EE4305: Fuzzy/ Neural Systems for Intelligent Robotics

Data-Driven Fuzzy Inference System for Stock Price Movement Prediction

Submitted by

Tey Kai Cong (A0203512Y)

Academic Year 2022/2023

1. INTRODUCTION

Stock market prediction is a highly challenging real-world problem due to the inherent uncertainties and fluctuations in financial data. Nevertheless, it remains an attractive area for investors as they strive to maximize their returns. Traditional prediction models often rely on historical financial performance, economic indicators, industry trends, and even the past performance of company directors. [1] However, the stock market is also influenced by noise generated by forecasters and economists who make market predictions about the next big boom or bust. [2]

Given the uncertainties, ambiguities, and complexities of stock trading, utilizing Fuzzy Logic can be a powerful tool to address these ill-posed problems, even in the presence of incomplete information. [3] Applying Fuzzy Logic to the stock market enables more accurate buy/hold/sell trading decision predictions for the listed securities of various exchanges, such as the NSE, by effectively handling the uncertainties, noise, nonlinearity, and ambiguity inherent in the financial domain. [3]

2. OBJECTTIVE

In this report, we propose a Fuzzy Inference System (FIS) to predict stock price movements by incorporating both traditional features and engineered features, such as moving averages, relative strength index, and rate of change of stock price. Our approach demonstrates the potential of Fuzzy Logic in tackling the challenging problem of stock market movement, by incorporating the predicted price difference into their strategies, investors can potentially improve their risk management, optimize their portfolio performance, and maximize their returns. All source code for this report can be found on https://github.com/kaicong12/FuzzyLogicStockPrediction

3. METHODOLOGY

3.1 Input and Output of the Fuzzy Inference System (FIS)

The dataset used in this experiment can be found on Kaggle. [4] The dataset used in this study comprises historical stock price data for Netflix Inc. (NFLX) spanning from February 5th, 2018, to February 5th, 2022. Besides, the dataset comprises of the following raw features:

1. Low: The lowest traded price of the stock during the specific trading day.
2. Open: The price at which the stock commenced trading at the beginning of a trading day.
3. High: The highest traded price of the stock during a specific trading day.
4. Close: The final price at which the stock was traded at the end of a trading day.
5. Volume: The total number of shares traded during a specific trading day.

To enhance the performance of the fuzzy inference system, several feature engineering techniques were applied to compute additional features that could contribute to better stock price prediction. [3] The features computed include:

1. 10-day Moving Average (MA10): The 10-day moving average is calculated by averaging the stock prices over the past 10 trading days. This feature smoothens the price series and helps identify the short-term trend in the stock prices.

$$\text{MA10} = \frac{P_1 + P_2 + \cdots + P_{10}}{10}$$

Figure 1: Equation for 10-days MA, where P_n represents the stock price on day n

2. 50-day Moving Average (MA50): Similar to the 10-day moving average, the 50-day moving average is computed by averaging the stock prices over the past 50 days. This feature is useful for identifying the intermediate-term trend in the stock prices.

$$\mathrm{MA50} = \frac{P_1 + P_2 + \cdots + P_{50}}{50}$$

Figure 2: Equation for 50-days MA

3. Relative Strength Index (RSI): The Relative Strength Index is a momentum oscillator that measures the speed and change of price movements. It ranges from 0 to 100 and is calculated using the average gain and average loss over a specified period (typically 14 days). The RSI can help identify overbought or oversold conditions in the stock price.

$$\mathrm{RSI} = 100 - \left(\frac{100}{1 + \left(\frac{\text{Avg Gain}}{\text{Avg Loss}} \right)} \right)$$

Figure 3: Equation for Relative Strength Index, where Avg. Gain and Avg. Loss are the average gain and loss over the specified period, respectively

4. Rate of Change (ROC): The Rate of Change is a momentum indicator that measures the percentage change in the stock price between the current day and a specified number of days prior (e.g., 10 days). This feature helps identify the momentum and trend strength in the stock price.

$$\mathrm{ROC} = \left(\frac{P_t - P_{t-n}}{P_{t-n}} \right) \times 100$$

Figure 4: Equation for Rate of Change, where P_t represents the stock price at time t and P_{t-n} is the stock price n days prior

The output variable for this system represents the predicted stock price movement for the next trading day. To calculate the stock price movement, we find the difference between the closing prices of two consecutive trading days, which can either be positive or negative. This continuous representation of stock price movement is used as the target variable for training and evaluating the FIS.

