

Software for a Novel Fuzzy Time Series Forecasting Model Based on Balanced Tree Data Structures

Thiago Henrique Barbosa de Carvalho Tavares¹, Bruno Pérez Ferreira²,
Eduardo Mazoni Andrade Marçal Mendes³

¹Departamento de Engenharia de Controle e Automação, Instituto Federal de Minas Gerais, Ibirité, Brasil

²Departamento de Ciências Administrativas, Faculdade de Ciências Econômicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

³Departamento de Engenharia Eletrônica, Universidade Federal de Minas Gerais, Escola de Engenharia, Belo Horizonte, Brasil

Email: thiago.tavares@ifmg.edu.br, brunoperez.bh@gmail.com, emmendes@ufmg.br

How to cite this paper: Tavares, T.H.B.C., Ferreira, B.P. and Mendes, E.M.A.M. (2025) Software for a Novel Fuzzy Time Series Forecasting Model Based on Balanced Tree Data Structures. *Journal of Computer and Communications*, 13, 13-28.

<https://doi.org/10.4236/jcc.2025.135002>

Received: January 28, 2025

Accepted: May 13, 2025

Published: May 16, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Predicting stock market trends remains a heavily researched topic due to the inherent complexity of market fluctuations, driven by subjective and often unpredictable factors. These challenges make accurate predictions difficult for traditional models. Fuzzy models address this issue by accounting for the uncertainty of market returns, thereby mitigating the influence of human emotions on predictions. A key component of these models is the use of data structures to define clusters effectively. In this context, this paper introduces a novel fuzzy model that leverages the red-black tree (RBT) data structure to enhance predictive capability. As a balanced binary search tree, the RBT plays a critical role in improving forecasting precision. A comparative analysis with existing fuzzy models from the literature demonstrates the superior performance of the proposed approach.

Keywords

Finance, Fuzzy, Statistics, Red-Black Tree, Machine Learning, Deep Learning

1. Introduction

Investing in the stock market remains a challenge, as both the economy and finance are not deterministic processes. Research in financial forecasting has explored various model types and techniques, such as time series analysis using

ARIMA [1] and fuzzy regression [2]. A model for predicting university enrollments based on fuzzy time series was proposed by [3], and [4] argues that the interval lengths associated with fuzzy sets (also known as clusters) significantly and directly impact the forecasting procedure's performance. A series of later works, including [5]-[7], propose alternatives for determining the best interval lengths. These works emphasize that the interval lengths need not be identical, as suggested by [8].

A more advanced model based on the work from [3] explores a fuzzy weighted approach, in which the forecast considers how frequently a particular pattern appears in the fuzzy time series, as discussed by [9]. The work from [10] proposes a model where the universe of discourse is partitioned into clusters based on Schrödinger's concepts from quantum mechanics. Additionally, the work from [11] suggests a model based on cumulative probability distributions, and the work from [12] introduces a model based on rough set rules. A central focus of many research efforts in fuzzy time series is the improved distribution of data mass into smaller groups (clusters) that share similar statistical characteristics. These efforts aim to develop tree-like data structures, as demonstrated by [10]. Red-black trees, which are balanced binary trees, have been extensively studied for their efficiency in node insertion and imbalance correction. A comprehensive discussion on red-black trees can be found in [13].

2. Background Literature

The work from [6] contributed to the development of a new method for clustering data, generating varying interval lengths for fuzzy sets. The work from [14] introduced a multi-variable method for fuzzy forecasting that combines fuzzy clustering and fuzzy interpolation techniques. The work from [15] proposed a method using a triangular membership function and extended autocorrelation functions. The GARCH (Generalized Auto-Regressive Conditional Heteroskedasticity) models are considered particularly suitable for economic and financial predictions, as they handle non-stationary data. A hybrid model combining GARCH and fuzzy linear regression is proposed in [16], with a fuzzy sliding window variance developed to estimate forecast weights. The work from [17] introduces a model that integrates genetic algorithms (GA) with adaptive neuro-fuzzy inference systems (ANFIS) to improve models such as ARIMA and GARCH. The work from [18] uses a fuzzy forecasting model that helps establish a relation between the main taxes fee in Brazil and the amount of credit provided for nations having intentions in investing in that country. It is important to note, however, that other fields of science also take advantage in fuzzy models, which can be observed in civil engineering from the paper published by [19].

Other works such as [20] show how a support decision systems can be developed using modern tools involving genetic algorithms. Different variables such as volume have its impacts analysed during forecasting procedures as in the work from [21].

Different works as [22]-[25] show various forecasting methods using fuzzy time series while the original work which described fuzzy mathematics can be seen at [26].

This paper proposes a new model for stock market forecasting, integrating fuzzy time series where the partitioning of the universe of discourse is based on transforming a red-black tree structure into clusters.

