# AI6124 - Neuro Evolution and Fuzzy Intelligence Project

Stock Trading with Fuzzy Support Vector Regression (SVR)

By: Anish Madhusudan Rao
Matric ID: G2102625J

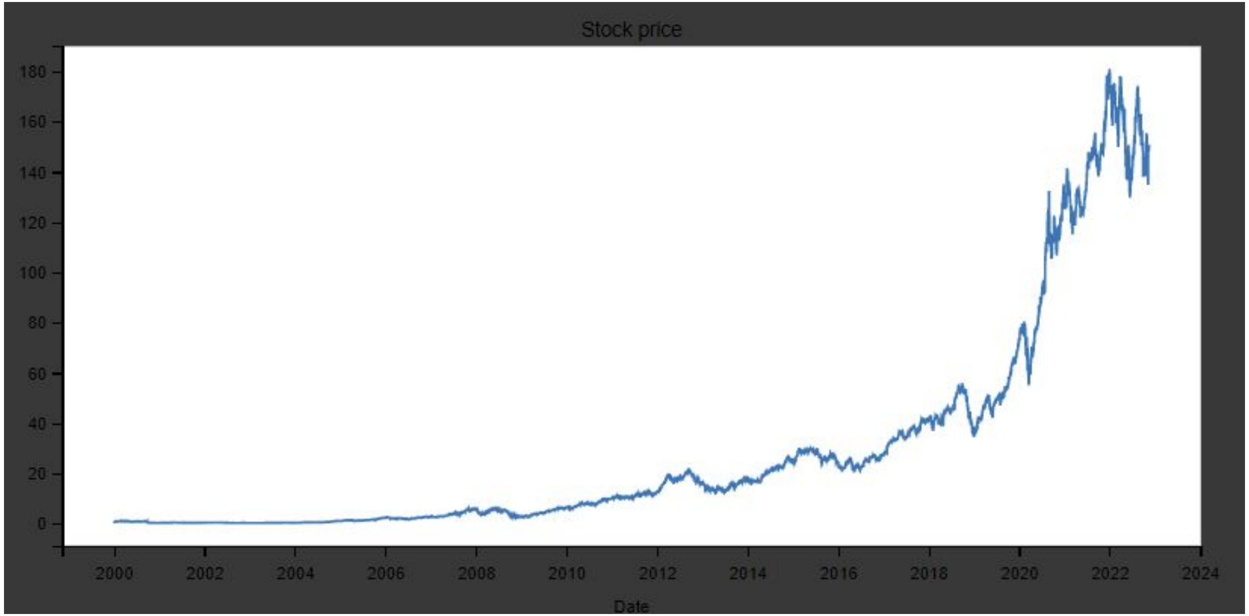
# Introduction

The objective of this project is to implement a fuzzy AI system for stock price prediction and to explore various algorithmic trading strategies.

To that end, the following are my choice of stock, method and baseline:

**Stocks**: AAPL and DBS
**Fuzzy Method**: Fuzzy SVR
**Baseline Method**: Vanilla SVR

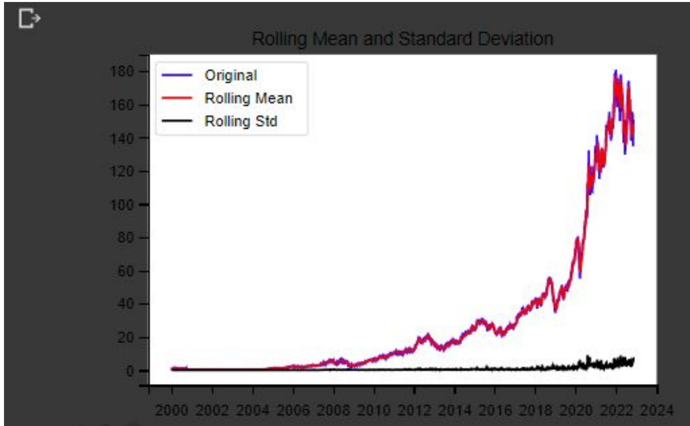
# AAPL Closing Price



# Preprocessing Steps - Stationarity

- Many time series forecasting or analysis methods need the data to be stationary (ex: ARIMA, regression, etc).
- In the case of regression, this is necessary to avoid spurious regression [1] and other related issues.
- So first, we check if the stock price data is stationary or not through the Dickey-Fuller test.