3.2 Fuzzification

Drawing on earlier studies on prediction systems using Fuzzy Inference Systems, it has been observed that when there is a lack of prior knowledge regarding the fuzzification process, Fuzzy C Means Clustering can be an effective approach to determine the appropriate threshold values for each linguistic variables. [5][6] This section outlines the fuzzification process employed in our study, with a focus on the use of Fuzzy C-Means clustering to determine the parameters. For the sake of simplicity, triangular membership functions are employed for fuzzifying the input variables.

3.2.1 Fuzzy C Means

In order to transform crisp values into fuzzy values, Fuzzy C Means clustering is used to compute the centroids for the membership functions. The resulting centroids are then used as the peak values for the membership functions. Taking into account the presence of edge values which lie beyond the range defined by the computed "Low" and "High" centroids, the Low and High membership functions are extended to ensure that values smaller than the Low centroid or larger than the High centroid would still be assigned a non-zero membership grade. Table below summarizes the fuzzfied input variables with their respective linguistic variables and range. (Table 1)

Table 1: Summary of Fuzzified Input Variables with their linguistic variables and range

Input Variable	Linguistic Variable	Range
Low	Low, Medium and High	Low is between 220.68 and 368.04 Medium is between 294.36 and 530.12 High is between 386.04 and 692.21
High	Low, Medium and High	Low is between 231.65 and 380.12 Medium is between 305.98 and 544.94 High is between 380.12 and 709.77
Open	Low, Medium and High	Low is between 226.41 and 374.64 Medium is between 300.54 and 537.94 High is between 374.64 and 701.25
Close	Low, Medium and High	Low is between 162.23 and 500.63 Medium is between 331.43 and 614.62 High is between 500.63 and 728.62
Volume	Low, Medium and High	Low is between -1622679.78 and 10100047.89 Medium is between 4390464.56 and 20134352.23 High is between 10100047.89 and 35868739.56
ShortTermMA	Low, Medium and High	Low is between 226.90 and 373.75 Medium is between 300.81 and 536.71 High is between 373.75 and 699.66
LongTermMA	Low, Medium and High	Low is between 160.25 and 504.70 Medium is between 332.47 and 620.42 High is between 504.70 and 736.14
Relative Strength Index	Negative, Neutral, Positive	Oversold is between 14.72 and 52.30 Neutral is between 33.01 and 71.72 Overbought is between 52.30 and 90.42
Rate of Change	Oversold, Neutral, Overbought	Negative is between -6.96 and -0.14 Neutral is between -3.55 and 3.08 Positive is between -0.14 and 6.30

Figure below illustrates the Mamdami Fuzzy Inference System with the given 9 inputs, their respective membership functions and parameters, and 1 output variable. (Figure 5)

