International Journal of Business Analytics
Volume 9 • Issue 3

Futuristic Prediction of Missing Value Imputation Methods Using Extended ANN

Ashok Kumar Tripathi, Jaypee University of Information Technology, India
Hemraj Saini, Jaypee University of Information and Technology, India*

https://orcid.org/0000-0003-2957-1491
Geetanjali Rathee, Netaji Subhas University of Technology, India

ABSTRACT

Missing data is a universal complexity for most of the research fields, which introduces uncertainty into data analysis. This can take place due to many mishandling, inability to collect an observation, measurement errors, aberrant value deleted, or merely being short of study. The nourishment area is not an exemption to the difficulty of missing data. Most frequently, this difficulty is determined by manipulative means or medians from the existing datasets which need improvements. The paper proposes hybrid schemes of MICE and ANN known as extended ANN to search and analyze the missing values and perform imputations in the given dataset. The proposed mechanism is efficiently able to analyze the blank entries and fill them with proper examination of their neighboring records in order to improve the accuracy of the dataset. In order to validate the proposed scheme, the extended ANN is further compared against various recent algorithms or mechanisms to analyze the efficiency as well as the accuracy of the results.

KEYWORDS

Extended KNN, Food Analysis, Food Consumption, K-Nearest Neighbor, MICE, Missing Data, Missing-Data Imputation, Nutrient Values

1. INTRODUCTION

Food intake is a periodic behaviour. It is induced at numerous moments of the day via way of means of some of converging elements like time of day, want state, sensory stimulation, social context, etc. Promoting healthful diets and life to lessen the worldwide burden of non communicable illnesses calls for a multi sect oral technique related to the numerous applicable sectors in societies. The agriculture and meals region figures prominently on this business enterprise and should take delivery of due significance in any attention of the promoting of healthful diets for people and populace groups. Food techniques should now no longer simply be directed at making sure meals protection for all, however should additionally obtain the intake of good enough portions of secure and properly high-satisfactory ingredients that collectively make up a healthful diet. Any advice to that impact may have implications for all additives within side the meals chain. It is therefore, beneficial at this juncture to

DOI: 10.4018/IJBAN.292055 *Corresponding Author

look at traits in intake styles international and planned at the capacity of the meals and agriculture region to fulfill the needs and challenges.

Economic improvement is typically followed through upgrades in a country's meals deliver and the sluggish removal of nutritional deficiencies, as a result enhancing the general dietary popularity of the country's population. Furthermore, it additionally brings approximately qualitative adjustments within side the production, processing, distribution and advertising of meals. Increasing urbanization can even have effects for the nutritional styles and life of individuals, now no longer all of which can be positive. Changes in diets, styles of labor and leisure, frequently known as the "nutrients transition" this is already contributing to the causal elements underlying non communicable illnesses even withinside the poorest countries. Moreover, the tempo of those adjustments appears to be accelerating, mainly within side the low-earnings and middle-earnings countries.

The FCD contained in presently to be had FCDBs in Europe and global are a variety of quality, which displays the diverse approaches wherein they're obtained and the origin they arrive from. In order to pick out the records there are codes and orientations of the records sorts and origin utilized by many countries (Bogoviz, 2018; Gurinović et al., 2016). The origins of records so as of preference are:

Original values analytical - records in use from posted literature or unpublished research laboratory reports (that have passed through the proper first-rate checks), even if or not organized openly used for the motive of accumulated the database.

Estimated values – Estimation resulting as of logical values acquired for the same foodstuff or for any other shape of the equal foodstuff.

Calculated values – statistics resulting from recipes, supposed as of the nutrient stuffing of the elements and accurate for grounding.

The actual-world data frequently have a group of the missing values. The motive of missing values may be information dishonesty or breakdown to document facts. The dealing with the missing data may be very critical throughout the pre-processing of the data set as a lot of gadgets gaining knowledge of algorithms do now no longer assist missing values. To dealing with missing values in the dataset there are certain method to including, Rows Deleting by missing values and Impute the missing values of incessant changeable, Missing values impute for definite variable, Imputation Other techniques, by means of Algorithms support that missing values, forecast of the missing values, and using Imputation Deep Learning Library — Data wig. It defined up the food consumption dataset for the evaluation of the missing value data imputation.