# Preprocessing Steps - Stationarity

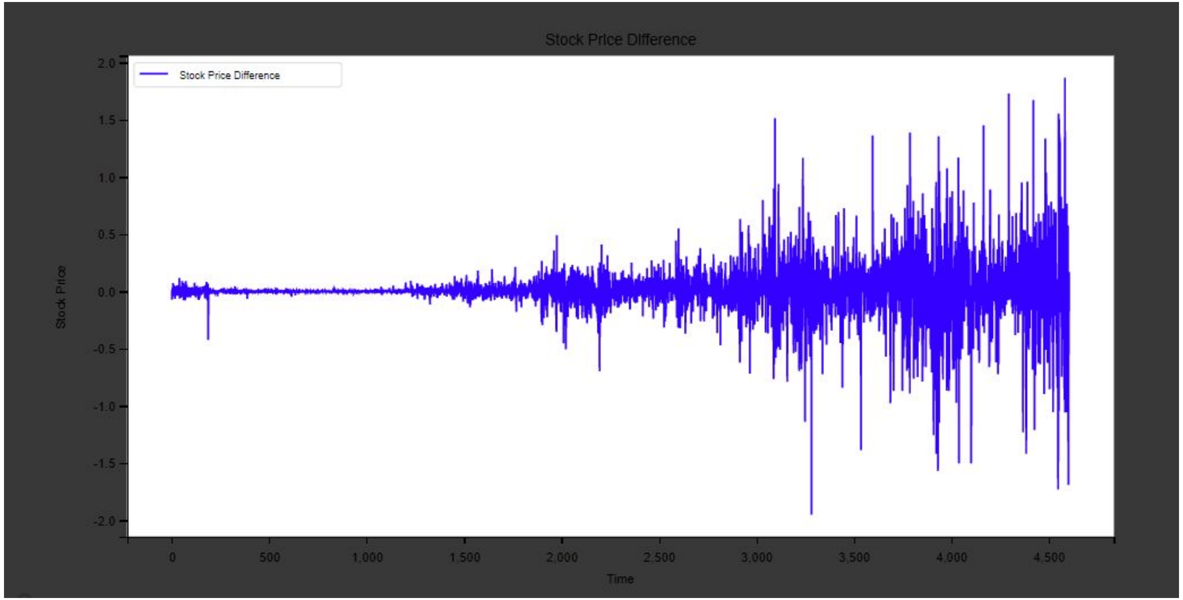



```
Results of dickey fuller test
Test Statistics                    1.582418
p-value                            0.997804
No. of lags used                  34.000000
Number of observations used     5723.000000
critical value (1%)               -3.431493
critical value (5%)               -2.862045
critical value (10%)              -2.567039
```

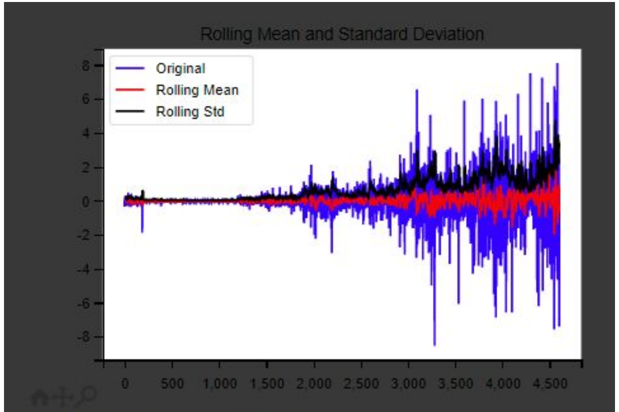
As seen from the p-value, the null hypothesis cannot be rejected. So the close price data is not stationary.

# Preprocessing Steps - Stationarity

To make the data stationary, consider price difference instead:



# Preprocessing Steps - Stationarity




```
Results of dickey fuller test
Test Statistics                  -1.316226e+01
p-value                           1.300108e-24
No. of lags used                  2.500000e+01
Number of observations used       4.580000e+03
critical value (1%)              -3.431779e+00
critical value (5%)              -2.862171e+00
critical value (10%)             -2.567106e+00
```

So the price differenced time series is stationary.

# Vanilla SVR

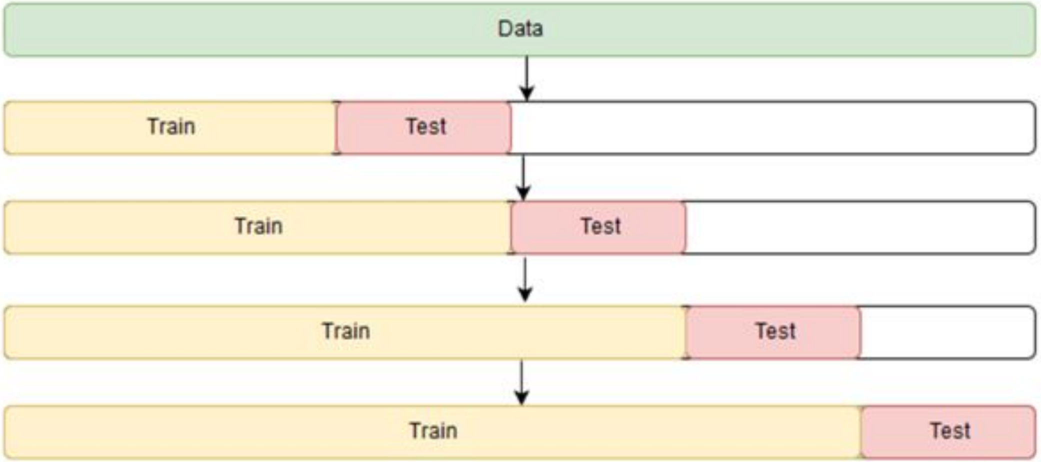
- Support Vector Regression has shown promise in forecasting stock data for well over a decade (ex [14]).
- Its inherent non-linearity and ability to handle the dynamic nature of the stock market allows it to work well for this task.
- In this approach, I use SVR as a baseline by doing windowed prediction on the price difference data.