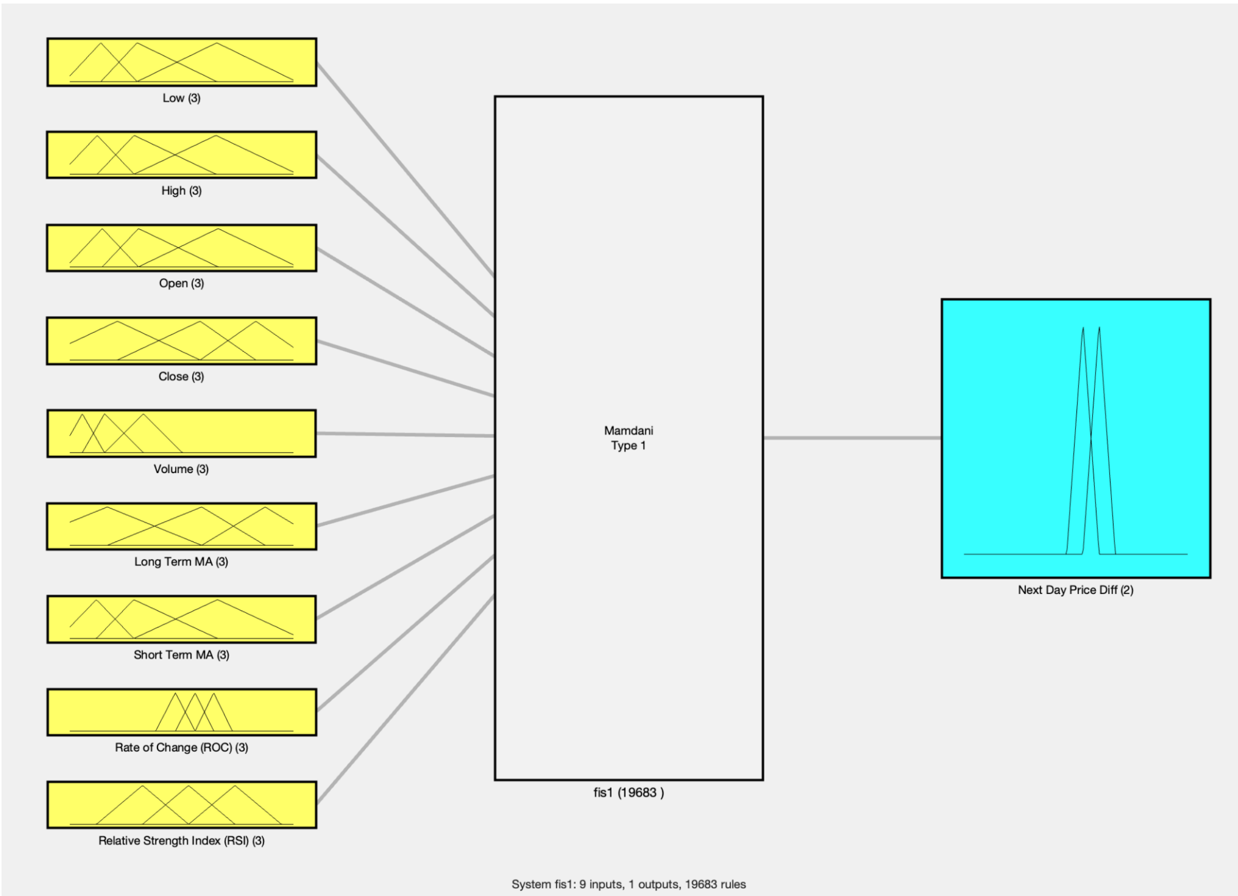

Figure 5: Mamfis 1 with 9 inputs, 1 outputs and 19683 rules

3.3 Inference Rule

In contrast to traditional expert-driven approaches, we employ a data-driven method to derive the inference rules for our Fuzzy Inference System (FIS). This approach leverages the wealth of available historical data to automatically generate rules that can capture the complex relationships and patterns present in the stock market. By using a data-driven approach, we can more effectively adapt our FIS to the dynamic nature of the financial markets, ensuring that our predictions remain relevant and accurate as new information becomes available.

To implement the data-driven approach, we tune our FIS using the training data to determine the most suitable rule bases. The tuning process will be demonstrated in the upcoming section.

3.4 Defuzzification

In the Mamdami Fuzzy Inference System in MATLAB, the centroid defuzzification is the default defuzzification method to convert the fuzzy output back to a crisp value.

The centroid defuzzification method involves calculating the weighted average of all the output membership function values. Each value is weighted by its corresponding membership function's degree of truth. The resulting crisp value represents the best compromise among all the activated output membership functions.

$$x^* = \frac{\int x \cdot \mu_A(x)dx}{\int \mu_A(x)dx}$$

Figure 6: Equation for Centroid Defuzzification Method

3.5 Train Test Split

In this analysis, the dataset is split into training and testing sets using an 80:20 ratio, which is a common approach for evaluating the performance of machine learning models. Columns from the dataset are reordered to match the specific input order required by the fuzzy inference system (FIS) tree.

