Data Science II

Project 8 - A stock market recommendation engine

Final Presentation

Group 10 – Tatu RE

Christian Leomil de Paula — Matriculation number: 2616162
Guilherme Fernandes — Matriculation number: 2971519
Rodrigo Mardegam Morais — Matriculation number: 2285470
Caio Victor Gouveia Freitas — Matriculation number :2328654

Our Team

Team division and roles

Rodrigo Morais

Developer

Hidden Markov Model and Presentation

Caio Freitas

Developer

Hidden Markov Model and Dashboard

Guilherme Fernandes

Developer

Linear regression model and Cloud deployment

Christian Leomil

Developer

Routines for performance testing and Presentation

Agoston Torok

Team supervisor

Technical consultancy and Mentor

Tatu RE

Index

Index

Problem Statement

Outperform the market to maximize return is a challenging task

Investing Money

Time consuming
Being regularly aware about the market and monitoring sec are necessary steps

Risk involved
The market has a high volatility and is sensitive to random economical events

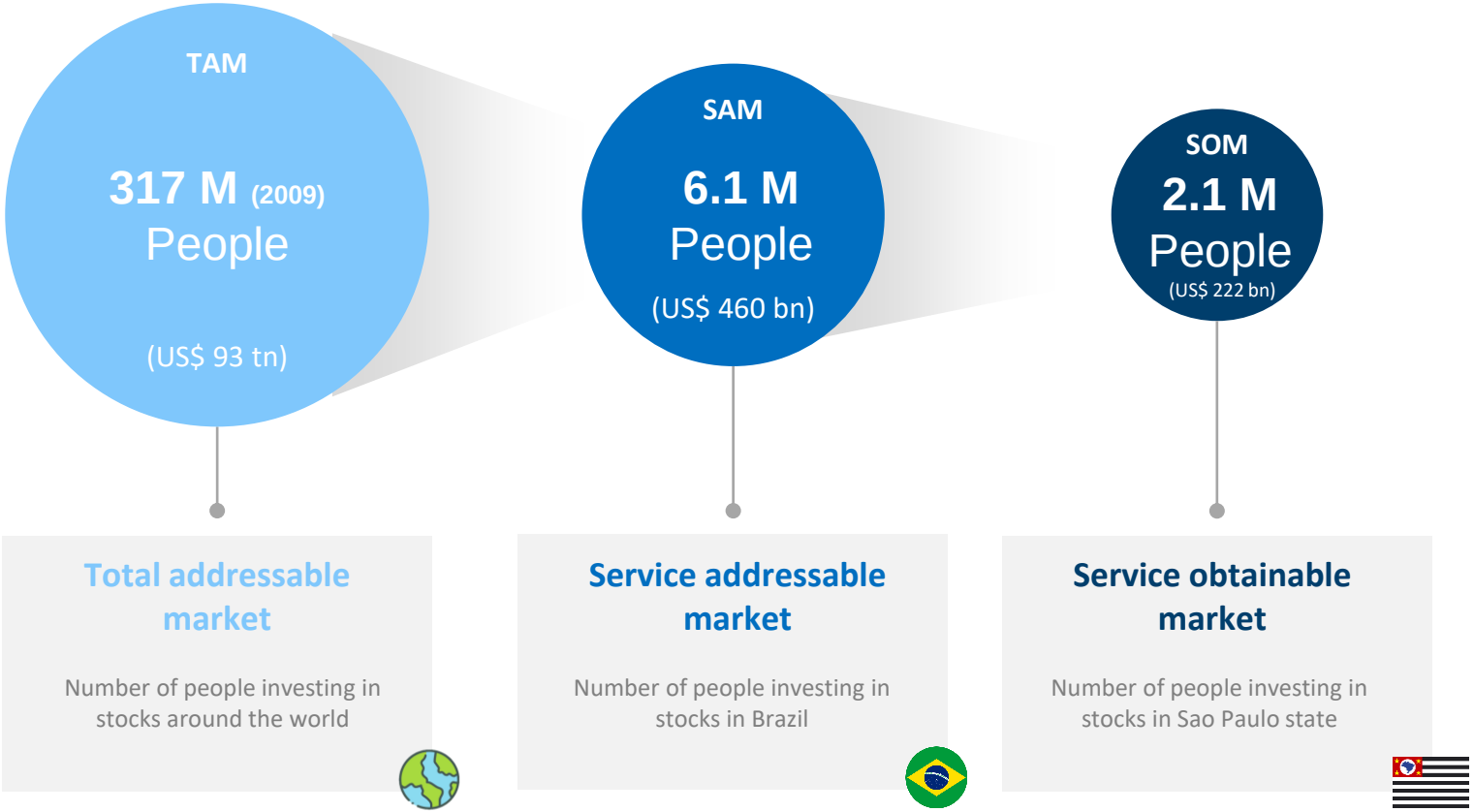
High complexity
The market is constantly changing and influenced by many factors

Knowledge is required
An economic understating of the market and of investing strategies is needed

Problem solving e.g. by investing in the SPDR S&P 500 ETF with automated recommendations

Market Size

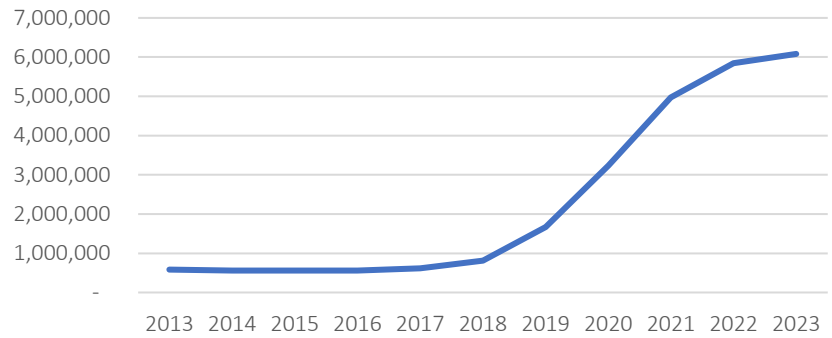
Obtainable market separated geographically and focused on number of customers, suitable for a subscription product



