ARTIFICIAL INTELLIGENCE STUDIES
2022, 5(2):47-55
www.aistudies.org

Aralık Değerli Veri Analizi: Bir İnceleme

Mustafa AL-ASADI[*,a]
[a] *Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, Selçuk Üniversitesi, 42130 KONYA, TÜRKİYE*

MAKALE BİLGİSİ
Alınma: 20.09.2022
Kabul: 11.11.2022

Anahtar Kelimeler
Aralık değerli veriler, Makine öğrenmesi, Veri madenciliği, doğrusal regresyon, Doğrusal olmayan regresyon.

[*]**Sorumlu Yazar:**
e-posta: masadi@lisansustu.selcuk.edu.tr

ÖZET

Aralık değerli veriler, tek değerler yerine aralıklar olarak gözlemlenmekte ve veri toplama süreçlerinde ve makine öğrenmesi tekniklerinde ileri teknolojilerle sıklıkla kullanılmaktadır. Aralık değerli veri tahmin görevleri için doğrusal ve doğrusal olmayan regresyon yöntemleri kullanılmış ve doğrusal olmayan yöntemlerin diğer yöntemlere üstünlüğü kanıtlanmıştır. Bununla birlikte, önceki literatür, aralık değerli verilerin analizinde çoğunlukla parametrik doğrusal regresyon modellerine odaklanmıştır. Sonuç olarak, aralık değerli verilerin analizi ve incelenmesi hala erken aşamalarında olup, üzerinde çalışılacak pek çok çevrilmemiş taş ve birçok benzer yöntem bırakmaktadır. Bu makale, özet içgörüler ve değerlendirme için aralık değerli verileri analiz etmek için kullanılan en önemli yöntemlerin bir incelemesini sunar.

DOI: 10.30855/AIS.2022.05.02.02

Interval-valued Data Analysis: A Review

ARTICLE INFO
Received: 20.09.2022
Accepted: 22.11.2022

Keywords
Interval-valued data, Machine learning, Data mining, linear regression, Nonlinear Regression.

[*]**Corresponding Authors**
e-mail: masadi@lisansustu.selcuk.edu.tr

ABSTRACT

Interval-valued data is observed as ranges rather than single values and is frequently observed in advanced data collection processes and machine learning techniques. The linear and non-linear regression methods were used for the interval valued data prediction (IVD) tasks, and the superiority of the non-linear methods over the other methods was proven. However, the previous literature primarily concerned with parametric linear regression models in the analysis of interval valued data. Thus, the analysis and study of interval-valued data are still under development, leaving many stones unturned and many similar methods to study and work with. This paper presents a review of the most important methods used to analyze interval-valued data to provide summary insights and evaluation.

DOI: 10.30855/AIS.2022.05.02.02

1. INTRODUCTION (*GİRİŞ*)

The process of cleaning, preprocessing, and modeling data in order to discover information that can be used to make appropriate decisions is known as data analysis. The goal of analysis is to extract insights from data and make decisions based on that information [1]. Based on business and technology, there are various types of techniques for data analysis. Data analysis techniques, on the other hand, are divided into two categories methods based on statistical and mathematical strategies (Statistical Analysis) and methods based on machines learning and artificial intelligence (Predictive Analysis). Unlike statistics, where models are used to understand data and identify correlations, predictive analytics is focused on developing models to predict future values based upon past and present data sets. Predictive analytics relies on **machine learning algorithms** to create predictive models to perform **classification and regression** operations [2]. Machine learning is an artificial intelligence subfield that broadly defines as a machine's ability to imitate human abilities. Artificial intelligence (AI) systems are used to solve complex problems in a manner similar to how humans solve problems [3, 4]. To identify data distributions, statistics employs basic mathematical formulas and concepts such as determining the median, mean, mode, hypothesis testing, standard deviation calculation, and variance [5]. However, predictive analytics are implemented based on statistical analysis.

