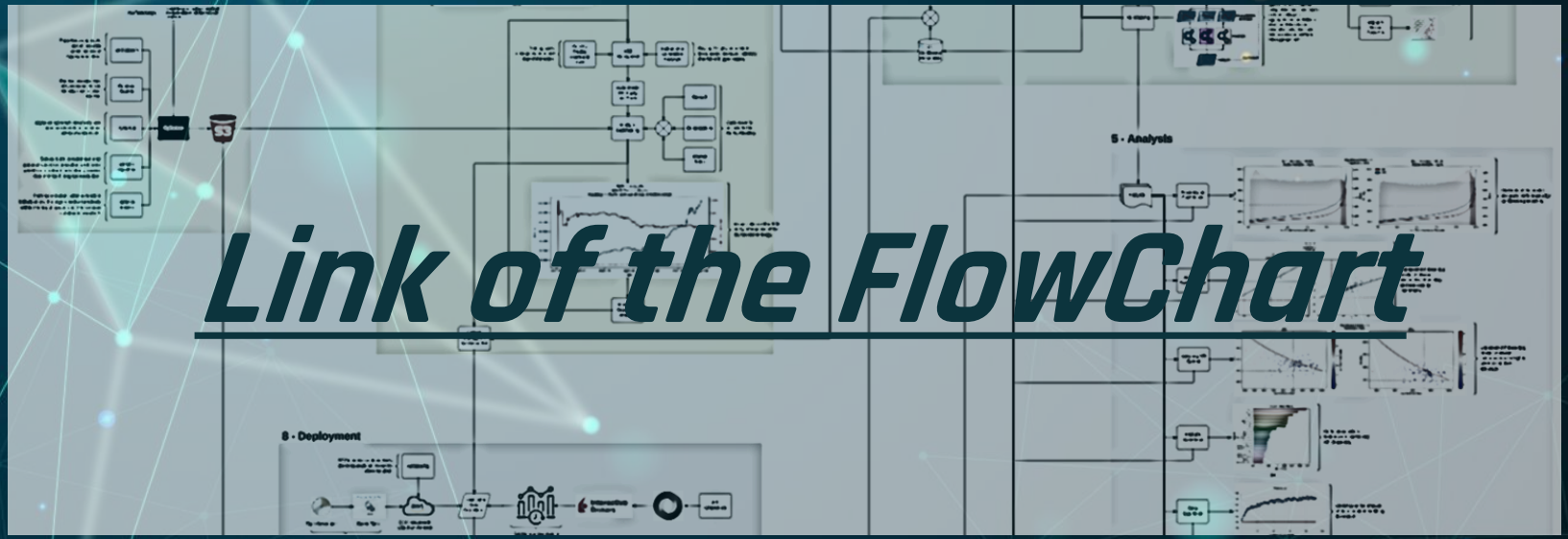
PERSONAL INTRODUCTION

Professional Website

OBJECTIVE

To generate profit by automate trading using several financial instruments (Forex, Stock, Index, Commodity, Crypto) through different brokers.

FLOWCHART



Link of the FlowChart

RUBIKS PROJECT STEPS

10 YEARS OF HISTORICAL DATA

1. HISTORICAL DATA

~10 years of hourly data downloaded from **YAHOO!** and other **brokers.**

Once extracted the data will be cleaned and anomalies will be removed.

yahoo! finance

OANDA SMARTER TRADING

Interactive Brokers

FXCM

datetime — DateTime of the start of the period

open — Price at the start of the period

high — Highest price recorded in the period

low — Lowest price recorded in the period

close — Price at the end of the period

volume — Volume exchanged in the period

spread — Difference between Ask and Bid price

THERE ARE TWO TARGETS: TARGET BUY & TARGET SELL

2. TARGET

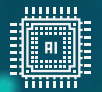
There are two different targets: **Target Buy** & **Target Sell** which are extracted using the **Take Profit** (**TP**) / **Stop Loss** (**SL**)* ratio:

- **1 - successful trade: Profit**
- **0 - unsuccessful trade: Loss**

*TP & SL could be chosen based on a fixed percentage or by looking at the recent volatility.