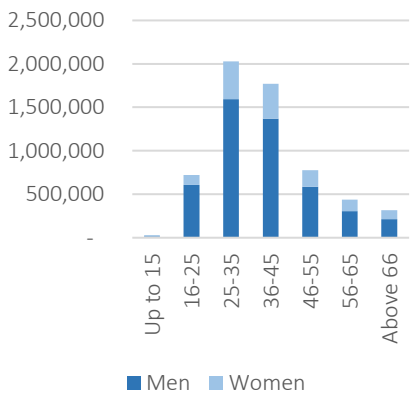
Market Size – Target Customer

Ideal market segmentation for product that increases return on investment

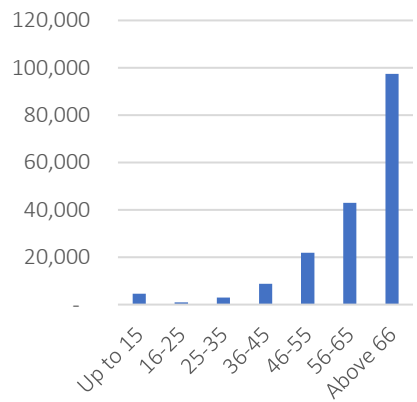
Number of individuals investing in B3



Investor age group (2023)



Average amount invested (US$)



Sharp market growth

Market with intense growth in the past and tendency to continue growing.

Young investors

Average age of investors in the age group with the highest monetary burden due to family construction.

Low investment amount

Investment made in safe options, due to the low value available for riskier investments.

Market Size – Competitors

No direct competitors found in research

World competitors

Analysis Tools

MarketSmith
TradingView
TD Ameritrade

Recomendation Tools

BENZINGA Pro
Robinhood
The Motley Fool

Brazilian competitors

Analysis Tools

?

Recomendation Tools

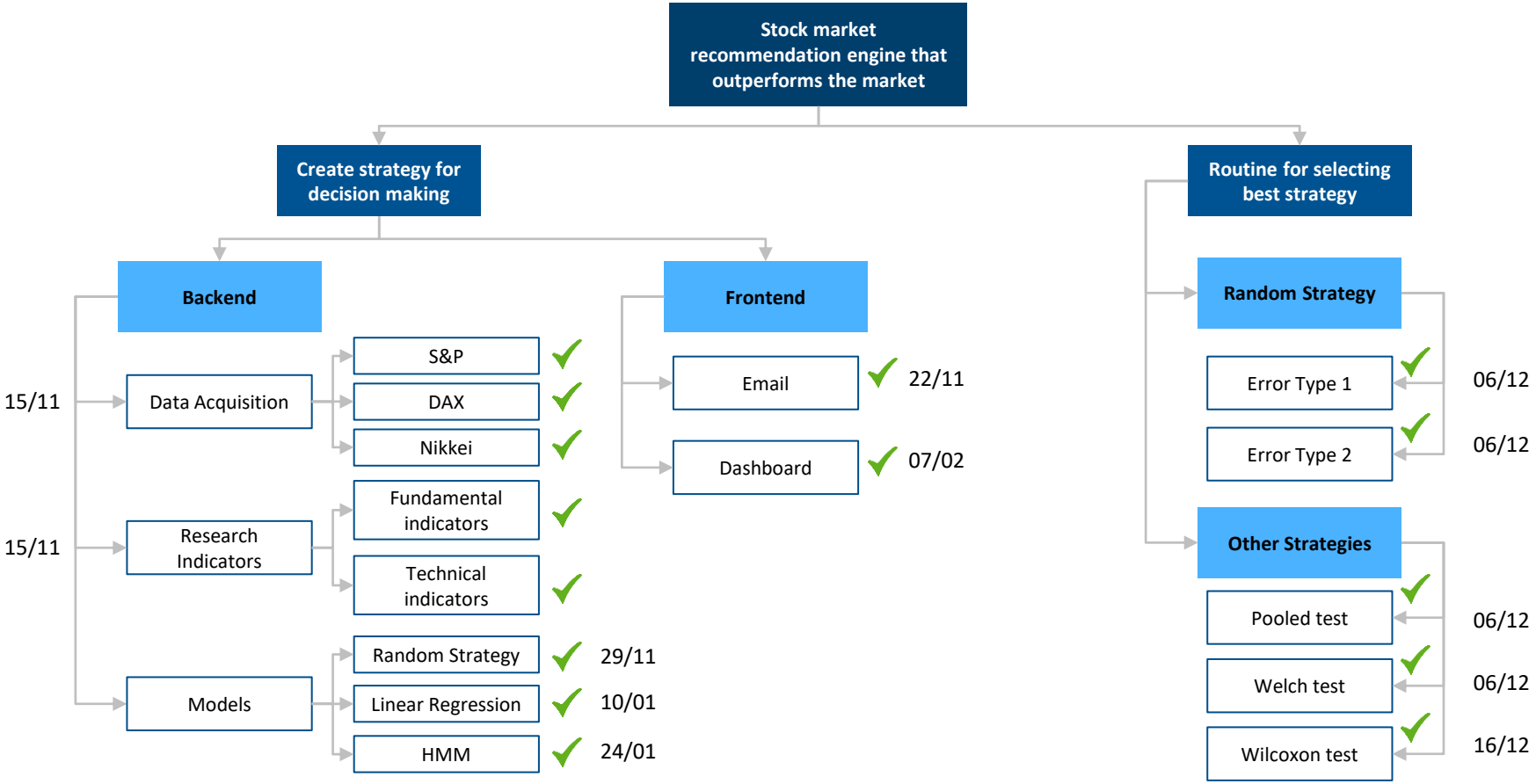
InfoMoney
TERRA INVESTIMENTOS
easynvest
xp investimentos
Guide

- No tool specialized in daily recommendations found in the Brazilian market
- Average ticket per subscription in the US market of 19 dollars per month, access to premium recommendations reaching 1499 dollars/year
- No products found working on our project model.