4. OPTIMIZATION

A fuzzy inference system (FIS) with 9 inputs and 1 output can have a large number of rules depending on the number of membership functions associated with each input variable. In the case study with 9 input variables and 3 membership functions each, the total number of rules in the rule base can be computed as follow:

3 (membership functions) ^ 9 (input variables) = 19683 rules

4.1 Aggregated Fuzzy Tree

According to the MATLAB manual on Fuzzy Tree, having such a large number of rules can increase the complexity of the system and may lead to longer computation times. [7] This can be optimized by reconstructing the Fuzzy Inference System into an Aggregated Fuzzy Tree.

An aggregated fuzzy inference system (FIS) can be designed by dividing the input features into groups based on their contextual relevance. In the case of stock market prediction, we can divide the features into three groups: Price-based features, Technical Indicators, and Moving Averages.

Table below illustrates the input features together with their respective groups in the Multi-Level Fuzzy Tree. (Table 2)

Table 2: Input Features and their respective groups

Group	**Input Variable**
Price-based features	Low
	High
	Open
	Close
Technical Indicators	Rate of Change (ROC)
	Relative Strength Index (RSI)
	Volume
Moving Averages	Short Term Moving Average (10 days)
	Long Term Moving Average (50 days)

Figure below (Figure 7) illustrates the optimized Multi-Level Fuzzy Inference System which segregates the input variables based on the categories mentioned in Table 2.

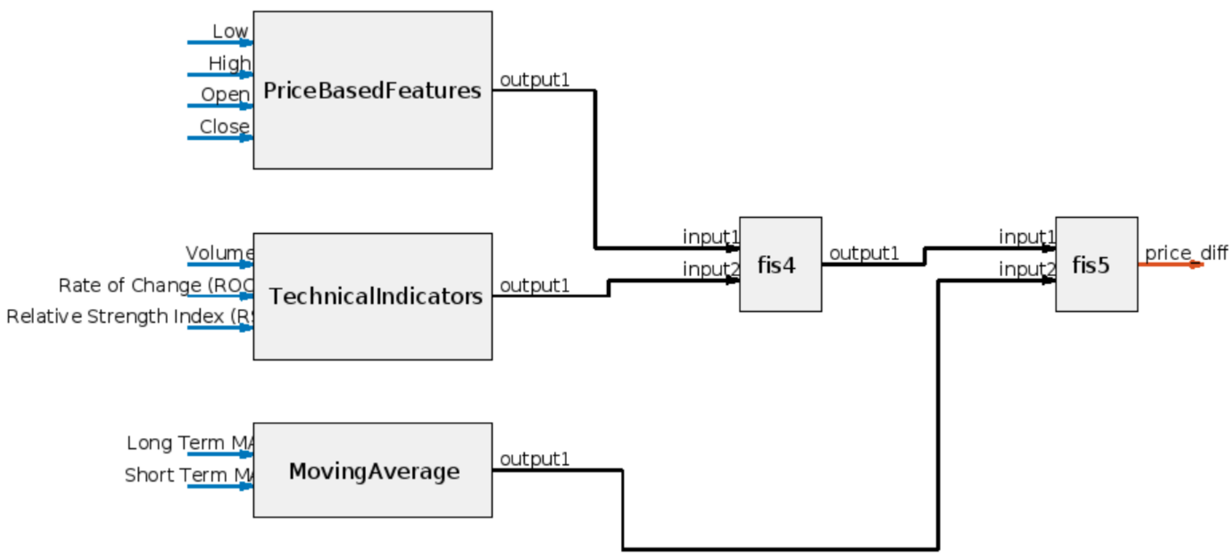

Figure 7: Optimized Aggregated Fuzzy Inference System

By reorganizing the FIS into an aggregated structure, we can simplify the system by reducing the number of inference rules to be searched for when tuning the FIS. The first level, which focuses on price-based features, has 3 (membership functions) ^ 4 (num inputs) = 81 rules. The second level, which focuses on technical indicators, has 3 ^ 3 = 27 rules, and the third level, which focuses on moving averages, has 3 ^ 2 = 9 rules. Additionally, fis4 and fis5 do not contribute any new rules, as they serve to connect and aggregate the outputs from the previous levels. In total, there are only 117 rules, which is much lower compared to the 19683 rules in the original single-level FIS. This hierarchical approach makes the FIS more manageable, computationally efficient, and contextually relevant.