SET STOP LOSS & TAKE PROFIT LEVELS

2. TARGET



Stop Loss Buy
Limit losses if I decide to buy the instrument

Take Profit Buy
Target of profits if I decide to buy the instrument

Stop Loss Sell
Limit of losses if I decide to sell the pair

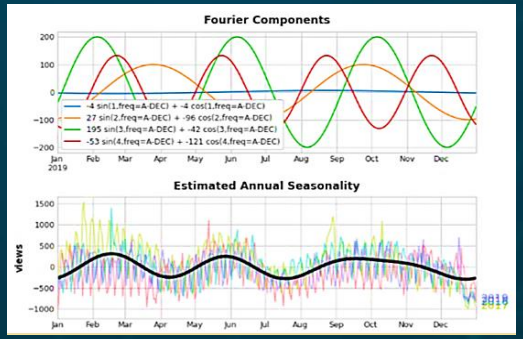
Take Profit Sell
Target of profits if I decide to sell the pair

EXTRACT MODEL FEATURES

3. FEATURE ENGINEERING

- We created many **financial indicators** considering various things: from the Fourier coefficients to Meta Prophet model, from Seasonality to Technical Trading Indicator.

- The two targets, Buy & Sell, and the model features are calculated for each instrument: the model is trained using the full combined dataset.



PROPHET

MODEL FEATURES DEEP DIVE i

3. FEATURE ENGINEERING

LEADING INDICATORS
Some instruments might act as early warning on others

CALENDAR
extracted from datetime index

FOURIER SERIES COEFF CALCULATOR FUNCTION
[10, 20, 30]

AVG TRUE RANGE
identify the market volatility

MACD
identifies moving averages that indicate a new trend

AVG DIRECTIONAL INDEX
identify whether the market is ranging or starting a new trend [4, 14, 24]

STOCHASTIC OSCILLATOR
speed or velocity of price changes [4, 14, 24]

COMMODITY CHANNEL INDEX
identify overbought and oversold levels

RELATIVE STRENGTH INDEX
evaluate the strength of the current market [4, 12, 24]

MODEL FEATURES DEEP DIVE ii

3. FEATURE ENGINEERING

TECHNICAL TRADING INDICATORS
using an external package

JAPANESE CANDLESTICK PATTERNS
method of visualizing charts

INSTRUMENT TYPE
macro group of the financial instrument

BOLLINGER BANDS
measure market volatility

SLOPE
help to understand direction and slope of the trend [4, 12, 24, 48]

ICHIMOKU KINKO HYO
determine future areas of support and resistance

SINE SERIES COEFF CALCULATOR FUNCTION
[4, 12]

HARMONIC PATTERNS
help to spot possible retracements of recent trends

HEIKEN ASHI OHLC CANDLES
charting technique used to display prices that, at a glance, looks similar to a traditional Japanese candlestick chart

MODEL FEATURES DEEP DIVE iii

3. FEATURE ENGINEERING

PROPHET
Meta's time series package

ACCUMULATION DISTR. OSCILLATOR
provide insight about the strength of the trend [4, 12, 24]

SEASONALITY
MODELLED CLASSIFIER LOOK AHEAD PREDICTION USING SEASONAL INDICATOR

MOMENTUM
identify when the price is moving upward or downward and how strongly [4, 12, 24]

WILLIAMS OSCILLATOR
momentum indicator to detect when the pair might be "overbought" or "oversold" → less popular and more sensitive version of Stochastic [4, 14, 24]

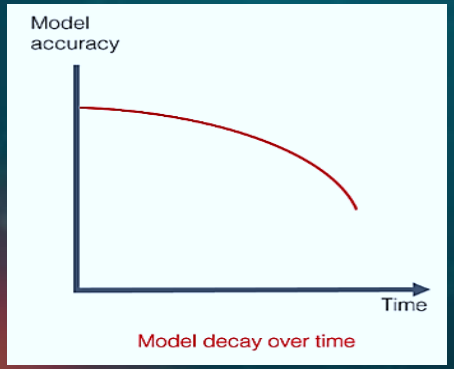
WILLIAMS ACCUMULATION DISRIBUTION
cumulative indicator that uses volume and price to assess whether a currency is being accumulated or distributed

TRAINING MACHINE LEARNING MODEL CLASSIFIER

4. MODEL

1. Model Inputs

Combined engineered features and 2 targets (buy/sell) for every instrument



2. Model Settings

- Train / Test split
- Anomaly detection
- Feature correlation analysis
- Manual recommendation
- NLP news analysis
- Model performance decay
- Cross validation for hyperparameter tuning
- SKLearn multioutput wrapper

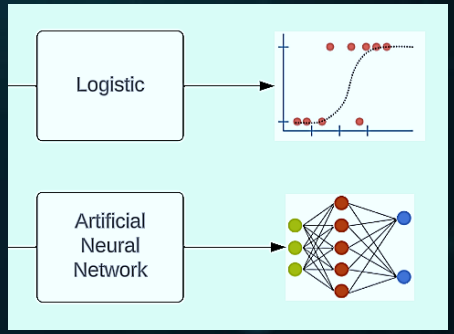
3. Model Classifier

- There are different model classifiers that can be chosen and need to be tested e.g., Logistic Regression
- Model stacking: ensembling different classifiers

4. Model Output

Confidence probability of success/unsuccess (profit/loss).

