

UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Mathematics and Statistics

Narmin Mammadli

Interpretable approaches for financial time series forecasting

Actuarial and Financial Engineering
Master's Thesis (30 ECTS)

Supervisors: Prof. Stefania Tomasiello
Dr. Toomas Raus

Tartu 2023

Interpretable approaches for financial time series forecasting

Abstract:

Nowadays complex and non-linear real-world challenges have led to the development of advanced machine learning models. In most machine learning techniques, the ability to interpret the model's predictions is a critical aspect. On the other hand, it is still difficult to understand if it is possible to provide a taxonomy of the various definitions of interpretability. In this thesis, a grey-box method, namely the Adaptive Neuro-Fuzzy Inference System with fractional Tikhonov regularization (ANFIS-T) was adopted to predict stock prices. Its performance was compared against white-box models, such as the autoregressive integrated moving average (ARIMA) and Naive, and the standard Adaptive Neuro-Fuzzy Inference System (ANFIS) model, with an emphasis on the interpretability factor. The root mean squared error (RMSE) and training time by the different approaches were compared.

Overall, in this study, our investigation focuses on interpretability, aiming to offer an overview of all the definitions and related matters, and to discuss the ANFIS-T model's performance for a forecasting problem, concluding that such performance is on par with white-box models. It also highlights the ANFIS-T model's interpretability, offering helpful insights for situations involving decision-making where transparency and understandability are essential elements.

Keywords:

Machine learning, Neuro-fuzzy systems, penalized least squares

CERCS: P170 Computer science, numerical analysis, systems, control

Tõlgendatavad lähenemisviisid finantsaegridade prognoosimiseks

Lühikokkuvõte:

Tänapäevased komplekssed ja mittelineaarsed praktilised väljakutsed on viinud arenenud masinõppe mudelite arenguni. Enamiku masinõppetehnikate puhul on kriitiliseks küsimuseks mudeli prognooside tõlgendatavuse küsimus. Samuti on endiselt raske aru saada, kuidas esitada erinevate tõlgendatavuse definitsioonide taksonoomiat. Magistri-töös kasutati aktsiahindade ennustamiseks halli kasti meetodit Adaptive Neuro-Fuzzy Inference System koos Tikhonovi murdarvulise regularisatsiooniga (ANFIS-T). Selle meetodi sooritusvõimet võrreldi valge kasti mudelitega (ARIMA mudel ja naiivne prognoosimudel) ning standardse Adaptive Neuro-Fuzzy Inference System (ANFIS) mudeliga, kusjuures rõhuasetus oli tõlgendatavuse küsimusel. Võrreldi erinevate meetoditega saadud ühesammuliste prognooside ruutkeskmist viga ja treeningaega.

Kokkuvõtteks keskendub see töö prognooside tõlgendatavuse probleemile, ning töö eesmärk on anda ülevaade tõlgendatavuse kõigist definitsioonidest ja nendega seotud küsimustest ning analüüsida ANFIS-T mudeli prognoosivõimet. Osutub, et selle mudeli prognoosivõime on võrdne valge kasti mudelite prognoosivõimega. Samuti pööratakse

tähelepanu ANFIS-T mudeli tõlgendatavusele, pakkudes kasulikku teavet olukordades, kus prognooside läbipaistvus ja arusaadavus on olulised tegurid.

Võtmesõnad:

Masinõpe, neuro-hajussüsteemid, Tihhonovi regularisatsioon

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine
(automaatjutimisteooria)

List of Tables

1	Number of Observations in Training and Test Sets for Years 2000-2006	27
2	RMSE Results of Naive, ARIMA, ANFIS, and ANFIS-T Models for Years 2000-2006	33
3	Training time of Naive, ARIMA, ANFIS, and ANFIS-T Models for Years 2000-2006	34

List of Figures

1	From the top, triangular, trapezoidal and bell-shaped membership functions	12
2	Uniform fuzzy partition by triangular membership functions.	12
3	An ANFIS architecture with two terms (i.e. fuzzy sets) for each input variable and two rules (adapted from mathworks.com)	13
4	The Interpretability Spectrum	20
5	Correlation between two unconnected variables	22
6	Daily time series data for stock prices between 2000 and 2006 years . .	28
7	Changing RMSE in different number of terms m	29
8	Training time in different number of terms m	30
9	Change in computational time in different values of γ	31
10	Change in computational time in different values of γ	32
11	Change in computational time in different values of α	32

Contents

1	Introduction	6
1.1	Problem Overview	6
1.2	Motivation and Main Goals	7
2	Financial time series forecasting	8
2.1	Forecasting Methods	9
2.2	Forecasting Stock Price	9
3	Considered approaches	11
3.1	Preliminaries	11
3.2	Adaptive Neuro-Fuzzy Inference System (ANFIS)	13
3.3	ANFIS with fractional Tikhonov regularization (ANFIS-T)	15
3.4	Naive and ARIMA model	16
4	Interpretability	18
4.1	Properties of Interpretability	21
4.2	Scope of Interpretability	23
4.3	Interpretability in Time Series	24
4.4	Interpretability and complexity analysis of Fuzzy Inference Systems	24
5	Numerical Experiments	26
5.1	Datasets	26
5.2	Error measures: RMSE	26
5.3	Numerical results	27
5.3.1	Preliminary study	27
5.3.2	Comparative study	33
6	Concluding Remarks	34
	I. Licence	40

1 Introduction

Bringing interpretability to predict future outcomes is not a new concept. According to Ozyegen and Cevik [1], the significance of explanations and comprehensibility regarding the learned models for industrial applications began in the 1980s. In the beginning, experts used a variety of terms, such as justifiability, transparency, understandability, comprehensibility, and so on, in order to present the explainability and interpretability, which are currently used by experts. Lipton [2] argued that interpretability is not a monolithic approach, as it is quite an elusive process due to the domain of certain notions.

Moreover, Lage et. al. [3] believed that interpretability is more linked with psychological understanding, which is more relevant to the degree of a human is able to evaluate models. As a whole, a variety of users comprehend the concept of interpretability in a different manner. The interpretability approach is mainly used in the machine learning methods, such as time series analysis, in order to understand the reason behind the outcomes as well as the actual results when resolving a problem. This is critical in the case of black-box models, which do not offer a transparent cause-effect relationship, as opposed to white-box models, which have a clear formulation. Time series analysis is more linked with identifying the trend and pattern based on data by considering a specific time period [4]. It also predicts future results based on historical data by considering variables, although along with the outcomes, the reason behind them is also needed.

Financial time series analysis plays a critical role in hedging market risks and optimizing investment decisions, although a variety of issues can be highlighted with respect to interpretability [5]. For instance, stock price movements reflect the condition of complex market conditions in different diffusion speeds, such as associated events, media news, historical price series, and so on. No doubt about the fact that the financial sector requires forecasting models to be compliant and interpretable for efficient decision-making in the future. Since the beginning of the twentieth century, the financial industry has performed various studies to search for suitable prediction models and has perfected such technologies ever since by making million-dollar investments in this field. In the past, traditional quantitative methods were used to rely on historical time-series price data for predicting stock price movements. Such models were focused only on technical indicators, while after the development of social media and the internet, unstructured news has also started playing a critical role to improve the prediction procedure [6].

1.1 Problem Overview

Predicting variables at a specific future time period is a crucial challenge in time series forecasting. The majority of the real-world datasets consider the time component, as forecasting the future can help to unlock great value [7]. For instance, retailers can

allow to consider future sales in optimizing the supply chain and promotions, healthcare institutions can consider the number of expected patients to perform adequate planning of human, assets, and capital resources, and investment managers can use forecasting techniques to determine the expected returns in a certain point of time to maximize their returns. While different models exist to predict the future, interpretability is considered the primary issue. Humans not only need to interpret the outcomes received but also the reasons behind the outcomes, and the interpretability being the subjective approach may vary as per humans. Some humans may understand the complexity, while others may need detailed and logical interpretation [8]. Moreover, the interpretation objective is also important, as businesses may desire to determine the reasons behind the trend, while in a low-risk environment, only the outcome is important. Very few researches have been performed to understand the interpretable approaches that can guarantee accuracy in machine learning methods, such as time series forecasting [9]. Moreover, different definitions have existed on interpretability, owing to the lack of a written universally accepted definition, which also creates confusion to understand the concept. Therefore, there is a high need to understand the interpretable approaches that can ensure accuracy, mainly in financial time series forecasting techniques.

1.2 Motivation and Main Goals

In this thesis, the research questions below will be covered.

- **Q1** Is it possible to provide a taxonomy of the diverse definitions of interpretability and a unified vision of the matter?
- **Q2** Considering that grey-box models represent a good compromise between black-box and white-box models, how do they perform when compared to white-box models in the context of financial time series forecasting?

2 Financial time series forecasting

In the last few decades, time series analysis has become a popular topic. In 2006, Linardatos and Kotsiantis [17] performed a rating on the mining information obtained from the time series and considered it one of the top ten challenging data mining issues because of its specific properties. One of the major areas of research is to identify and assess the similarities between the time series analysis, such as the degree to which a specific time series resembles another one, such as classification, clustering, retrieval, or even forecasting.

A time series is a set of observations $y(t)$, referred to a specific time t [37]. In real-world applications, we consider discrete time series, i.e. the set of times at which observations are made is a discrete set.

The classical analysis of time series data is based on the decomposition of time series into components [38]. Each component represents a factor which can affect any time series. Three major components in time series may be identified. Trend refers to the long-term tendency of a time series to rise or fall. Seasonality refers to the periodic behavior of a time series within a specified period of time. The fluctuation in a time series after the trend and seasonal components have been removed represents the irregular component.

The prediction of time series signals is based on their past values. Therefore, it is necessary to get a data set, which should be as much as possible informative and with an adequate number of samples for prediction purposes. The future values of a time series $x(t)$ can be predicted as a function of past values [32]:

$$y(t + \tau) = f(y(t - 1), y(t - 2) \dots y(t - p)) \quad (1)$$

where τ refers to the number of prediction steps ahead, and p is the number of past observations taken into consideration.