5. TUNING

The tuning process involves optimizing the fuzzy rules to minimize the prediction error on the training set. Global optimization techniques, such as the Genetic Search Algorithm is generally preferred over other local optimization techniques, such as the 'Pattern Search" algorithm because it can effectively explore the complex and nonlinear solution space, avoiding getting trapped in local optima and leading to more accurate and robust solutions. For this experiment, the tuning process is performed using the genetic search algorithm.

5.1 Cost Function

The Fuzzy Inference System (FIS) in this study is tuned using Root Mean Square Error (RMSE) as the distance metric. RMSE is a popular cost function for regression problems, as it measures the average squared difference between the predicted and actual values. The objective of the tuning process is to minimize the RMSE, resulting in more accurate predictions from the FIS. The cost function for RMSE is given by the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Figure 8: Equation for RMSE, where n is the number of data points, y_i is the actual value and $\hat{y}_i$ is the predicted value for the i-th data point

While learning the rule base for the Inference System, the input and output membership function parameters remain constant, this is achieved by setting the options for the tuning function to 'learning'. By configuring the tuning function in this manner, the learning process focuses solely on the rule base without altering the membership functions. This approach enables the system to establish a solid foundation for discerning the underlying patterns in the data, before subsequent fine-tuning of parameters to optimize performance.

6. RESULTS

6.1 Features Characteristics

Table below describes the mean and standard deviation of each feature used to tune the Fuzzy Inference System.

Table 3: Mean and Standard Deviation of each feature

Feature Name	**Mean += std**
Low	412.38 ± 107.60
High	425.33 ± 109.32
Open	419.07 ± 108.59
Close	419.00 ± 108.34
Volume	7570475 ± 546820
ShortTermMA	419.01 ± 107.91
LongTermMA	418.98 ± 106.73
Relative Strength Index	52.45 ± 16.43
Rate of Change	0.083 ± 2.66
Price Difference	0.15 ± 11.05

6.2 Tuning Results

In order to prevent over-tuning the FIS tree to the training data, which increases the validation error, a maximum limit of 20 generations has been established. Tuning is a time-consuming process, it is worth noting that tuning the FIS for 20 generations using the Genetic Search algorithm took 25 hours. Hence, no further experiment was conducted to validate the performance of the FIS given more iterations.

When learning the rule base, The algorithm continuously tries to improve the fitness (minimize the RMSE) by evolving the population of solutions. As the generations progress, it can be observed that the Best f(x) and Mean f(x) values generally decrease, indicating that the optimization process is converging towards better solutions. In other words, this suggests that during the first phase of tuning, the genetic search algorithm is effectively refining the inference rules to better model the relationship between input and output variables. Table below documents the first 20 generations of the genetic search algorithm, when learning the rule base for the FIS.

Table 3: First 20 generations of the Genetic Search Algorithm when learning the rule base

Generation	Func-Count	Best f(x)	Mean f(x)	Stall Generation
1	400	11.29	25.84	0
2	590	11.22	22.12	0
3	780	11.16	22.49	0
4	970	11.16	20.15	1
5	1160	11.14	19.93	0
6	1350	11.05	18.19	0
7	1540	11.05	17.48	0
8	1730	11.03	14.88	0
9	1920	11	14.13	0