Index

How do we do it? – Organization

Agile epic structure was laid out to define milestones and tasks



How do we do it? – Research

Initial research indicated that the path of technical analysis was the most promising

General knowledge about stock market

- Stocks
- Commodities
- Mutual funds
- Options
- Bonds
- REITs
- **Index Funds (e.g. ETF)**

Strategies for investing in ETFs

- Dollar Cost Averaging
- Asset Allocation
- **Swing Trading**
- Sector Rotation
- **Short Selling**
- **Betting on seasonal trends**
- Hedge

Swing Trading

- **Technical Analysis**
  - Trend Indicators
  - Momentum indicators
  - Volatility Indicators
  - Volume indicators
  - ...
- **Fundamental Analysis**
  - Earnings per share
  - Price to Earnings ratio (P/E)
  - Projected Earnings Growth (PEG)
  - Free Cash Flow (FCF)
  - Price-to-Book ratio (P/B)
  - Return on Equity (ROE)
  - Dividend Payout Ratio (DPR)
  - Price to Sales Ratio (P/S)
  - Dividend Yield Ratio
  - Debt-to-Equity Ratio (D/E)
- **Alternative Data**

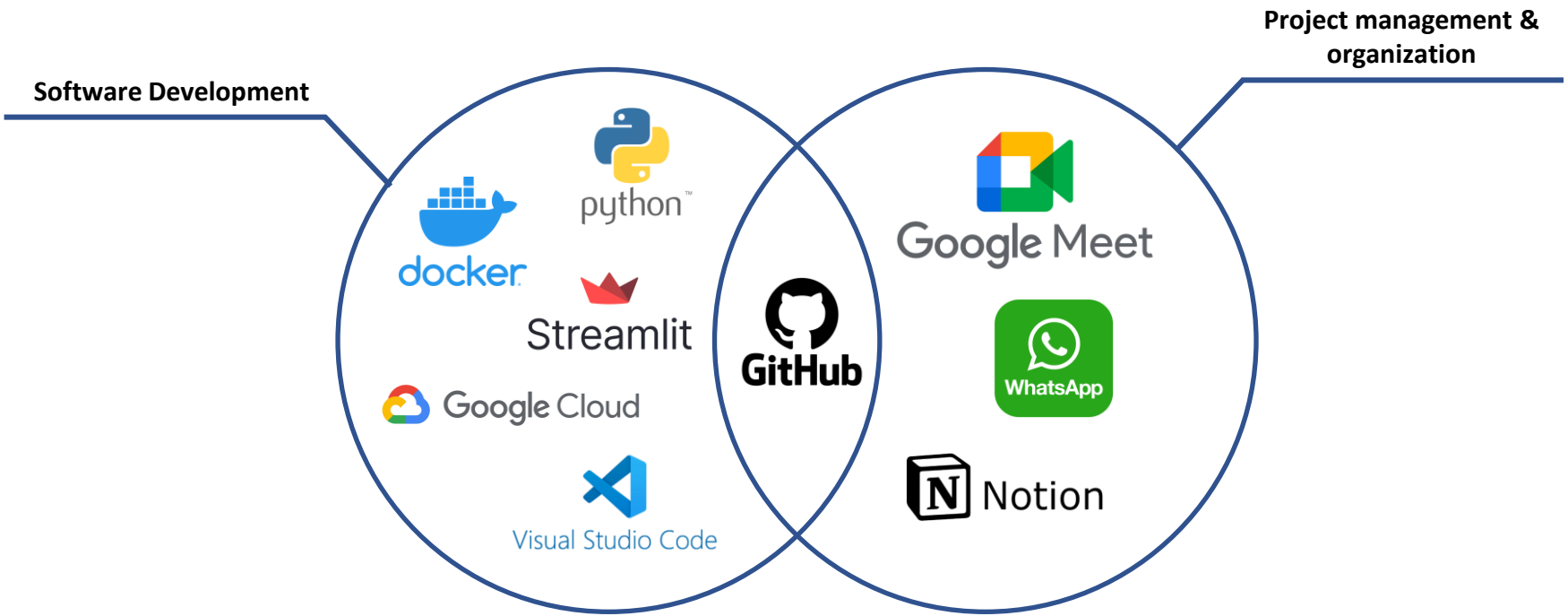
Other Terms

- Candle Stick Indicators
- Hidden Markov Model
- T-Test
- H-Estimation
- SK Learn
- Flask

- Our group understands the tendency to focus on technical analysis

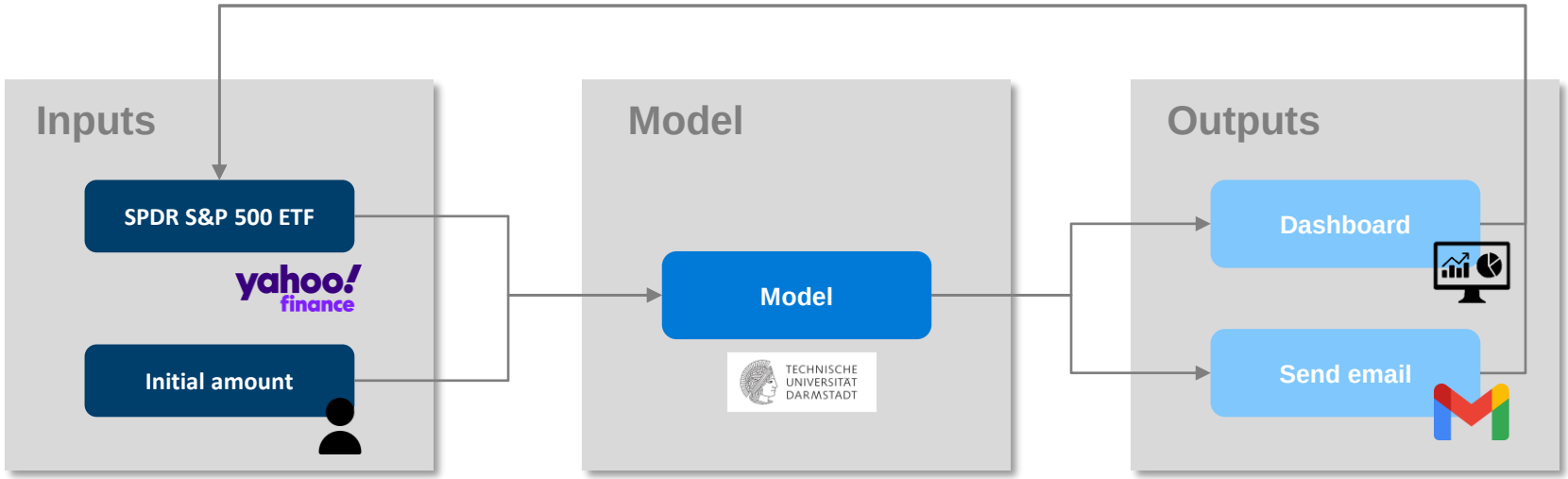
How do we do it? – Technology Stack

Tools available for free online are sufficient to meet fulfill objective with excellence



How do we do it? – Simplified Flow Diagram

Macro structure of the project was built in first phases to increase work organization



- yFinance Library provides real-time multi-stock data
- Input of initial amount available for investment is given by the customer.