3. Materials and Methods

This section presents the background literature on the main topics discussed in this paper: Fuzzy Time Series and Red-Black Trees. The objective is to provide the reader with a fundamental understanding of the mathematical principles underlying this work. Additionally, relevant references are included for those interested in exploring these topics in greater depth.

3.1. Fuzzy Time Series

A conventional time series is composed by real numbers whereas a fuzzy time series is composed by fuzzy sets [27], where a fuzzy set is a class with fuzzy boundaries. In order to define a fuzzy time series let U be the universe of discourse as defined in Equation 1 and a fuzzy set A of U is defined by Equation (2).

$$U = \{u_1, u_2, \dots, u_n\} \quad (1)$$

$$A = \left\{ \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \dots + \frac{f_A(u_n)}{u_n} \right\} \quad (2)$$

Based on [3], f_A is the membership function of A , $f_A : U \rightarrow [0, 1]$. $f_A(u_i)$ indicates the degree of membership of u_i in A , where $f_A(u_i) \in [0, 1]$ and $1 \leq i \leq n$.

The Equations (1) and (2) presented define key concepts in fuzzy time series. Equation (1) establishes the **universe of discourse** U , which represents the set of possible values in the time series. Equation (2) then defines a **fuzzy set** A over U , where each element u_i is assigned a **degree of membership** $f_A(u_i)$ within the interval $[0, 1]$, as determined by the membership function f_A . These fuzzy sets serve as the foundation for defining a fuzzy time series, which, unlike conventional time series composed of precise numerical values, consists of fuzzy sets representing linguistic variables. The definitions provided further formalize fuzzy time series by describing their structure and classification, distinguishing between time-invariant and time-variant cases based on how values evolve over time.

The background definitions of fuzzy time series can be reviewed as follows.

Definition 2.1 [14]: Let $Y(t)$ ($t = 0, 1, 2, \dots$), a subset of \mathbb{R} , be the universe of discourse on which fuzzy sets $f_i(t)$ ($i = 1, 2, \dots$) are defined, and let $F(t)$ be a collection of $f_i(t)$ ($i = 1, 2, \dots$). Then $F(t)$ is called a fuzzy time series on $Y(t)$ ($t = 0, 1, 2, \dots$).

$F(t)$ can be regarded as a linguistic variable [28], and $f_i(t)$ ($i = 1, 2, \dots$)

can be viewed as possible linguistic values of $F(t)$ where $f_i(t)$ ($i=1,2,\dots$) are represented by fuzzy sets. Note that $F(t)$ is a function of time since its values can be different at different times. According to [27], if $F(t)$ is caused by $F(t-1)$ only, then this relationship is represented by $F(t-1) \rightarrow F(t)$.

Definition 2.2 [3]: Let $F(t)$ be a fuzzy time series. If for any time t , a displacement at t causes a displacement at $F(t)$ and $F(t)$ only has finite elements, then $F(t)$ is called a time-invariant fuzzy time series. Otherwise, it is called a time-variant fuzzy time series.

The above definitions will be used throughout the paper.

3.2. Red-Black Tree

A Red-Black tree is a self-balancing binary search tree that ensures balance by performing rotations and recoloring of nodes, following specific rules [13]. The relevance of this data structure can be observed in the work developed by [29], where a concurrent RBT version is presented. In this paper, we focus only on the insertion function, which proceeds as follows:

- 1) If the tree is empty, create a new node as the root and color it black.
- 2) If the tree is not empty, insert the new node as a red leaf.
- 3) If the parent of the new node is black, no further action is needed, and the insertion is complete.
- 4) If the parent of the new node is red, check the color of the parent's sibling (the new node's "uncle"):
 - *If the uncle is black or absent, perform a suitable rotation and recolor the nodes to restore balance.
 - *If the uncle is red, recolor both the parent and the uncle to black, and check whether the grandparent (parent of the new node's parent) is not the root. If the grandparent is not the root, recolor it and recheck for any red-red conflicts, treating the grandparent as the new node.

4. Fuzzy Model Based on Red-Black Tree

This section outlines the structure and implementation details of the proposed fuzzy model. The diagram of the model is shown in **Figure 1**, followed by an explanation of each phase. The algorithm's flow is depicted in **Figure 2**.

The core script (*main.m*) initializes the variables and calls the *fuzzyoutOfSample* function. This function then calls additional functions responsible for constructing the tree, recoloring the nodes, and generating the clusters. The function *RedBlackTree* handles the tree construction, while *TCARtree* creates the matrix representation of the Red-Black Tree. Fuzzification and defuzzification are carried out by the *FuzzyRBTModel* function, which generates clusters and performs forecasts, returning the results to *fuzzyoutOfSample*.