With regard to the stock markets, the practical and theoretical analysis is the critical part of the time series analysis. Initially, the empirical research was started in 1933 [39], where the major focus was on the stock market analysis as an individual independent time series, normally referred to as univariate time series analysis. In that case, the financial time series analysis mainly included individual observations related to recorded sequentially over equal time increments. While since the last few years, a variety of factors have started affecting the stock market, owing to increased globalization and interdependence of global economies. Therefore, different phenomena, such as trade wars between two countries, global pandemics, social media platforms, politics, and many others, affect the financial time series analysis.

Many scholars believed that it is challenging to interpret linkages between changes during a specific time since there is a high-level statistical uncertainty and ever-changing environmental conditions. Therefore, Novak and Mirshahi [40] believed that similarity

between two-time series should not only rely on the values but also on the corresponding slopes. Mainly, in the case of stocks, the relative changes significantly impact the stock performances in the future. The upward slope reflects the positive trend, while the downward slope represents the negative trend. While in order to provide further explanations on the slopes, a variety of models have been presented, such as fuzzy transform (F-transform), where the average values are provided of a specific area to distinguish degrees of the F-transform for reflecting the estimation of the given function.

2.1 Forecasting Methods

Financial time series forecasting involves stock price predictions about the future. It is considered an effective application to demonstrate the value of time-series forecasting. Success in stock price forecasting can lead to monetary rewards, but there are a variety of political, economic, and social factors that influence the stock markets. Several traditional time series models are proposed and applicable to economic forecasting. The Autoregressive Moving Average model (ARMA) plays a pivotal role in combining the moving averages with a linear difference in the stationary process [43]. Models that can be used to explain such homogenous nonstationary behavior can be designed by assuming adequate differences in the procedure that is stationary. For this purpose, the Autoregressive Integrated Moving Average (ARIMA) model was suggested, having the assumption of linearity between variables, in order to handle the nonstationary behavior datasets [42]. As a whole, traditional time series forecasting techniques demand more past data and the data must have a normal distribution for better forecasting. The Autoregressive conditional heteroscedasticity (ARCH) model was proposed by Engle [41], which has been considered by many financial professionals; the GARCH (Generalized ARCH) is a generalized type of ARCH [42]. Owing to this situation, the original fuzzy time series model was suggested, and further research was conducted to explain two major processes, such as the formation of fuzzy relationships in forecasting and fuzzification [44]. Moreover, Chen and Chung [45] suggested a technique, where the length of linguistic intervals is formed by genetic algorithms. In order to establish forecasting and fuzzy relationships, Yu [46] claimed that recurrent fuzzy relationships should be determined to forecast and recommend that different weights are assigned to several fuzzy connections. As a whole, there is a variety of forecasting methods with respect to time-series analysis.

2.2 Forecasting Stock Price

A stock is considered a security that applies for proportionate sole ownership in the issuing organization. It generates returns or profits for the investors. In general, stocks can be purchased or sold in the stock markets, however, there are some private deals and the formation of each portfolio [47]. Time series analysis can be used on any series

of variables that change over a specific time period. In the case of the share price or stocks, time series is considered the common approach to track the prices of stocks over time. This can be represented over a specific time period, such as the opening and closing stock prices of a business over time. Such flow showcases the highlights of the stock market. According to Kumbure, Luukka, and Porras [48], the time series analysis on stock prices helps to reflect how a company's capital changes over time, with respect to the selected data points in comparison to the shifts in other variables over the same time period. While forecasting is used to predict future outcomes by analyzing the historical data for many areas, such as finance, environmental science, economics, and business and industry. The time-series data forecasting allows companies to provide useful information that is important for making effective decisions. Major forecasting techniques involve the analysis of time [49]. In time-series analysis, forecasting can be referred to as the chronological sequence of observations for a chosen variable. It can be either multivariate or univariate. Multivariate data contains stock prices of more than one company, while univariate data contains information from one particular stock.

3 Considered approaches

In this section, the considered approaches for the comparative analysis are described. A Preliminaries section is needed to introduce the neuro-fuzzy systems.

3.1 Preliminaries

A *fuzzy set* is a generalization of the set. While the latter is defined by a characteristic function, which can assume only the values 0 or 1, a fuzzy set is defined by a membership function, a real function which can assume any value in the interval $[0, 1]$. In Figure 1, some typical membership functions (representing continuous fuzzy sets) are depicted. A fuzzy set is uniquely identified by its membership function

$$A : U \rightarrow [0, 1] \quad (2)$$

where U is a subset of the real line, which is called *universe of the discourse* or simply *universe*. For any element $x \in U$, $A(x)$ represents the degree of membership of x in the given fuzzy set. A general class of intersection operators for fuzzy sets is provided by the t-norm. It aggregates two membership grades in different ways, namely by the minimum operator, algebraic product, bounded product, drastic product.

It is also useful to recall the notion of *fuzzy partition*. Let $I = [x_0, x_{m+1}]$ be a closed interval and $\{x_0, x_1, \dots, x_{m+1}\}$, with $m \geq 3$, be points of I , called nodes, such that $x_0 < \dots < x_{m+1}$. A fuzzy partition of I is a sequence $\{A_1, A_2, \dots, A_m\}$ of continuous normal and convex fuzzy sets $A_i : I \rightarrow [0, 1]$, with $i = 1, \dots, m$ such that $\sum_{i=1}^m A_i(x) = 1$, $\forall x \in I$. The fuzzy sets form a *uniform fuzzy partition*, when the nodes are equidistant. Figure 2 shows the uniform fuzzy partition obtained by triangular membership functions.

The approximate reasoning uses fuzzy sets, linguistic variables, and different connectives, with several types of implications. A linguistic variable in the universe X is a variable whose values are fuzzy sets of X . More precisely, for a linguistic variable there is a term set, where each term is a fuzzy set of X , and represents an attribute of the linguistic variable. For instance, let x be a linguistic variable, labelled as "age", with $X = [0, 100]$ defining the universe and $x \in X$ being the age in years. A term set of the variable x could be:

$T(\text{age}) = \{\text{very young, young, middle-aged, old, very old}\}.$

Every term is a fuzzy set, uniquely represented by a membership function. Such fuzzy sets could be arranged to obtain a uniform fuzzy partition.

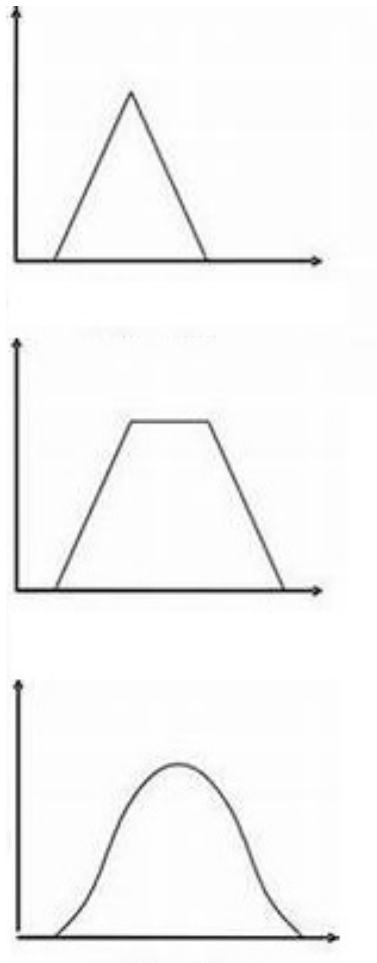


Figure 1. From the top, triangular, trapezoidal and bell-shaped membership functions

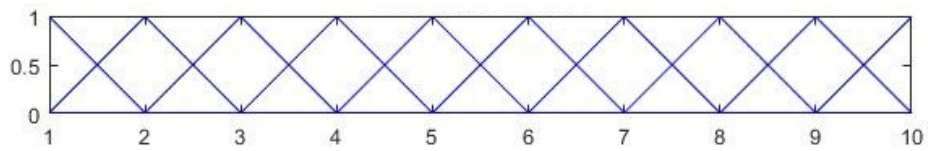


Figure 2. Uniform fuzzy partition by triangular membership functions.

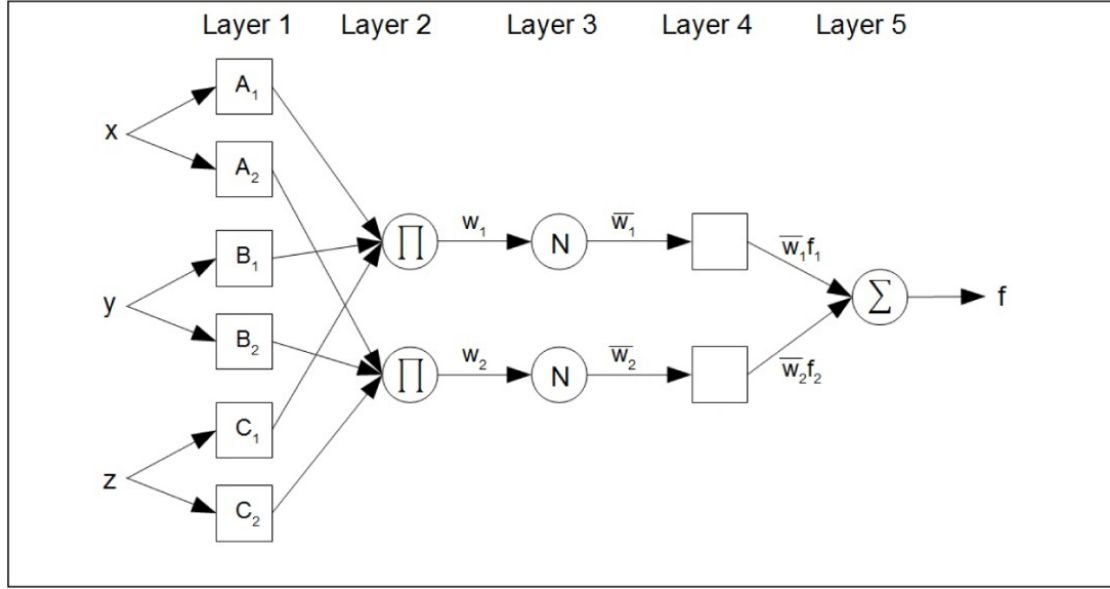


Figure 3. An ANFIS architecture with two terms (i.e. fuzzy sets) for each input variable and two rules (adapted from mathworks.com)

3.2 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The Adaptive Neuro-Fuzzy Inference System (ANFIS) was introduced by Jang [52]. It is a five-layer network architecture which represents the first-order Sugeno (or TSK) fuzzy inference system. The name adaptive means that the nodes in this network are adaptive, i.e. the outputs of these adaptive nodes depend on the parameter(s) of that node. The common learning algorithm is based on gradient descent. To minimize the error measure, parameters should be tuned and this tuning is specified by the learning. Backpropagation with least-squares method is usually used to tune the membership function parameters for this kind of fuzzy inference system (FIS). An ANFIS architecture is represented in Figure 3.