In predictive analytics, predictive models use available data to create or train a model that predicts values for new or different data. Statistics, on the other hand, summarizes data for public consumption. There are two major statistical techniques: inferential statistics and descriptive statistics [6].

Single data has been the paradigm until now in statistics, machine learning for work, and data analysis. However, new kinds of data are arisen called "**Interval data**". This data is a typical interval with lower and upper bounds such as weekly highs and lows in daily temperatures, heart tension values and many other magnitudes that do not give the variables as a single value as we are used to [7-9].

Table 1 shows the cardio logical interval dataset taken from [10]. The dataset contains 44 rows, and five attributes were recorded as intervals for each patient: pulse rate (Pulse), diastolic blood pressure (Diast), systolic blood pressure (Sys), Art1 and Art2. The goal was to predict (Pulse).

Table 1 shows a dataset of cardiological intervals. Syst means systolic; Diast means diastolic; Art means artificial.

	Pulse	**Syat**	**Diast**	**Art1**	**Art2**
1	[45, 69]	[91, 100]	[51, 70]	[5, 9]	[1, 7]
2	[60, 73]	[91, 135]	[72, 92]	[3, 9]	[5, 6]
3	[57, 91]	[145, 182]	[91, 101]	[10,11]	[5, 7]
4	[73, 115]	[112, 140]	[82, 107]	[7, 9]	[5, 6]
5	[55, 74]	[93, 100]	[60, 75]	[5,8]	[7, 9]

Prediction uncertainty refers to the variability in prediction due to input values [11]. Uncertainty is quantified using a probability distribution that is based on our current knowledge of the likelihood of what the single, true value of the uncertain quantity is [12]. Both **confidence intervals** and **prediction intervals** express uncertainty in statistical estimates [13]. The prediction interval in interval data predicts where a future individual observation will fall**.** As a result, prediction intervals provide a method for quantifying and communicating the degree of uncertainty in a prediction. A confidence interval, on the other hand, in the single data shows the likely range of values associated with some statistical parameter of the data, such as the mean or standard deviation. It determines uncertainty in population parameters in this manner. Prediction intervals, on the other hand, describe the uncertainty surrounding a single specific outcome [14].

A prediction or estimate is a single actual result value given some input variables in predictive modeling. As an example:

yhat = model.predict (X) (1)

Where *yhat* is the trained model's estimated outcome or prediction for the given X input data. This is a point forecast. It is an approximation or estimate with some uncertainty, according to the definition.

Interval-valued data (IVD) is data in which each feature is an interval. [15]. Interval-valued data can be expressed as intervals with lower and upper limit values for the observed values of the variables. [16]. The predictive model for interval-valued data is more difficult to capture than point prediction because it must capture the link between the response variables $Y = (Y^L, Y^U)$ and $X = (X^L, X^U)$ the explanatory variables [17]. A prediction interval expresses the degree of uncertainty in a prediction. It provides predictive upper and lower bounds on the estimate of an outcome variable [14]. An interval of prediction for a single future observation, on the other hand, is an interval that will contain a future randomly selected observation from a distribution with a specified degree of confidence [18]. Prediction intervals are most widely used when predicting or forecasting a quantity using a regression model. Figure 1 depicts the relationship between prediction, actual value, and prediction interval.

Given a prediction of 'y' and a result of 'x,' there is a 95% chance that the range 'a' to 'b' covers the true outcome [14]. The two most common methods for representing IVD are the midpoint and the boundary. However, because only midpoint or endpoint is used, their structure information (such as size and location) may be incomplete, resulting in poor data processing results [15]. This paper provides an overview of the most important methods for analyzing interval-valued data in order to provide summary insights and estimates.

The remainder of this paper is organized as follows. Section II provides information about Interval-valued Data Analysis Approaches. Section III and Section IV explain the linear and nonlinear regression models used for the Interval-valued Data models and the most important related studies. Finally, Section V provide the conclusions.

