Neutrosophic Sets and Systems {Special Issue: NeutroGeometry, NeutroAlgebra, and Their Applications}, Vol. 71, 2024

University of New Mexico

Integrating Neutrosophic Theory into Regression Models for Enhanced Prediction of Uncertainty in Social Innovation Ecosystems

Juan Carlos López Gutiérrez[1], Isarelis Pérez Ones[2], Jesús Estupiñán Ricardo[3], and Jean Pierre Ramos Carpio[4]

[1]Higher Institute of Technology for Scientific Research and Innovation, Quito. Ecuador. E-mail: direccionacademica@istici.edu.ec
[2] Latin American Faculty of Social Sciences (FLACSO), Quito. Ecuador. E-mail: isaperezfl@flacso.edu.ec
[3] Higher Institute of Technology for Scientific Research and Innovation, Quito. Ecuador. E-mail: canciller@istici.edu.ec
[4] Higher Institute of Technology for Scientific Research and Innovation, Quito. Ecuador. E-mail: direccionejecutiva@istici.edu.ec

Abstract. This paper focuses on a crucial issue in data analysis: the incorporation of neutrosophic theory into regression prediction to accurately characterize and depict uncertainty in the social innovation ecosystem. The study centers on the limitations of conventional regression methods in modeling intricate and uncertain events related to social innovation, a subject of growing importance in a dynamic global landscape. This paper aims to address the gap in the literature by utilizing neutrosophic theory to provide a more comprehensive and dynamic representation of innovation processes, which are characterized by indeterminacy and ambiguity. The methodology employed for this study involves including neutrosophic numbers in the regression models, therefore enabling a more comprehensive and intricate assessment of the variables associated with social innovation. Through empirical analysis and simulations, the results demonstrate that the neutrosophic approach enhances predictive capability by more effectively capturing the intricacies and uncertainty of the data. This work makes a theoretical contribution to the area by presenting a novel viewpoint on the modeling of social innovation and its inherent difficulties. It also has practical consequences by offering more accurate tools for evaluating and designing innovation practices in social settings. Furthermore, the results enhance the comprehension of how uncertainty can be efficiently controlled in the prediction and decision-making processes of social innovation.

Keywords: Neutrosophic Statistical Prediction, Regression Analysis, Predictive Modeling, Social Innovation.

1 Introduction

The depiction of uncertainty in forecasting is a subject of growing importance in the scientific study of social innovation. In a dynamic and complex world, conventional predictive models are typically inadequate in capturing the extensive dynamics and uncertainties of the innovation ecosystem. The primary objective of this work is to include neutrosophic theory in regression analysis to overcome the existing constraints. This novel approach holds the potential to enhance the precision and comprehension of predictions within dynamic social environments. The significance of this method is in the necessity to create more resilient instruments for decision-making in a setting marked by data ambiguity and variability [1]. The modeling of uncertainty in prediction has traditionally progressed from basic statistical methods to intricate strategies rooted in probability theory and sophisticated regression models. Nevertheless, with the increasing importance of the social innovation area, there has been a demand to integrate approaches that not only take into account numerical data but also manage qualitative and contextual uncertainty. Neutrosophic theory, a development of set theory that addresses indeterminacy and contradictions, presents a novel viewpoint that has the potential to transform our approach to these types of problems [2,3].

The issue addressed by this work is the absence of efficient techniques to depict uncertainty in predictive models inside the social innovation ecosystem. This study aims to achieve two objectives: The first objective is to investigate the application of neutrosophic theory in enhancing the predictive ability of regression models within social innovation contexts. An effective integration of uncertainty into predictive modeling is sought in this work through a series of empirical investigations and simulations [4]. Essentially, this study seeks to address a significant gap in the existing research by proposing a methodology that integrates neutrosophic theory with regression techniques to tackle the difficulties associated with prediction in a social innovation ecosystem. The findings hold the potential to not only enhance theoretical knowledge but also provide practical implementations that can enhance the precision of forecasts and the effectiveness of innovation initiatives in real-life situations.

2 Preliminaries
2.2 Social Innovation Ecosystem

The current panorama is characterized by a complex and multifaceted social innovation ecosystem, which represents the complicated interactions of actors, resources, and dynamics aiming to revolutionize society through innovative solutions. Fundamentally, this ecosystem encompasses not just social innovators and entrepreneurs, but also a diverse array of entities including governments, corporations, non-governmental organizations, and local communities. Each of these components offers a distinct and perhaps incompatible viewpoint on what qualifies as a suitable "solution" to current societal issues. Comprehending and examining these relationships is essential, not only to enhance the efficiency of social programs but also to cultivate a culture of cooperation and reciprocal knowledge acquisition that enables a more holistic approach to tackling global complexities [5].

