An Intelligent Data Classification System for Micronutrient Analysis of Rural Women

Dr. V V R Maheswara Rao* and N Silpa**
*Professor, Department of CSE, Shri Vishnu Engineering College for Women, Andhra Pradesh, India
mahesh_vvr@yahoo.com
**Assistant Professor, Department of CSE, Shri Vishnu Engineering College for Women, Andhra Pradesh, India
nrusimhadri.silpa@gmail.com

Abstract: The health and nutrition status of marginalized women with constraint economic resource in rural setting remains a great challenge. In addition, poor nutritional knowledge, attitudes and dietary practices significantly contribute to the rapid growth of malnutrition among the rural women. In an attempt to re-architect the present scenario, it is necessary to employ the interventions of science and technology in rural women community. To adapt the interventions of science and technology practically it is very much required to collect, analyze and interpret the data on health profile, family history, dietary habits and blood profile of an individual. Naturally this data is huge in volume, potentially high dimensional, absolutely diversified and structurally complex. To extract and study the insights of such data and segment the rural women efficiently it is necessary to employ a capable data mining technique.

Towards this goal, the investigators in the present paper present an Intelligent Data Classification System (IDCS). The IDCS is designed with four stages on the basis of classification technique with a comprehensive approach. Initially, the first stage of IDCS focusses on collecting the data from various resources on nutritional knowledge, attitudes and dietary practices of the rural women systematically. Later it rightly preprocesses the data with the standard practices and prepare in suitable structure to apply classification technique. Subsequently an intelligent classifier is modeled using learning algorithm which can classify the rural women in desired number of groups with respect to nutritional levels. Finally, this system identifies suitable interventions of science and technology like creating awareness about nutritional knowledge, arranging training workshops to prepare the food without losing nutrients & ready to cook nutritious complementary food, promoting homestead gardens and backyard poultry etc. appropriately to the respective rural women groups. This system helps to identify the under nutritional women accurately and efficiently. This model reduces the error rate and shows the significant performance. To validate the proposed system several experiments are conducted and results proven this are claimed in this paper.

Keywords: Classification, Data Mining Applications, Learning Algorithm and Micronutrient analysis etc.

Introduction

The rapid advancement in the past decades has witnessed an unprecedented change in the lifestyle of rural women affecting the knowledge on nutrition, food attitudes, eating practices and health consciousness led to negative impact on the health of rural women. On the other hand the modern agricultural practices [12] resulted in producing fewer nutrients food. The rural women are consuming unbalanced diet, nutrition transition consequences of occurrence of chronic diseases like increased blood pressure, glucose, abnormal blood lipids, and obesity at exponential rate. To study the complete insights of such data and segment the rural women it is essential to employ a capable data mining technique.

The rural women from 30 villages of 3 mandals in the west Godavari district of Andhra Pradesh, India, are selected for the present study with the support of DST-SEED Division. The data is collected using the standard survey form designed under the supervision of nutrition expert, pertaining series of questions to retrieve the information regarding the nutritional knowledge, attitudes, dietary practices, family history and present health status of an individual rural women. To understand the insights of this data, the present investigators are triggered and motivated towards the advancements of data mining techniques [13, 14] in order to study and explore the needs of different people.

Among all techniques of data mining [6, 7, 18], the present problem promotes the necessity of applying the classification technique to segment the rural women in desired groups. A classification is a systematic approach with a task of assigning data objects to one of several predefined classes. Each classification technique builds on a learning algorithm to decide the fitness of relationship between the data object and the class. As conventional classification techniques require more domain experts' intervention thus, the investigators in the present work, pay attention towards learning classification techniques which classify the desired rural women groups with reduced human intervention.

With this background the present investigators reported in this paper, the necessity of learning algorithm [2, 10, 11, 15, 17] to classify the rich nutrition group and poor nutrition group among the rural women. It helps to provide right interventions of science and technology like creating awareness about nutritional knowledge, arranging training workshops to prepare the food without losing nutrients & ready to cook nutritious complementary food, promoting homestead gardens and backyard poultry etc among the rural women.
The remaining paper is organized as follows: Initially, it gives a brief related work. Then, the proposed Intelligent Data Classification System is presented. Subsequently, it provides mathematical model in detail. Later the experimental analysis is showcased. Finally the conclusions are made.

Related Work

In the literature, many authors [1, 6, 11, 15, 17] expressed that applying of the learning algorithms are more suitable for dynamically growing data. The experimental results of classification techniques [2, 4, 7] have been proposed to improve the overall success rate in segmenting the health care data and its management. Some more authors [2, 3, 10, 13, 18] also expressed the relevance of the data mining in the field of health care, analysis of steady growth of related health survey data and predict human deceases. In the literature [8], the classification technique is used on the health data sets having more features for the diagnosis of heart deceases and they also expressed in their future work, leaning classification techniques lead to better performance. The research provided by some more authors [5, 8, 9] in the literature, found the applications of data mining are growing in analyzing health care data for better health policy making and detection of decease outbreak with root causes. They also mentioned the relevance of learning system that classify the data accurately and alerts the people about the dietary habits promptly.