1.1. Paper Contribution

The missing value in any database may leads to certain inaccurate and false outputs while analysing the records. In case where values are zero or blank at some entries corresponding to a particular attribute, then it may be very difficult to analyse the actual output of the result. In addition, the overall accuracy of examining or analysing the data may get affected. Therefore, it is needed to further focus on some missing value imputation schemes/techniques to fill the blank entries either by recognizing the neighbouring columns or calculate the average of particular column. In this paper we address the hassle of missing information in food consumption databases. Due to missing records in food intake make the records incomplete which end up in have its restrained utilization due to the fact any nutritional evaluation may be achieved most effective on an entire dataset. Most often, this hassle is determined via way of means of manipulative means/medians from the prevailing dataset of the equal data base or borrowing information from different missing dataset from food consumption databases. These answers introduce great error. We attention on missing data imputation strategies primarily based totally on techniques for substituting lacking values with statistical prediction. Imputation of Missing Value is the usage of Recurrent Neural Network (RNN), Iterative KNN imputation method, K-Nearest Neighbours (KNN), and Artificial Neural networks (ANN) and in comparison, them with

generally used approaches - fill-in with the mean and median. The statistics used is from countrywide FCDBs accrued via way of means of Open MV .internet Datasets. The relative intake of sure meals gadgets in European and Scandinavian international locations are depicted via way of means of the dataset. The outcomes display that the state-of the-art techniques for imputation provide way higher outcomes than the conventional approaches.

2. RELATED WORK

This section elaborates the number of techniques that can be used while imputing the missing values in a database at a particular blank entry. This section discussed the number of schemes or techniques proposed by various scientists in order to fill the blank entries of a particular column in a given database.

This segment affords all of the associated paintings had to apprehend the techniques utilized in our research. We introduce the conventional strategies for borrowing facts used for imputation of missing FCD within side the context of FCDBs, and superior contemporary technology for missing value imputation (Fao, 2012; Vaughn et al., 2019).

Missing information are unnoticed values in a dataset. This information may be of various kinds and may be lacking for distinct reasons. Knowing the character of the missing values can assist become aware of the maximum suitable approach for coping with missing information. There are three forms of missing data that could arise in distinct datasets, consistent with the mechanism of missing ness, which describes the connection among the possibility of a value being missing and the alternative variables within side the dataset.

- Missing at random (MAR) - The first mechanism of missing ness is MAR, this means that the opportunity so as to a value is missing relies upon most effective on located values and now no longer on unobserved values. The missing records are only a random subset of the records. A easy instance of MAR in FCDBs is that if there has been an evaluation performed for sure vitamins in sure foods and this data is entered, however within side the equal FCDBs there's missing records approximately different vitamins for the equal foods. In this example the analyzed foods are located covariate.
- Missing completely at random (MCAR) - The 2D mechanism of lackingness is a unique case of MAR called MCAR. The opportunity of lackingness isn't always depending on any located or unobserved values. One instance of MCAR is probably if after the compositional evaluation there has been a pc malfunction even as coming into the records that arbitrarily deleted a number of the records values
- Missing not at random (MNAR)- The 3rd mechanism of missingness is called MNAR, and the facts is neither MAR nor MCAR. This mechanism of missingness happens whilst the situations of MAR are violated in order that the possibility of missingness relies upon the unobserved factors. One example of MNAR would possibly be - if a FCDB is lacking facts for ingredients that don't develop with inside the vicinity of the use of a wherein the FCDB originates from, due to the weather of that vicinity. In this case, the missingness is depending on the unobserved response - climate situations, i.e. weather.

2.1. Conventional Techniques for Borrowing Food Consumption Databases

There are a few techniques for managing missing data primarily based totally on statistical prediction. In this section, we explain 4 techniques for missing data imputation primarily based totally on statistical prediction decided on to be able to cowl strategies extensively carried out within side the place of missing value imputation.

