International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Sample Data Prediction using Regression Model on ANN

Sijin P[1], Vidya A[2], Hanumantharayappa T A[3]
[1,2,3]*Vivekananda Institute of Technology*

Abstract— **It is essential for a classification algorithm to predict the appropriate place for the user data in the spatial data space. It is possible by focusing on the contextual attributes and measuring their impact on the target variable to identify the nature of data flow in the dataset. The proposed neural network classification model predicts the spatial location of the dependent target attributes by using contextual variables and the hidden relationship among attributes. The model uses the characteristics of central tendency of data in order to identify the relationship among data and to know the direction and flow of data. The learning system targets the instance prediction of dependency of attributes with respect to contextual variables. The experimental analysis on synthetic and real datasets shows promising results with the proposed model.**

Keywords**—Classification, neurons, regression, attribute, hidden layer, active learning**

I. Introduction

The data analysis in a dataset or a normalized data frame deals with causes of relationships. The causes help to target relationships among variables. The relationship among variables can be used in comparison, representation, plotting the data direction, identify the flow of data, and the overall exploratory data analysis of user needs. Most of the time these approaches are used as the prerequisite of data mining applications. In multivariate data analysis such as Principle Component Analysis (PCA), regression analysis, and path analysis, more than two dependent attributes are considered to identify the attribute relationship and its influence on data flow.

Prediction is an important aspect of any classification model. The more accuracy the models can offer, the more accuracy will be gotten in data prediction. The proposed work predicts the dependent variables from multiple attributes using a regression pipeline using slope and y-intercept. User can adopt a built-in model or can design their own in order to implement and deploy the design.

II. Literature Review

The input layer of the Artificial Neural Network (ANN) reads the input and passes it to the nodes of the hidden layer. The connection weights determine the weight of each Dence of the linking network.

The computing nodes perform some computation based on input information and if the weighted some of the input information on the node is above some threshold, it immediately fires the values to the next level. Most of the time, activation functions are used to transform the data into a specified format. Biasing agents are used to fine-tune the data. When the logical sequencing of the data is over, it provides the required results.

The works in [1], [2], [3] proposed an Active Learning (AL) framework for regression, model change maximization etc.,. The AL system uses the unlabeled data instances that result in the maximum change of the current model once labeled. The concept of random neural networks with state-dependent firing neurons [4] shows the importance of identifying the data flow change in the environment. The works in [5] proposed the indoor water supply pipeline model experimental results which can be used to locate the burst point in urban water supply systems. It also deals with the topological structure and learning parameters that are used for optimization.

Pattern aided regression (PXR) models proposed in [6] provide an accurate and interpretable prediction models. Which reflects the logical and behavioral characterizations of distinct predictor-response relationships. The works of [7] present a new learning formulation for multiple model estimation. Here training data samples are produced by several individual built-in models. Regression models predict continuous variables like price. This model takes input x and produces a model in the form of a fitted line of the form y=f(x). Here x is the independent variable that may have one or more attributes and y is the dependent variable.

According to the works in [8], [9] classification is a supervised learning method in which input attributes of the classification algorithms are called independent variables. The target attribute is called a dependent variable. The relationship between the dependent and independent variable is represented in the form of a structure called classification model. In Machine Learning (ML) and generative AI applications also the data relationship, data direction, and data flow can make a considerable impact on information processing and computing [10], [11].

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

III. METHODOLOGY

The proposed approach explains the variability in a dependent variable and uses that to predict the future values. It is widely used in Market analysis, real estate, sales, advertising, health care, and education, for determining causal effects, such as how the independent attributes influence the forecasting of dependent attributes.

TABLE 1:
Sale Data table

Sl. No	Weeks	Sales in Lakhs
1	1	1.2
2	2	1.8
3	3	2.6
4	4	3.2
5	5	3.8

The TABLE 1 gives the five weeks' sales of a textile organization. The columns of the table maintain a relationship of data points but it is vague. The proposed model fit a line between the data points in the table. Each row of the table corresponds to a data point with a tuple representation called (x, y). Each line is fitted by the equation $y=a_0+a_1x+error$, where a_0 is the y-intercept and a_1 is the slope. This representation does not remove the vagueness in relationship, but aligns the data point about to a line and gives the trends in data in various modes such as positive or negative, monotone or anti-monotone etc. The squared distance of each data point to the line of fit can be found out and added up to get the total error in data.