- Model developed by the group receives market information and performs analysis.
- The project was structured so that different strategies could be tested and their performance compared

- After calculations, the action is sent via email along with indicators to measure the strategy's performance.
- Link to dahboard with more information is also sent by email.

How do we do it? – Data Processing

Indicators are calculated for the use of the models in later steps

Trend Indicators

Trend indicates the price movement direction, upward or downward, over a certain period of time.

Moving averages
- Trend of the average price
- SMA or EA depending on the purpose

MACD
- Identify reversal of trend
- Lagging indicator, usually paired with RSI

PSAR
- Identify reversal of trend
- Looks at extreme highs and lows

OBV
- Measures how much a stock is traded
- higher the OBV is, higher the stock interest

VWAP
- Relative strength/weakness of a stock
- Stock above VWAP → Bullish and vice verse

AD
- Divergences between price and volume flow
- Helps confirm a rising price trend

Volume indicators

Volume indicates the amount of shares traded in a certain period of time.

Volatility Indicators

Volatility indicates the degree of price fluctuation of a security over a certain period.

BBands
- Envelopes at a standard deviation level above and below a SMA of the price
- Prices tendency to bounce within the bands

ATR
- Measures stock moves, on average, during a given time frame

RSI
- Identify overbought or oversold stock
- Usually paired with MACD

Stochastic RSI
- Stochastic oscillator formula applied to RSI
- Better for sideways or choppy markets

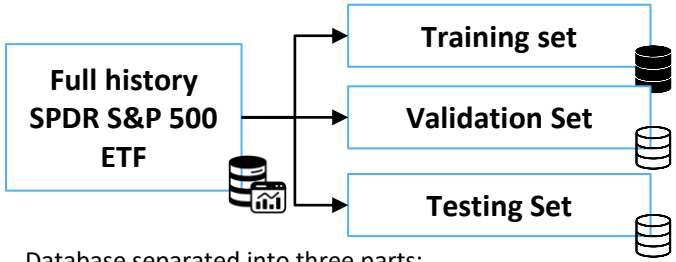
Momentum Indicators

Momentum indicates the rate of price change of a security over a certain period of time.

How do we do it? – Training

Training was done with care to avoid overfitting

Training



Database separated into three parts:

- 1993-2011 (training set),
- 2012-2021 (validation set),
- 2022 (test set)

	S&P	Nikkei	DAX
Training Set	✓	✗	✗
Validation Set	✓	✗	✗
Test Set	✓	✓	✓

- Other actions were used to verify plausibility of results.
- With more time, we could optimize performance also with an eye on other actions.

Selected indicator

Mean portfolio value

$$\frac{\sum_{day\ 1}^{day\ n} (invested\ \$ + not\ invested\ \$)}{n}$$

- In stock markets with low volume, there are effects of volatility and low liquidity, both for buying and selling
- Within this context, mean portfolio value is an appropriate variable, due to low liquidity

How do we do it? – Strategies

Random Strategy (monkey)

Model

Model

TECHNISCHE UNIVERSITÄT DARMSTADT

- Model developed by the group receives market information and performs analysis.
- The project was structured so that different strategies could be tested and their performance compared

Random Model

- Random model was the first to be tested and was set to fill the "model" space while better strategies were developed.

Coin Toss

- Random value between 0 and 1 is generated.
- Buy, sell, and hold action are taken on a "thirds" basis in this range.

Action

- More elaborate strategies were also compared in relation to the monkey, to measure effectiveness.
- The average value of this strategy was expected to fluctuate around the natural course of the SPDR S&P 500 ETF

How do we do it? – Strategies

Linear Regression Model

Benchmark

Linear Regression provides a good benchmark for more complex models. It gives an idea of what can be achieved with simple models and helps to compare more complex models with it.

Interpretability

Linear Regression provides clear and interpretable coefficients that can be used to understand the relationship between the variables.

Simplicity

Simple and easy to implement: Linear Regression is a straightforward and simple approach to modeling. It is easy to understand and implement.

Works better when...

Linear Regression is better when the relationship between the dependent and independent variables is linear: Linear Regression assumes a linear relationship between the variables, and it is most effective when this assumption holds.

Model:

Linear Regression: Single Variable

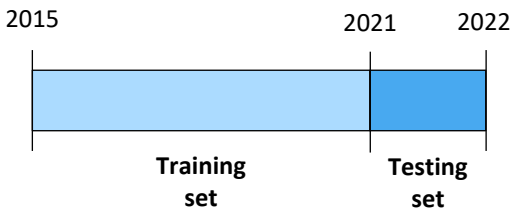
$$\hat{y} = \beta_0 + \beta_1 x + \epsilon$$

Predicted output — Coefficients — Input — Error

Linear Regression: Multiple Variables

$$\hat{y} = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$$

Data split (EoP):



- Assumes a linear relationship between the independent and dependent variables
- Can be sensitive to outliers, and a single outlier can have a significant impact on the model coefficients.

How do we do it? – Strategies

Hidden Markov Model

Flexibility

HMM can model a wide range of dynamic systems with hidden states and observed outputs, making it a versatile tool for many applications.

Temporal Modeling

HMM is particularly useful for modeling sequences of observations over time, making it suitable for speech recognition, handwriting recognition, part-of-speech tagging, and other sequence-based tasks.

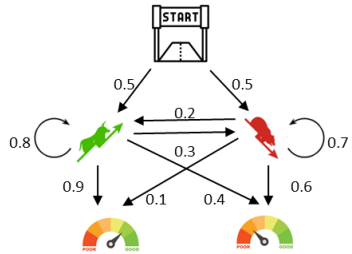
Simple assumptions

HMM makes relatively simple assumptions about the underlying process, making it easier to understand and interpret the results.