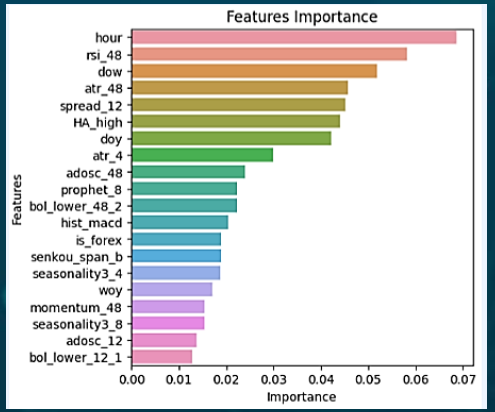
Classifiers example

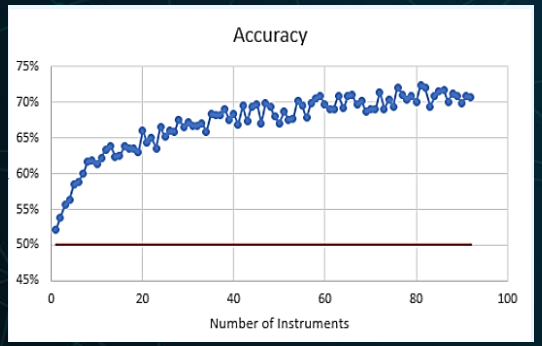


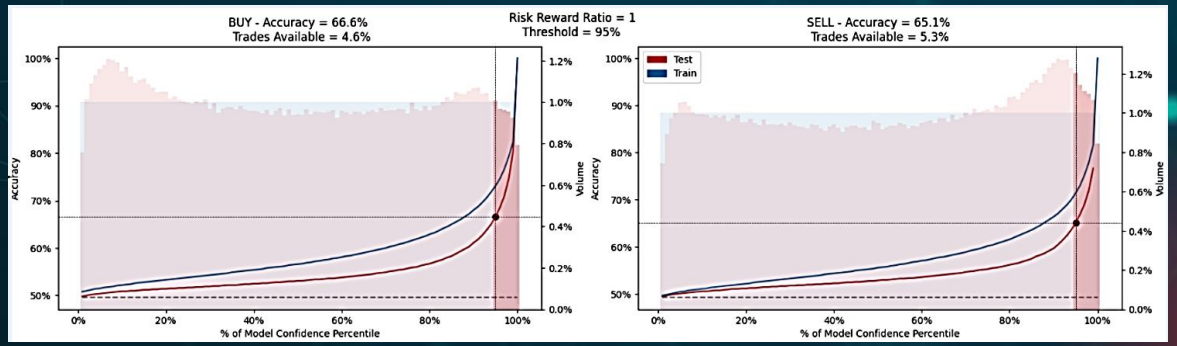
INSIGHTS TO DRIVE MODEL IMPROVEMENTS

5. ANALYSIS

We worked on several analysis done in different areas (model accuracy, model features, model data, trades). They are created to improve the model i.e. the feature importance has been created to reduce the number of features and mitigate overfitting.