Step 1: Retrieve Stock Index Data

In this step, stock index data is collected. The dataset should consist of a time series representing the closing prices of an index or stock. For the examples in this

paper, the data was sourced from historical records on the <http://www.investing.com> website, which is widely trusted by traders for its accurate stock market information.

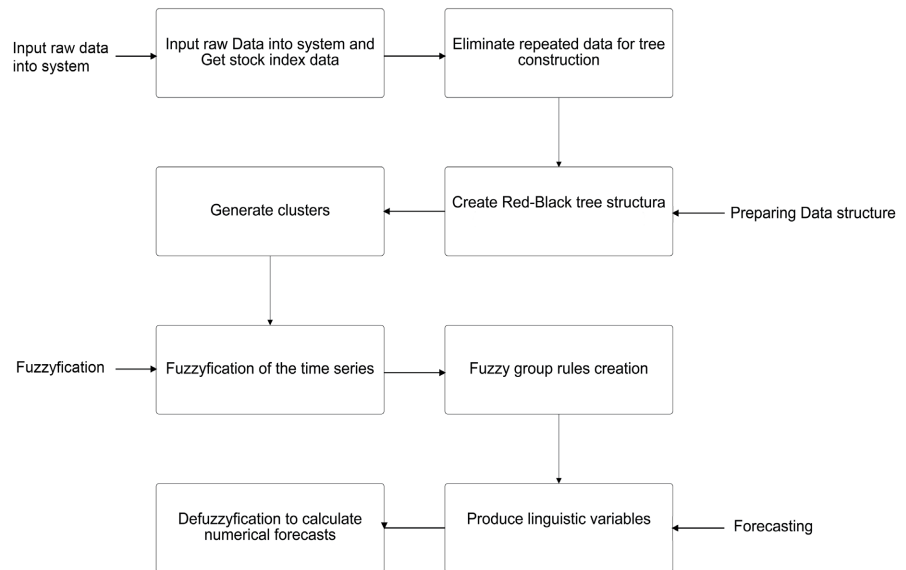


Figure 1. Structure of the proposed fuzzy model based on red-black trees.

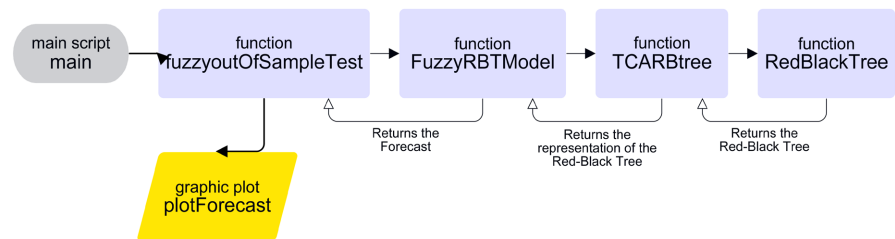


Figure 2. Flowchart of the implemented algorithm.

The time series data is input into the model in array format, as shown in Equation (3):

$$x(t) = \{x_1, x_2, \dots, x_n\} \tag{3}$$

In Equation (3), t is the time index, where x_1 is the first value and x_n is the last.

Step 2: Remove Duplicates from Data for Tree Construction

In this step, a new array is created from the time series, where duplicate elements are removed, resulting in $x_u(t)$. This step is crucial for the subsequent red-black tree construction.

Step 3: Build Red-Black Tree Structure

Here, the red-black tree is built from the array $x_u(t)$, ensuring that no duplicate nodes are included, as required by the model. A detailed example of constructing a red-black tree is provided in [30].

For example, given the following time series:

$$x(t) = \{6, 7, 15, 7, 10, 2, 8, 20, 5\} \quad (4)$$

After constructing the red-black tree from the data in Equation (4), the resulting structure is shown in **Figure 3**.

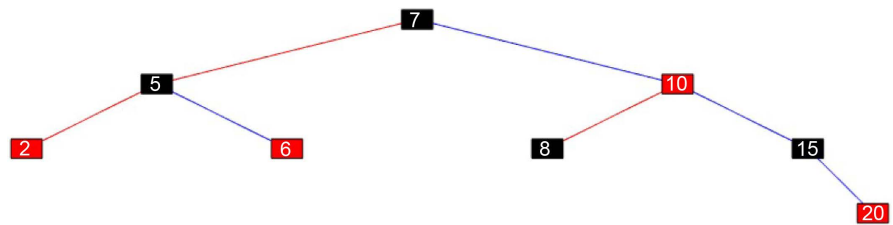


Figure 3. Example of a red-black tree.

Step 4: Cluster Generation

In this step, clusters are generated to define fuzzy sets, with each cluster corresponding to a distinct fuzzy set.