Works better when...

HMM is designed to model systems with hidden states and observed outputs, making it a good choice for problems where the hidden states are important for understanding the system.

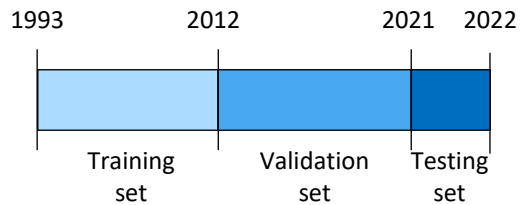
Model:



Hidden variables:
Market system

Observed variables:
Indicators

Data split (EoP):



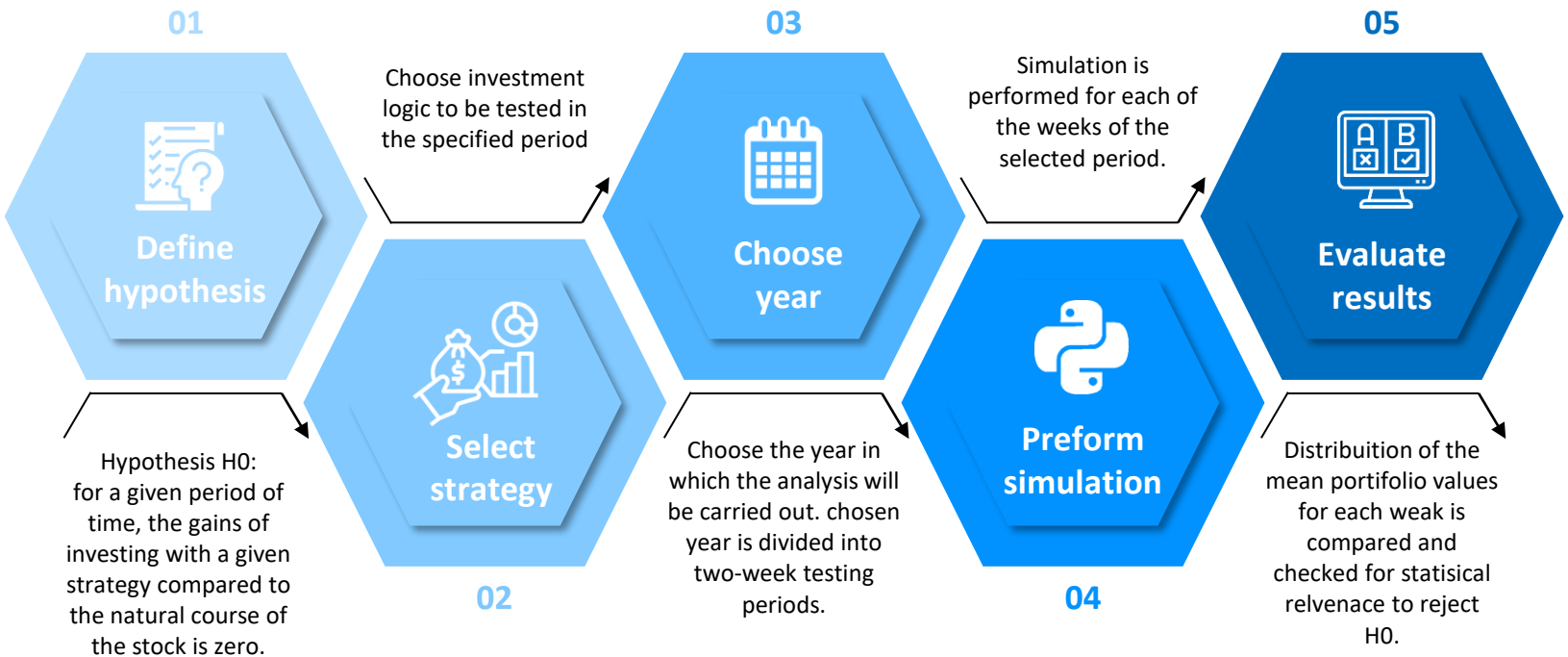
Hyperparameters:

- Latency days = 10
- Hidden States=10

- Markov property
- HMM is susceptible to overfitting, high number of hidden states or the data is limited

How do we do it? – Evaluation

The process of evaluating the strategies was carried out systematically



How do we do it? – Evaluation

The process of evaluating the strategies was carried out systematically

05

Evaluate results

Distribuition of the mean portifolio values for each weak is compared and checked for statisical relvenace to reject H0.

Period splitting

2022 → 26 x 2 weeks → → KPi

1. Year is divided into periods of 2 weeks

2. Simulation with strategy is performed for each period

3. Distribution of mean portfolio value for each period is compared with the natural course of stock

Levene Test

- The Levene test is a statistical test that assesses the equality of variances in different groups.

Welch test

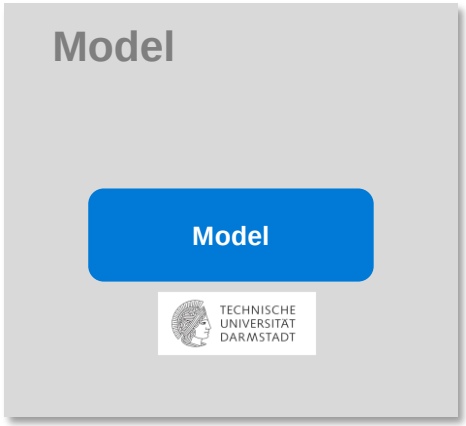
- This test is a modified version of the t-test. It compares the means of two populations when the variances are unequal and/or the sample sizes are different.