From a historical standpoint, the notion of social innovation has witnessed substantial evolution. In its initial manifestations, social innovation primarily emphasized problem-solving through community initiatives and social movements aimed at effecting changes in public policies. However, as time has passed, the approach has become more varied, including more organized and cooperative techniques that engage several stakeholders. In large part, this transformation has been propelled by the increasing acknowledgment that intricate social issues cannot be resolved independently, but necessitate a comprehensive strategy that combines various knowledge, resources, and capacities [6].

The problem within this ecosystem is defined as the task of enhancing the efficacy of social innovation efforts. Notwithstanding the enthusiasm and increasing investment in this domain, notable obstacles remain, including insufficient collaboration among stakeholders, issues in quantifying the actual influence of innovations, and reluctance to embrace change in established societal frameworks. The primary inquiry is: How can participants in the social innovation ecosystem enhance their collaboration to surmount these obstacles and attain significant positive outcomes? To uncover avenues for increased integration and effectiveness, this question directs the investigation and evaluation of present practices and tactics [7].

To tackle intricate social problems and attain sustained social progress, the social innovation ecosystem is essential. Thus, current literature fails to provide a comprehensive perspective that incorporates all components of the ecosystem. This paper presents a hybrid approach that combines qualitative and quantitative approaches to comprehensively analyze the dynamics and perspectives of many participants in the ecosystem. Network analysis tools will graphically represent the connections and patterns of resource allocation among participants, therefore offering a more lucid perspective on how cooperation and coordination can be enhanced. The results may guide the development of policies and strategic allocation of resources. Furthermore, the report provides practical suggestions to enhance the execution of social innovation projects, assisting stakeholders in identifying areas that need development and adopting more efficient strategies. This methodology enhances the trajectory of social innovation and facilitates the resolution of worldwide social issues.

a. Machine learning

Machine learning (ML) [8] is the methodology of using mathematical formulas to develop models capable of learning from data in order to generate predictions or judgments, without the need for explicit programming of such tasks. Statistical interval prediction in machine learning is the method of forecasting a range of potential results for a given input, rather than a single estimate. Through the provision of intervals, these approaches not only provide forecasts but also offer valuable understanding of the dependability and unpredictability of the predictions, which is essential for making decisions in uncertain circumstances.

For a dataset consisting of independent variables $X = [x_1, x_2, \dots, x_n]$ and a dependent variable, the objective of regression analysis is to precisely represent the connection between X and Y. The mathematical expression for this relationship is as follows[9]:

$$y = f(X;\ \theta) + \epsilon \tag{1}$$

where:

y is the dependent variable or the objective to be predicted.

X represents the independent or explanatory variables that are used in prediction.

f is the regression function, which can vary depending on the type of regression model used (linear, polynomial, logistic, etc.)

θ are the parameters or coefficients of the model, adjusted during the training process to minimize a loss function, classically the mean square error (MSE) for regression [10].

ϵ is the error term or noise, which represents the deviation or error that cannot be entirely explained by the model.

The representation of predictions as prediction intervals in regression analysis offers a more comprehensive perspective on the uncertainty linked to the predictions. A prediction interval yields a range in which we anticipate the actual value of the dependent variable to lie with a certain likelihood, usually 95% or 99%. The utility of this approach lies in its consideration of data variability that may not be adequately reflected by prediction alone [11].

Computation of a prediction interval requires consideration of both the uncertainty in the estimate of the regression model and the intrinsic variability of the data. The interval is delineated around the projected value and typically exhibits symmetry, encompassing a specific extent both above and below the predicted value. This interval is established by considering the standard error of the forecast and the residual standard deviation, which indicates the spread of the residuals or errors of the model.

For example, in a logistic regression model, the prediction interval for a new observation is given by [1]:

$$\hat{y}_0 \pm t_{\propto /2,n-2} \cdot SE \tag{2}$$

Where

$\hat{y}_0$ is the predicted probability of the outcome of the new observation

$t_{\propto /2,n-2}$, is the critical value from the t-distribution for a given confidence level $\propto$ and $n-2$ degrees of freedom.

SE is the standard error of the prediction, which quantifies the uncertainty around the predicted value.

Using prediction intervals in regression analysis is beneficial because they offer a realistic spectrum of possible outcomes, which aids in the decision-making process. This recognizes that a single predicted value is not absolute but rather a likely scenario within a range of potential outcomes. This forecasting method effectively incorporates the inherent uncertainties associated with future predictions, providing a more accurate description of what to expect [12].