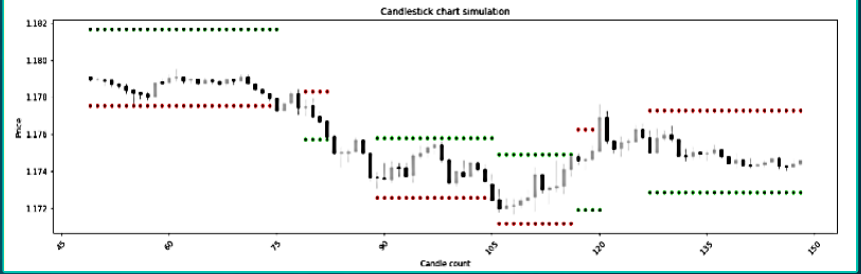




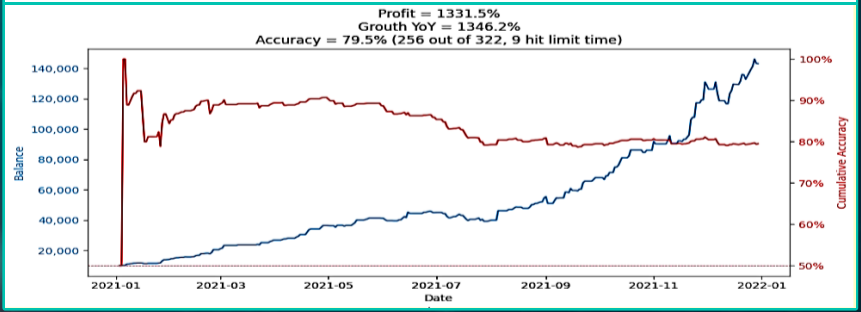
ITERATIVE PROCESS TO ESTIMATE HYPOTHETICAL ROI

6. BACKTESTING

Visualization of the backtesting process



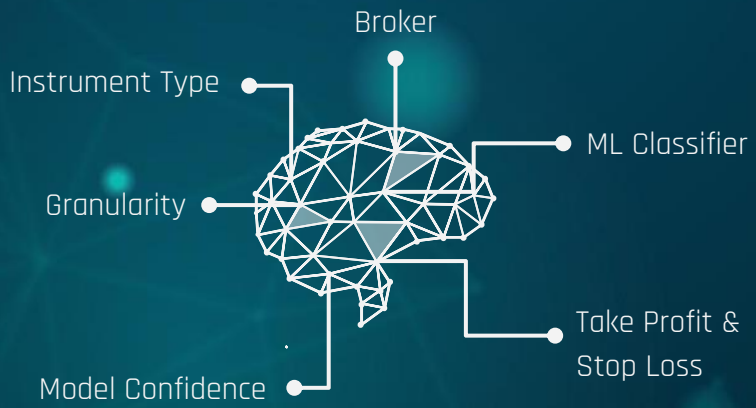
Cumulative accuracy & expected ROI (including costs) of a certain strategy



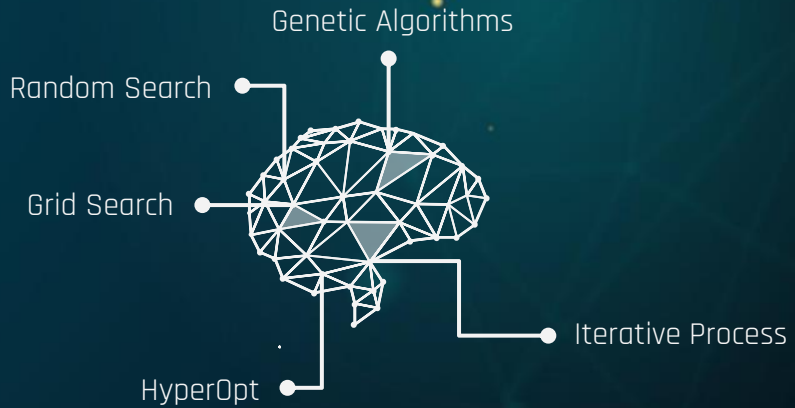
Backtesting to check what the Return on Investment would have been by applying a certain trading strategy.
It is done to assess profit and losses of a strategy by including the costs.

SEARCH FOR OPTIMAL COMBINATION OF MODEL STRATEGIES

7. OPTIMIZER



Hyperparameters



Optimizer

FULLY AUTOMATE THE PROCESS IN AWS CLOUD

8. DEPLOYMENT

Scheduler runs the script (python code) on AWS environment

Trading Algorithm

In real time, save and process in a database the incoming data from yahoo! and / or brokers

Estimate the expected ROI to decide to sell, to buy or to not take any action. Finally, monitor the overall performance.

THANKS!