10	2110	10.96	13.2	0
11	2300	10.96	12.97	1
12	2490	10.96	12.61	2
13	2680	10.64	11.95	0
14	2870	10.64	11.44	1
15	3060	10.64	11.31	2
16	3630	10.63	11.03	0
17	3440	10.63	11.03	0
18	3630	10.63	10.94	0
19	3820	10.63	10.87	0
20	4010	10.63	10.8	0

6.3 Inference Performance

After the tuning process, the FIS is evaluated on the test set. The Root Mean Square Error (RMSE) of the system if found to be 10.9889. Figure below illustrates the scatter plot which displays the relationship between actual and predicted stock price movement. (Figure 8)

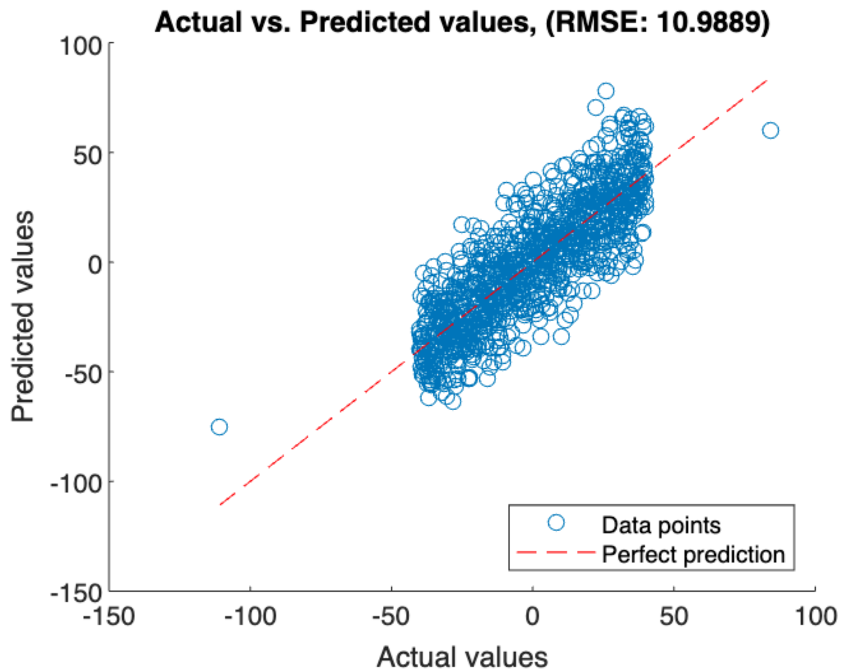

Figure 8: Actual vs Predicted value by the Aggregated FIS

6.4 Intermediate Output

An advantage of a modular fuzzy tree is that we are able to gain insights into the operation of the fuzzy inference system by inspecting the intermediate FIS outputs. By evaluating the correlation coefficients between the outputs of individual FIS modules and the final output, it is possible to ascertain the necessity of each FIS. A high correlation between a specific FIS output and the final output indicates that similar results can be achieved even after removing the FIS, and by doing so, this will help reduce the complexity of the inference system.

The correlation analysis reveals that the outputs from the PriceBasedFeatures modular FIS exhibit a strong negative correlation with the final outputs. Consequently, we proceed to eliminate fis1 and re-execute the tuning process.

```
ans =

   -0.7662    0.1494   -0.0181    1.0000
```

Figure 9: Correlation between outputs from PriceBasedFeatures, TechnicalIndicator and MovingAverage modular FIS against the final output from the aggregated FIS

6.5 Simplify and Retrain FIS

The figure presented below demonstrates the updated Aggregated Fuzzy Inference System following the removal of the PriceBasedFeatures Fuzzy Inference System. Fis4 has also been removed, as its primary function was to aggregate the outputs from the PriceBasedFeatures and TechnicalIndicators systems. Given that the PriceBasedFeatures system has been removed, Fis4 no longer provides any valuable contribution to the overall system.