The user can bring the data points to the line of fit, can call the regression line by generalizing the variable values, but it may penalize with the accuracy of the model. The computation and data analysis work is also an overhead. Better option is to use a fully connected feed forward ANN network as shown in FIGURE1.

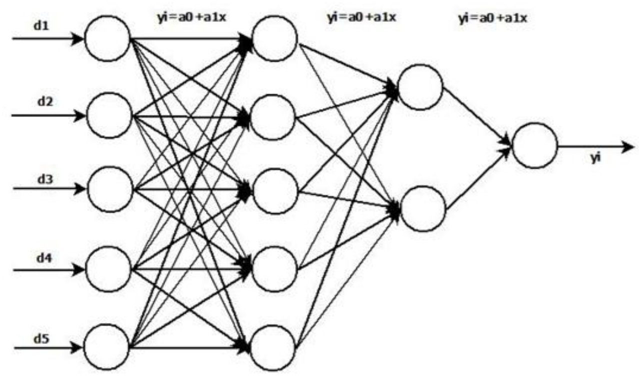

FIGURE 1: Forward neural network

This paper aims to find out the data relation so advanced ANN design is not practiced here. Initially five nodes can be set to receive the inputs and the weighted input can be given to the hidden layers. The hidden layer is dedicated to working on the fed data with a multi-column regression algorithm to run on, set on the biasing agent obtained from the domain knowledge. Any linear activation function can be used to hold the data in a calculatable range during computation. The hidden layer can pass the generated new y values to the output layer. More layers can be included and the dense of the network can vary accordingly to the output range if needed. The user can predict the data by using the many levels of forecaster evidence available in the neural network.

A. Data analytics used by the neural network model

The hidden layer algorithm's computation table gives for the data of TABLE 1; the mean of dependent variable is 2.52 and the mean of independent variable obtained is 3. The mean of the squared independent variable is 11. The mean of the product of dependent and independent variables obtained is 8.88. The y-intercept a_0 is 0.54 and the slope of the line of fit obtained is 0.66. The model gives a positive data flow, the attributes are influencing the data frame, and it is slightly monotone.

While working on the Boston Housing Dataset the results obtained are like this. The mean of dependent variable is 6.8788 and the mean of independent variable obtained is 5.4954.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

The mean of the squared independent variable is 32.119. The mean of the product of dependent and independent variables obtained is 35.54. The y-intercept a_0 is 13.54 and the slope of the line of fit obtained is -1.19. The model gives a positive data flow; the attributes are influencing the data frame, and it is slightly monotone.

IV. Performance Analysis

During experimental analysis, one synthetic and one real data set are used. The results are promising. The attributes are related to the dataset and are more visible with the proposed model. It shows very accurate results for predictor-based forecasting.

The Python data frame of Boston Housing dataset for a head value of five is given here (FIGURE 2).

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2

FIGURE 2: Portion of Boston Housing dataset

The results for Boston Housing Dataset for the DIS value of 5 predict the INDUS value 7.4739 a very promising value. It is given in FIGURE 3. While comparing FIGURE2 and FIGURE3 it is clear that the DIS value finds a suitable position for the dependent variable INDUS.

The result is more promising in FIGURE 4 in which the dependent variable TAX is predicted with PTRATIO. For the PTRATIO value of 16, the predicted TAX value obtained is 277.97 and placed in a further better location.