To further refine this model, neutrosophic statistics can be applied, which excel in managing ambiguity and indeterminacy in data. By converting the interval to a neutrosophic number, the traditional interval is enhanced to include a component of indeterminacy. This addition captures the uncertainty and imprecision that is typically present in real-world data, offering a more nuanced understanding of data variability. The neutrosophic treatment of the interval is as follows [1]:

$$\hat{y}_0 - t_{\propto/ 2,n-2} \cdot SE + (\hat{y}_0 + t_{\propto /2,n-2} \cdot SE)I \tag{3}$$

Here, $I_N$represents the indeterminacy factor associated with the prediction, where $I_N \in [I_l, I_u]$. This notation introduces the limits of indeterminacy. I_l(lower indeterminacy) and I_u (upper indeterminacy), which defines the range of possible deviations due to uncertain elements that affect the prediction [12,13].

3 Methods

Regression analysis of a data set and representation of uncertainty using neutrosophic numbers can be divided into four essential stages, as outlined in reference [1]:

1. Data Segmentation: The initial stage involves dividing the data into separate sets for training and testing. This separation is essential as it enables the verification of the model using never-before seen data, therefore guaranteeing that the performance of the model is not solely due to overfitting the training data. Data scientists commonly employ a 70-30 or 80-20 partition arrangement, whereby 70% or 80% of the data is allocated for training purposes and the remaining portion is reserved for testing.
2. Model Optimization: Subsequently, each model undergoes training using the training set. This entails fine-tuning the model parameters to more accurately correspond to the facts. The training procedure entails identifying model parameters that minimize a loss function, therefore effectively capturing the fundamental structural characteristics of the dataset.
3. Estimation of Prediction Ranges: After training the models, the subsequent task is to quantify the prediction intervals for new observations. This is the point at which neutrosophic numbers become relevant. Neutrosophic intervals differ from conventional sharp intervals by include measurements of truth, indeterminacy, and falsehood, therefore enabling a more sophisticated depiction of uncertainty in predictions. Depending on the features of the model and the nature of the data, each model may necessitate distinct approaches to compute these ranges.

4. Uncertainty Analysis: The final stage involves analyzing the uncertainty represented by neutrosophic values. This step examines how indeterminacy varies across models and what it implies for data complexity and variability. A higher indeterminacy level may indicate greater external constraints or inherent unpredictability within the dataset [15, 16].

In this work, we employ neutrosophic methods to combine interval predictions with other approaches as part of a fusion theory in regression analysis. Neutrosophic approaches offer significant benefits in effectively merging several predictive models by allowing the smooth incorporation of uncertainty, indeterminacy, and contradicting information that may arise from separate data sources or model outcomes. The suggested methodology improves the robustness and reliability of prediction models by providing a complete framework that takes into account several aspects of uncertainty. The neutrosophic mean, represented as X_n, is computed by taking into account the neutrosophic inclusion I_N that falls inside the interval. $[I_l, I_u]$.This mean comprises two primary components: X_l, representing the average of the lower section of the neutrosophic samples, and X_u, representing the average of the upper section. The corresponding definitions are [17, 18]:

$$X_l = \frac{\sum_{i=1}^{n_l} x_{il}1}{n_l} \quad (4)$$

$$X_u = \frac{\sum_{i=1}^{n_u} x_{iu}}{n_u} \quad (5)$$

where $n_l$and n_u represent the number of elements in the lower and upper parts of the neutrosophic samples, respectively. Therefore, the neutrosophic mean $X_n$is expressed as the sum of $X_l$and X_u, adjusted by the interval of indeterminacy I_n [19, 20]:

$$X_N = X_l + X_u I_N;\ I_N \in [I_l, I_u] \quad (6)$$

$I_{l,}$=0, andI_u

$$I_u = (X_u - X_l)/X_u \quad (7)$$

Hence, the neutrosophic mean can be regarded as a versatile depiction that encompasses both specified values (the minimum and maximum limits) and a component of uncertainty, represented by the indeterminacy interval I_N [21]

4 Results

The dataset used contains 60 records and includes the following variables:

- Investment (thousands of $): The financial resources allocated to the project.
- Number of Employees: The staff working on the project.
- Duration (months): The length of the project.
- Direct Beneficiaries: The number of people directly impacted by the project.
- Community Involvement (%): The level of community involvement as a percentage.
- Social Impact (0-100): The target variable, represents the social impact based on various criteria such as quality of life improvements, access to services, and poverty reduction.