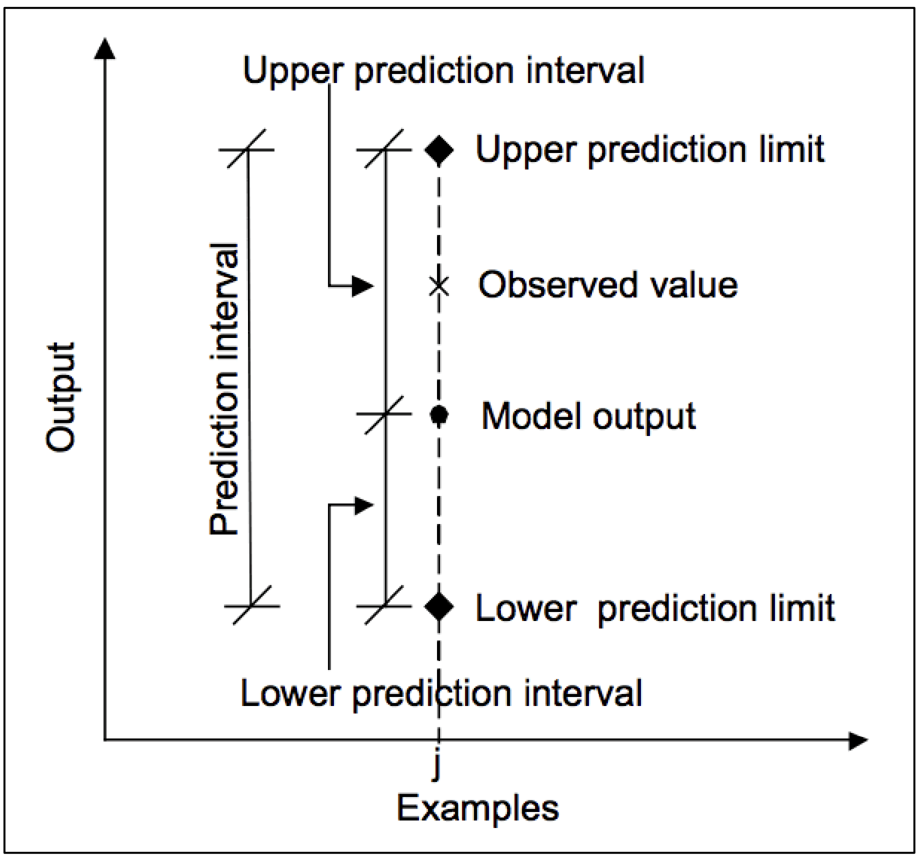

Figure 1: Relation between actual value, prediction, and prediction interval.

2. INTERVAL-VALUED DATA ANALYSIS APPROACHES *(ARALIK DEĞERLİ VERİ ANALİZİ YAKLAŞIMLARI)*

Many statistical data are imprecise as a result of measurement errors, a lack of information, and computation errors. In such cases, intervals are preferable to single numbers for representing data. Two existing methods for analyzing interval-valued data are regressions in the metric space of intervals and symbolic data analysis (SDA). There has, however, been a paucity of literature on parametric modeling and distribution-based inferences for interval-valued data [19].

Symbolic data analysis (SDA) is a new area of statistics concerned with understanding and modeling data in the form of symbols, such as intervals and histograms. It was built on the assumption that the statistical unit of interest is the symbol and that inference is required at this level [20]. By extending the standard input to a set of classes of individual entities, symbolic data analysis (SDA) in general opens up a new way of thinking in Data Science. As a result, classes within a given population are regarded as units of a higher-status population to be studied. Such classes are frequently the actual units of interest. Intervals, distributions, sets of categories, or numbers are used to describe classes, sometimes weighted, and the like to account for variability among members of each class. As a result, we obtain new types of data, dubbed "symbolic," due to they cannot be reduced to numbers without sacrificing a great deal of information

SDA begins with the creation of a symbolic data table with rows representing classes and variables that can take symbolic values. The second step is to investigate and extract new knowledge from these new kinds of data, preferably using a symbolic data extension of Computer Statistics and Data Mining. SDA is a new paradigm that provides results that are complementary to conventional methods applied to standard data, thereby opening up a vast domain of applications and research [21].