2.1.1. K-nearest-neighbor (KNN)

K-nearest-neighbor (KNN) classifications is one of the maximum essential and easy categorization techniques and have to be single of the 1st pick for a categorization examine whilst there's very small previous know-how approximately the sharing of the data. K-nearest-neighbor classification changed into advanced as of the want to carry out discriminate evaluation whilst dependable parametric estimates of opportunity thickness are not recognized or hard to decide. In an un-published Air Force School of Aviation Medicine US file in the 1951, Fix and Hodges delivered a nonparametric approach for sample categorization that has for the reason that come to be regarded the k-nearest neighbor regulation (Silverman & Jones, 1989). Later in 1967, a number of the formal residences of the k-nearest-neighbor rule had be work out; for example it's changed into proven that for k=1 and n®¥ the k-nearest-neighbor categorization fault is surrounded above via way of means of two times the Bayes fault rate (Cover & Hart, 1967). just the once such proper residences of k-nearest-neighbor type had been recognized, a protracted line of research ensued which include new negative response techniques (Carleial & Hellman, 1975), refinements by means of admire to Bayes fault rate (Fukunaga & Hostetler, 1975), distance weighted techniques (Dudani, 1976; Pan et al., 2017), tender computing (Bermejo & Cabestany, 2000) strategies and fuzzy techniques (Jóźwik, 1983; Keller et al., 1985).

2.1.2. Recurrent Neural Network (RNN)

RNN is a simplification of feed ahead neural network this is the internal memory. RNN is recurring in natural world because it plays the identical characteristic for each enter of data on the equal time because the outcomes of the contemporary enter relies upon at the beyond single calculation. After generating the production, it's far copied and sends off again into the recurring society. For creating a decision, it considers the contemporary enter, the outcome that it has discovered as of the preceding input. Different feed ahead neural networks, RNN can be using their internal memory to procedure series of inputs. These create them significant to responsibilities inclusive of unregimented, connected script reputation or verbal communication reputation. In different neural networks, all of the inputs are unbiased of every different. But in RNN, each and every one of the inputs are associated with every different.

2.1.3. Iterative KNN Imputation Method

Missing values are an inescapable trouble in some of actual international programs and a way to impute those missing values has turn out to be a hard difficulty in built-up production. Even aleven though there are a few famous imputation techniques projected, those techniques carry out poorly within side the assessment of missing values within side the trash pickup logistics management system (TPLMS). The trouble of missing values within side the TPLMS is important and might bring about illogical decision making. Within the paper we describe an iterative KNN imputation approach which is connections with weighted k-nearest neighbor imputation and the gray relational analysis (GRA). That approach is example of primarily depend on totally imputation approach that takes advantage of association of attributes with the aid of using the use of a gray relational grade in place of Euclidean detachment or different similarity measures to look k-nearest neighbour instances. The manageable values for the missing values are expected from those nearest neighbour instances iteratively. In adding, the iterative imputation permits everyone to be had values which includes the characteristic values within side the times with the missing values and imputed values from preceding new release to be applied for estimation of the missing values. Specially, the imputation approach is able to fill in all of the missing values by means of dependable records regardless of the lacking price of the TPLMS dataset.

2.2. Merits and Demerits of Existing Schemes/Techniques

ANN, KNN, RNN and iterative KNN are some standard ways to impute the values in blank entries of a database. Various KNN, RNN based schemes have been provided by the scientists in existing work on various types of datasets. However, very few of them have focused on food consumption database (Bhalaik et al., 2020; Fan & Sharma, 2021; Rathee, Ahmad, Kurugollu et al, 2020; Rathee et al., 2021; Rathee, Garg, Kaddoum et al, 2020; Rathee, Sandhu, Saini et al, 2020; Ren et al., 2021; Sharma et al., 2021; Sun et al., 2021; Xu et al., 2021). The existing KNN, RNN, ANN techniques are efficient of identifying the homogenous and structured way of database upon analysing and filling with other values in blank entries. The existing mechanism used the clustering, grouping and certain set of specified rules, however, the time to analyse and way to fill the blank entries may further be enhanced by merging the various schemes. In this paper, in order to improve the accuracy and efficiency of analysing, and filling the missing values, ANN is merged with MICE technique. The proposed mechanism is able to efficiently analyse the dissimilar datasets in a database and further provide the accurate way to impute the correct value in the blank entries.

In this paper, we test our proposed technique on numerous TPLMS datasets at dissimilar missing rates in compare with a few current imputation techniques. The investigational outcomes recommend that the planned technique receives a higher overall presentation than different techniques in phrases of imputation correctness and convergence speed.