In general, ANFIS is based on IF-THEN rules with a functional consequent part. ANFIS aims at adjusting the antecedents and the consequent parameters. It is worth mentioning that in ANFIS, the number of rules, the number of terms and a number of parameters are important characteristics that describe the complexity of the model.

The number of rules in ANFIS refers to the number of fuzzy if-then rules that are used to describe the relationship between the input variables and the output variable. Each rule consists of a set of membership functions that describe the degree to which each input variable belongs to a particular fuzzy set, as well as a consequent function that describes the output variable based on the input variables. A larger number of rules can result in a more complex model that is able to capture more complex relationships

between the input and output variables but can also increase the risk of overfitting and make the model more difficult to interpret. On the other hand, a smaller number of rules can result in a simpler model that is easier to interpret and less prone to overfitting but may be less accurate in capturing complex relationships.

The number of terms refers to the linguistic labels or fuzzy sets defined for each input variable in the fuzzy inference system. For example, if we have an input variable "Weather," the terms could be "Cold", "Mild", and "Hot" representing different levels of weather. These terms represent different membership functions and capture the linguistic relationship between the inputs and outputs. In ANFIS computational time highly depends on a number of terms. Meaning that if it uses m terms for each of n input variables, the number of generated rules in ANFIS is m^n , which affects the complexity of the model. As a result, the computational time is affected accordingly.

The number of parameters in ANFIS refers to the total number of adjustable parameters in the model, which includes the parameters of the membership functions and the consequent functions. The number of parameters increases as the number of rules or the complexity of the membership functions and consequent functions increases.

There are five layers ($S_1, S_2, \dots, S_5$) in the ANFIS network. All the outputs from layer S_4 are summed together to get the final output f . The output of S_1 is the membership degree of the input ξ to the fuzzy set B_i . Layer S_2 has all the nodes fixed and the output of each node represents the firing strength of a rule. The nodes in S_2 are computed using product-type t-norms (to model AND). For each layer S_i one has:

$$\begin{aligned} S_1 : & O_i = \mu_{B_{ir}}(x_i); \\ S_2 : & \omega_r = \prod_{i=1}^n \mu_{B_{ir}}(x_i); \\ S_3 : & \bar{\omega}_r = \omega_r / \sum_{j=1}^R \omega_j; \\ S_4 : & \rho_r = \bar{\omega}_r f_r = \bar{\omega}_r (\delta_{0r} + \delta_{1r}x_1 + \dots + \delta_{nr}x_n); \\ S_5 : & f = \sum_{r=1}^R \rho_r, \end{aligned}$$

where δ_{ir} are the consequent parameters, with $i = 1, \dots, n$ and $r = 1, \dots, R$. Any suitable parameterized membership function (MF) can be a membership function for the i th term B_i . One of them is the generalized bell-shaped function:

$$\mu_{B_{ir}}(x) = \left(1 + \left| \frac{x_i - c_{ir}}{a_{ir}} \right|^{2b_{ir}} \right)^{-1}, \quad (3)$$

where $\{a_i, b_i, c_i\}$ is the parameter set. The value-changing parameters are referred to as premise parameters. An alternative is the classical bell-shaped function:

$$\mu_{B_{ir}}(x) = \exp \left(- \left(\frac{x_i - c_{ir}}{a_{ir}} \right)^2 \right). \quad (4)$$

In ANFIS model generalized membership function is used. Backpropagation and least-squares are part of the hybrid learning algorithm used to train ANFIS. The output from all the nodes go forward until layer S_4 , and least-squares method is used to determine

the consequent parameters for fixed antecedent parameters. Matrix equation $\mathbf{M}\delta = \mathbf{Y}$ is obtained using the training data, where δ collects the unknown parameters and $\mathbf{Y}$ the target values. The least-squares (LS) method is formulated as

$$\min_{\delta} \|\mathbf{M}\delta - \mathbf{Y}\|^2 \quad (5)$$

with the solution

$$\delta = \overline{\mathbf{M}}\mathbf{Y}, \quad (6)$$

where $\overline{\mathbf{M}} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$ is the pseudoinverse of $\mathbf{M}$. Backpropagation is then used to adjust the antecedent parameters.

3.3 ANFIS with fractional Tikhonov regularization (ANFIS-T)

In the considered approach ANFIS is trained only by LS, with Tikhonov regularization, both in standard and fractional versions. The properties of ANFIS with fractional Tikhonov regularization (FTR) have been recently discussed in [53]. The standard Tikhonov regularization (STR), as a generalization of the standard LS method, is a well-known method for solving discrete ill-posed inverse problems. The minimization problem is formulated as follows:

$$\min_{\delta} \|\mathbf{M}\delta - \mathbf{Y}\|^2 + \gamma\|\delta\|^2, \quad (7)$$

where γ is a positive regularization parameter. The solution is:

$$\delta = (\mathbf{M}^T\mathbf{M} + \gamma\mathbf{I})^{-1}\mathbf{M}^T\mathbf{Y}, \quad (8)$$

where $\mathbf{I}$ is the identity matrix.

The FTR formalizes the following minimization problem

$$\min_{\delta} \|\mathbf{M}\delta - \mathbf{Y}\|_Q^2 + \gamma\|\delta\|^2, \quad (9)$$

where $\|\delta\|_Q = (\delta^T\mathbf{Q}\delta)^{\frac{1}{2}}$ and $\mathbf{Q}$ is a symmetric positive semi-definite matrix defined as [54]

$$\mathbf{Q} = (\mathbf{M}^T\mathbf{M})^{\frac{\alpha-1}{2}}, \quad (10)$$

with $\alpha \in (0, 1)$.

The solution can be written as [54]:

$$\delta = (\mathbf{P}^{\frac{\alpha+1}{2}} + \gamma\mathbf{I})^{-1}\mathbf{P}^{\frac{\alpha-1}{2}}\mathbf{M}^T\mathbf{y}, \quad (11)$$

where $\mathbf{P} = \mathbf{M}^T \mathbf{M}$. It is clear that when $\alpha = 1$ then (11) reduces to (8), that is the standard regularization. The case without any regularization is obtained by setting $\gamma = 0$ and $\alpha=1$.

This variant of ANFIS with FTR, called ANFIS-T after [53], has been successfully tried on a number of benchmarks [53]. The FTR allows for a small number of rules R and terms m (usually $R = m \in \{2, \dots, 5\}$), by keeping a good accuracy. Thanks to such a small number of terms and rules the interpretability greatly improves. For instance, when $m = 2$, the typical terms are "Low" and "High" and the rules would be as below, with output z :

IF x_1 is Low ... AND x_n is Low
 THEN $z_1 = \delta_{01} + \delta_{11}x_1 + \dots + \delta_{n1}x_n$,

IF x_1 is High ... AND x_n is High
 THEN $z_2 = \delta_{02} + \delta_{12}x_1 + \dots + \delta_{n2}x_n$.

From the complexity perspective, ANFIS-T has generally a low number of rules and terms. The fuzzy sets are one-dimensional, continuous, normal and convex. Regarding the semantics, since the membership functions are pre-defined for each input variable (based on expert opinions or in general forming a uniform fuzzy partition of the considered domain) and remain fixed during the learning process, their well-defined semantics is retained. For the same reason, the distinguishability constraint is not violated. The rule base meets the high level requirements.

3.4 Naive and ARIMA model

One of the basic forecasting model is Naive model which makes the assumption that future value of a time series is equal to the last observed value. Mathematically, it can be expressed as:

$$\hat{y}(t+1) = y(t) \quad (12)$$

Where $y(t)$ represents the last observed value and $y(t+1)$ represents the predicted value for the next time period. This means that any trends, seasonality, or other factors that can have an impact on the time series are not taken into consideration by the model. The Naive model is frequently used as a comparison model to assess how well more complex models function.

The ARIMA model is a popular time series model that stands for Autoregressive Integrated Moving Average. It is a combination of three components: autoregression (AR), differencing (I), and moving average (MA).

The ARIMA model has the distinct advantage of being applicable to time series data with non-stationary behavior. Stationarity refers to a time series that has a constant

mean and variance over time. On the other hand, non-stationarity describes a time series with a changing mean or variance over time. ARIMA models can be used to transform non-stationary time series data into stationary data through the process of differencing.

Another unique aspect of the ARIMA model is that it can be extended to incorporate seasonal patterns in the data through the use of seasonal differencing and seasonal autoregression. This allows the model to capture seasonal trends in the data and make more accurate forecasts.

The general form of an ARIMA(p, d, q) model is:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d y(t) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \varepsilon(t) \quad (13)$$

Where $y(t)$ is the time series, $\varepsilon(t)$ is the error term, $\phi_1, \phi_2, \dots, \phi_p$ are the AR coefficients, $\theta_1, \theta_2, \dots, \theta_q$ are the MA coefficients, d represents the degree of differencing required to make the time series stationary and B is the lag operator. B is frequently used to change a sequence's time index. The moving average and autoregressive parts of the ARIMA(p,d,q) model are represented by the lag operator. The MA component utilizes B to indicate the lagged values of the error term, whereas the AR component uses B to denote the lagged values of the dependent variable.