# Preprocessing steps

- The closing price difference data is split into training and testing sets via a 80:20 split across the whole period.
- The resulting training set is normalized to have zero mean and unit standard deviation. MinMax scaling can also be used for normalization.
- Next, the training set was windowed with 3 time-steps and the 4th was taken as the target for the SVR. This was done as a rolling window.
- Further, the training set is split into validation sets for 3-fold cross validation. But this is not as simple to create since causality needs to be preserved.
- Test data was also normalized using the training mean and standard deviation.

# Preprocessing Steps - Cross Validation Strategy

Rolling Cross Validation is used to preserve causality. In particular, 3-fold cross validation was used in the SVR experiments.



# Vanilla SVR - Choice of hyperparameters

- The choice of hyperparameters greatly influences the final result in case of SVR.
- Choosing the three hyperparameters - C, gamma and epsilon is not trivial and often requires a long process of trial and error.
- Hence, I use Genetic Algorithm to tune the hyperparameters by using the 3-fold-cross validation scores as the fitness of the solution.

# Results and Analysis - Vanilla SVR

Cross Validation Results (in terms of negative root mean square error):

[-0.14211513, -0.48616426, -1.07839629]

RMSE on test set:

| Method | RMSE |
|---|---|
| Baseline SVR | 2.2533855062801282 |
| Pure Random Walk | 2.2541601966960347 |

# Results and Analysis - Vanilla SVR

Empirical results from experiments:

- For the given search space and number of generations given for GA, polynomial kernel slightly outperforms rbf.

- MinMax scaling performs worse than mean and standard deviation normalization.

- There is no appreciable improvement when using up to 10 timesteps in the past instead of 3.

# Fuzzy SVR

- There are a few approaches in literature [13]. Idea is to make the outputs of the SVR interpretable.
- This is done by two parts:  First, we need to make the SVR operate in the input and output space of the linguistic labels.
- Next, we need corresponding fuzzy rules. I have chosen to do this through Mamdani Fuzzy Inference.
- This allows the inference to be done through the SVR for unseen samples and the rules can be used to interpret the result, the best of both worlds.

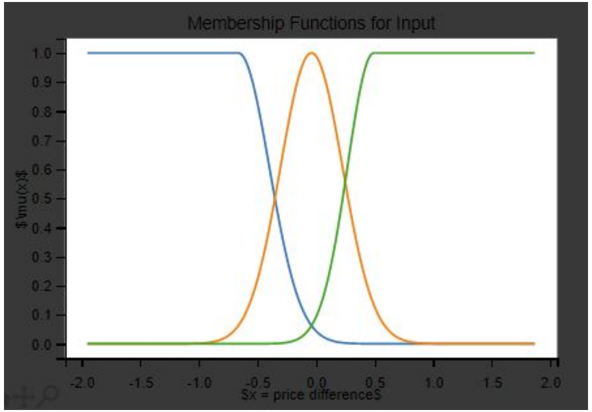
# Fuzzy SVR

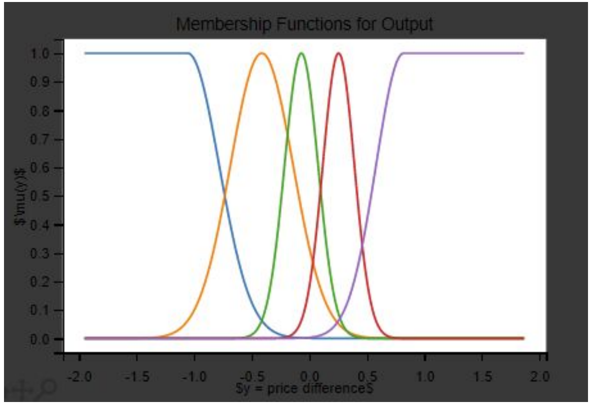
- Same preprocessing steps are needed.
- In addition, the input and output spaces (price difference) need to be mapped into membership functions.
- I have chosen 3 membership functions corresponding to linguistic labels of Decrease, No change and Increase.
- Similarly, I choose 5 membership functions for the output, corresponding to Decrease, Small decrease, No change, Small increase and Increase.

# Input and output membership functions

- Gaussian membership functions are chosen to model the input and output.

- Learning Vector Quantization is used to determine the centroids. It is trained on the training data for the same. The following are the resulting membership functions for input and output:



# Fuzzy SVR

- Now the input to the SVR is the fuzzified (degree of membership) of each of the input membership functions. Correspondingly, the output is the from one of the membership functions for the target.

- So, 5 SVRs are trained, one for each of the output membership functions.

- The result of all the 5 SVRs then need to be combined and defuzzified to give the final answer for the predicted price difference.