3. PROPOSED PHENOMENON

This segment begins off evolved with a proof of the facts utilized in our research, how it's miles obtained, how it's miles established and the layout wherein it's miles used after which follows an outline of the techniques. This segment deliberates one-of-a-kind groups of missing data imputation techniques: conventional techniques and techniques primarily based totally on statistical prediction (modern cutting-edge procedures for imputing missing values) consisting of K-Nearest Neighbors (KNN) (Altman, 1992). KNN is a set of rules that suits a factor with its closest k neighbors in a multi-dimensional space. The concept at the back of the KNN imputation technique is that k neighbors are selected primarily based totally on a ways measure and their common is used as an imputation estimation, i.e. the missing value may be approximated with the aid of using the values of the factors that it's miles closest to. This imputation technique may be used for nonstop, discrete, ordinal and express facts, which makes it especially beneficial for handling all varieties of missing data. The KNN technique can utilized in variety of schemes like food, health, coverage and so forth for facts looking and analysis.

3.1. Data Searching and Analysis

When attempting to find statistics on a specific food and searching through numerous databases and the clinical literature, a few matters to remember are the meals names and descriptions, data excellence, and data variability. Food's names and their descriptors are extremely vital to ensure that the related data corresponds to the food for which facts is required. Data base compilers attempt to create a few consistencies for the meals names and descriptors to assist customers discover the ingredients they need. Though, in doing this, supposition can be made while meals names are badly or unclearly defined within side the authentic sources. Table 1 depicts the dataset of countrywide FCDBs ingredients to decide the supply of potassium.

The analysis of the mentioned dataset to determine the availability of potassium in foods can be determined through various machine learning algorithms. In this paper, for better understanding and imputation of missing values, a hybrid approach is used that is a combination of two different algorithms such as MICE and ANN which further represented as Extended-ANN for the analysis of missing value and imputes the missing data.

Table 1. Dataset with values for Potassium in several foods from several national FCDBs

Country	Germany	Italy	....	Norway	Finland	Spain	Ireland
Real coffee	90	82		92	98	70	30
Instant coffee	49	10		17	12	40	52
Tea	88	60		83	84	40	99
Sweetener	19	2		13	20	62	11
Biscuits	57	55		62	64	43	80
Powder soup	51	41		51	27	2	75
Tin soup	19	3		4	10	14	18
Potatoes	21	2		17	8	23	2
Frozen fish	27	4		30	18	7	5
Frozen veggies	21	2		15	12	59	3
Apples	81	67		61	50	77	57
Oranges	75	71		72	57	30	52
Tinned fruit	44	9		34	22	38	46
Jam	71	46		51	37	86	89
Garlic	22	80		11	15	44	5
Butter	91	66		63	96	51	97
Margarine	85	24		94	94	91	25
Olive oil	74	94		28	17	16	31
Yoghurt	30	5		2	64	13	3
Crisp bread	26	18		62			9

3.2. Multivariate Imputation by Chained Equations (MICE)

Missing data are an ordinary difficulty in research. However, multivariate imputation by chained equations (MICE), from time to time called "fully conditional specification" or "sequential regression multiple imputation" Has come out within side the arithmetical article as single righteous approach of address the missing data. Generating a couple of imputations, in place of single imputations, debts for the statistical indecision within side the imputations. The chained equations technique is very flaxy and may manage variables of various kinds in addition to difficulties along with bounds or survey bypass outlines. though, regardless of those benefits, a lot of psychiatric researchers have now no longer but discovered approximately this technique; there are a few sensible assets to be had to help in its implementation and till currently software program boundaries inhibited well known researchers and practitioners from the usage of the MICE process.

There are many dissimilar methods to deal with the missing records and the primary query in researchers may ask is "why use multiple imputations?" In sure state of affairs whole case evaluation can be an appropriate technique to addressing missing records (Graham, 2009). In practice, those instances hardly ever occur, at the same time as whole case evaluation can be smooth to put into effect it is based upon more potent missing records assumptions than multiple imputations and it could bring about biased estimates and a discount in power. Single imputation events, which include mean imputation, are an development however do now no longer account for the uncertainty within side the imputations; as soon as the imputation is completed, analyses continue as though the imputed values had been the known, actual values instead of imputed. This will cause overly specific consequences

and the capacity for wrong conclusions. Maximum probability techniques are every now and then a feasible technique for coping with missing data (Graham, 2009); however, those techniques are often to be had most effective for convinced styles of models, which include longitudinal or structural equation models, and may typically be run most effective the use of unique software program which include Amos (Arbuckle, 2011) and Lisrel (Jöreskog & Sörbom, 1996).