If we define $\tilde{y}$ as follows:

$$\tilde{y} = y - E(y) \quad (14)$$

Then the AR component can be written as:

$$\tilde{y}(t) = \phi_1 \tilde{y}(t-1) + \phi_2 \tilde{y}(t-2) + \dots + \phi_p \tilde{y}(t-p) + \varepsilon(t) \quad (15)$$

The MA component can be written as:

$$\tilde{y}(t) = \theta_1 \varepsilon(t-1) + \theta_2 \varepsilon(t-2) + \dots + \theta_q \varepsilon(t-q) + \varepsilon(t) \quad (16)$$

Where $\varepsilon(t)$ is a white noise error term.

The I component involves differencing the time series d times to make it stationary. This can be expressed as:

$$(1 - B)^d y(t) \quad (17)$$

Where d is the number of times the time series needs to be differenced to make it stationary.

Overall, in terms of time series analysis and forecasting, the ARIMA model and the Naive model are two crucial tools. While the Naive model is a simple baseline model, the ARIMA model is a more sophisticated model that can account for non-stationary data and seasonal patterns.

4 Interpretability

The term “interpret” refers to presenting or explaining in understandable terms, and interpretability is considered as the ability for a piece of information to become understandable for humans. According to [10], interpret refers “to explain” or “to tell the meaning”. In general, interpretability is critical for humans, as they are born curious [1]. Being a subjective approach, information that is interpretable relies on the knowledge of the perceiver. The visual information, such as smell, human language, visual representation, and so on, are interpretable, while non-expert may not be able to interpret time series analysis, due to complexity. Normally, the concept of interpretability is linked with the black-box model. This model refers to a system in which inputs and outputs are used to create useful information without having an idea about the internal processes and workings [10]. Therefore, the interpretation concept is considered to determine the internal processes and workings. In general, considering the scope of interpretability, the current models are divided into two classes: local and global. Local interpretability techniques mainly give focus on describing the reasons and cause behind a certain decision, while on the other hand, the global interpretability techniques aim to define the whole reasoning and logic behind a model [1]. In addition to this, machine learning models only provide information about the patterns of some specific data, while people also become curious about the factors that drive that patterns or outcomes. With regard to predictive modeling, the major focus is not to determine what is predicted, such as the probability of an increase in sales or how much customer ratio will improve, but to comprehend why the specific prediction was made in terms of interpretation to determine predictive performance. Molnar [11] believed that this behavior helps to learn more about the problems incurred, data, and reasons behind the increase or decrease. Mainly, in the high-risk environment, where the business or human life is based on prediction, a correct prediction along with the factors behind it is highly equally important [11].

The interpretation of the outcomes in the machine learning models is moved by the human-centricity goal of modern computer science, i.e. to help human beings for a better life. For instance, if a person is sick, then they may need to know why he is feeling sick, in order to avoid such things to remain healthy. In the case of opaque machine learning models, the scientific results remain a mystery, as only results are given without any explanation [11]. As few humans prefer to focus on outcomes for making decisions, while the majority believe that it is important to determine the reasons behind the outcomes. For instance, if a business trend shows a sharp increase in future profits, then it may require knowing the reasons, in order to further improve the profits, rather than only the increase, which may not be much useful for any future decision-making in the company.

Furthermore, research showed that there are huge benefits in higher interpretability. To begin with, it can formulate trust by reflecting on and explaining the diverse factors

that contribute to decision-making. The trust will lead to more acceptance of all the machine learning systems, including time series forecasting. Secondly, interpretability techniques help to showcase the incompleteness in the problem formation, which can help to debug the issue and design the models in a better way [1]. At last, interpretability techniques can play a significant role in improving scientific understanding, as a better interpretation will help in enhancing the knowledge of the subject matter in a better way. In addition, trust is considered an important concept of interpretability for machine learning (ML) and time series analysis [13]. Few researchers believe that interpretability is closely related to trust, such as the strong trust in the model will leave no room for interpretation. But the concept of trust is also subjective and varies as per humans, such as some models may consider well-understood by few people having no specific objective [31], while others may denote trust as confidence regarding the outcome of the model with regards to some specific situation or objectives. As, it is significant to trust a prediction from a model before considering it, and a person must have faith in the predictions. Moreover, Linardatos and Kotsiantis [17] believed that the interpretable models follow the unbiased approach, as they are highly ethical and regulated, and help to determine the causes or causality of the interpretability. While most scholars consider that trust plays an important role and driving factor to make the time series models interpretable, people can hardly trust such models due to their complexity and inability to understand the parameters used. Therefore, interpretability also refers to techniques to resolve trust issues, and if there are no interpretable approaches, one can hardly determine whether the learned features make sense or not. In addition to this, such predictive models provide critical predictions that can raise safety measures [18].

For this purpose, the system must be robust and secure, otherwise, it may cause severe damage. For instance, Uber, a well-known multinational company, introduced a self-driving Uber feature in 2018, which resulted in the killing of a pedestrian. It is considered the first death due to self-driving cars. In the self-driving feature of Uber, a variety of time series analysis and machine learning programs are used to determine the prediction, perception, and motion learning, in order to drive the vehicle. Later, it was identified that the self-driving vehicle could not detect pedestrians unless they were near a crosswalk. This shows that the prediction techniques can cause severe harm to people.

In addition, various research works have embedded the concept of explainability with machine learning models. Usually, interpretability and explainability terms are mixed. Linear regression, and decision tree models can be good examples of interpretable models. But there are some models which are very difficult to interpret, like random forest, neural networks, convolutional neural networks. These are less interpretable models, but at the same time explainable models. We can make them explainable with the help of various methods like agnostic and specific. These methods provide a general explanation of the model's decision-making process, and are not specific to a particular type of model [20]. Some well-known agnostic methods are LIME (Local Interpretable

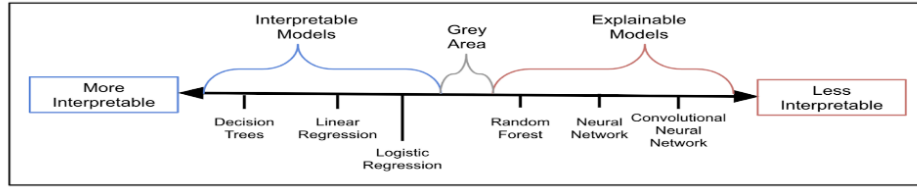


Figure 4. The Interpretability Spectrum

Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanation). In contrast to agnostic methods, specific methods are designed particularly to work with a given set of models or data. An agnostic and specific both offer benefits and drawbacks. Choosing an approach depends on the precise requirements and goals of the explainability task at hand. Figure 1 [21] shows examples of interpretable and explainable model concepts in a graphic way.

There are different ways to define interpretable models. Various authors use different ways to explain interpretability but these are the five main features: transparency, simplicity, comprehensibility, meaningfulness, and intelligibility. Transparency sometimes can be referred to as readability, human readability. This is the property of white boxes in contrast to black boxes which are completely opaque [22]. For example, decision trees are white boxes because they provide all the steps that lead to a final choice. Artificial neural networks are black boxes: no link can be made between the inputs and the output of the models [23]. Some authors defined intelligibility as the capacity to show clearly how each model input influences the final choice [24], [25]. According to them, Generalized Additive Models are good examples for intelligible models. The ability to comprehend a model's global and synthesized vision is a crucial aspect of its interpretability. Another important factor in interpretability is simplicity. According to Occam's razor, simplicity must be found in any process. In interpretability, it may mean reducing the number of variables [26], which also creates faster decision inference [2]. Obviously, it is likely that a non-simple model is not comprehensible, but simplicity is not a guarantee of comprehensibility [27]. Simplicity can be assessed by the number of variables in models, while comprehensibility can be more difficult to define. Another aspect of interpretability is meaningfulness. As reported by Mencar, experts may not understand the model if it is expressed in non-domain-specific language: any part, any parameter or feature of the model must make sense, i.e. refer to the usual vocabulary/knowledge of the experts [28].

In the past, humans had a monopoly on society, as they were responsible for all decision-making processes. For instance, if a person applied for a loan or a job, people used to make decisions about approval, and, in the case of consequential decisions, there would have been an explanation to justify it, such as why the loan application has been denied. As a whole, the reason behind the decisions is normally considered important. However, currently, the predictive models only focus on providing the

outcomes considering the inputs without any reasoning.

Over the last few decades, most of the decision-making is shifted to automated procedures. Algorithms are formulated to consider inputs and predict the related outcomes. For instance, considering a set of features, including the financial figures, the machine learning algorithms will calculate the potential return of investment (ROI). This will ease the businesses to predict the future ROI, but the factors that drive that figure is still unknown. Similarly, in major fields, such as health, financial markets, the criminal justice system, and medicine, the human's inability to fully understand such models can be problematic. In such a situation, interpretability is considered the solution. While the concept of interpretability is not much studied and researched, therefore there is no universally written definition accepted. According to Benk and Ferrario [30], the term interpretability is ill-defined, showing that the interpretability of a variety of models showcases a quasi-scientific character. Being a subjective approach, it does not focus on a single concept but on distinct ideas.

4.1 Properties of Interpretability

According to Lipton, there are five properties of interpretability that must be considered: trust, causality, transferability, informativeness, and fair and ethical decision-making [2]. He stated that a person will feel secure and at ease by using a well-understood model, which can be built through training and deployments.

Trust. By saying trust, what do we really mean? Should we just trust models with good performance? Trust is useful because people do not only expect high accuracy from standard metric models. As of now, there are totally different areas where ML algorithms can be applied. People need to understand and trust models that are not working in a biased way. As an example, consider the increasing use of machine learning models to forecast crime rates in order to assign police officers. We may trust the model to make accurate predictions but not to account for racial biases in the training data for the model's own effect in perpetuating a cycle of incarceration by over-policing some neighborhoods [2].