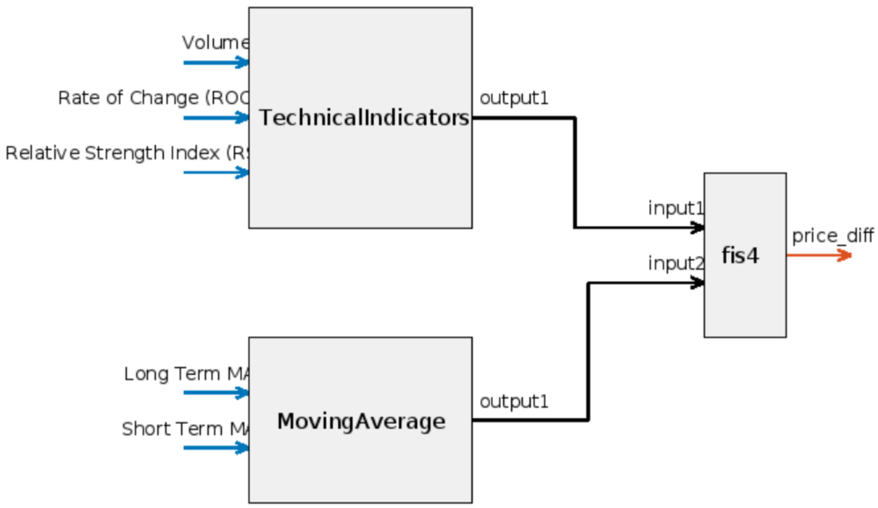

Figure 9: Simplified Aggregated FIS after removing PriceBasedFeatures

The updated FIS was retuned on the training set. In the learning phase, the existing rule parameters are tuned to fit the new configuration of the FIS tree. Upon evaluating the performance of the updated Fuzzy Inference System using the same test dataset, it can be observed that its performance is inferior to that of the initial FIS. The Root Mean Square Error (RMSE) for the updated system is found to be 13.97, which indicates a decrease in predictive accuracy compared to the original system. Figure below illustrates the scatter plot between the actual and predicted price movement from the FIS, which resulted in a RMSE of 13.97. (Figure 10)

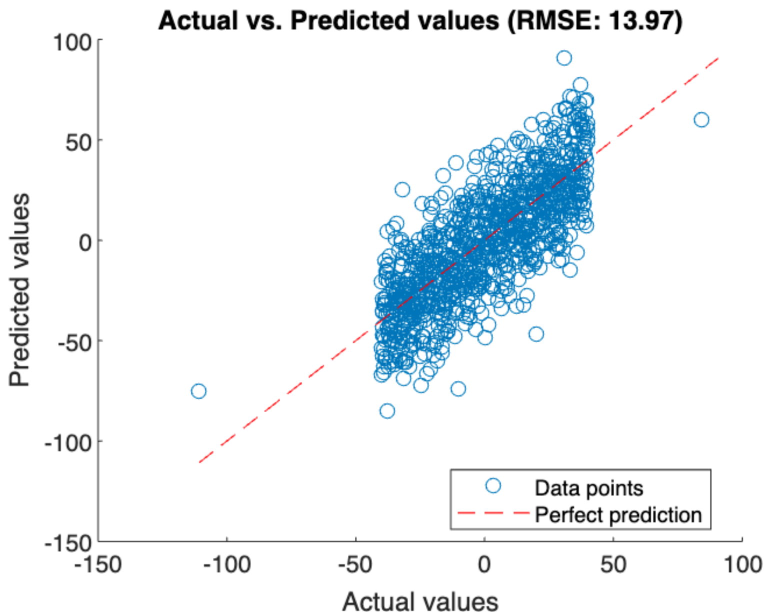

Figure 10: Scatter plot between actual and predicted price movement

7. DISCUSSION

The results of the present study demonstrate the effectiveness of the Aggregated Fuzzy Inference System (FIS) in predicting stock price movements for Netflix Inc. (NFLX). The optimized FIS achieved a Root Mean Square Error (RMSE) of 10.9889 on the test set, indicating its ability to provide reasonably accurate predictions for the target variable. The data-driven approach employed in this study to derive the inference rules for the FIS allowed the model to adapt more effectively to the dynamic nature of the financial markets.