Algorithm-I:

Input: **Number of rows containing missing values in a dataset.**
Output: **Whether the missing values are imputed using MICE algorithm**
MICE = []
data2 = data
diff_mat = np.subtract(data2,data1);
mean2 = diff_mat.mean()
mean1 = 1000
while (mean2<=mean1):
mean1 = mean2
data2 = data1;
for i in test_missing:
y_imp = process_fn(data1,i)
data1[i[0],i[1]] = y_imp
diff_mat = np.subtract(data2,data1)
mean2 = diff_mat.mean()
for i in test_missing:
mice.append(data2[i[0]][i[1]])
Func process()
def process_fn(data, i):
x_train = np.delete(data, (i[0]), axis=0)
x_train = np.delete(x_train, (i[1]), axis=1)
x_test = data[:,i[1]]
x_test = np.delete(x_test,i[0])
x_missing = data[i[0],:]
x_missing = (np.delete(x_missing,i[1]))
x_missing = x_missing.reshape((1,19))
y_imp = model_fn(x_train,x_train,x_missing)
return y_imp

Algorithm 1 presents the method to impute the missing values in the given data set using MICE algorithm. As depicted in Algorithm 1, the missing information is processes by computing the difference and mean of each data for n number of rows. Upon identifying the missing values in a given dataset, the missed entries can be further filled with reshaping or other missing imputation schemes.

Similarly, Algorithm 2 proposed an extended version of ANN where after applying the MICE for searching out the missing data; the extended ANN can be used to further process the missing data efficiently. In the given algorithm 2, a matrix is initialized with zero values where the data is subtracted column wise. An absolute mean value from the diff_mat matrix is computed by analyzing the maximum integer values. In addition, fitting models are used to further impute the missed values in dataset.

Algorithm-II: Proposed Algorithm

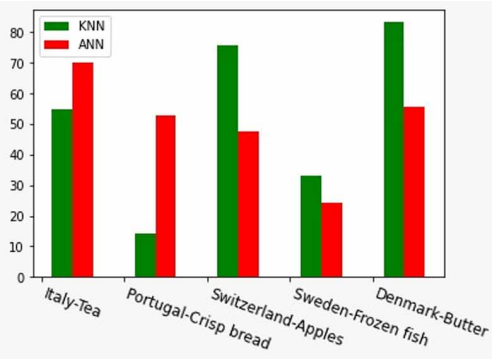
Input: **Number of rows containing missing values in a dataset.**
Output: **Whether the missing values are imputed using extended-ANN algorithm**
Extended_ANN ® [] //Initialize empty array
data2 = data //data is the value of dataset before applying imputation
data1 ® [0] // initiaze matrix with zero value.
diff_mat = data2 – data1 //take difference of data2 and data1
mean2 = average(diff_mat) //mean2 is absolute average of the matrix diff_mat
mean1 ® INT_MAX //initialize mean1 as max integer possible
while (mean2 <= mean1): //iterate until error is reducing.
mean1 = mean2 //make mean1 as previous value for the next iteration
data2 = data1; // make data2 as previous value for the next iteration
for i ® Missing: // iterate for every missing value
imputed_value ® process_fn(data1,i) //Function for fitting the model
data1[index][index2] = imputed_value // update the current value
diff_mat ® data2-data1 // updating difference matrix
mean2 = diff_mat.mean() //updating current mean

A sample dry run of the conference is represented in APPENDEX-A.

4. EXPERIMENT AND PERFORMANCE ANALYSIS

The implementations every of these approach are done using PYTHON having various limitations that require to be set. Despite the reality that each one of them has a default putting for the corresponding limitations, those settings aren't usually best for the dataset of matter. Our first method for putting the parameters for every approach become to do grid-seek and pick out those that yield the smallest mistakes for the cutting-edge dataset. This, however, introduces a massive over-fitting of the models. Thus, we determined to remember the optimum parameter putting primarily based totally on the scale of every dataset. For the NMF approach, the requirement is to set the factorization rank decrease than the wide variety of columns on your dataset (wide variety of nations within side the first sort of datasets or wide variety of foods within side the 2nd sort of datasets).