# Fuzzy SVR - Defuzzification

Centre of Area (CoA) defuzzification is performed to produce a crisp output from the 5 degrees of membership. It follows the formula below:

$$CoA = \frac{\int_{x_{min}}^{x_{max}} f(x) * x \, dx}{\int_{x_{min}}^{x_{max}} f(x) dx}$$

A discrete approximation of the above formula is used to estimate the defuzzified output in the range of the training values.

# Results and Analysis - Fuzzy SVR

```
Cross validation scores: [0.08739814 0.17509793 0.41338116]
Index 0 SVR RMSE: 1.1537384233837504
Cross validation scores: [0.35569884 0.62504759 0.76492055]
Index 1 SVR RMSE: 0.4536067692644715
Cross validation scores: [0.56588213 1.06905509 1.41490801]
Index 2 SVR RMSE: 1.07607843244563
Cross validation scores: [0.51370495 0.83355902 0.92292848]
Index 3 SVR RMSE: 0.647781724997428
Cross validation scores: [0.03947344 0.27473904 0.57224594]
Index 4 SVR RMSE: 1.2602378756854051
```

| Method | RMSE |
|---|---|
| Fuzzy SVR | 2.2531795407599042 |
| Pure Random Walk | 2.2541601966960347 |

# Results and Analysis - Fuzzy SVR

Empirical results from experiments:

- Using span membership functions instead of the ones made through LVQ did not result in appreciable change in performance.

- Polynomial kernel performed slightly better for estimation compared to rbf.

# Fuzzy Rules

The Fuzzy Rules are structured like the following [week 9 slides]:

$$R_k: \text{IF } x_1 \text{ is } A_{k,1} \text{ AND } \cdots \text{AND } x_{n_1} \text{ is } A_{k,n_1} \text{ Then } y \text{ is } B_k$$

For this system, x means the timestep price difference input, A is the input linguistic label, y is the target and B is the output linguistic label.

# Fuzzy Rules

- Since there are 3 input linguistic labels and 3 timesteps used for prediction, there are 27 rules corresponding to every combination of the input.

- The rules are weighted according to their distribution in the data via a Hebbian Learning process similar to POP learning [week 11 slides]:

$$f_{fw} = min_i(\mu_{C_i}(x_i)),$$

$$f_{bw} = \mu_D(y),$$

$$f_{rule} = f_{fw} * f_{bw}$$

Here, the Cs are the input linguistic terms in the rule in consideration, D is the output linguistic term of the rule. As before, x and y have usual meanings.

# Fuzzy Rules

- Since there is a high number of samples that have no change, the pseudo weights are skewed to be the largest for rules where the consequent is no change.

- There are many sophisticated ways to overcome this issue (see [2] for example). But I've gone with a simple approach of obtaining the pseudo weights through a subset of the data that more representative of the other linguistic labels.

- The obtained fuzzy rules can also be pruned using some heuristics on the values of the pseudo weights.

# Fuzzy Rules

Example rules obtained:

IF x1 is Decrease and x2 is Decrease and x2 is Increase then y is Small Decrease.

IF x1 is Decrease and x2 is Increase and x2 is Decrease then y is Small Decrease.

IF x1 is Decrease and x2 is No change and x2 is No change then y is No change.

Please check the notebook for all rules obtained before and after pruning.

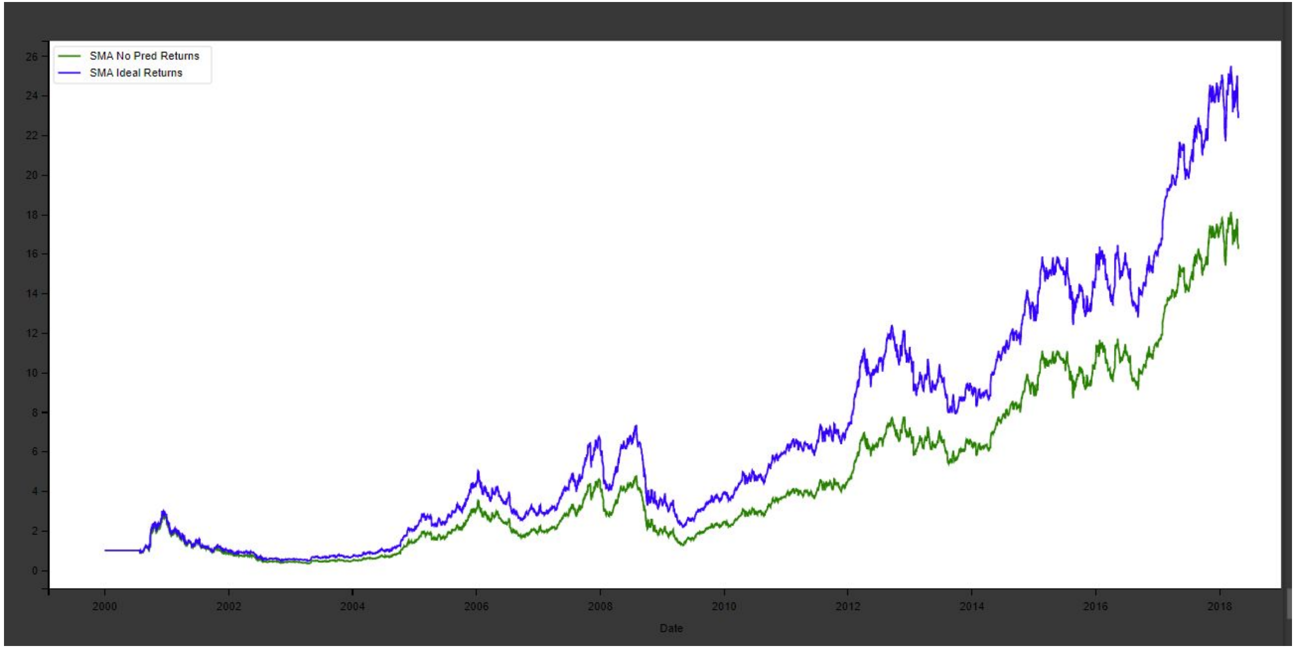
# Algorithmic Trading With SMA Crossover