Wilcoxon Test

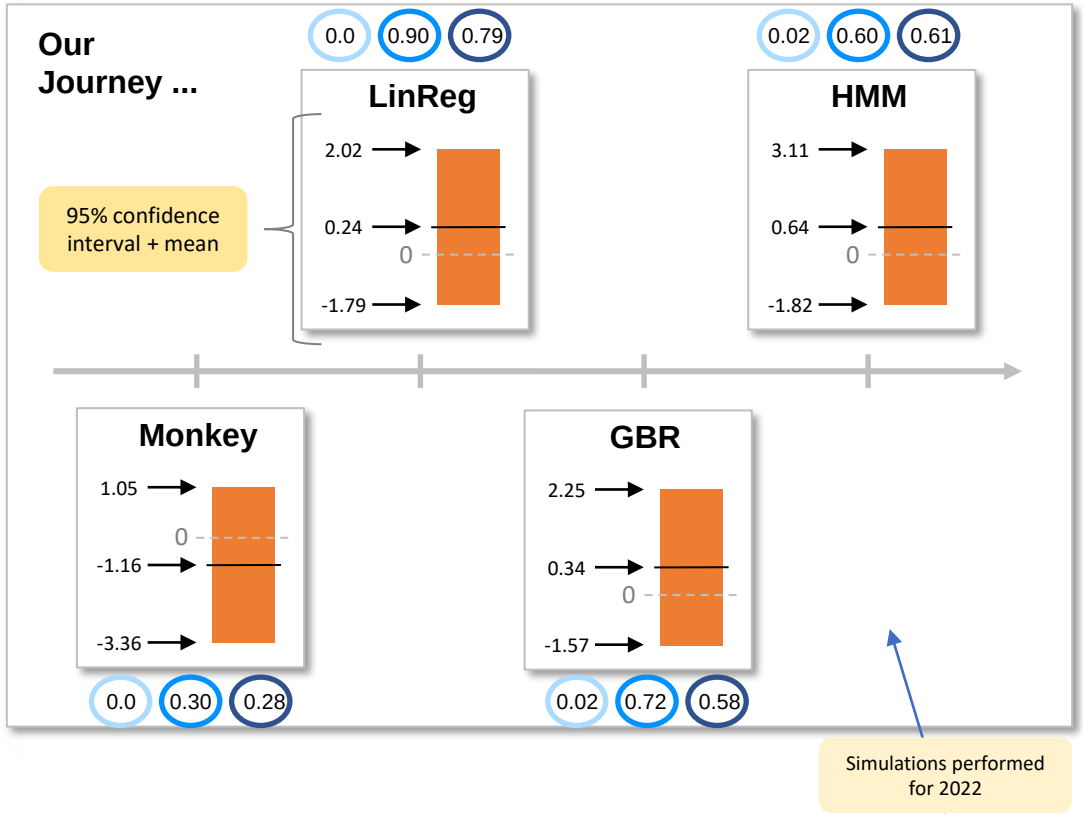
- The Wilcoxon test is a non-parametric statistical test that is used to compare the medians of two related or paired groups, often when the data is not normally distributed.

How do we do it? – Selecting Best Model

Through the KPIs generated by the evaluation, it was possible to select the best model for the deployment



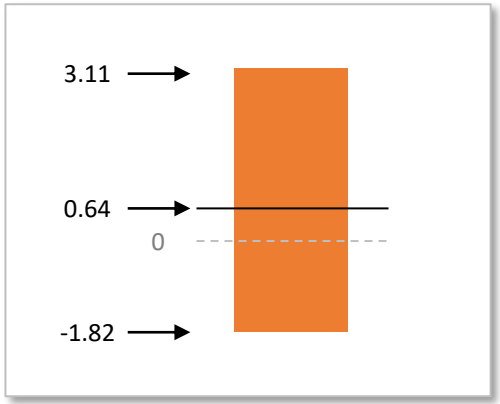
- Model developed by the group receives market information and performs analysis.
- The project was structured so that different strategies could be tested and their performance compared



How do we do it? – Selecting Best Model

The process of evaluating the strategies was carried out systematically

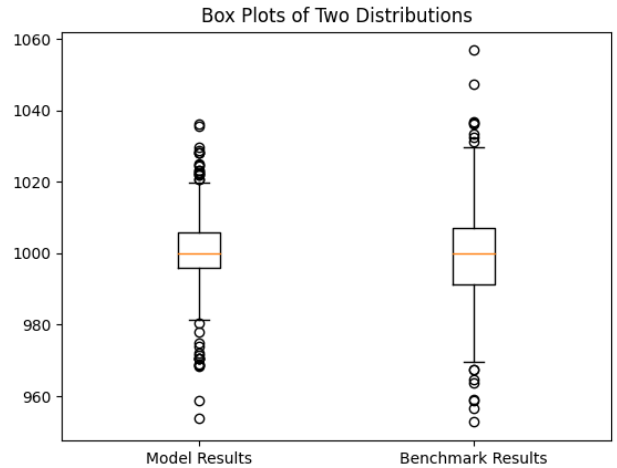
95% prob. interval
(AUC strategy - AUC stock)