Causality. Another important aspect of interpretability is causality. The term "causality" refers to a machine learning model's capacity to recognize cause-and-effect connections between the input variables and the output variable. But one must be careful when defining it, as we can find countless correlated stuff but not all of them are really connected. So that again shows the importance of causality, because it may be feasible to pinpoint the most important characteristics or inputs that are influencing the model's decisions by studying the causal linkages between the variables via causality. For example, a bank wants to reduce the default rate on business loans. Our goal is to reduce default rates by understanding what changes need to be made, which requires understanding not only which loans default but also why those loans default. In general, the number of small loans with non-payment is higher than larger loans. In such a case, we can decide

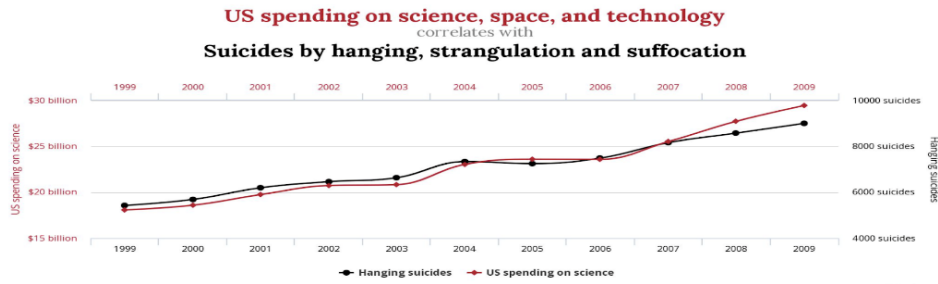


Figure 5. Correlation between two unconnected variables

that we should stop giving small loans. However, if we look deeper, we can see that small loans usually are taken by small businesses. Those have higher chances to fail than bigger companies. Therefore we should look into the correlation between business sizes. Sometimes, we can even find correlation between two totally unconnected variables. A good example can be the following. In Figure 2.[12], it can be easily thought that there is a correlation between US spending on science and technology and the number of suicides. Normal human beings can easily understand that even though there is a correlation it is not causality. One of them did not trigger another.

The two examples above show us the importance of causality. It is possible to interpret causal relationships based on certain correlations. However, ML algorithms do not guarantee that the associations they learn reflect causal links. One might hope, however, that by interpreting supervised learning models, we could generate hypotheses that scientists could then test experimentally [2].

Transferability. In the context of machine learning, transferability is essential as models must be able to perform well on tasks they have already learned from and, in addition, be able to apply that learning to slightly altered tasks. It is common that before implementing the model separate data into 2 parts: training and test. Basically a model learns how to make decisions according to training data, then it is tested by using test data. But what happens if we change test data to a slightly unfamiliar environment? Is the model still performing well enough? For humans it is easy to recognize the difference, nevertheless, for algorithms, we cannot say the same. Even though they can work very well in cases where humans can make a mistake, in the cases they did not learn, it can be very challenging to get the algorithm working correctly. For example, if we use pictures of a wolf in the snow for predicting a wolf pic, when we apply the same model for defining wolf photos without snow it will give us a very wrong answer. Another good example is the model described by Caruana [14] to figure out which of pneumonia patients are low-risk or high-risk. Patients with pneumonia who also had asthma were categorized as low-risk which is not correct. From the examples, we are able to spot that a model can make a mistake in a situation that people never can do. We will have a better grasp of potential problems and can take measures to avoid them if we understand how

models work. When a model is not interpretable, it is difficult to define what is the root cause of the problem.

Informativeness. According to Kim and Huysmans sometimes we use models not for decision making, but to get output - information from the model and make a decision by ourselves [15], [16]. In this kind of models, accuracy is not the main factor, since the model will not be used for making decision. On the other hand, without having enough information about the model, it can be tough to draw any conclusion. Interpretability is the key to understand model's output. You cannot have informativeness without an interpretable model.

Fair and ethical decision-making. At present, politicians, journalists, and researchers have expressed concern that interpretations must be produced to assess whether decisions produced automatically by algorithms conform to ethical standards [2]. Recidism prediction could be one of the examples. Let us imagine a model to decide if a person with a criminal record should go to jail or be free. Because criminal offenses can vary, it is important that an ML model does not make decisions based on characteristics such as a person's race, gender, age, etc. Without interpretability it is difficult to understand if the model's decision is fair enough.

4.2 Scope of Interpretability

In order to make a prediction, algorithms are mainly used by different models. Each step can be assessed with regard to interpretability and transparency. The concept of algorithm transparency is mainly used to define the procedure of how the algorithm learns the relationship from the data [17]. For instance, if the convolutional neural networks are used to classify the images, then the algorithm identifies from the filters and edge detectors on the lowest layers. This whole idea reflects the understanding of about the algorithm works, but it cannot be used for all the models.

While on the other hand, O'Sullivan [21] believes that a model is interpretable if the entire model can be comprehended at once. In the case of the global approach, there is a need for a trained model and an understanding of the data and algorithm. Such interpretability level refers to the understanding of how the model makes decisions on the basis of a holistic review of the features and each component, like structures, parameters, and weights. In general, global model interpretability supports understanding the distribution of the target outcomes based on features. Although in the real world this is highly difficult to achieve, as humans cannot remember multiple features and prefer to give focus on a few features in order to make predictions.

Instead, Aas and Loland [29] believes that a single feature can be focused to determine the prediction from a model along with its reason, and the individual prediction is more effective to understand the behavior of complex models than the multiple ones. For instance, if the area of the house is considered as a feature and data is gathered regarding it, then the models can be used to predict how the price of a house can differ due to

changes in the house area, which will result in more accurate predictions than global ones, where all features are considered and complexity is created. As a whole, more scholars indicate that local interpretation shows higher interpretation results than global ones.

4.3 Interpretability in Time Series

From the beginning, the task of interpretation in most machine learning applications is considered one of the critical tasks. Mainly, it is important to determine why and how a model presents specific predictions and outcomes. Owing to the recent introduction and advancements in artificial intelligence (AI), a major focus of scholars is on its interpretation, while few studies are present to understand the interpretability in time series forecasting. Time series analysis refers to the way of studying the characteristics of the variable responses with regard to time, being an independent variable. For the target variable estimation, the time variable is used as the point of reference. According to Brockwell [32], time series is only a series of a variety of data points that happened in a successive order for a specific period. The main outcome of time series analysis is to determine how time series actually works and which factors are influencing certain variables at different points of time. It aims to predict the future values of the time series variable. While the major assumption used in time series analysis is that the origin of time does not affect the properties of the process based on statistical factors.

In general, time series analysis is considered the prediction and forecasting tool, specific to issues related to time-based [30]. It analyses the historical trends and the patterns, in order to understand the existing situation with the patterns obtained from the past stage. While the major limitation of time series analysis is that it does not consider the missing values, which may result in some inaccurate predictions. Also, it requires the data points to be linear, which might not be possible. Therefore, the interpretation of models, such as time series, must have some margins for such limitations. Moreover, there can be some seasonal effects, such as while reviewing the sales of a retail sector [33], it may show a sharp increase at the time of Christmas, and also during COVID-19, e-commerce sales were increased significantly, while this effect has been subsidized due to the reduction in COVID-19 spread. Therefore, the interpretation of time series analysis can be quite difficult due to temporary and seasonal effects, as it may affect the overall prediction, and thus decision-making at large.

4.4 Interpretability and complexity analysis of Fuzzy Inference Systems

Fuzzy Inference Systems (FISs), in several variants, have been mostly designed to ensure high accuracy, often making them similar to black-box models [34]. Anyhow,

a FIS is supposed to show high interpretability and low complexity. Interpretability or transparency can be meant as human readability [34]. Accuracy maximization and interpretability maximization (or complexity minimization) seem to be two conflicting goals.

Complexity is usually measured in terms of the number of operations, and in this case, the number of elements in the rule base, including the number of membership functions and rules. Assuming a small number of suitable linguistic terms is beneficial to interpretability, also resulting in a less complex system [35].

The interpretability assessment involves three levels: low level, abstraction level, and high level [34, 36]. Low-level interpretability is related to the fuzzy sets, which should be one-dimensional, continuous, normal and convex [34]. The abstraction level is related to fuzzy partitions and fuzzy rules. A fuzzy partition is formed by a number of fuzzy sets defined in the same universe, and it expresses the semantics of a linguistic variable. A fuzzy partition should ensure meaningfulness, i.e. each fuzzy set in the partition should have a proper meaning. For this reason, the number of fuzzy sets should be small enough (e.g. 3-5) [34]. Fuzzy rules should have a reasonable length. The length of a fuzzy rule is the number of linguistic variables in the rule. The smaller the number of linguistic variables, the more interpretable and transparent rules. A rule should not systematically include all input variables: it should include just important variables in its context. [34] Besides, distinguishability and coverage requirements should be met. Regarding the first one, since fuzzy sets are associated with distinct linguistic terms, the fuzzy sets of a partition should be separated to some extent, allowing a certain degree of overlap because concepts are not completely disjoint. Regarding coverage, each element of the universe of discourse should belong to at least one fuzzy set of the partition with a nonnull membership degree.

High-level interpretability involves fuzzy rule bases. In order to ensure interpretability, the rule base should be [34]: compact, i.e. with a small number of rules allowing simple representation of knowledge and easy reading and understanding, complete, such that for each input at least one rule is activated with a certain firing strength, consistent, i.e. without contradictory rules.