- Using SMA crossover for the trading signal is a well known method used in algorithmic trading.
- The basic idea is that the crossover between the fast and slow moving average provides the trading signal.
- This method is extremely sensitive to the choice of fast and slow windows. So they need to be tuned on the training set.

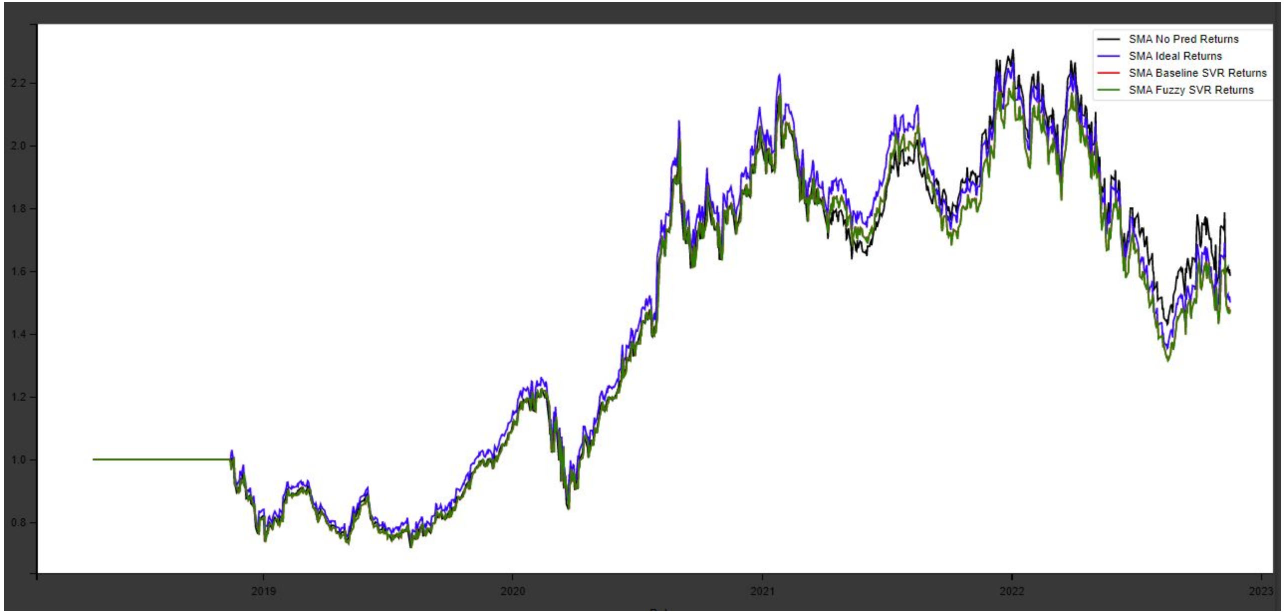
# SMA with GA window values

- Using GA makes choosing the window lengths easier and allows the method to work for a variety of stocks and periods.
- GA picks the two parameters - fast and slow moving average window lengths - by using the average return on training set as the fitness.