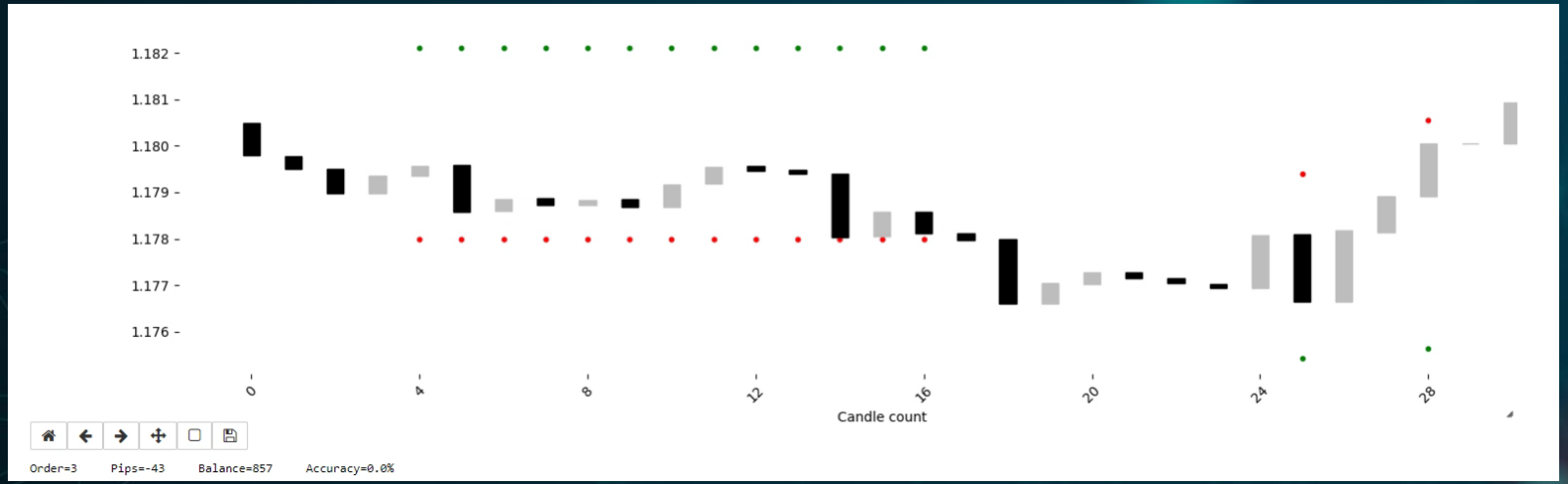
Do you have any questions?

diego.bettega.90@gmail.com
+39 3247728599
Linkedin

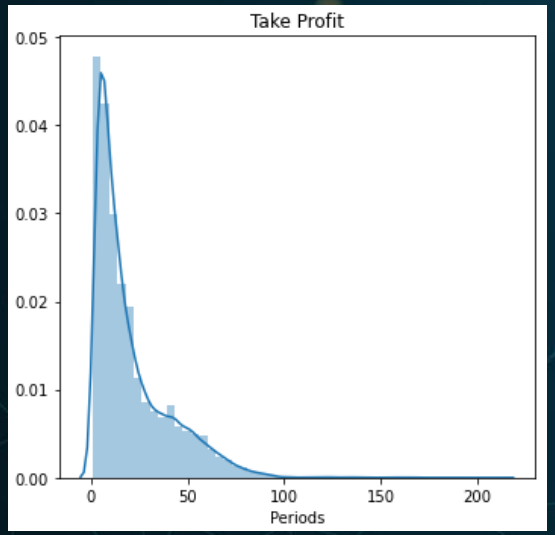
APPENDIX

BACK-TESTING ENGINE

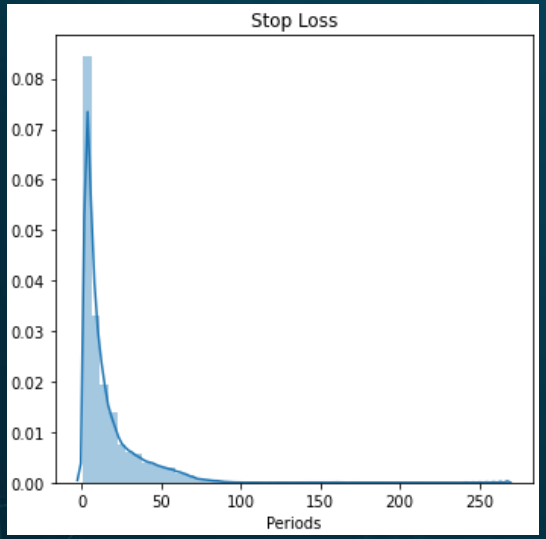
Simulation of the Signal applied to historical data.
The dotted lines represent the Stop Loss and Take Profit levels.



HOW MANY PERIODS DO WE NEED TO HIT A PROFIT?



50% of the Take Profit are hit within 13 periods



50% of the Stop Loss are hit within 8 periods

→ Focus the analysis to predict the price N = 10 periods ahead.

FEATURES CORRELATION

- The Pearson’s Correlation coefficient reveals that many features are correlated
- To prevent overfitting, the correlated features are dropped