The clustering procedure is based on the parent-child difference in node values. Its advantages stem from ensuring that similar values are grouped within the same cluster. This way, the clusters maintain elements with values that preserve closer statistical measurements. Additionally, since the method relies on differences rather than absolute values, it is resistant to global shifts or biases in the data. Finally, it is also advantageous because it reduces noise in the data, provided there is a trend in the related values.

The clustering algorithm follows these steps:

- **Step 1.** As shown in **Figure 2**, the first task is to verify if the tree has a root (referred to as Troot). The following actions are taken:
 - If Troot has no children, it forms a new cluster. For instance, if the last cluster was C2, the root will belong to C3, and so on. The first node belongs to cluster C1.
 - If Troot has only one child, the child is assigned to the same cluster as Troot.
 - If Troot has both children, the difference between the Troot and its left and right children is computed. The child with the smaller difference is assigned to the same cluster as Troot, while the other child is assigned to a new cluster.
- **Step 2.** For each level of the tree, the same procedure is repeated for all nodes at that level, treating each node as a new Troot.
- The process ends when all nodes have been assigned to clusters.

For the tree shown in **Figure 3**, the clusters formed are as follows:

- Cluster 1: 7, 5, 6
- Cluster 2: 8, 10
- Cluster 3: 2
- Cluster 4: 15, 20

Step 5: Fuzzification of Time Series

Fuzzification refers to the process of converting the time series, which consists of real numbers (x_1, x_2, \dots, x_n) , into a fuzzy time series composed of fuzzy sets (A_1, A_2, \dots, A_c) . Here, n is the number of time series elements, and c is the

number of fuzzy sets (which is equal to the number of clusters from Step 4). In this example, $n = 9$ and $c = 4$.

The fuzzification process is as follows:

- For each element in $x(t)$:
- For each cluster, if the value belongs to a cluster c , assign the fuzzy value A_c to the time series element.

The result of this process is a fuzzy time series, as shown in **Table 1**.

Table 1. Example of fuzzification: The second column represents the fuzzy time series, denoted as $A(t)$.

Time Series	Fuzzy Time Series
6	A1
7	A1
15	A4
7	A1
10	A2
2	A3
8	A2
20	A4
5	A1

Step 6: Creation of Fuzzy Group Rules

In this step, fuzzy rules are created based on the fuzzy time series. A typical fuzzy dependency rule is represented as:

$$A(t) \rightarrow A(t+1) \quad (5)$$

here, $A(t)$ is the antecedent (precedent) and $A(t+1)$ is the consequent.

For each fuzzy time series value, a rule is generated to predict the next fuzzy time series value. For example, if the first fuzzy time series value is A1 (corresponding to the original time series value of 6), this implies the next fuzzy time series value will also be A1 (corresponding to 7 in the original time series).

The fuzzy rules corresponding to the fuzzy time series from Equation (4) are shown in **Table 2**.

Table 2. Fuzzy rules derived from the fuzzy time series.

Time Series (t)	Time Series ($t+1$)	Fuzzy Time Series	Fuzzy Time Series ($t+1$)	Fuzzy Rule
6	7	A1	A1	A1→A1
7	15	A1	A4	A1→A4

Continued

15	7	A4	A1	A4→A1
7	10	A1	A2	A1→A2
10	2	A2	A3	A2→A3
2	8	A3	A2	A3→A2
8	20	A2	A4	A2→A4
20	5	A4	A1	A4→A1

Next, fuzzy rules are grouped. If multiple rules share the same antecedent, they are combined into a single group rule with the common antecedent and all their consequents. For instance, the following rules

$$\begin{aligned}
 &A1 \rightarrow A1 \\
 &A1 \rightarrow A2 \\
 &A1 \rightarrow A7
 \end{aligned}
 \tag{6}$$

can be grouped as:

$$A1 \rightarrow A1, A2, A7 \tag{7}$$

Table 3 shows the fuzzy group rules (FGRs) derived from **Table 2**. Note that the fuzzy group rules include all consequents, even when repeated.

Table 3. Generated fuzzy group rules.

Fuzzy Group Rules
A1→A1, A4, A2
A2→A3, A4
A3→A2
A4→A1, A1

Step 7: Linguistic Forecasting

In this step, linguistic forecasting is performed using the fuzzy group rules (FGRs) from Step 6. Given the fuzzy time series $A(t)$, the forecast for the next time step is obtained by applying the appropriate fuzzy rule.

For example, consider the fuzzy time series from Equation (8):

$$A(t) = A1, A1, A4, A1, A2, A3, A2, A4, A1 \tag{8}$$

To forecast the second sample, we first identify the fuzzy value for $t = 1$, which is A1. Referring to the FGRs in **Table 3**, the corresponding rule is $A1 \rightarrow A1, A4, A2$. Thus, the forecast for $t = 2$ is A1, A4, A2.