5 Numerical Experiments

Evaluations and comparisons have been done in this part, using the RMSE as the assessment criterion, as described in the following. The considered dataset is described in the next subsection. There are two subsections in the numerical experiments. The first one (preliminary part), mainly focuses on how parameters (α, γ) and the number of terms m affect the results by ANFIS-T. In the second one (comparative part), the results of the four considered approaches (i.e. ANFIS, ANFIS-T, Naive and ARIMA) were compared. The experiments were conducted using Scilab 6.0.2 for ANFIS-T, and Python 3.10.8 for the ARIMA and Naive approaches. The experiments were run on a 2.3 GHz Dual-Core Intel Core i5 processor with 8GB of RAM. For numerical experiments with ANFIS-T, the code from the textbook *Contemporary Fuzzy Logic: A Perspective of Fuzzy Logic* [51] and numerical experiment results for the ANFIS model were taken from the work by Lian-Yang. [55]. In the work by Lian-Yang, $m = 3$ was used. It must be recalled that for the standard ANFIS and ANFIS-T the generalized and the classical bell-shaped membership functions were adopted, respectively.

5.1 Datasets

To perform some experiments, we used the TAIEX dataset [50] from 2000 to 2006. The dataset is given by time series data, and it shows the price during the mentioned time.

TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index) is a stock market index that tracks the performance of the stock market in Taiwan. It is one of the most widely recognized indicators of the Taiwanese stock market and is often used as a benchmark for investors and traders. In the analysis, TAIEX data from 2000 to 2006 [50] is used. It contains daily data and in total 7 time series will be used, one for each year. For each year, first 10 months data was used as training data, last 2 month data was used as test data. Table 1. contains information about number of observations bot in training and test datasets for each year. For all the models (Naive, ARIMA, ANFIS and ANFIS-T), one-steap ahead forecasting was used.

5.2 Error measures: RMSE

There are various tests that can be utilized to assess the effectiveness of the predictive model. According to this test results decision can be made on which model is better. Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) are examples of these kinds of tests. These are just a few of the many tests and metrics that can be used to evaluate the performance of a predictive model, and the choice of which test to use will depend on the specific problem being addressed and the type of model being used.

Year	Training Observations	Test Observations
2000	223	46
2001	200	42
2002	204	42
2003	205	42
2004	204	44
2005	202	43
2006	224	42

Table 1. Number of Observations in Training and Test Sets for Years 2000-2006

A frequently used metric for measuring the performance of a predictive model is RMSE. It is particularly useful in assessing the accuracy of a model's predictions in relation to the actual observed values.

The RMSE is calculated by taking the square root of the average of the squared differences between the predicted and actual values. Mathematically, the RMSE can be expressed as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y(t) - \hat{y}(t))^2}$$

where n is the number of observations, $y(t)$ is the actual value at time t , and $\hat{y}(t)$ is the predicted value at time t .

The RMSE is helpful in a variety of ways. First, it offers a straightforward measurement of how successfully a model can forecast future values from historical data. Second, It enables us to evaluate the performance of various models., and choose the one that provides the lowest RMSE. Third, it helps us to identify which parts of the time series are difficult to predict, which can be useful in deciding how to refine the model.

In general, a lower RMSE denotes better forecasting accuracy. However, it is crucial to remember that the RMSE must always be viewed in the context of the particular issue being addressed and that there may be situations in which other metrics are more appropriate for evaluating a model's performance.

5.3 Numerical results

5.3.1 Preliminary study

Figure 6 shows the stock price data for the years between 2000 and 2006. The x axis contains daily data and the y axis contains RMSE. As mentioned before, in the first stage, the main goal was to find out which values of the model's parameters can lead to the

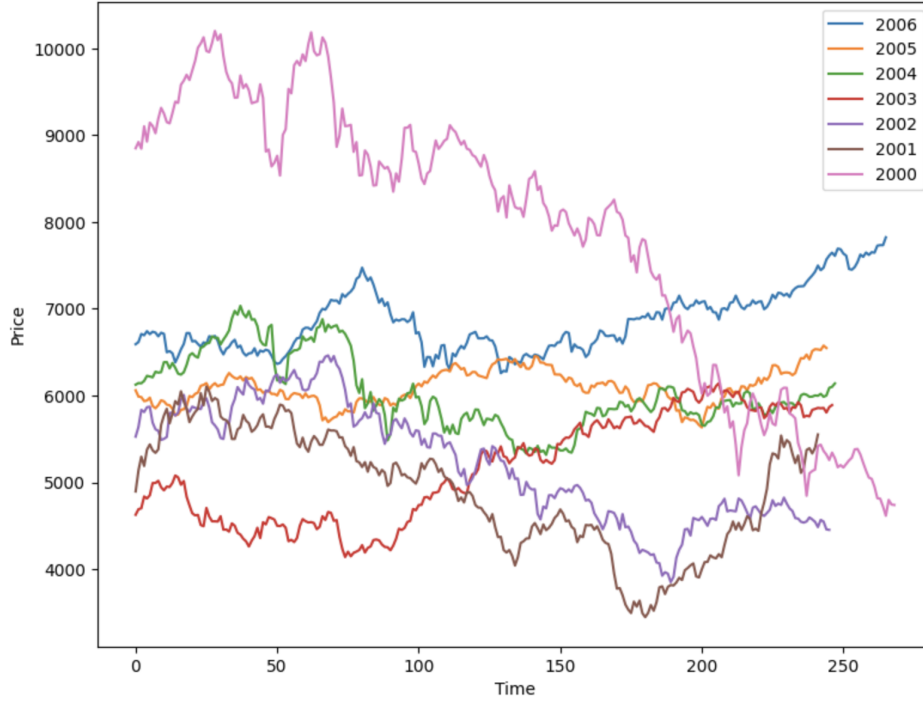


Figure 6. Daily time series data for stock prices between 2000 and 2006 years

least RMSE. It is worth mentioning that having a small RMSE means having a better model. Overall, there is not a single triplet of values working for all the years, in fact for each year, the lowest RMSE is achieved for different values of parameters α , γ and m . The parameters α and γ influence the learning and convergence behavior of the model. The optimal values of these parameters depend on the specific problem being addressed and the nature of the input and output variables. For running experiments, various values were adopted, namely $\gamma \in \{0, 1, 0.1, 0.01, 10\}$, $\alpha \in \{0.1, 0.3, 0.9, 1\}$ and $m = 2, 3$ was used.

In the next stage, we separate values and try to compare and see for which parameters' values the model gets rather better results. As a consequence of comparing results, the experiments were performed by dividing some parts. Firstly, the results regarding the number of terms m were analyzed. In Figure 7 and 8, the plots show RMSE and time according to the different values of m (2 and 3). From the plot in Figure 7, it can be seen that we have a slightly higher RMSE when the number of terms $m = 3$ is chosen, while $m = 2$ shows a lower RMSE in the model. As our model has a training time, results can be analyzed regarding this as well. Figure 8 shows correlation between the training time and the number of terms. It is clear from the box plot that we have a longer training time when $m = 3$, Therefore, for the second step $m = 2$ is chosen as the better number of terms.