Figure 1. Imputation of missing values using ANN and KNN approach

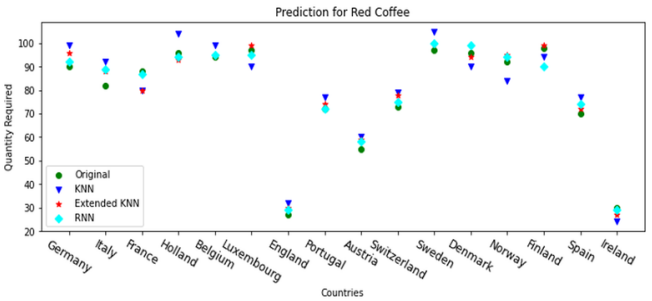


4.1. Results and Comparison

If we are set the factorization rank simply under the wide variety of columns, the approach is able to manipulative the missing values simply to a positive percentage, after which due to wide variety of missing values it cannot be applied. Because of this issue we determined to decrease the factorization rank with the growing of the percentage of missing data, as a consequence the very best viable factorization rank is selected for every percentage.

Figure 1 presents the imputation of missing values against KNN and ANN values. The presented figure shows better results of KNN as compare of ANN because of efficient data searching and analysis with better comparison as compare to KNN approach. Further, the ANN problem is used or further comparison as compare to other schemes.

Figure 2. Prediction of red coffee consumption by various countries



When the factorization rank is simply too excessive the NMF set of rules returns an mistakess, we lower the factorization rank till the NMF set of rules is implemented and no mistakess is returned. We are capable of try this due to the fact we're handling small sized datasets, and this will be relevant

Figure 3. Imputed values using KNN, KNN extended and RNN

```
Original Values:  [90, 82, 88, 96, 94, 97, 27, 72, 55, 73, 97, 96, 92, 98, 70, 30]
Imputed Values using KNN:  [99, 92, 80, 104, 99, 90, 32, 77, 60, 79, 105, 90, 84, 94, 77, 24]
Imputed Values using KNN Exyended:  [96, 88, 80, 93, 95, 99, 30, 74, 59, 78, 100, 94, 95, 99, 72, 27]
Imputed Values using RNN:  [92, 89, 87, 94, 95, 95, 29, 72, 58, 75, 100, 99, 94, 90, 74, 29]
```

in all instances whilst handling FCDBs. As represented in Figure 2, the consumption of red coffee by various citizens of the countries can be presented using various algorithms. The prediction of data through various algorithms across various countries can be analyzed having various result values before and after imputing the values using KNN, KNN extended and RNN upon original dataset.

The presented Figure 3 determines the result values after imputing the data in missing columns from original dataset using KNN, KNN extended and RNN. As determined in Figure 2, the results of KNN are better as compare to other schemes which can be further compared against our proposed extended ANN that is a combination of MICE and ANN.

Figure 4. Imputation value using original, extended ANN and ANN

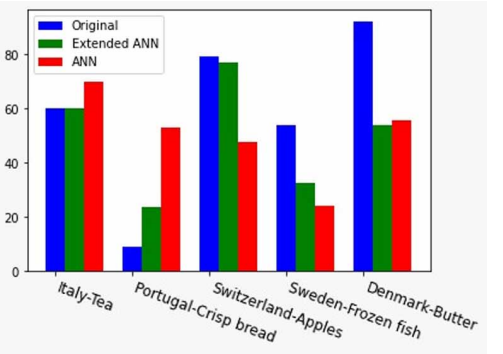


Figure 4 presents the missing values imputation on the given dataset to determine the consumption of products over original data having missing values, ANN and extended ANN. The presented Figure 4 presents better results in case of extended ANN as compare to original and ANN.