Step 8: Defuzzification for Numerical Forecasts

In the final step, the linguistic forecast is converted back to a numerical value using defuzzification. To explain this, consider the fuzzy forecast for the second

sample, as shown in Equation (9):

$$\text{fuzzy_forecast}(2) = \{A1, A4, A2\} \quad (9)$$

Each forecast set is assigned a weight, as shown in **Table 4**.

Table 4. Weights assigned to each forecast set.

Forecast Sets	Weights
A1	1
A4	2
A2	3

Next, the average of each cluster center is computed, as shown in **Table 5**.

Table 5. Cluster centers and their averages.

Cluster	Cluster center (average)
1	6
2	9
3	2
4	17.5

Finally, the numerical forecast is computed as the weighted average of the cluster centers. For example, for the second sample, the forecast is:

$$\text{forecast}(2) = \frac{6 \times 1 + 17.5 \times 2 + 9 \times 3}{1 + 2 + 3} = 11.333 \quad (10)$$

This completes the process for forecasting subsequent samples. Note that Steps 1 through 6 represent the training phase, and Steps 7 and 8 are used repeatedly to make forecasts for future samples.

5. Performance and Application of the Proposed Model

This section illustrates the proposed method using two distinct examples and compares its forecasting performance with that of four existing models from the literature: (CHEN, 1996) [3], (YU, 2005) [4], (TEOH *et al.*, 2008) [11], and (SINGH; DHIMAN; KAUR, 2018) [10]. The comparison involves using these models on the same dataset, and their forecasting results are compared with the proposed model.

The models are evaluated using the following two performance metrics:

- RMSE—Root Mean Square Error
- RMSE relative to the *naive model* (*i.e.*, using the last observed value as the forecast), also known as normalized RMSE.
- R²—R-squared that represents the proportion of the variance in the dependent

variable which is explained by the linear regression model.

- MAE—Mean Absolute Error that represents the average of the absolute difference between the actual and predicted values in the dataset.

The comparison between the models is based on the RMSE and RMSE relative to the naive model, as defined by the equations below.

$$RMSE_{\text{model}} = \sqrt{\frac{\sum_{k=1}^n (y(k) - \hat{y}(k))^2}{n}} \tag{11}$$

$$RMSE_{\text{relative_naive}} = \frac{\sqrt{\sum_{k=1}^n (y(k) - \hat{y}(k))^2}}{\sqrt{\sum_{k=1}^n (y(k) - \hat{y}_{\text{naive}}(k))^2}} \tag{12}$$

$$R^2_{\text{model}} = 1 - \frac{\sum_{k=1}^n (y(k) - \hat{y}(k))^2}{\sum_{k=1}^n (y(k) - \bar{y}(k))^2} \tag{13}$$

$$MAE_{\text{model}} = \frac{\sum_{k=1}^n (y(k) - \hat{y}(k))^2}{n} \tag{14}$$

In these equations, $y(k)$ refers to the actual observed value, $\hat{y}(k)$ is the forecasted value by a specific model, and $\hat{y}_{\text{naive}}(k)$ represents the forecasted value using the “naïve” approach.

5.1. In-Sample Test

The in-sample test was conducted using the closing prices of the Ibovespa (Brazilian stock index) from 03/29/2018 to 10/11/2018, obtained from <http://investing.com/>. A sample of the data used for the model forecasts is presented in **Table 6**.

Table 6. Ibovespa daily closing index data used for in-sample testing, showing forecasts from various models, including the proposed one.

Date	Ibovespa Real Data	Chen 1996	Yu 2004	Teoh 2008	Teoh 2008	Tavares 2022	Proposed Model
03/29/2018	85,365	-	-	-	-	-	-
04/02/2018	84,666	84,790	85,148	84,721	85,141	84,760	84,820
04/03/2018	84,623	83,484	84,315	84,721	84,553	84,760	84,820
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/09/2018	86,088	84,790	85,115	84,721	84,858	84,833	84,737
10/10/2018	83,679	84,790	85,115	84,721	84,858	84,833	84,737
10/11/2018	82,921	83,484	84,481	83,302	83,859	84,106	83,936

The performance of the proposed model relative to existing models is summarized in **Table 7**, based on four different performance metrics.

Table 7. Performance comparison of the proposed model and existing models over the Ibovespa training dataset.

Model	RMSE	RMSE relative Naive	R-squared	MAE
Chen 1996	1235	1.070	0.0921	1248
Yu 2004	1090	0.946	0.0497	877
Teoh 2008	1370	1.189	0.1090	1383
Singh 2018	1032	0.896	0.0456	836
Tavares 2022	1043	0.9108	0.0415	809
Proposed Model	1001	0.869	0.0388	753

The next section discusses the out-of-sample test to further validate the effectiveness of the proposed model.