The use of an aggregated FIS contributed to the overall success of the model by enabling a more efficient and contextually relevant organization of the input variables. By grouping the input variables into categories, such as Price-based Features, Technical Indicators, and Moving Averages, the FIS could better capture the complex relationships among the variables and produce more accurate predictions. Furthermore, the aggregated FIS significantly reduced the number of rules, from 19683 in the single-level FIS to 117 in the multi-level FIS, making it more computationally efficient and manageable.

However, it is also observed that the system suffers from an increment in RMSE following the simplification procedures. Specifically, the RMSE has deteriorated from its initial value of 10.9889 to 13.97 after the simplification and retraining process. This degradation in performance can be primarily attributed to the loss of complexity in the FIS. This is because simplifying the FIS reduces the number of rules or membership functions, which in turn results in a loss of complexity. When this complexity is reduced during simplification, the FIS may lose some of its ability to model the underlying patterns in the data, leading to a decrease in predictive accuracy.

8. FURTHER ACTIONS

In this experiment, our primary focus was on learning the rule base between the input and output data. However, the modeling capabilities of the Fuzzy Inference System (FIS) can be further enhanced by fine-tuning certain parameters associated with the rules. These parameters includes membership function parameters, scaling factors, fuzzification and defuzzification methods,. By optimizing these parameters, the overall performance and accuracy of the FIS can be significantly improved.

In the current approach, the tuning algorithm optimizes the membership function parameters to minimize the RMSE. However, users with prior knowledge of the stock market or experience in trading strategies may have insights into certain rules that are likely to hold true. In such cases, predefined rules can be incorporated into the FIS, and the tuning algorithm can be prevented from modifying these rules. This approach can potentially improve the performance of the FIS by leveraging the user's domain knowledge and expertise. To achieve this, users can create an initial rule base and mark specific rules as fixed, preventing the optimization algorithm from changing them during the tuning process. This hybrid approach can help users find a balance between using their domain knowledge and leveraging the optimization algorithm's power.

9. CONCLUSION

In conclusion, this study demonstrates the potential of the Aggregated Fuzzy Inference System as a viable tool for predicting stock price movements. The optimized FIS, with its reduced rule base and contextually relevant organization of input variables, achieved reasonably accurate predictions and provided valuable insights into the relationships among the input variables. These findings suggest that the aggregated FIS could be a valuable addition to the toolbox of financial analysts, traders, and investors for making more informed decisions in the stock market.

REFERENCE

1. J. Shen and M. O. Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system - journal of big data," *SpringerOpen*, 28-Aug-2020. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00333-6. [Accessed: 28-Apr-2023].
2. A. Hayes, "Behavioral finance: Biases, emotions and financial behavior," *Investopedia*, 13-Dec-2022. [Online]. Available: https://www.investopedia.com/terms/b/behavioralfinance.asp. [Accessed: 28-Apr-2023].
3. I. N. Davies, D. Ene, I. B. Cookey, and G. F. Lenu, "Implementation of a type-2 fuzzy logic based prediction system for the Nigerian Stock Exchange," *arXiv.org*, 04-Feb-2022. [Online]. Available: https://arxiv.org/abs/2202.02107. [Accessed: 28-Apr-2023].
4. J. Shah, "Netflix stock price prediction," *Kaggle*, 05-Feb-2022. [Online]. Available: https://www.kaggle.com/datasets/jainilcoder/netflix-stock-price-prediction. [Accessed: 28-Apr-2023].
5. C. Mangale, S. S. Meena, and P. Purohit, "A mathematical model of social media popularity with standard incidence ..." [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/546/5/052086. [Accessed: 28-Apr-2023].
6. S. Handoyo and H. Kusdarwati, "Implementation of fuzzy inference system for classification of Dengue fever on the villages in Malang," *NASA/ADS*. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2019MS%26E..546e2026H/abstract. [Accessed: 28-Apr-2023].

7. “Fistree,” *MathWorks*. [Online]. Available: https://www.mathworks.com/help/fuzzy/fuzzy-trees.html. [Accessed: 28-Apr-2023].