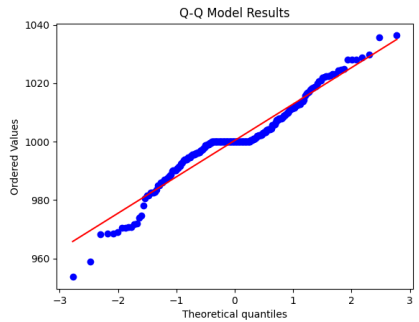


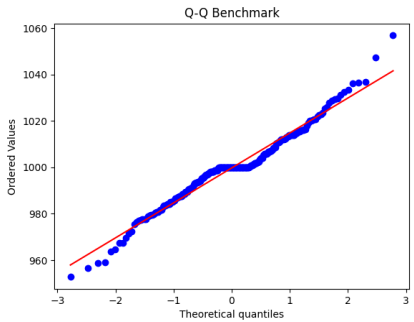
0.02 P-value Levene

0.60 P-value Welch

0.61 P-value Wilcoxon







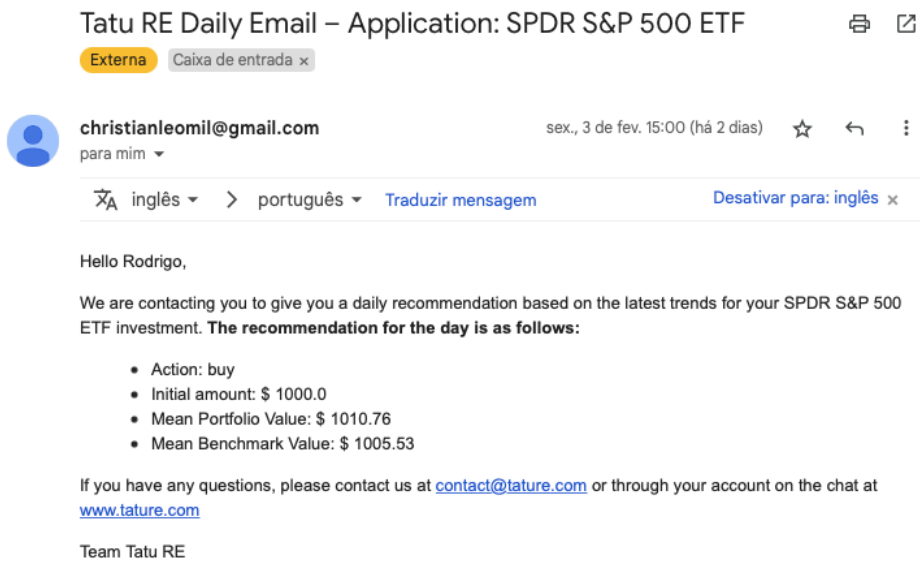
- Levene test shows a big difference in variances, not possible to use pooled test.
- 95% confidence interval show distribution shift slightly above the zero mean
- Although the p value of the tests is not low, there is a numerical difference between the means. The model with the greatest difference was considered the best.

How do we do it? – Output

Recommendation is given directly via email, more details about the dashboard will be shown next

Dashboard

Email



Tatu RE Daily Email – Application: SPDR S&P 500 ETF

Externa | Caixa de entrada ×

christianleomil@gmail.com — sex., 3 de fev. 15:00 (há 2 dias)

para mim

inglês > português | Traduzir mensagem | Desativar para: inglês ×

Hello Rodrigo,

We are contacting you to give you a daily recommendation based on the latest trends for your SPDR S&P 500 ETF investment. **The recommendation for the day is as follows:**

- Action: buy
- Initial amount: $ 1000.0
- Mean Portfolio Value: $ 1010.76
- Mean Benchmark Value: $ 1005.53

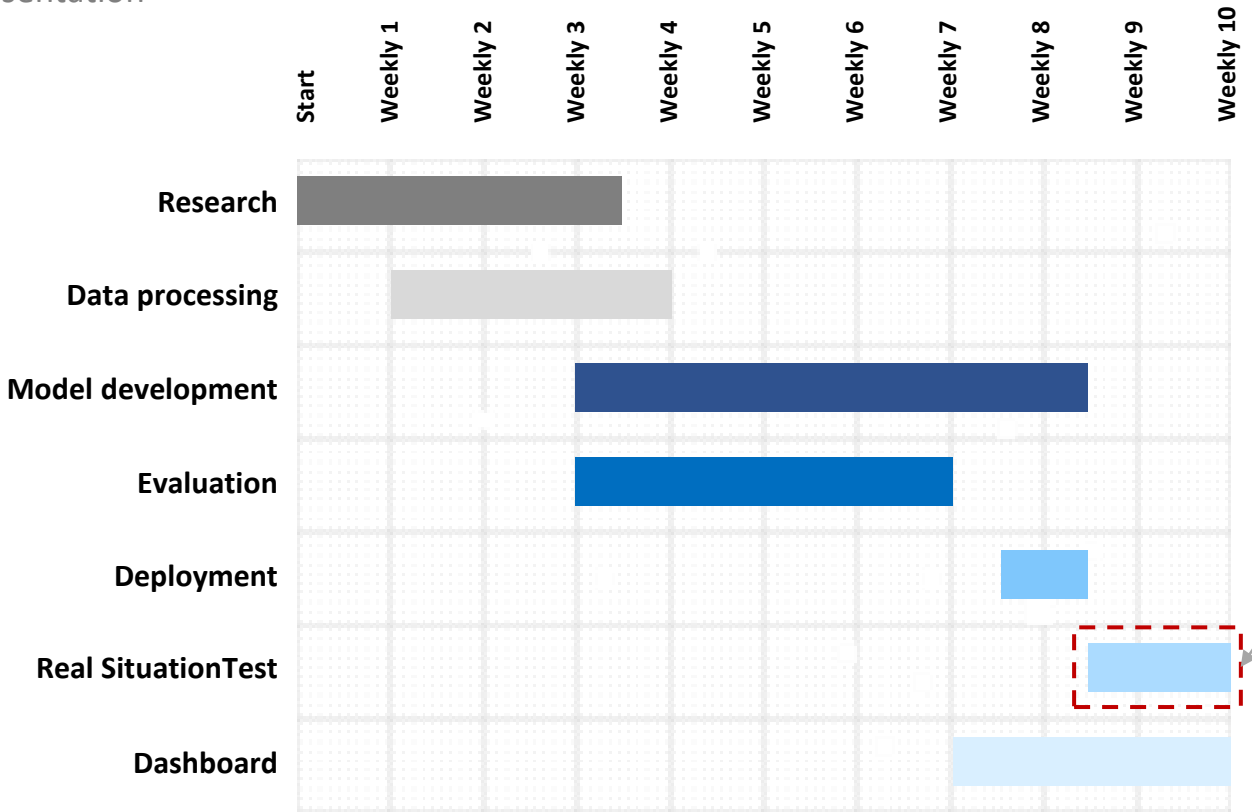
If you have any questions, please contact us at contact@tature.com or through your account on the chat at www.tature.com