Regression is one of the **SDA** techniques in which research has advanced recently [20]. Intervals, unlike single data, have many significant values (center, radius, lower limit and upper limit) making classical regression methods complicated to apply. The first-ever basic methods were created for interval regression whereby Billard and Diday in 2000 [22] and 2002. Later they were picked up and improved by Neto and Carvalho in 2008 [23] and 2010.

Many of these methods, especially the earliest ones, use single-value regression analysis applied to different parameters of the interval; the most recent ones, however, like Sinova et al in 2012 [24] or Souza et al in a paper in 2017 [25] try to go deeper and use techniques like the arithmetic of intervals or its parameterization.

A prediction interval is determined as some combination of the estimated variance of the model and the variance of the outcome variable. In straightforward cases, such as linear regression, we can directly estimate the prediction interval. It is much more difficult in the case of nonlinear regression algorithms and necessitates the selection and implementation of specialized techniques [14].

3- INTERVAL LINEAR REGRESSION MODELS *(ARALIK DOĞRUSAL REGRESYON MODELLERİ)*

For usual single data, regression is something that has been worked with for a long time. Throughout the twentieth century many articles, books or manuals have already established a great foundation of regression methods for simple or multiple variables, one example of this is Rencher's Methods of Multivariate Analysis [26] published in 1995 which includes methods for simple or multiple, multivariate or univariate regression. In interval-valued data, however, something as basic as a simple linear regression has barely been worked within the last 15-20 years and no clear method has been established as the norm.

3.1 Related Works

According to the literature, there are eight different basic methods of ILR. The basis of most of them is the application of classic crisp linear regression to different interval parameters.

3.1.1 Centre Method (CM)

The Centre Method was introduced by Billard and Diday in 2000 [22]. This method obtains a regression model from the centers of the intervals and applies such a model to the upper and lower limits to obtain [Yˆ]. This model was the first one and because of that, it was still very rudimentary and did not take into account the range part of the intervals.

3.1.2 MinMax method

The *MinMax* method was introduced by Billard and Diday in 2002 [23] as an improved alternative to the *Centre Method* developed by the same authors a couple of years earlier. This method states the fact of using different parameters to obtain the lower and upper limits of [Yˆ]. This is the same as supposing independence between the values of the upper and lower bounds and treating them as if they were different variables. This model is a clear upgrade from the Centre method as it takes into account the singularities that interval-valued data have and by taking into account both upper and lower limits independently it makes a much more reliable estimation for [Yˆ]. It was the first real interval regression model that provided a decent estimation to work with and an example to look to by models developed later.

3.1.3 Centre and Range Method (CRM)

The Centre and Range Method were introduced by Lima Neto and De Carvalho in 2008 [24] also as an improved alternative to the Centre Method developed by Billard and Diday [22]. This method takes the idea of the Centre Method and decides to add the ranges or radii information to the model to improve the prediction performance. This way follows the idea of the Minmax method to add more information than the CM to meet the necessities of interval-valued data complexity; in this case the range and center. This model uses the same idea as the MinMax method to include more information in the CM method to improve its accuracy. It produces a similar performance as that of the MinMax method and improves that of the CM method adding the extra information needed for intervals.

3.1.4 Constrained Centre and Range Method (CCRM)

The Constrained Centre and Range Method were introduced by Lima Neto and De Carvalho in 2010 [22] as an improved version of their CRM presented a couple of years before [24]. The CCRM Method uses the same principle of the CRM of using the center and radii of [X1, . . . , Xn] to predict [Yˆ] but it includes a needed restriction that the radii cannot be negative; something that was possible with the CRM and that makes no mathematical sense. This model is a step forward from CRM as, not only does it provide a good prediction, but also it accounts for cases that CRM or even MinMax would create an illogical result like intervals with negative radii or lower limits bigger than upper limits.