5. CONCLUSION

This paper presents number of schemes or machine learning algorithms to determine the imputation of missing values as compare to other approaches. The paper proposed hybrid schemes of MICE and ANN known as extended ANN to search and analyze the missing values and perform imputation in given dataset. The proposed mechanism is efficiency able to analyse the blank entries and fill with proper examining their neighbouring records in order to improve the accuracy in the dataset. In order to validate the proposed scheme, the extended ANN is further compared against various recent algorithms or mechanisms to analyse the efficiency as well as the accuracy of the results. The simulated results represent proposed mechanism outperformance as compare to other approaches in various graphs such as prediction of red coffee consumption by various citizens of country and imputation of missing values on consumption of various foods in several countries.

REFERENCES

Altman, I. (1992). A transactional perspective on transitions to new environments. *Environment and Behavior*, *24*(2), 268–280. doi:10.1177/0013916592242008

Arbuckle, J. L. (2011). *IBM SPSS Amos 20 user's guide*. Amos Development Corporation, SPSS Inc.

Bermejo, S., & Cabestany, J. (2000). A batch learning vector quantization algorithm for nearest neighbour classification. *Neural Processing Letters*, *11*(3), 173–184. doi:10.1023/A:1009634824627

Bhalaik, S., Sharma, A., Kumar, R., & Sharma, N. (2020). Performance modeling and analysis of WDM optical networks under wavelength continuity constraint using MILP. *Recent Advances in Electrical & Electronic Engineering, 13*(2), 203-211.

Bogoviz, A. V. (2018). *A critical review of Russia's energy efficiency policies in agriculture*. Academic Press.

Carleial, A., & Hellman, M. (1975). Bistable behavior of ALOHA-type systems. *IEEE Transactions on Communications*, *23*(4), 401–410. doi:10.1109/TCOM.1975.1092823

Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-6*(4), 325–327. doi:10.1109/TSMC.1976.5408784

Fan, M., & Sharma, A. (2021). Design and implementation of construction cost prediction model based on SVM and LSSVM in industries 4.0. *International Journal of Intelligent Computing and Cybernetics*.

Fao, F. (2012). Agriculture organization of the United Nations. 2012. *FAO Statistical Yearbook*.

Fukunaga, K., & Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, *21*(1), 32–40. doi:10.1109/TIT.1975.1055330

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. doi:10.1146/annurev.psych.58.110405.085530 PMID:18652544

Gurinović, M., Milešević, J., Kadvan, A., Djekić-Ivanković, M., Debeljak-Martačić, J., Takić, M., Nikolić, M., Ranković, S., Finglas, P., & Glibetić, M. (2016). Establishment and advances in the online Serbian food and recipe data base harmonized with EuroFIR™ standards. *Food Chemistry*, *193*, 30–38. doi:10.1016/j.foodchem.2015.01.107 PMID:26433284

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Scientific Software International.

Jóźwik, A. (1983). A learning scheme for a fuzzy k-NN rule. *Pattern Recognition Letters*, *1*(5-6), 287–289. doi:10.1016/0167-8655(83)90064-8

Keller, G., Paige, C., Gilboa, E., & Wagner, E. F. (1985). Expression of a foreign gene in myeloid and lymphoid cells derived from multipotent haematopoietic precursors. *Nature*, *318*(6042), 149–154. doi:10.1038/318149a0 PMID:3903518

Pan, Z., Wang, Y., & Ku, W. (2017). A new general nearest neighbor classification based on the mutual neighborhood information. *Knowledge-Based Systems*, *121*, 142–152. doi:10.1016/j.knosys.2017.01.021

Rathee, G., Ahmad, F., Kurugollu, F., Azad, M. A., Iqbal, R., & Imran, M. (2020). CRT-BIoV: A cognitive radio technique for blockchain-enabled internet of vehicles. *IEEE Transactions on Intelligent Transportation Systems*. Advance online publication. doi:10.1109/TITS.2020.3004718

Rathee, G., Balasaraswathi, M., Chandran, K. P., Gupta, S. D., & Boopathi, C. S. (2021). A secure IoT sensors communication in industry 4.0 using blockchain technology. *Journal of Ambient Intelligence and Humanized Computing*, *12*(1), 533–545. doi:10.1007/s12652-020-02017-8

Rathee, G., Garg, S., Kaddoum, G., & Choi, B. J. (2020). Decision-making model for securing IoT devices in smart industries. *IEEE Transactions on Industrial Informatics*, *17*(6), 4270–4278. doi:10.1109/TII.2020.3005252