5.2. Out-of-Sample Test

The out-of-sample test was conducted as follows:

- 1) Train the model using the Ibovespa dataset from 03/29/2018 to 10/11/2018.
- 2) Predict the price for the next sample (10/15/2018). (Only business days are considered).
- 3) On 10/12/2018, Brazil observed a holiday, and the stock market was closed. Since 10/13/2018 and 10/14/2018 were non-business days, there were no trades on these dates.
- 4) The forecast for 10/15/2018 was recorded as forecast (1).

For subsequent forecasts, the training dataset was updated by shifting it one day ahead, using data from 04/02/2018 to 10/15/2018 (real prices, not forecasts). This process was repeated for four steps, yielding forecasts forecast (1) to forecast (4). The results for all methods are shown in **Table 8**.

Table 8. Forecasting results for the out-of-sample test.

Date	Ibovespa Real Data	Chen 1996 Forecast	Yu 2004 Forecast	Teoh 2008 Forecast	Tavares 2022 Forecast	Proposed Model Forecast
10/15/2018	83,359	83,484	83,448	83,144	83,201	82,789
10/16/2018	85,717	83,484	82,815	83,121	83,156	82,978
10/17/2018	85,763	84,790	85,315	84,559	84,602	85,389
10/18/2018	83,847	84,790	85,315	84,575	85,698	85,465

The performance comparison among all methods in the out-of-sample test is shown in **Table 9**.

Table 9. Performance comparison of the proposed model and existing models over the Ibovespa testing dataset.

Model	RMSE	RMSE relative Naive	R-squared	MAE
Chen 1996	1308	0.851	1.007	890
Yu 2004	1642	1.070	0.801	760
Teoh 2008	1480	0.964	1.254	1000
Singh 2018	1627	1.060	1.166	948
Tavares 2022	1305	0.849	0.776	754
Proposed Model	1293	0.842	0.722	741

5.3. Results Discussion

As observed, the forecasts generated by the proposed model are closer to the actual Ibovespa data compared to those of the other models, demonstrating superior predictive performance.

The RMSE value serves as a measure of model performance, where lower RMSE values indicate higher forecasting accuracy. The proposed model exhibits the lowest RMSE among the models analyzed, signifying its superior performance.

The RMSE relative to the Naïve model provides an additional benchmark. A value greater than one indicates that the model performs worse than the Naïve model, while a value below one signifies better performance. The Naïve model assumes that the forecast for the next period is equal to the actual value of the current period. The proposed model not only achieves an RMSE relative to the Naïve model below one but also outperforms all other models in this regard.

Finally, the R-squared value reaches its highest level in the proposed model, indicating its superior ability to explain the variability in the original data.

6. Conclusions

This paper presents a hybrid fuzzy time series model that incorporates the red-black tree data structure to define clusters from the universe of discourse. In terms of performance, the proposed model outperforms other models from the literature. Additionally:

- It surpasses the naive model, proving its efficacy, particularly in stock market applications.
- RMSE values confirms the superior performance of the proposed model.
- Proposed model consistently provides accurate predictions in both the training and testing phases.
- The use of the red-black tree structure, while requiring more computational resources, significantly improves forecasting accuracy.
- The red black trees highlights the crucial role of partitioning the universe of discourse into clusters for enhancing forecasting performance.

It is important to note, however, that this study is limited to working with qualitative variables, as it is necessary to measure numerical differences between values during the red-black tree construction phase. A concern may also arise when the data contains too many repeated values, as these are ignored during the clustering procedure.

Future work will focus on enhancing the proposed model by exploring the following avenues: 1) applying the model to alternative datasets such as currency exchange rates, individual stocks, metals, and financial futures; 2) developing software to simulate the profits generated by the system regularly; and 3) experimenting with novel methods for partitioning the universe of discourse.

Impact Overview

The software discussed in this paper is one of the outcomes developed during a postdoctoral program in Electrical Engineering. From a technical standpoint, the primary impact lies in the innovative organization of data using red-black trees, as well as the application of quantum mechanics principles to cluster the trees prior to the fuzzification process for the forecasting model. While the results presented in this paper demonstrate the application of the forecasting model in the field of finance, its underlying logic can be extended to other datasets from diverse domains such as engineering and biology. The software is also accessible in Matlab, as detailed in the code metadata section. From a literature perspective, three papers which used this software as the foundational methodology for their development were published in high impact journals, making a considerable impact in the Artificial Intelligence research field. One of the papers was derived from a project entitled “Fuzzy Time Series Model Based on Red-Black Trees for Stock Index Forecasting”, which resulted in a publication in 2022 in the journal “Applied Soft Computing”. One of the softwares modules, related to the fuzzification and defuzzification of the time series input data allowed the development of two different papers entitled: “Relationship Between Selic Rate and Basel III Parameters: A Statistical Approach and a Fuzzy Forecasting Model”, published in 2021 in the “Journal of Intelligent & Fuzzy Systems” and “Clustering-Based Fuzzy Model for Predicting Anchor Bearing Capacity” published in 2022 in the “International Journal of Geomechanics”.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Kendall, S. and Ord, K. (1990) Time Series. Oxford University Press.
- [2] Huarng, K.-H. and Yu, T.H.-K. (2012) Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series. *International Journal of Innovative Computing, Information and Control*, 8, 7415-7426.
- [3] Chen, S.-M. (1996) Forecasting Enrollments Based on Fuzzy Time Series. *Fuzzy Sets*