3.1.5 Arithmetic-based simple linear regression (ABSLR)

The *Arithmetic-based simple linear regression* was introduced by Sinova et al in 2012 [27]. This method is different from the previous ones exposed as it uses an arithmetic approach for the regression. It also differs in the fact that is the only method in the tool only for simple regression.

It should be mentioned that a linear model based on standard interval arithmetic considers both the center and radius of IVD (i.e., each interval data as a whole), ensuring that the interval arithmetic-based model is always well-defined. Instead, we could examine the center and radius separately, or we could treat them as separate random variables. However, one cannot assure that a separate linear model for radius, as discussed in previous models, always makes sense.

3.1.6 Linear regression based on Lasso technique

Paolo Giordani introduced *the Linear regression based on Lasso technique in* 2015 [28]. Because the problem is addressed as an optimization problem involving the constrained minimization of an objective function, this method may resemble the CCRM approach. It will try to find a shared set of regression coefficients for the midpoint and radius. This will be accomplished by adding specific regression coefficients for the radii in order to properly deal with all situations in which the slope differs from the propagation of the imprecision.

3.1.7 Constrained center and range joint method (CCRJM)

The *Constrained center and range joint method* was introduced by Hao and Guo in 2017 [29]. This method is a step forward from the CCRM method, it looks for a coefficient for the center and the radius like CCRM does but does it jointly, it uses the radii values for the center model and the centers for the radius model. It goes a step ahead because not only does it account for the fact that interval-valued data have multiple components like the previous methods, but it also acknowledges the relationship among those components.

3.1.8 Parameterized Method for Linear Regression of Interval Data

The *Parameterized Method for Linear Regression of Interval Data* was introduced by Souza et al in 2017 [30]. Just like CCRJM, this method is a step forward from the MinMax method, it looks for a coefficient for the lower limit and the upper limit like the MinMax method does but does it jointly, it uses the lower limit values for the upper limit model and the upper limits for the lower limit model. Like CCRJM It goes a step ahead because not only does it account for the fact that interval-valued data have multiple components like the previous methods, but it also acknowledges the relationship among those components.

Aside from the eight major linear regression methods mentioned above, Sun and Ralescu proposed a linear model in 2015 to optimize the balance between linear model flexibility and interpretability. The authors presented a general classifier for multiple predictors in matrix, from which the LS predictions of the model parameters are directly derived with a number of nice linear model theory properties [31].

4- INTERVAL NONLINEAR REGRESSION MODELS (*ARALIK DOĞRUSAL OLMAYAN REGRESYON MODELLERİ)*

As we discussed in the previous section, regression methods for interval value variables have recently been proposed. The majority of these contributions, however, have taken into account a linear relationship between the response IVD variable and the set of explanatory IVD variables. As a result, nonlinear methods capable of dealing with interval-valued data have received relatively less attention. However, because many real-world problems have a nonlinear relationship between variables, the association between them cannot be represented by a linear regression model. Two analytics methods, **Predictive Modeling** and **Predictive Analytics,** were used for interval-valued data.

Predictive modelling involves running a machine learning algorithm on data to predict the probability of an outcome and can be applied to any unknown event based on a specified amount of input data. There are two classes of predictive models: **Parametric Model** and **Non-Parametric Model** [32].

In predictive modelling, data analysis techniques based on machine learning algorithms are used to model linear and nonlinear problems. Simple and multiple linear regression were used to solve linear problems, and complex algorithms such as neural networks and decision trees were used to solve nonlinear problems [33]. Furthermore, several optimization techniques can be used to obtain parameter estimates for a nonlinear model using machine learning [34].

Predictive analytics includes a wide range of statistical techniques, including data mining techniques (such as feature selection) and machine learning algorithms (like Regression). Tools from both areas are applied to existing large data sets to: Identify patterns trends and improve prediction quality [2, 35].