The social impact was predicted using many regression models based on the characteristics indicated above. Statistical prediction intervals were computed for each model to measure the level of uncertainty in the forecasts. In addition, neutrosophic forms were developed to depict the uncertainty or indeterminacy linked to each model, shown as a spectrum of potential results of the indeterminate parameter.

The following table provides a summary of the prediction intervals and their accompanying neutrosophic forms for the various regression models used on the dataset.

Table 1: Neutrosophic Forms and Prediction Intervals for Regression Models

Model	Prediction Interval [Lower Bound, Upper Bound]	Neutrosophic Form
Linear Regression	[77.54, 95.23]	77.54+95.23I;$I \in [0,0.186]$
Random Forest	[80.12, 97.65]	80.12+97.65I;$I \in [0,0.180]$
Support Vector Machine	[75.34, 94.12]	75.34+94.12I;$I \in [0,0.200]$
Mean	[77.67,95.67]	77.67+95.67I;$I \in [0,0.188]$

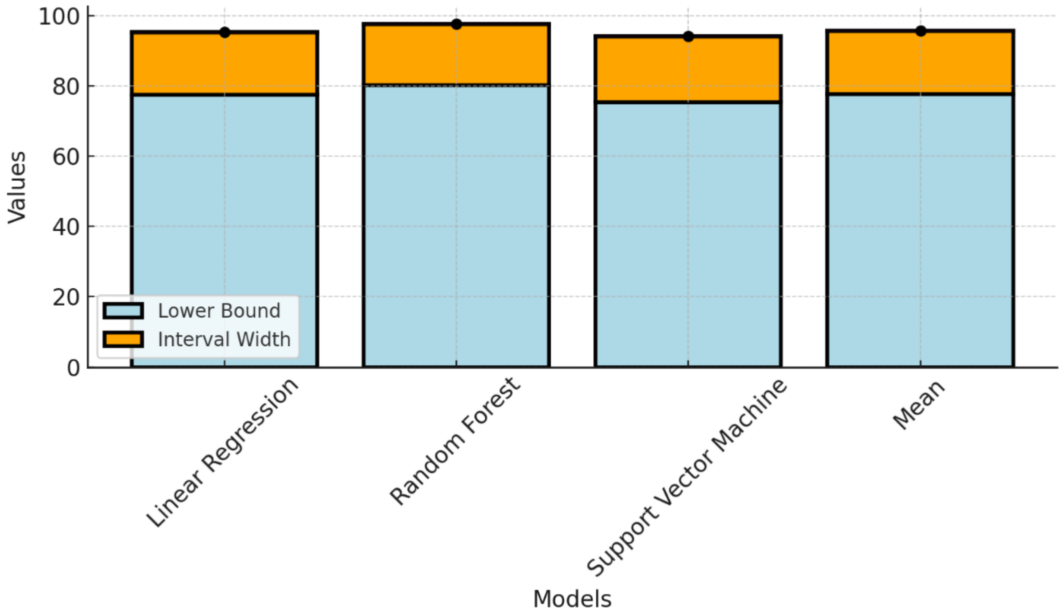

Figure 1. Model Prediction Intervals Comparison

The table presents prediction intervals along with their neutrosophic forms, illustrating the varying levels of indeterminacy for each model. Linear Regression, Random Forest, Support Vector Machine, and the Mean-based approach were evaluated, with Random Forest demonstrating the smallest indeterminacy. The neutrosophic forms provide a deeper understanding of the uncertainty inherent in each model's predictions, which is crucial for decision-making processes that rely on accurate social impact assessments.

Conclusion

This paper explores the application of neutrosophic theory to enhance the representation of uncertainty in regression models used for predicting social effect. The application of neutrosophic numbers enabled us to address the intrinsic uncertainties and variations in the social innovation ecosystem, therefore providing a more advanced and flexible understanding of prediction intervals. The results suggest that incorporating neutrosophic components into traditional regression models, such as Linear Regression, Random Forest, and Support Vector Machines, can enhance their capacity to effectively capture prediction uncertainty. In this specific configuration, Random Forest exhibited the least amount of uncertainty among the analyzed models, suggesting a more reliable prediction performance. The synthesis of prediction intervals and neutrosophic forms offers a comprehensive viewpoint on possible outcomes, thereby enabling better-informed decision-making in social innovation initiatives.

Moreover, this work offers prospects for further exploration of the potential of neutrosophic theory in novel machine learning models and broader applications. Future study should give priority to improving the methodology in order to more precisely quantify the degree of uncertainty in various data-driven scenario, such as economic forecasting or evaluation of public policy. Moreover, the integration of neutrosophic theory with real-time data analysis and dynamic modeling has the capacity to enhance the adaptability of prediction models in rapidly changing social contexts. Implementing this approach not only improves the accuracy of forecasts but also enables the creation of more resilient techniques for addressing complex social issues.