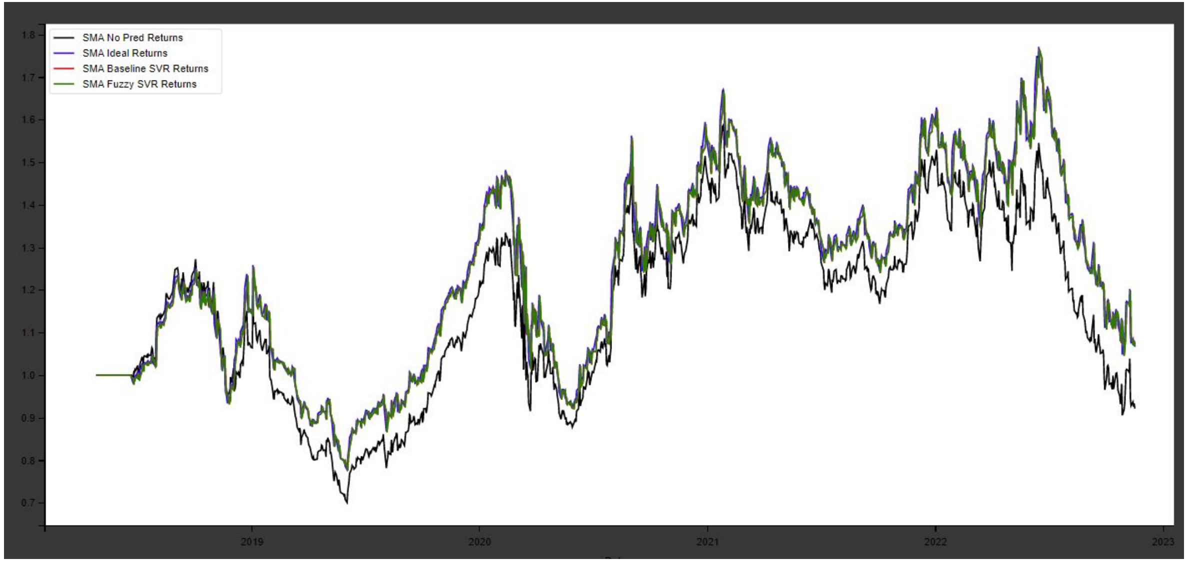
# GA tuned SMA returns on training set



# GA tuned SMA returns on test set



# Handcrafted SMA on test set



# SMA final returns

No forecast final return: 0.93

Ideal forecast final return: 1.08

Baseline SVR final return: 1.06

Fuzzy SVR final return: 1.06

# Analysis - SMA

- The difference in returns between train set and test set with the same window parameters indicates that the train set is not representative of the test period.

- Test area includes the COVID years and in general has a lot more volatility day to day as observed from the initial graph of price differences.

- The difference between ideal and predicted returns are similar because of low RMSE which is smoothed away and does not show up in the training signal as much. However, If the test period was larger, this effect will have grown and there might be a significant difference.

# MACD for algorithmic trading

- Moving Average Convergence Divergence is another well-known method for algorithmic trading.
- The trading signal is obtained by the crossover of the MACD line with its smoother version, called MACD signal.
- The MACD line is the difference between the fast and slow exponential moving average (EMA).

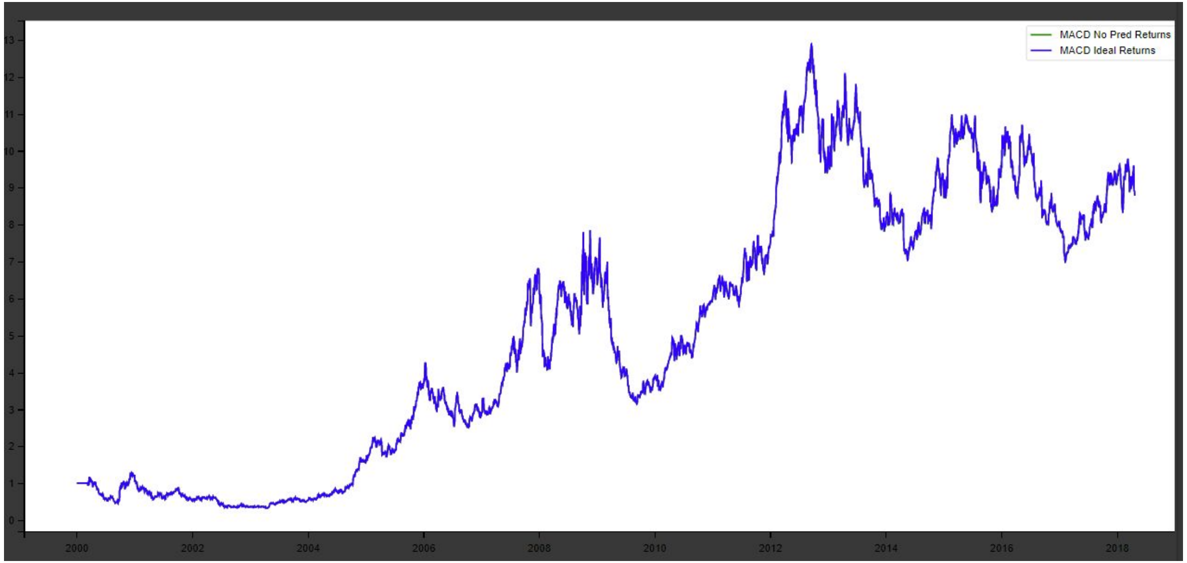
# GA tuning for MACD

- Similar to the SMA crossover, the MACD trade signals are highly susceptible to the values of the 3 smoothing window lengths.