4.1 Related works

Because the function describing the relationship between the response and explanatory variables is known, linear and nonlinear models are examples of parametric regression models.

The shape of the functional relation between the response and explanatory variables is not predetermined in nonparametric regression, but can be adjusted to capture unusual. Parametric regression models should be used when the relation between the dependent and independent variables is known. Nonparametric regression models should be used if the relation is unknown and nonlinear [36].

In real-life applications, data rarely follows a linear structure. Based on the fact the toolkit of regression models for interval value data sets (which focused on linear methods) has been extended to nonlinear methods.

Lima Neto and De Carvalho proposed developing the first nonlinear regression model for IVD in 2017. Their model similar to the regression analysis for each of the previous approaches center points and ranges, except that they use a nonlinear function instead. The new method takes into account two independent nonlinear regression models over the range and midpoint of the intervals, allowing for the fitting of different nonlinear functions for the range and midpoint [37].

Yan et al. (2019) proposed a regularized deep artificial neural network (RANN) method for predicting IVD. This method incorporates a non-crossing regularizer in neural nets to control the interval-crossing issue. The proposed method solves complex nonlinear issues and can flexible balance prediction accuracy and interval crossing. According to experimental results on both simulation and real-world data, the suggested RANN model is an effective tool for interval-valued prediction tasks, particularly for complicated non-linear datasets. In general, the RANN model outperforms its counterparts in the linear case, and it is competitive with the CCRM and the Lasso-IR [17].

Jang and Kang presented a nonparametric regression model based on the kernel function and a nonlinear regression model for IVD data in 2020. The authors also propose using a nonparametric method, the local linear regression model, on interval-valued data. Simulations are run using various distributions of the center point and range. The results show that the proposed local linear estimator outperforms the other linear methods [16].

Chacón and Rodrguez presented new approaches to fitting regression models for symbolic interval-valued variables in 2021, which improved and extended the center method proposed by Billard and Diday and the center and range technique proposed by Lima-Neto. The proposed regression models, like the previously mentioned methods, include the midpoints and half the length of the intervals as additional variables. The authors considered tree-based models, K-NN, SVM, and ANN to fit the regression models. During the experimental evaluation, the authors discovered that using nonlinear methods greatly improved prediction results in regression problems. Furthermore, when compared to other methods, particularly those based on linear methods, a simple ANN was able to significantly improve predictions [10].

In addition to regression methods, distribution-based inferences were used to improve prediction quality for interval data. In this regard, Sun and Ralescu introduced a new random set model in 2015. The authors broaden Lyashenko's concept of normality for random sets and propose a Normal hierarchical model for random intervals. Furthermore, the authors create a minimum contrast estimator (MCE) for the model parameters that is both consistent and asymptotically normal. Theoretical findings are supported by simulation studies, which yield very promising results [19].

5- CONCLUSIONS (SONUÇLAR)

Intervals have gained a lot of traction in the research world, especially in the 21st century. The analysis and study of Interval-valued Data are still in development and many breakthroughs have occurred only in the early twenty-first century or even in our current decade, leaving still many stones unturned and many similar methods to study and work with. The conclusions are summarized as follows:

1- The traditional statistical methods and the machine learning-based methods were used for the IVD prediction tasks. Linear and nonlinear regression were used for prediction tasks. However, it has been proven that nonlinear regression methods are superior to linear methods [10, 17].

2- Symbolic Data Analysis (SDA) analyzes complex data by aggregating it at the individual level into group-based distributional summaries by intervals, histograms, lists, etc. However, as shown in the field of data mining, it is useful to process and clean data to improve prediction quality and gain new knowledge. However, many efforts have focused on extending traditional statistical methods to respond to issues related to interval-valued data such as the study presented by [19]. Therefore, there is still a need to discover additional, more flexible and efficient methods for analyzing the IVD.