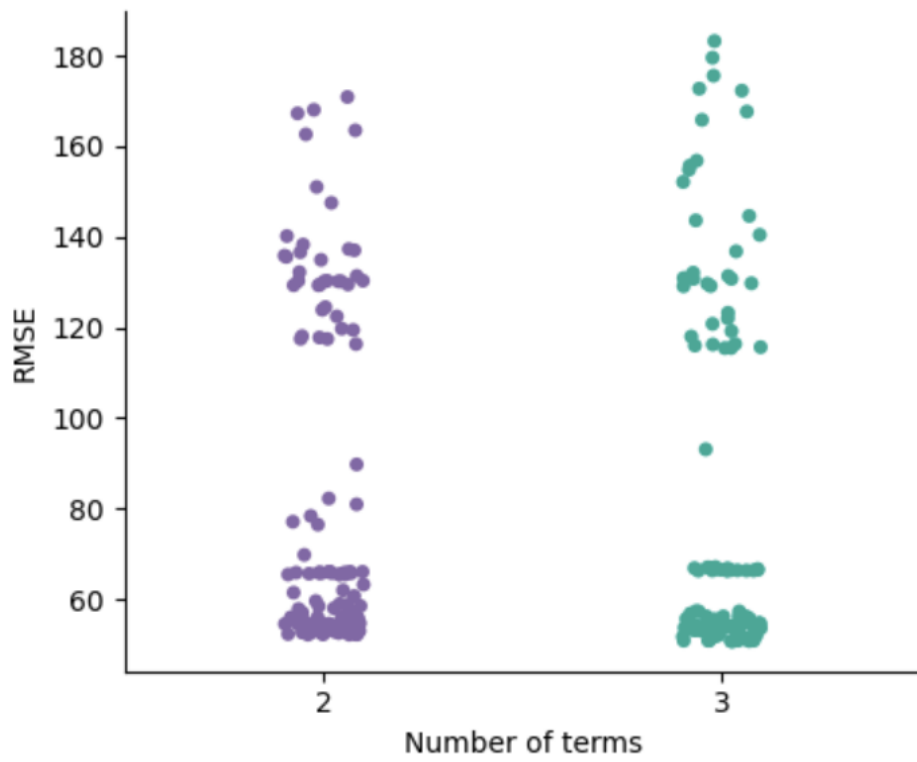


Figure 7. Changing RMSE in different number of terms m

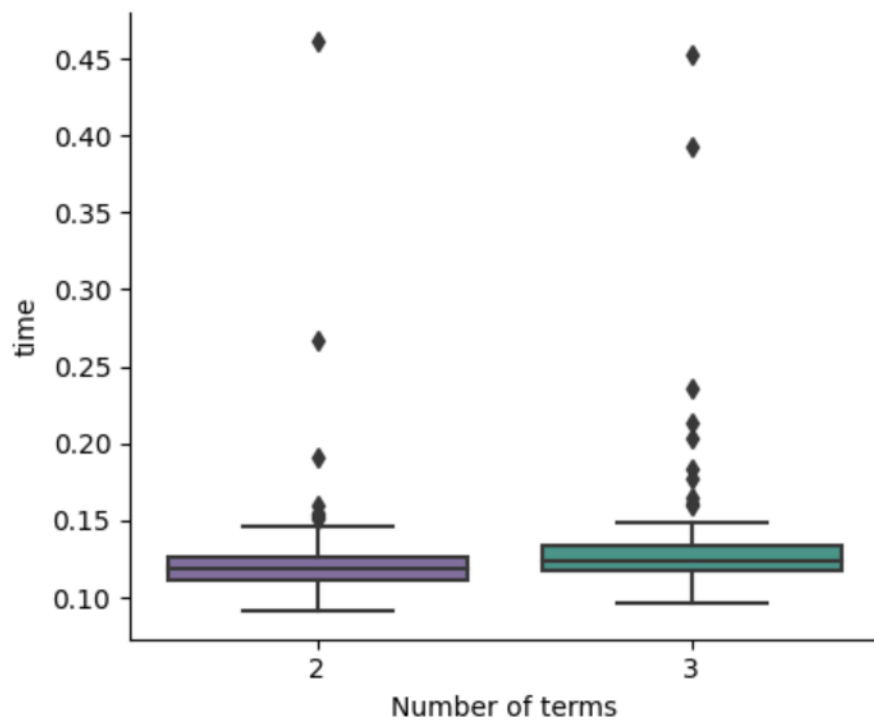


Figure 8. Training time in different number of terms m

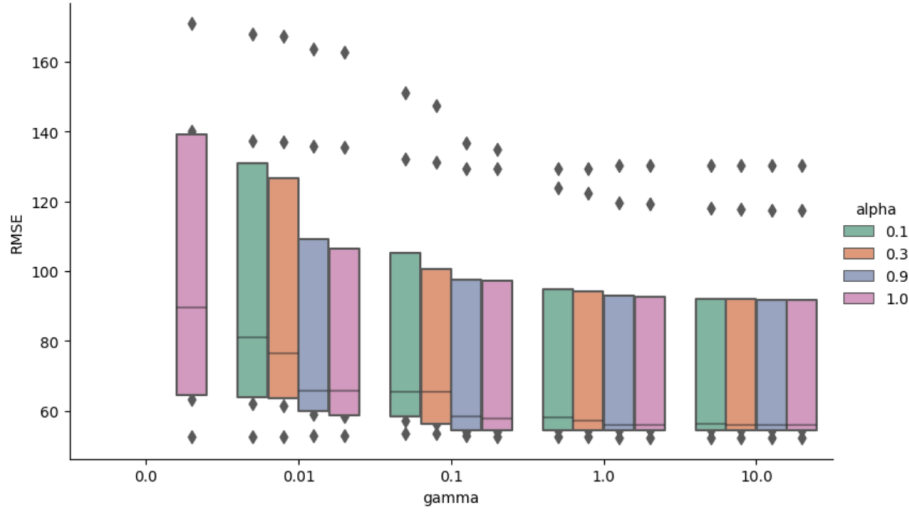


Figure 9. Change in computational time in different values of γ

After choosing $m = 2$, moving forward comparing outcomes regarding different values of γ is the next step. Figure 9 shows the relationship between different values of gamma and RMSE and as mentioned above, in ANFIS-T when $\gamma = 0$, only $\alpha = 1$ case is possible (case without regularization). From the graph, it can be easily noticed that for $\gamma = 0$, the RMSE result is the highest. It can be seen that increasing gamma makes the RMSE results better; $\gamma = 1$ and $\gamma = 10$ show the best results in terms of RMSE. Nevertheless, it is worth checking training time for several values of gamma, since it also affects our decision. Figure 10. shows a correlation between time and gamma.

According to Figure 10, the smaller time corresponds to $\gamma = 10$. For this reason, $\gamma = 10$ is chosen.

In the last stage of the preliminary study, the correlation between α and RMSE was checked. It can be seen from Figure 9 as well. Since $\gamma = 10$ was chosen, when $\gamma = 10$ there is no big difference in RMSE results for various values of alpha. In that case, to find a better value for α , the training time can be checked. From Figure 11, it can be seen that both $\alpha = 0.9$ and $\alpha = 1$ show rather better results in terms of RMSE. However, when $\alpha = 1$, there is an outlier. Hence, $\alpha = 0.9$ is chosen as the best value.

Taking all these into account, according to numerical experiments with the model, when we choose $m = 2$, $\alpha = 0.9$ and $\gamma = 10$, the results look better in most cases, although this is not a general rule. In other words, both RMSE and training time are rather less when these values of parameters are chosen.

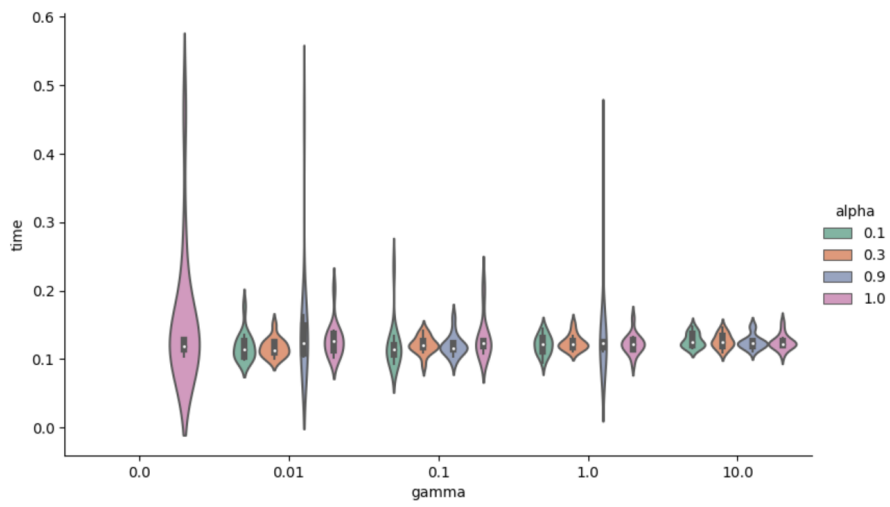


Figure 10. Change in computational time in different values of γ

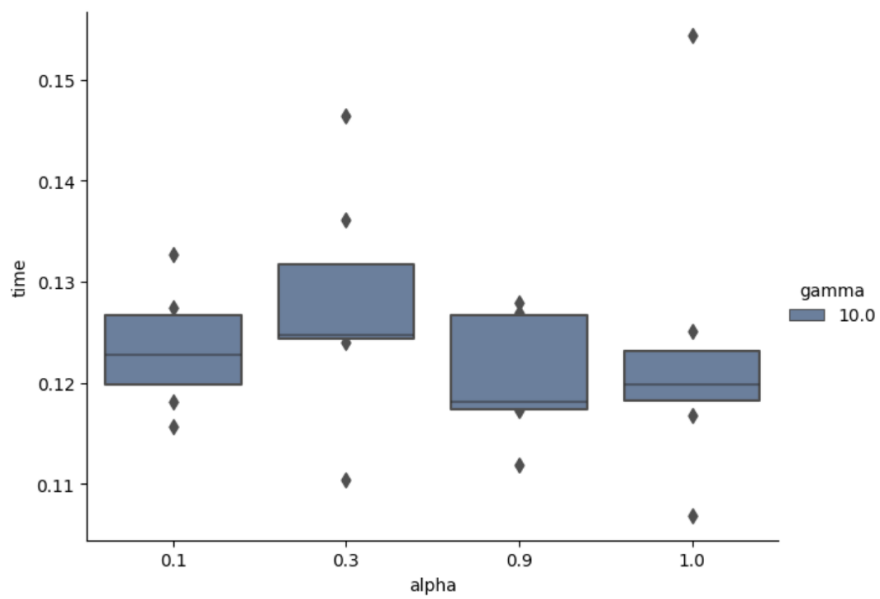


Figure 11. Change in computational time in different values of α

5.3.2 Comparative study

In this part of the analysis, RMSE and training time will be compared for different models. In Table 3, one can see the results in terms of RMSE. To get the best parameters for ARIMA model **pm.auto arima** from pmdarima library in Python was used. It helps to find the best parameters according to data. As a result of this function for all years order(0,1,0) - AR parameter - $p = 0$, differencing $d = 1$, MA parameter $q = 0$ parameters were chosen as the best. ARIMA and Naive model results can be the same, since using (0,1,0) order for ARIMA model can give similar result to Naive model. However, when we look at the training time results in Table 3, it can be seen that less time was spent in Naive model.

If we compare grey-box models - ANFIS and ANFIS-T, RMSE results are slightly better in most cases in ANFIS-T model. If we look at training time, there is a big difference between these two models. It is worth mentioning that $m = 3$ was used in the ANFIS model [55]. ANFIS-T has a much smaller training time compared to ANFIS model. If we compare results between white-box and grey-box model, even though regarding training time ANFIS-T has less training time than ARIMA model, Naive model has the smallest training time among the 4 models. In respect of RMSE results, white-box models gives slightly better result as well.

Year	Naive	ARIMA	ANFIS	ANFIS-T
2000	127.84	127.84	130	129.08
2001	115.06	115.06	120	115.41
2002	64.82	64.82	65	65.38
2003	54.16	54.16	55	52.49
2004	53.4	53.4	55	54
2005	54.93	54.93	57	50.69
2006	53.24	53.24	85	51.98

Table 2. RMSE Results of Naive, ARIMA, ANFIS, and ANFIS-T Models for Years 2000-2006

Year	Naive	ARIMA	ANFIS	ANFIS-T
2000	0.004	0.03	1.86	0.11
2001	0.007	0.04	3.09	0.12
2002	0.01	0.03	2.22	0.11
2003	0.01	0.04	1.83	0.1
2004	0.01	0.05	1.94	0.12
2005	0.003	0.05	1.95	0.13
2006	0.004	0.04	1.49	0.11

Table 3. Training time of Naive, ARIMA, ANFIS, and ANFIS-T Models for Years 2000-2006

6 Concluding Remarks

In this study, for the purpose of predicting stock prices with a focus on interpretability, the recently proposed grey-box method ANFIS-T was adopted. Developing trust in a model's output depends on its ability to explain how it makes predictions. The main contributions of this thesis are to check if it is possible to provide a general definition for interpretability and compare grey-box and white-box models. Results suggest that the ANFIS-T model can make precise predictions while facilitating a more perceptual understanding of the model's decision-making procedure. It can be seen that with ANFIS-T better results can be achieved compared to the standard ANFIS model, but regarding the comparison of white-box and grey-box models, white-box models outperform others. Literature shows that ANFIS-T was successfully used for high-dimensional datasets but it could not perform well with relatively small datasets as in the examined case. Regarding interpretability, research shows that even though there are different ways to define interpretability, making one unique definition is still tricky.