Team Tatu RE

Index

Cloud Development– Testing Period

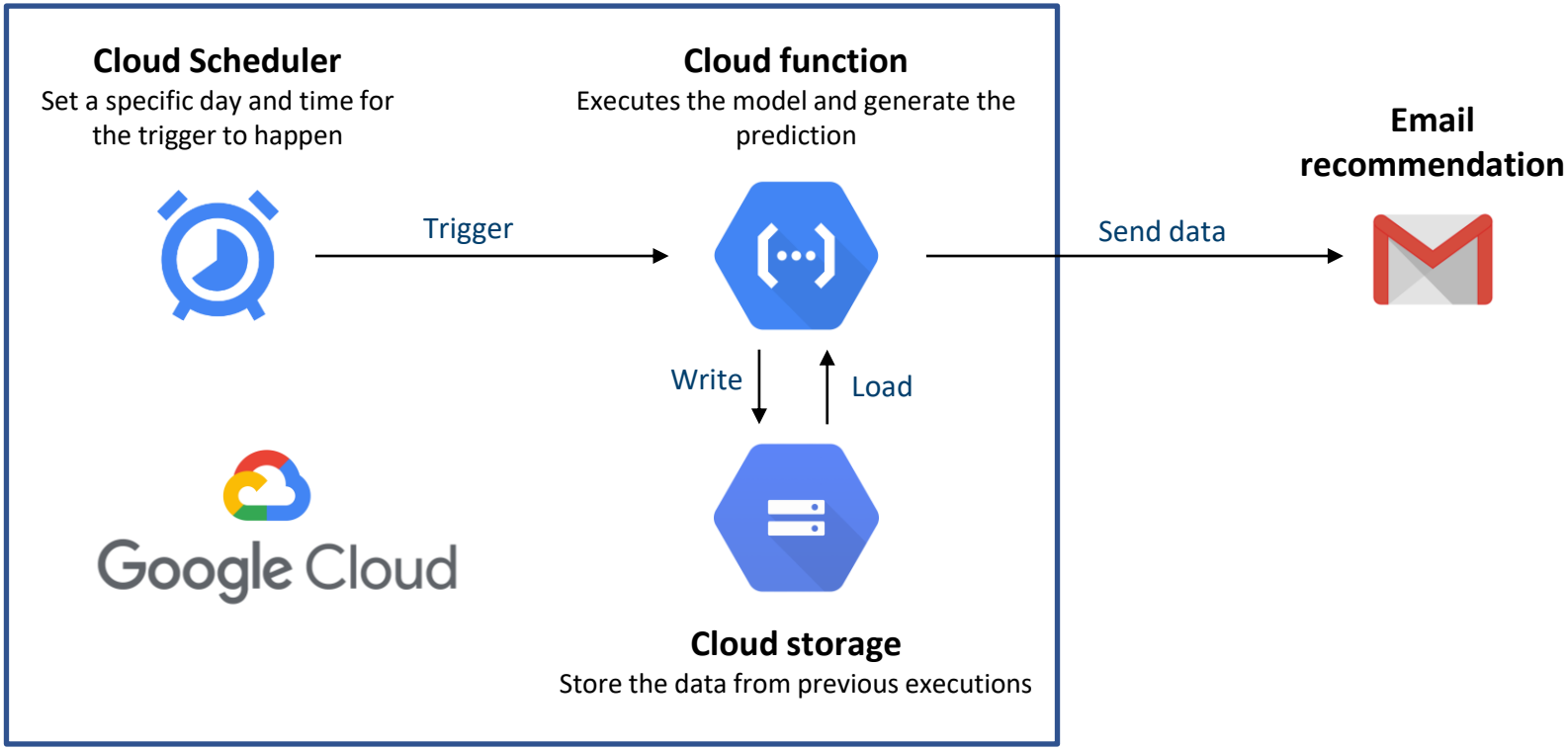
To put the model to real test conditions, it was deployed in a period of two weeks before the presentation



- In order to simulate the routine in a real-world situation, it was deployed into a controlled environment.
- Deployed model was not altered throughout two weeks

Cloud Development– Deployment

Google tools offer the basis we need for deployment



Index

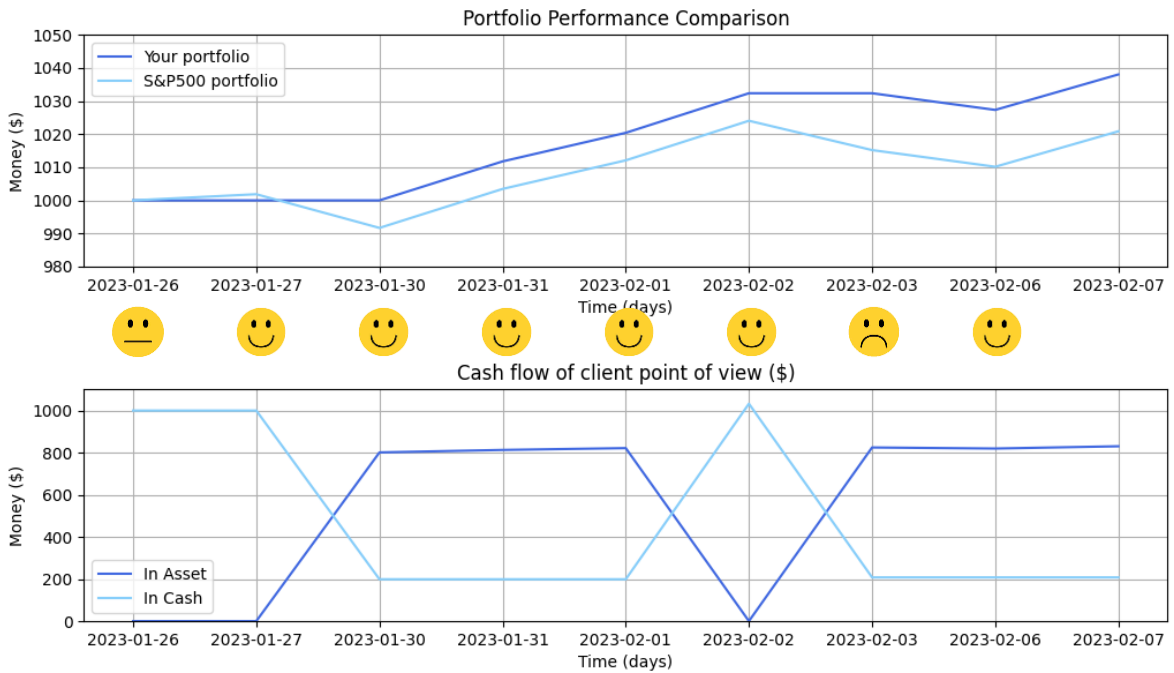
Result – Tatu RE

During the period of ten days, we had very positive results.

Good decision

Neutral decision

Bad decision



	Mean portfolio value	Final value
HMM	1015.54	1038.34
S&P	1008.82	1020.88

76% Greater return on mean value

82% Greater return on final last day

3.8% (264%)* Return on two weeks

Demo

Access QR code to see our dashboard!

Tatu RE thanks you for the atention!

Scan this QR code to have look at our dashboard→