References

[1] Vázquez, M. L., & Torres, L. C. (2024). Integrating Neutrosophic Numbers in Regression Analysis for Enhancing Predictive Modelling through Uncertainty Representation. Plithogenic Logic and Computation, 1, 120-126.

[2] El-Douh, A. (2023) "A Neutrosophic Multi-Criteria Model for Evaluating Sustainable Soil Enhancement Methods and their Cost Implications in Construction", Sustainable Machine Intelligence Journal, 5, pp. (1):1–11.

[3] Vázquez, M. L., Franco, P. E. D. P., & Palacio, A. J. P. (2022). Neutrosophic dematel in the analysis of the causal factors of youth violence. International Journal of Neutrosophic Science, 18(3), 199-99

[4] Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning.

Journal of Applied Science and Technology Trends, 1(2), 140-147.

[5] Audretsch, D. B., Eichler, G. M., & Schwarz, E. J. (2022). Emerging needs of social innovators and social innovation ecosystems. International Entrepreneurship and Management Journal, 1-38.

[6] Phillips, W., Lee, H., Ghobadian, A., O'regan, N., & James, P. (2015). Social innovation and social entrepreneurship: A systematic review. Group & Organization Management, 40(3), 428-461.

[7] Chueri, L., Vasconcelos, A., & dos Santos, R. P. (2019, November). An observational study on the challenges faced by actors in a social innovation ecosystem. In Proceedings of the 11th International Conference on Management of Digital EcoSystems (pp. 219-223).

[8] Domingos, P. (2015). The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books.

[9] Westfall, P. H., & Arias, A. L. (2020). Understanding regression analysis: a conditional distribution approach. Chapman and Hall/CRC.

[10] Robeson, S. M., & Willmott, C. J. (2023). Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. PloS one, 18(2), e0279774.

[11] Delgado Estrada, S. M., Rocio Ruiz Molina, K. D., Ponce Orella-na, F. E., & Chabusa Vargas, J. L. (2024). Neutrosophic Statistics for Enhanced Time Series Analysis of Unemployment Trends in Ecuador. Neutrosophic Sets and Systems, 67(1), 14.

[12] Olive, D. J. (2007). Prediction intervals for regression models. Computational statistics & data analysis, 51(6), 3115-3122.

[13] Aslam, M., & Alamri, F. S. (2023). Algorithm for generating neutrosophic data using accept-reject method. Journal of Big Data, 10(1), 175.

[14] Smarandache, F., & Aslam, M. (Eds.). (2023). Cognitive intelligence with neutrosophic statistics in bioinformatics. Elsevier.

[15] Smarandache, F. (2023). Introduction and advances to neutrosophic probability and statistics and plithogenic probability and statistics and their applications in bioinformatics and other fields (review chapter). In Cognitive Intelligence with Neutrosophic Statistics in Bioinformatics (pp. 1-23). Academic Press.

[16] Alanazi, B. and Alrashdi, I. (2023) "Anomaly Detection in Smart Agriculture Systems on Network Edge Using Deep Learning Technique", Sustainable Machine Intelligence Journal, 3, pp. (4):1–31.

[17] Alomair, A. M., & Shahzad, U. (2023). Neutrosophic mean estimation of sensitive and non-sensitive variables with robust hartley–ross-type estimators. Axioms, 12(6), 578.

[18] Kandemir, H. Ş., Aral, N. D., Karakaş, M., & Et, M. (2024). Neutrosophic Statistical Analysis of Temperatures of Cities in the Southeastern Anatolia Region of Turkey. Neutrosophic Systems with Applications, 14, 50-59.

[19] Delcea, C., Domenteanu, A., Ioanăș, C., Vargas, V. M., & Ciucu-Durnoi, A. N. (2023). Quantifying Neutrosophic Research: A Bibliometric Study. Axioms, 12(12), 1083.

[20] Haq, R. S. U., Saeed, M., Mateen, N., Siddiqui, F., & Ahmed, S. (2023). An interval-valued neutrosophic based MAIRCA method for sustainable material selection. Engineering Applications of Artificial Intelligence, 123, 106177.

[21] Razak, S. A., Rodzi, Z. M., Ahmad, N., & Ahmad, G. (2024). Exploring The Boundaries of Uncertainty: Interval Valued Pythagorean Neutrosophic Set and Their Properties. Malaysian Journal of Fundamental and Applied Sciences, 20(4), 813-824.

Received: June 27, 2024. Accepted: August 19, 2024