

FINTECH BOOTCAMP - PROJECT 2

Group Members

- | | |
|--------------------|-----------|
| • Asha Devi | Slide 1-2 |
| • Peter Tran | Slide 3-4 |
| • Roshan Paudyal | Slide 5-6 |
| • Stephen Bridgens | Slide 7-8 |

INTRODUCTION

- Our project is about developing the most efficient stock prediction model and enhance the accuracy of the predictions.
- Stock markets were typically projected by financial specialists in the past. However, with the development of machine learning techniques this is now possible to achieve with the automated prediction models.
- We have chosen technical analysis method, which focuses on previous stock prices and values. This analysis uses historical charts and patterns to predict future prices.
- The main reason behind this prediction is to buy & sell stocks at the optimum entry points in order to gain profits.
- Some of the advantages of using prediction model includes:
 - Efficiency
 - Emotion-free trading
 - 24/7 trading
 - Backtesting and optimization
 - Diversification



Data Source: Alpaca



File type: Jupyter Notebook

Questions Posed:

How do we determine the success of our machine learning model and the metrics used to evaluate its performance?

Can we create a successful and profitable machine learning model?



DATA EXPLORATION AND PREPARATION

Data Exploration:

- Google search - Alpaca Portals (API integration)
- Online Articles:
 - a. Simplilearn** - Top ten machine learning Algorithms you need to know in 2023.
<https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article>
 - b. Kaggle** - Stock prediction using linear regression and SVM.
<https://www.kaggle.com/code/mohammedmeligy/stock-prediction-using-linear-regression-and-svm/notebook>
 - c. Medium** – SVM (Support Vector Machine) theory. <https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72>

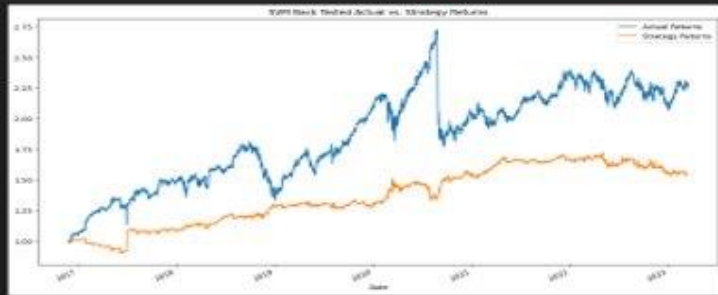
Data Clean-up:

Clean-up in the Jupiter Notebook:

- Reset Index
- Retrieve DataFrame
- Slice dataset
- Drop NAN
- Created new DataFrame with relevant information

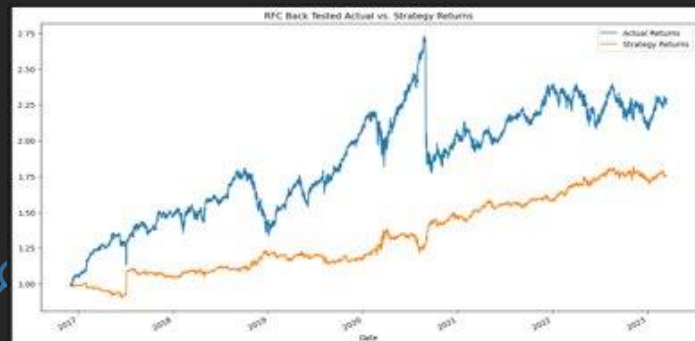
SVM Classification Report on Testing data

	precision	recall	f1-score	support
-1.0	0.80	0.68	0.73	7884
0.0	0.95	0.97	0.96	85206
1.0	0.76	0.69	0.72	5485
accuracy			0.93	98575
macro avg	0.83	0.78	0.80	98575
weighted avg	0.93	0.93	0.93	98575



RFC Classification Report on Testing data

	precision	recall	f1-score	support
-1.0	0.64	0.63	0.64	7884
0.0	0.94	0.94	0.94	85206
1.0	0.60	0.60	0.60	5485
accuracy			0.90	98575
macro avg	0.73	0.73	0.73	98575
weighted avg	0.90	0.90	0.90	98575



ANALYSIS PROCESS

Q1 - How do we determine the success of our machine learning model and the metrics used to evaluate its performance?

Answer: The success of our machine learning model was determined by comparing its performance metrics to the project's goals and the performance of alternative models. We compared two models as a part of our evaluation. Our alternate model had slightly lower accuracy score than the original model.

Steps taken to develop the model:

- Data Import, Slicing : The data was imported from Alpaca using Alpaca API and pre-processed. We sliced the data and created new Data frames as per our requirement.
- Feature engineering: Technical indicators were calculated and added as features to the dataset.
- Splitting the data: The data was split into training and testing datasets.
- Model building: The following classification models were trained and evaluated using the testing data:
 - Support Vector Machine (SVM)
 - Random Forest Classifier (RFC)
- Compared the Strategy returns to Actual returns using line graph.

Analysis continues...

ANALYSIS PROCESS

Q2 - Can we create a successful and profitable machine learning model?

Answer -

- Success and profitability of a machine learning model depend on various factors such as data quality and quantity, feature selection, algorithm choice, and evaluation metrics.
- The SVM and Random Forest models in this project appeared to achieve some level of success and profitability in predicting trading signals and generating returns based on the metrics and backtesting results provided.
- However, past performance is not a guarantee of future results, and there are many unpredictable factors that can influence the success of a trading strategy, such as market conditions and transaction costs.
- It is important to continuously monitor and evaluate the performance of any trading strategy to ensure its continued success and profitability.



Analysis continues...

DISCUSSION

FINDINGS:

SVM Model vs RFC Model Outcome

- Accuracy – SVM 93% and RFC 90% - Good outcome.
- The number of signals points are exactly same for both the buy & sell signal points.
- Sell - 969
- Buy - 974

Actual Returns vs Predicted Strategy Returns

- Plots shows that the strategy returns did not performed well as compared to Actual returns. However, it is a good performance as the accuracy is way above 50%.

POSTMORTEM

- CHALLENGES & RESOLUTION

1. Struggled to find market indicators that would predict profitable outcome.

We calculated the indicators using correlations metrics and used indicators with least correlation to predict our buy and sell signals.

2. Two weeks is not enough for the project like this. We need more time to research on the indicators and perform an evaluation using those indicators.

- ADDITIONAL QUESTIONS AROSE BUT UNABLE TO ATTEMPT DUE TO TIME CONSTRAINT.

Would have researched more market indicators that would have helped us perform more robust parameters to buy and sell stock.

- WHAT WOULD YOU RESEARCH NEXT IF YOU HAD TWO MORE WEEKS?

We would have conducted additional research into utilizing neural networks to enhance the accuracy using sentiment analysis.





QUESTIONS?