Future research in this area might examine methods to depict the ANFIS-T model's decision-making process and identify important variables that influence the model's predictions in order to further improve the model's interpretability. It would be beneficial to compare the ANFIS-T model's interpretability and precision with other machine learning models.

Overall, our research demonstrates that adding interpretability to financial forecasting models can produce predictions that are clearer and more reliable. In this aspect, the ANFIS-T model is a promising tool, but more investigation is required to fully explore its potential and overcome its shortcomings.

References

- [1] O. Ozyegen and M. Cevik, "Evaluation of interpretability methods for multivariate time series forecasting," *Applied Intelligence*, vol. 52, no. 1, pp. 4727-4743, 2022.
- [2] Z. C. Lipton, "The Mythos of Model Interpretability," *ICML Workshop on Human Interpretability in Machine*, 2016.
- [3] I. Lage, E. Chen, J. He, M. Narayanan and B. Kim, "Human Evaluation of Models Built for Interpretability," *National Library of Medicine*, vol. 7, no. 1, pp. 59-67, 2019.
- [4] O. D.Anderson, "More effective time-series analysis and forecasting," *Journal of Computational and Applied Mathematics*, vol. 64, no. 1-2, pp. 117-147, 1995.
- [5] D. Cheng, F. Yang, J. Liu and S. Xiang, "Financial time series forecasting with multi-modality graph neural network," *Pattern Recognition*, pp. 1-10, 2022.
- [6] A. Dingli and S. Fournier, "Financial Time Series Forecasting – A Deep Learning Approach," *International Journal of Machine Learning and Computing*, vol. 7, no. 5, pp. 1-5, 2017.
- [7] J. M. V. Barberan, "Adding external factors in Time Series Forecasting," *KTH Royal Institute of Technology*, pp. 1-68, 2020.
- [8] M. Krishnan, "Against Interpretability: a Critical Examination of the Interpretability Problem in Machine Learning," *Philosophy Technology* , vol. 33, no. 1, 2020.
- [9] W. J. Murdoch, C. Singh, K. Kumbier and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *PNAS*, vol. 116, no. 44, pp. 22071-22080, 2019.
- [10] R. Patir, S. Singhal and V. Goyal, "Interpretability of Blackbox Machine Learning Models through Dataview Extraction and Shadow Model creation," *Artificial intelligence*, pp. 1-13, 2020.
- [11] C. Molnar, *Interpretable Machine Learning - A Guide for Making Black Box Models Explainable*, Munich, Germany: paperback, 2022, pp. 1-329.
- [12] <http://tylervigen.com/spurious-correlations>
- [13] A. Mukherjee, "Trust and interpretability in machine learning," 30 03 2019. [Online]. Available: <https://towardsdatascience.com/trust-and-interpretability-in-machine-learning-b7be41f01704>. [Accessed 04 05 2022].

- [14] A. Mukherjee, "Trust and interpretability in machine learning," 30 03 2019. [Online]. Available: <https://towardsdatascience.com/trust-and-interpretability-in-machine-learning-b7be41f01704>. [Accessed 04 05 2022].
- [15] Kim, Been, Glassman, Elena, Johnson, Brittney, and Shah, Julie. ibcm: Interactive bayesian case model empowering humans via intuitive interaction. 2015.
- [16] Huysmans, Johan, Dejaeger, Karel, Mues, Christophe, Vanthienen, Jan, and Baesens, Bart. An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. Decision Support Systems, 2011.
- [17] P. Linardatos and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," Entropy, vol. 23, no. 18, pp. 1-45, 2021.
- [18] Y. Wang, "Interpretable time series classification," Neural and Evolutionary Computing, vol. 1, pp. 1-119, 2021.
- [19] Ian J. Goodfellow, Jonathon Shlens, Christian Szegedy, "Explaining and Harnessing Adversarial Examples"
- [20] Nesma Settouti, Meryem Saidi, Preliminary analysis of explainable machine learning methods for multiple myeloma chemotherapy treatment recognition
- [21] C. O'Sullivan, "Interpretability in Machine Learning," 21 10 2020. [Online]. Available: <https://towardsdatascience.com/interpretability-in-machine-learning-ab0cf2e66e1>. [Accessed 08 05 2022].
- [22] Baesens, B., Mues, C., Martens, D., Vanthienen, J. (2009). 50 years of data mining and OR: Upcoming trends and challenges. Journal of the Operational Research Society, 60(1), S16–S23.
- [23] Bellazzi, R., Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics, 77(2), 81–97.
- [24] Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1721–1730). ACM.
- [25] Lou, Y., Caruana, R., Gehrke, J. (2012). Intelligible models for classification and regression. In 18th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 150–158). ACM.
- [26] Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. Artificial Intelligence in Medicine, 23(1), 89–109.

- [27] Domingos, P. (1999). The role of Occam's razor in knowledge discovery. *Data Mining and Knowledge Discovery*, 3(4), 409–425.
- [28] Mencar, C., Castellano, G., Fanelli, A. M. (2007). On the role of interpretability in fuzzy data mining. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(05), 521–537.
- [29] K. Aas, A. Loland and M. Jullum, "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values," *Artificial Intelligence*, vol. 1, no. 1, 2021.
- [30] M. Benk and A. Ferrario, "Explaining Interpretable Machine Learning: Theory, Methods and Applications," *SSRN Electronic Journal*, vol. 1, 2020.
- [31] M. W. Shen, "Trust in AI: Interpretability is not necessary or sufficient, while black-box interaction is necessary and sufficient," vol. 1, no. 1, pp. 1-21, 2022.
- [32] P. J. Brockwell, "Time Series Analysis," *International Encyclopedia of Education*, 2010.
- [33] ABS, "Time Series Analysis: The Basics," Australian Bureau of Statistics, 2020.
- [34] C. Mencar, Interpretability of Fuzzy Systems. In: Masulli F., Pasi G., Yager R. (eds) *Fuzzy Logic and Applications. WILF 2013. Lecture Notes in Computer Science*, 2013, vol 8256. Springer, Cham.
- [35] H. Ishibuchi, Y. Nojima, Discussions on Interpretability of Fuzzy Systems using Simple Examples, 2009, IFSA-EUSFLAT 2009, 1649-1654.
- [36] A. Riid, E. Rustern (2014). Adaptability, interpretability and rule weights in fuzzy rule-based systems. *Information Sciences*, 2014, 257 , 301–312 .
- [37] Anderson, O. D. (1976). *Time series analysis and forecasting*. Butterworths.
- [38] Box, G. E. P., Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden-Day.
- [39] R. Sharia, A THEORETICAL AND EMPIRICAL STUDY OF STOCK MARKET DEVELOPMENT, ECONOMIC REFORM AND ECONOMIC GROWTH: A CASE STUDY OF ARAB COUNTRIES, University of Western Sydney, 2005.
- [40] V. Novak and S. Mirshahi, "On the Similarity and Dependence of Time Series," *Mathematics*, vol. 9, no. 1, pp. 1-14, 2021.

- [41] R. F. Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, vol. 50, no. 4, pp. 987-1007, 1982.
- [42] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307-327, 1986.
- [43] G. Box and F. Jenkins, "Time Series Analysis: Forecasting Control," Holden-Day, 1976.
- [44] K. Huarng, "Effective lengths of intervals to improve forecasting in fuzzy time series," *Fuzzy Sets and Systems*, vol. 3, no. 1, pp. 872-394, 2001.
- [45] S. Chen and N. Chung, "Forecasting enrollments using high-order fuzzy time series and genetic algorithms," *International Journal of Intelligent Systems*, vol. 21, no. 5, pp. 485-501, 2006.
- [46] H.-K. Yu, "Weighted fuzzy time series models for TAIEX forecasting," *Physica A: Statistical Mechanics and its Applications*, vol. 349, no. 3-4, pp. 609-624, 2005
- [47] E. Hoseinzade and S. Haratizadeh, "CNN-based stock market prediction using a diverse set of variables," *Expert Systems with Applications*, vol. 129, no. 1, pp. 273-285, 2019.
- [48] M. Kumbure, P. Luukka and J. Porras, "Machine learning techniques and data for stock market forecasting: A Literature Review," *Expert Systems With Applications*, vol. 1, pp. 1-41, 2022.
- [49] S. Chouksey, "Stock Price Prediction Using Time Series Models," *Dublin Business School*, pp. 1-43, 2019.
- brockwell Brockwell, P. J., Davis, R. A. (1991). *Time series: Theory and methods*, 2nd Edition. Springer.
- [50] https://www.twse.com.tw/en/page/trading/indices/MI_5MINS_HIST.html
- [51] S. Tomasiello, W. Pedrycz, V. Loia, (2022) *Contemporary Fuzzy Logic: A Perspective of Fuzzy Logic with Scilab*. Springer.
- [52] J. S. R. Jang. Anfis: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems Man & Cybernetics*, 23:665–685, 1993.
- [53] S. Tomasiello, W. Pedrycz, V. Loia, On Fractional Tikhonov Regularization: Application to the Adaptive Network-Based Fuzzy Inference System for Regression Problems, *IEEE Transactions on Fuzzy Systems*, 2022, in press

- [54] M. E. Hochstenbach, L. Reichel, Fractional Tikhonov regularization for linear discrete ill-posed problems, BIT Numerical Mathematics, vol. 51, pp. 197–215, 2011
- [55] Liang-Ying Wei, A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting

I. Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, Narmin Mammadli,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,
Interpretable approaches for financial time series forecasting,
supervised by Prof. Stefania Tomasiello and Dr. Toomas Raus.
2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work.
3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.
4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Narmin Mammadli
16.05.2023