- and Systems*, **81**, 311-319. [https://doi.org/10.1016/0165-0114\(95\)00220-0](https://doi.org/10.1016/0165-0114(95)00220-0)
- [4] Huarng, K. (2001) Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series. *Fuzzy Sets and Systems*, **123**, 387-394. [https://doi.org/10.1016/s0165-0114\(00\)00057-9](https://doi.org/10.1016/s0165-0114(00)00057-9)
- [5] Liu, H. and Wei, M. (2010) An Improved Fuzzy Forecasting Method for Seasonal Time Series. *Expert Systems with Applications*, **37**, 6310-6318. <https://doi.org/10.1016/j.eswa.2010.02.090>
- [6] Chen, S. and Tanuwijaya, K. (2011) Multivariate Fuzzy Forecasting Based on Fuzzy Time Series and Automatic Clustering Techniques. *Expert Systems with Applications*, **38**, 10594-10605. <https://doi.org/10.1016/j.eswa.2011.02.098>
- [7] Wang, L., Liu, X. and Pedrycz, W. (2013) Effective Intervals Determined by Information Granules to Improve Forecasting in Fuzzy Time Series. *Expert Systems with Applications*, **40**, 5673-5679. <https://doi.org/10.1016/j.eswa.2013.04.026>
- [8] Pal, S.S. and Kar, S. (2018) Fuzzy Time Series Model for Unequal Interval Length Using Genetic Algorithm. In: Chandra, P., et al., Eds., *Information Technology and Applied Mathematics*, Springer, 205-216. https://doi.org/10.1007/978-981-10-7590-2_15
- [9] Yu, H. (2005) Weighted Fuzzy Time Series Models for TAIEX Forecasting. *Physica A: Statistical Mechanics and Its Applications*, **349**, 609-624. <https://doi.org/10.1016/j.physa.2004.11.006>
- [10] Singh, P., Dhiman, G. and Kaur, A. (2018) A Quantum Approach for Time Series Data Based on Graph and Schrödinger Equations Methods. *Modern Physics Letters A*, **33**, Article ID: 1850208. <https://doi.org/10.1142/s0217732318502085>
- [11] Teoh, H.J., Cheng, C., Chu, H. and Chen, J. (2008) Fuzzy Time Series Model Based on Probabilistic Approach and Rough Set Rule Induction for Empirical Research in Stock Markets. *Data & Knowledge Engineering*, **67**, 103-117. <https://doi.org/10.1016/j.datak.2008.06.002>
- [12] Cheng, C. and Yang, J. (2018) Fuzzy Time-Series Model Based on Rough Set Rule Induction for Forecasting Stock Price. *Neurocomputing*, **302**, 33-45. <https://doi.org/10.1016/j.neucom.2018.04.014>
- [13] Guibas, L.J. and Sedgewick, R. (1978) A Dichromatic Framework for Balanced Trees. *19th Annual Symposium on Foundations of Computer Science (SFCS 1978)*, Ann Arbor, 16-18 October 1978, 8-21. <https://doi.org/10.1109/sfcs.1978.3>
- [14] Chen, S. and Chang, Y. (2010) Multi-Variable Fuzzy Forecasting Based on Fuzzy Clustering and Fuzzy Rule Interpolation Techniques. *Information Sciences*, **180**, 4772-4783. <https://doi.org/10.1016/j.ins.2010.08.026>
- [15] Carvalho, J.G. and Costa, C.T. (2017) Identification Method for Fuzzy Forecasting Models of Time Series. *Applied Soft Computing*, **50**, 166-182. <https://doi.org/10.1016/j.asoc.2016.11.003>
- [16] Mohamad Hanapi, A.L., Othman, M., Sokkalingam, R., Ramli, N., Husin, A. and Vasant, P. (2020) A Novel Fuzzy Linear Regression Sliding Window GARCH Model for Time-Series Forecasting. *Applied Sciences*, **10**, Article No. 1949. <https://doi.org/10.3390/app10061949>
- [17] Alameer, Z., Elaziz, M.A., Ewees, A.A., Ye, H. and Jianhua, Z. (2019) Forecasting Copper Prices Using Hybrid Adaptive Neuro-Fuzzy Inference System and Genetic Algorithms. *Natural Resources Research*, **28**, 1385-1401. <https://doi.org/10.1007/s11053-019-09473-w>
- [18] Tavares, T.H.B.d.C., Ferreira, B.P. and Mendes, E.M.A.M. (2022) Relationship be-