Rathee, G., Sandhu, R., Saini, H., Sivaram, M., & Dhasarathan, V. (2020). A trust computed framework for IoT devices and fog computing environment. *Wireless Networks*, *26*(4), 2339–2351. doi:10.1007/s11276-019-02106-3

Ren, X., Li, C., Ma, X., Chen, F., Wang, H., Sharma, A., Gaba, G. S., & Masud, M. (2021). Design of multi-information fusion based intelligent electrical fire detection system for green buildings. *Sustainability*, *13*(6), 3405. doi:10.3390/su13063405

Sharma, A., Kumar, R., & Bajaj, R. K. (2021). On Energy-constrained Quickest Path Problem in Green Communication Using Intuitionistic Trapezoidal Fuzzy Numbers. *Recent Advances in Computer Science and Communications, 14*(1), 192-200.

Silverman, B. W., & Jones, M. C. (1989). E. fix and jl hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation: Commentary on fix and hodges (1951). *International Statistical Review/Revue Internationale de Statistique*, 233-238.

Sun, H., Fan, M., & Sharma, A. (2021). Design and implementation of construction prediction and management platform based on building information modelling and three-dimensional simulation technology in industry 4.0. *IET Collaborative Intelligent Manufacturing*.

Vaughn, A. E., Bartlett, R., Luecking, C. T., Hennink-Kaminski, H., & Ward, D. S. (2019). Using a social marketing approach to develop Healthy Me, Healthy We: A nutrition and physical activity intervention in early care and education. *Translational Behavioral Medicine*, *9*(4), 669–681. doi:10.1093/tbm/iby082 PMID:30107586

Xu, X., Li, L., & Sharma, A. (2021). Controlling messy errors in virtual reconstruction of random sports image capture points for complex systems. *International Journal of System Assurance Engineering and Management*, 1-8.

Ashok Kumar Tripathi is currently PhD scholar in the Department of Computer Science & Engineering, Jaypee University of Information Technology, Waknaghat- Solan Himachal Pradesh.

Hemraj Saini is currently working as Associate Professor in the Department of Computer Science & Engineering, Jaypee University of Information Technology, Waknaghat. His working experience is almost 20 years in Academics, Administration, and Industry. He has obtained PhD (Computer Science) Utkal University, VaniVihar, Bhubaneswar; M.Tech. (Information Technolgy) Punjabi University, Patiala; and B.Tech. (Computer Science & Engineering) Regional Engineering College, Hamirpur (H.P.), now NIT. Five (05) Ph.D. degrees have been awarded under his valuable guidance and five more are likely to complete. He is an active member of various professional technical and scientific associations such as IEEE, ACM, IAENG etc. Presently he is providing his services in various modes like, Editor, Member of Editorial Boards, Member of different Subject Research Committees, reviewer for International Journals and Conferences including Springer, ScienceDirect, IEEE, Wiley, IGI Global, Benthem Science etc. and as a resource person for various workshops and conferences. He has published more than 100 research papers in International/National Journals and Conferences of repute.

Geetanjali Rathee is currently working as an Assistant Professor (Senior Grade) in the Department of Computer Science and Engineering with Jaypee University of Information Technology (JUIT), Waknaghat, Himachal Pradesh, since 2017 till today. She received her Ph.D. in Computer Science and Engineering from JUIT, in 2017. She has done her M.Tech in Computer Science and Engineering from JUIT, Waknaghat, in 2014 and B. Tech. From B.M.I.E.T. under MDU, Rohtak in 2011. Her research interests include handoff security, cognitive networks, blockchain technology, resilience in wireless mesh networking, routing protocols, networking, and industry 4.0. She has published 06 national and international patents. She has approximately 10 transaction papers in IEEE with an impact factor of 9.1, and 8.0, 20 SCI papers in Springer and Elsevier journals with an impact factor of more than 2.0. In addition, she has published around 40 Scopus index journals and more than 15 publications in international and national conferences and book chapter. She has also published one book on "Large-Scale Data Streaming, Processing, and Blockchain Security" IGI Global, U.K. She is also a reviewer for various journals such as IEEE Transactions on Vehicular Technology, Wireless Networks, Cluster Computing, Ambience Computing, Transactions on Emerging Telecommunications Engineering, and the International Journal of Communication Systems.