3- Data mining techniques such as feature selection are an important process for selecting parameters of a predictive model and can improve prediction quality. Balancing and feature extraction (like PCA) are also important techniques for improving prediction quality. The studies presented by [38, 39] have demonstrated the effectiveness of these techniques, which have not been widely used with interval-valued data.

Publication Ethics and Conflict of interest statement *(Yayın Etiği)*

Mustafa A. Al-ASADI certifies that the submission is original and that it is not currently being reviewed by another publication. There are no financial interests to disclose.

REFERENCES (KAYNAKLAR)

[1] Islam, M., Data Analysis: Types, Process, Methods, Techniques and Tools. International Journal on Data Science and Technology, 2020. 6(10).

[2] Al-Asadi, M.A. and S. Tasdemır, Predict the value of football players using FIFA video game data and machine learning techniques. IEEE Access, 2022. 10: p. 22631-22645.

[3] Ben-Ari, D., et al., Artificial Intelligence in the practice of law: an analysis and proof of concept experiment. Rich. JL & Tech., 2016. 23: p. 2.

[4] Al-Asadi, M.A. and S. Tasdemír, Empirical comparisons for combining balancing and feature selection strategies for characterizing football players using FIFA video game system. IEEE Access, 2021. 9: p. 149266-149286.

[5] Kumar, V. and M. Garg, Predictive analytics: a review of trends and techniques. Int. J. Comput. Appl, 2018. 182(1): p. 31-37.

[6] Vergura, S., et al., Descriptive and inferential statistics for supervising and monitoring the operation of PV plants. IEEE Transactions on Industrial Electronics, 2008. 56(11): p. 4456-4464.

[7] Maceda Dal-Re, P., Analytics system with interval-valued data. 2018.

[8] Chang, T.-J., et al., An Interval-Valued Time Series Forecasting Scheme With Probability Distribution Features for Electric Power Generation Prediction. IEEE Access, 2022. 10: p. 6417-6429.

[9] Dhibi, K., et al., Interval-Valued Reduced Ensemble Learning Based Fault Detection and Diagnosis Techniques for Uncertain Grid-Connected PV Systems. IEEE Access, 2022. 10: p. 47673-47686.

[10] Chacón, J.E. and O. Rodríguez, Regression Models for Symbolic Interval-Valued Variables. Entropy, 2021. 23(4): p. 429.

[11] McKay, M.D., Evaluating prediction uncertainty. 1995, Nuclear Regulatory Commission.

[12] Begg, S.H., M.B. Welsh, and R.B. Bratvold. Uncertainty vs. Variability: What's the Difference and Why is it Important? in SPE Hydrocarbon Economics and Evaluation Symposium. 2014. OnePetro.

[13] Mazloumi, E., et al., Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction. Engineering Applications of Artificial Intelligence, 2011. 24(3): p. 534-542.

[14] Brownlee, J., Statistical methods for machine learning. Discover how to transform data into knowledge with Python. Machine Learning Mastery. https://machinelearningmastery. com/statistics_for_machine_learning, 2020.

[15] Qi, X., et al., An Interval-Valued Data Classification Method Based on the Unified Representation Frame. IEEE Access, 2020. 8: p. 17002-17012.

[16] Jang, J. and K.-H. Kang, Local linear regression analysis for interval-valued data. Communications for Statistical Applications and Methods, 2020. 27(3): p. 365-376.

[17] Yang, Z., D.K. Lin, and A. Zhang, Interval-valued data prediction via regularized artificial neural network. Neurocomputing, 2019. 331: p. 336-345.

[18] Shrestha, D.L. and D.P. Solomatine, Machine learning approaches for estimation of prediction interval for the model output. Neural Networks, 2006. 19(2): p. 225-235.

[19] Sun, Y. and D. Ralescu, A normal hierarchical model and minimum contrast estimation for random intervals. Annals of the Institute of Statistical Mathematics, 2015. 67(2): p. 313-333.