- tween Selic Rate and Basel III Parameters—A Statistics Approach and a Fuzzy Forecasting Model. *Journal of Intelligent & Fuzzy Systems*, **42**, 5681-5694. <https://doi.org/10.3233/jifs-212128>
- [19] Tavares, T.H.B.C., Porto, T.B., Ferreira, B.P. and Marçal Mendes, E.M. (2023) Clustering-Based Fuzzy Model for Predicting Anchor Bearing Capacity. *International Journal of Geomechanics*, **23**, 1-12. <https://doi.org/10.1061/ijgnai.gmeng-6961>
- [20] Kuo, R.J., Chen, C.H. and Hwang, Y.C. (2001) An Intelligent Stock Trading Decision Support System through Integration of Genetic Algorithm Based Fuzzy Neural Network and Artificial Neural Network. *Fuzzy Sets and Systems*, **118**, 21-45. [https://doi.org/10.1016/s0165-0114\(98\)00399-6](https://doi.org/10.1016/s0165-0114(98)00399-6)
- [21] McMillan, D.G. (2007) Non-Linear Forecasting of Stock Returns: Does Volume Help? *International Journal of Forecasting*, **23**, 115-126. <https://doi.org/10.1016/j.ijforecast.2006.06.002>
- [22] Miller, G.A. (1956) The Magical Number Seven, plus or minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, **63**, 81-97. <https://doi.org/10.1037/h0043158>
- [23] Song, Q. and Chissom, B.S. (1993) Fuzzy Time Series and Its Models. *Fuzzy Sets and Systems*, **54**, 269-277. [https://doi.org/10.1016/0165-0114\(93\)90372-o](https://doi.org/10.1016/0165-0114(93)90372-o)
- [24] Takiyar, S., Upadhyay, K.G. and Singh, V. (2015) Fuzzy ARTMAP and GARCH-Based Hybrid Model Aided with Wavelet Transform for Short-Term Electricity Load Forecasting. *Energy Science & Engineering*, **4**, 14-22. <https://doi.org/10.1002/ese3.105>
- [25] Wang, L.-X. (1997) A Course in Fuzzy Systems and Control. Prentice-Hall International, Inc.
- [26] Zadeh, L.A. (1965) Fuzzy Sets. *Information and Control*, **8**, 338-353. [https://doi.org/10.1016/s0019-9958\(65\)90241-x](https://doi.org/10.1016/s0019-9958(65)90241-x)
- [27] Song, Q. and Chissom, B.S. (1993) Forecasting Enrollments with Fuzzy Time Series—Part I. *Fuzzy Sets and Systems*, **54**, 1-9. [https://doi.org/10.1016/0165-0114\(93\)90355-1](https://doi.org/10.1016/0165-0114(93)90355-1)
- [28] Zadeh, L.A. (1975) The Concept of a Linguistic Variable and Its Application to Approximate Reasoning—I. *Information Sciences*, **8**, 199-249. [https://doi.org/10.1016/0020-0255\(75\)90036-5](https://doi.org/10.1016/0020-0255(75)90036-5)
- [29] Besa, J. and Eterovic, Y. (2013) A Concurrent Red-Black Tree. *Journal of Parallel and Distributed Computing*, **73**, 434-449. <https://doi.org/10.1016/j.jpdc.2012.12.010>
- [30] Tavares, T.H.B.d.C., Ferreira, B.P. and Mendes, E.M.A.M. (2022) Fuzzy Time Series Model Based on Red-Black Trees for Stock Index Forecasting. *Applied Soft Computing*, **127**, Article ID: 109323. <https://doi.org/10.1016/j.asoc.2022.109323>

Code Metadata

Nr.	Code metadata description	Please fill in this column
C1	Current code version	v1.1
C2	Permanent link to code/repository used for this code version	https://github.com/PJbourne/FuzzyRBTeeModel
C3	Permanent link to Reproducible Capsule	https://codeocean.com/capsule/0098421/tree/v1
C4	Legal Code License	MIT License
C5	Code versioning system used	git
C6	Software code languages, tools, and services used	MATLAB, Bash
C7	Compilation requirements, operating environments & dependencies	
C8	If available Link to developer documentation/manual	
C9	Support email for questions	thiago.tavares@ifmg.edu.br