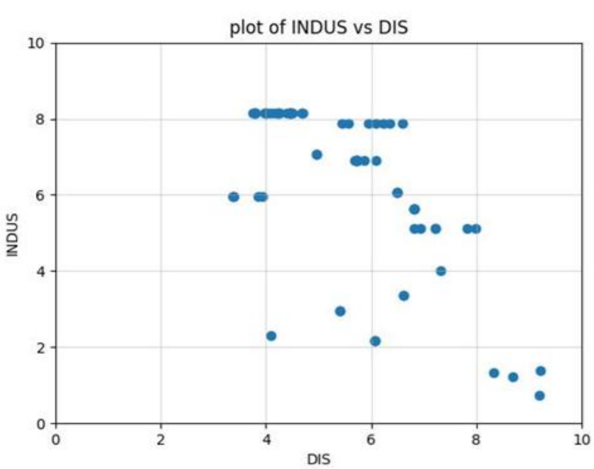

FIGURE 3: INDUS data placement

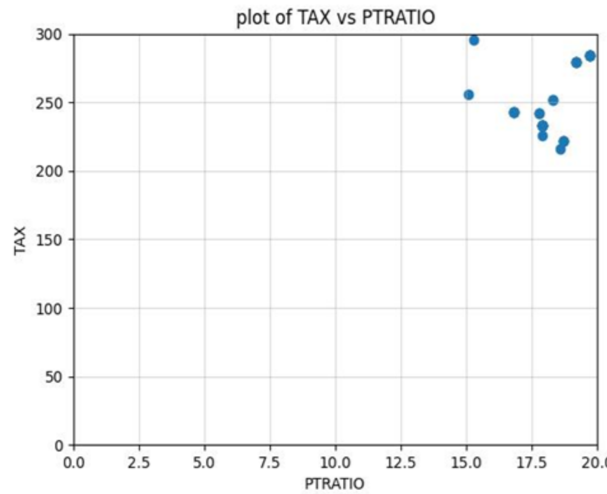

FIGURE 4: Tax data placement

V. Conclusion

Every dataset has data and all this data has some relation. Some are visible and clear but some are vague and not visible. The interrelationship of attributes in an attribute-based dataset can show some trends and patterns of data. The proposed neural network model defines a learning system which can predict the independent variables by using the data dependencies available in the dataset. This model uses the regression analysis of the data to bring the data points to the region of interest determined by the trend of the data. It is represented by the line of fit. The experimental results with both synthetic and real datasets show better prediction ways.

Acknowledgement

The work is acknowledged to Janatha Education Society (JES), Bangalore, Vivekananda Institute of Technology, Bangalore. Visvesvaraya Technological University (VTU), Bangalore University, and IEEE, Elsevier, Springer.

REFERENCES

[1] Wenbin Cai; Muhan Zhang; Ya Zhang, "Batch Mode Active Learning for Regression with Expected Model Change", IEEE transactions on neural networks and learning systems, vol. 28, no. 7, 2016.

[2] Stuart Russel, Peter Norvig, "Artificial Intelligence, A Modern Approach", Prentice-Hall, vol. 25, no. 27, 1995.

[3] Zidong Wang; D.W.C. Ho; Xiaohui Liu, "State Estimation for Delayed Neural Networks", IEEE Transactions on Neural Networks, vol. 16, no. 1, 2005.

[4] Sungho Jo; J. Yin; Zhi-Hong Mao, "Random neural networks with state-dependent firing neurons", IEEE Transactions on Neural Networks, vol. 16, no. 4, 2005.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

[5] Donghai Zhu; Tuqiao Zhang; Genhai Mao, “Back-propagation artificial neural networks for water supply pipeline model”, Tsinghua Science and Technology, vol. 7, no. 5, 2002.

[6] Guozhu Dong; Vahid Taslimitehrani, “Pattern-Aided Regression Modeling and Prediction Model Analysis”, IEEE Transactions on Neural Networks, vol. 16, no. 4, 2005.

[7] V.Cherkassky; Yunqian Ma, “Multiple model regression estimation”, IEEE Transactions on Neural Networks, vol. 16, no. 4, 2005.

[8] Andriy Burkov, “The Hundred-page Machine Learning Book”, IEEE Transactions on Neural Networks, vol. 1, 2019.

[9] S Sridhar, M Vijayalakshmi, “Machine Learning”, Oxford Press, vol. 1, 2021.

[10] Jamal, Suhaima, **“**Applications of Predictive and Generative AI Algorithms: Regression Modelling, Customized Large.

[11] Ding, Sherry and Raman, Veda, Harness the Power of Generative AI in Healthcare with Amazon AI/ML Services, IEEE, 2024.