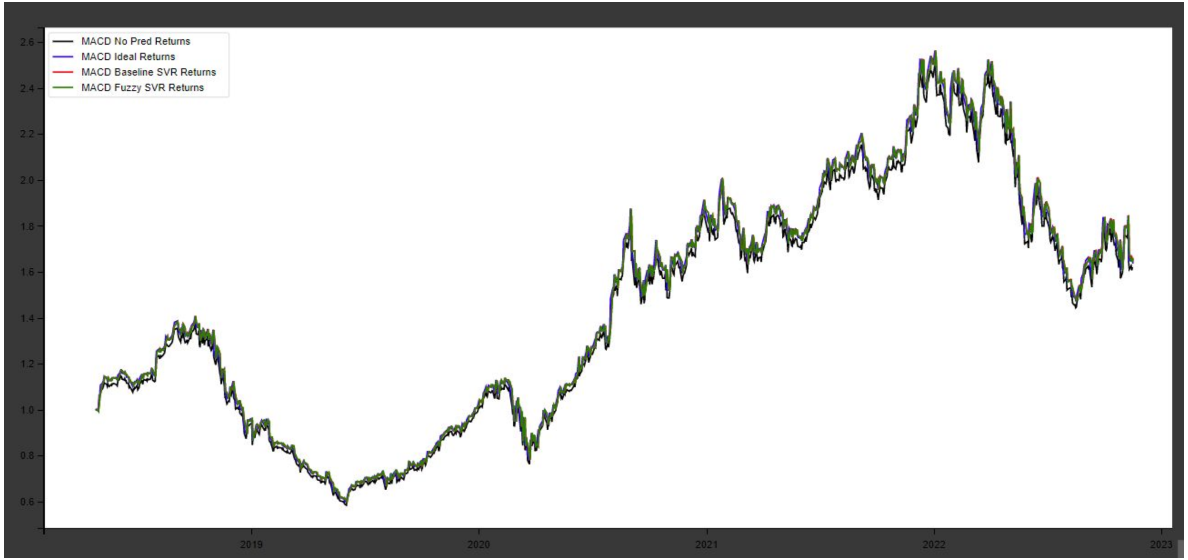
- Using GA will allow this trading mechanism to be applicable to a wide variety of stocks without pain staking manual hyperparameter tuning.

- The GA picks the best window lengths for fast, slow and smoothed MACD lines using the mean return on the training set as the fitness.

# GA tuned MACD returns on training set



# GA tuned returns for testing set



# MACD Final Returns

No forecast final return: 1.60

Ideal forecast final return: 1.65

Baseline SVR final return: 1.64

Fuzzy SVR final return: 1.64

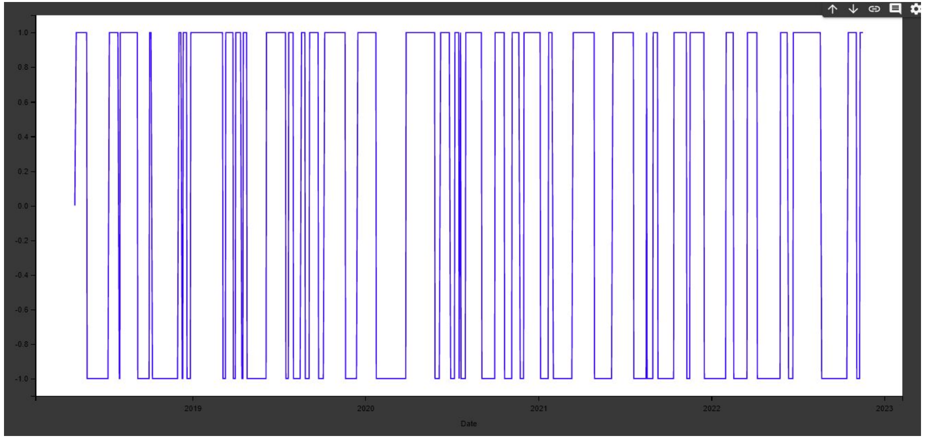
So the Fuzzy SVR matches the baseline but is also interpretable.

# Analysis - MACD


Source: Investopedia

# Analysis - MACD

- Lower windows can lead to more false signals and quick trades (whipsaw effect) which are penalised if there are no checks in place and effect the overall return negatively. Example with [26,12,9] - a typical set of window lengths for MACD:



# Analysis - SMA vs EMA

| | Summary | When to use | What to bear in mind |
|---|---|---|---|
| SMA | The slower-moving average, usually used to confirm a trend rather than predict it. | Good for longer-term trades. Can also be used to calculate EMA – though charts can do this for you. | The slow pace could mean missing a good trade entry point if you rely on SMA alone. |
| EMA | A faster-moving average that places more emphasis on recent price data. | Good for short-term trades where the most current price data is the most relevant. | The average is very reactive to new prices, so you could read too much into price spikes – they might not be part of a trend. |

Source: Nadex

# Advantages of Approach

- SVR is a highly flexible method that comes with many practical advantages like being easy and quick to train.
- In addition, it can be used in conjunction with other methods via a better embedding. For example, LSTM embeddings can be used with SVR to better encode the input and output space [4, 5].
- The Fuzzy SVR can also be fine tuned for the individual linguistic labels based on expert knowledge, which will improve its interpretability as well. This level of tuning is not possible with the vanilla SVR.
- Such possibilities make it an ideal candidate for ensemble models as well, since the underlying architecture doesn't need to change and specialized SVRs can be build whose individual outputs can be combined for the final prediction [6].