[20] Beranger, B., H. Lin, and S.A. Sisson, New models for symbolic data analysis. arXiv preprint arXiv:1809.03659, 2018.

[21] Diday, E., Thinking by classes in data science: the symbolic data analysis paradigm. Wiley Interdisciplinary Reviews: Computational Statistics, 2016. 8(5): p. 172-205.

[22] Billard, L. and E. Diday, Regression analysis for interval-valued data, in Data Analysis, Classification, and Related Methods. 2000, Springer. p. 369-374.

[23] Billard, L. and E. Diday, Symbolic regression analysis, in Classification, Clustering, and Data Analysis. 2002, Springer. p. 281-288.

[24] Neto, E.d.A.L. and F.d.A. de Carvalho, Centre and range method for fitting a linear regression model to symbolic interval data. Computational Statistics & Data Analysis, 2008. 52(3): p. 1500-1515.

[25] Neto, E.d.A.L. and F.d.A. de Carvalho, Constrained linear regression models for symbolic interval-valued variables. Computational Statistics & Data Analysis, 2010. 54(2): p. 333-347.

[26] Stroud, T., Methods of Multivariate Analysis. 1995, JSTOR.

[27] Sinova, B., A. Colubi, and G. González-Rodrı, Interval arithmetic-based simple linear regression between interval data: Discussion and sensitivity analysis on the choice of the metric. Information Sciences, 2012. 199: p. 109-124.

[28] Giordani, P. Linear regression analysis for interval-valued data based on the Lasso technique. in Proceeding of 58th World Statistical Congress, Dublin. 2011.

[29] Hao, P. and J. Guo, Constrained center and range joint model for interval-valued symbolic data regression. Computational Statistics & Data Analysis, 2017. 116: p. 106-138.

[30] Souza, L.C., et al., A parametrized approach for linear regression of interval data. Knowledge-Based Systems, 2017. 131: p. 149-159.

[31] Sun, Y. and D. Ralescu, A linear model for interval-valued data. arXiv preprint arXiv:1506.03541, 2015.

[32] Smith, B.L., B.M. Williams, and R.K. Oswald, Comparison of parametric and nonparametric models for traffic flow forecasting. Transportation Research Part C: Emerging Technologies, 2002. 10(4): p. 303-321.

[33] Mansouri, M., et al., Interval-Valued Reduced RNN for Fault Detection and Diagnosis for Wind Energy Conversion Systems. IEEE Sensors Journal, 2022.

[34] Li, S., et al., Decentralized kinematic control of a class of collaborative redundant manipulators via recurrent neural networks. Neurocomputing, 2012. 91: p. 1-10.

[35] Li, W., et al., Interval dominance-based feature selection for interval-valued ordered data. IEEE Transactions on Neural Networks and Learning Systems, 2022.

[36] Mahmoud, H.F., Parametric versus Semi and Nonparametric Regression Models. arXiv preprint arXiv:1906.10221, 2019.

[37] Neto, E.d.A.L. and F.d.A. de Carvalho, Nonlinear regression applied to interval-valued data. Pattern Analysis and Applications, 2017. 20(3): p. 809-824.

[38] Khosravi, A., et al., Comprehensive review of neural network-based prediction intervals and new advances. IEEE Transactions on neural networks, 2011. 22(9): p. 1341-1356.

[39] Gharsellaoui, S., et al., Interval-valued features based machine learning technique for fault detection and diagnosis of uncertain HVAC systems. IEEE Access, 2020. 8: p. 171892-171902.

Dr. Mustafa AL-ASADI[*a]

Dr. Mustafa AL-ASADI received her master's and PhD degrees from Selçuk University in 2018 and 2022, respectively. His master's thesis and doctoral dissertation were on the subject of decision support systems for football team management by using machine learning and deep learning techniques. He has also developed new methodologies for handling unbalanced data. His research interests include machine learning, deep learning, predictive models, data mining, and pattern recognition.