# Advantages of Approach

- The use of GA for hyperparameter tuning allows the entire pipeline to be used for any stock or period with no changes.
- Even if a new trading method needs to be tried, the optimal hyperparameters considering returns will be chosen by the GA. This requires only a single line change in the fitness function.
- In all, the Fuzzy SVR matches baseline performance and provides close to ideal returns with the given constraints in terms of the trading algorithm, stock and period chosen.

Note: LSTM and ARIMA have also been explored. Experiments and results can be found in the jupyter notebook.

# Disadvantages of Approach / Improvements

- Prediction RMSE can be lower. Right now, it is only slightly better than random walk for a small period. It might improve with a larger test period.
- Trading method works only for individual stock and not in a portfolio management setting.
- The GA parameter search space can be widened and run for more generations.
- TS Fuzzy inference system can be tried instead of Mamdani [3].

# Future Work

- Deep Learning methods are the major type of stock prediction algorithms in research at the moment [7,8,9]. Incorporating fuzzy rule sets with these models can lead to improved interpretability.
- Fuzzy modelling for stock portfolio selection [11] as well as management [10] is another interesting direction.
- Applying these techniques to trading bots can lead to improved debugging and understanding of how they perform and how to correct for particular issues.
- Incorporating other real-world indicators and social media information can lead to improved results. This entire system can be modelled in the fuzzy setting to improve interpretability. Of course, it is always a tradeoff [week 9 slides].

# References

[1] David E. A. Giles (2007) Spurious Regressions with Time-Series Data: Further Asymptotic Results, Communications in Statistics - Theory and Methods, 36:5, 967-979, DOI: 10.1080/03610920601041499

[2] Fernández, A., del Jesus, M.J., Herrera, F. (2009). Improving the Performance of Fuzzy Rule Based Classification Systems for Highly Imbalanced Data-Sets Using an Evolutionary Adaptive Inference System. In: Cabestany, J., Sandoval, F., Prieto, A., Corchado, J.M. (eds) Bio-Inspired Systems: Computational and Ambient Intelligence. IWANN 2009. Lecture Notes in Computer Science, vol 5517. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02478-8_37

[3] Juang, C. F., & Hsieh, C. D. (2009). TS-fuzzy system-based support vector regression. *Fuzzy Sets and Systems*, *160*(17), 2486-2504.

[4] Ye, Y., Wang, L., Wang, Y., & Qin, L. (2022). An EMD-LSTM-SVR model for the short-term roll and sway predictions of semi-submersible. *Ocean Engineering*, *256*, 111460.

[5] Guo, J., Xie, Z., Qin, Y., Jia, L., & Wang, Y. (2019). Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM. *IEEE Access*, *7*, 42946-42955.

[6] Deng, Y. F., Jin, X., & Zhong, Y. X. (2005, August). Ensemble SVR for prediction of time series. In *2005 International Conference on Machine Learning and Cybernetics* (Vol. 6, pp. 3528-3534). IEEE.

[7] Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015, June). Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*.

[8] Karim, M., Foysal, M., & Das, S. (2023). Stock Price Prediction Using Bi-LSTM and GRU-Based Hybrid Deep Learning Approach. In *Proceedings of Third Doctoral Symposium on Computational Intelligence* (pp. 701-711). Springer, Singapore.

[9] Polamuri, S. R., Srinivas, K., & Mohan, A. K. (2022). Multi-model generative adversarial network hybrid prediction algorithm (MMGAN-HPA) for stock market prices prediction. *Journal of King Saud University-Computer and Information Sciences*, *34*(9), 7433-7444.

[10] Yang, X. Y., Chen, S. D., Liu, W. L., & Zhang, Y. (2022). A Multi-period Fuzzy Portfolio Optimization Model with Short Selling Constraints. *International Journal of Fuzzy Systems*, 1-15.

[11] Rasoulzadeh, M., Edalatpanah, S. A., Fallah, M., & Najafi, S. E. (2022). A multi-objective approach based on Markowitz and DEA cross-efficiency models for the intuitionistic fuzzy portfolio selection problem. *Decision Making: Applications in Management and Engineering*, *5*(2), 241-259.

# References

[12] Fung, K. F., Huang, Y. F., & Koo, C. H. (2019). Coupling fuzzy–SVR and boosting–SVR models with wavelet decomposition for meteorological drought prediction. *Environmental Earth Sciences*, *78*(24), 1-18.

[13] Chang, P. C., Wu, J. L., & Lin, J. J. (2016). A Takagi–Sugeno fuzzy model combined with a support vector regression for stock trading forecasting. *Applied soft computing*, *38*, 831-842.

[14] Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. The Journal of finance and data science, 4(3), 183-201.