Proposed Intelligent Data Classification System (IDCS)

The micro nutrient deficiency analysis of rural women is a great socio economic importance, responsible not only in India across the world. This is not always clinically apparent but more associated with food consumption habits and family history data. To understand the significant knowledge in the insights embedded in nutritional knowledge, attitudes and practices of rural women and segmentation of this data accurately becomes a central problem. In addition, handling of such data lie not only high volume, potentially high dimensional, absolutely diversified and also structurally complex, is one of the main motivation for the present investigators. To adapt the right interventions of science and technology for improving the nutritional levels of right rural women, endorsed as a basic motivation factor for the present central problem of a classification.
In order to overcome the inherited challenges of segmenting rural women, the investigators in the present paper propose an intelligent data classification system as shown in Fig 1, which follows fundamentally a comprehensive approach and modeled as a four stage model. The IDCS helps to classify the rural women with respect to the nutritional status accurately and efficiently. Initially, the first stage of IDCS concentrates on collecting the data through various possible sources like nutritional knowledge, dietary and practices of rural women under the guidance of nutritional expert. In the second stage, the IDCS mainly focuses on the preparation of data by employing necessary preprocessing techniques, suitable to classification algorithms. The IDCS in third stage, classify the rural women namely poor & rich nutritional groups by adapting an intelligent classifier which is modeled using a learning algorithm. The final stage of IDCS, identifies right interventions of science and technology to the right groups based on classification of the previous stage under the supervision of nutritionist.

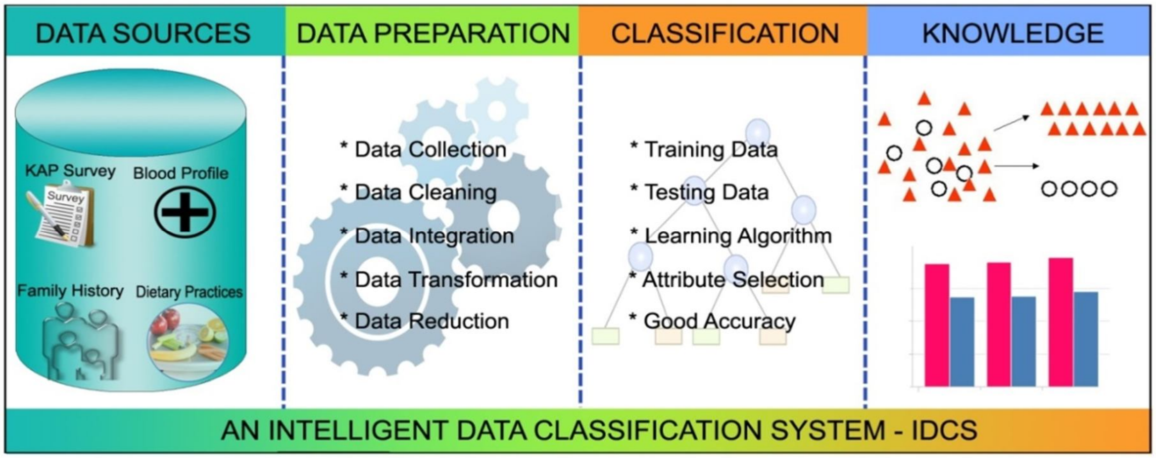

Figure 1. Architecture of IDCS

Data Sources

In order to accurately classify the rural women with respect to nutritional levels is highly dependent on the selection of data resources. Towards this, the investigators in the present work elect survey form as a primary resource to estimate the nutritional levels of rural women. The questions in the survey form are designed under the guidance of nutritional expert. The questions intrinsically retrieve the facts about nutritional knowledge, attitudes, dietary practices and family history of rural women. To adapt interventions of science and technology, and classification of rural women accurately the investigators select the blood sample as a secondary resource.

Data Preparation

The initial stage of IDCS focuses on preparation of desired data under the guidance of a knowledge expert as raw data, highly predisposed to noise, missing values and inconsistency. This data preparation stage is the most important phase in the knowledge discovery process and is critical in successful extraction of desired data. The data preparation helps to improve the efficiency and ease of classification technique. The task of data preparation consumes a bulk amount of effort in the entire investigation. The data preparation of IDCS covers all the activities including data collection, data cleaning, data integration, data transformation and data reduction to prepare data suitable to classification algorithm from the initial raw data.

Data Collection: The data collection task is performed on the basis of input drawn from the nutritional status of the rural women. The possible ways of collecting the data of rural women are knowledge Attitude and Practices (KAP) survey, family history, dietary habits and blood profile under the supervision of nutritional expert. This activity includes significance to the investigation goal, quality and technical limitations such as limits on data volume or data types. It is very important, however, to understand how data collection affects the results of investigation, since such a prior knowledge is also useful for the final interpretation of results.

Data Cleaning: Data cleaning deals with detecting and removing the incomplete, noisy and inconsistent data in order to improve the quality of data. This data preparation activity of IDCS is particularly required when integrating heterogeneous databases. Initially, to fill the missing values, IDCS employs a popular strategy “Use the most probable value to fill in the missing value”. This strategy uses most of the information from the present data to predict missing values that are determined with Bayesian approach. Subsequently, IDCS designates a linear regression model to smoothing the noisy data. The mathematical equation derived using linear regression model fit the data and helps to smooth out the noise. Finally, IDCS adopts the concept of functional dependencies between attributes to resolve the inconsistencies.

Data Integration: Data integration is a process that combines data from multiple distributed sources into a coherent data store. It specifically, aims at increasing the completeness, conciseness and correctness of the data which is fed to the classification techniques. The completeness measure concentrates on the number of attributes while the conciseness identifies the uniqueness of attribute in the integrated data. Additionally, correctness measure focuses on confirmation of integrated data to the real world. The data integration activity of IDCS primarily resolves heterogeneity and schema level by establishing semantic mapping among contents of multiple data sources.

Data Transformation: The data transformation consolidates the data into a single desired form which is readily fed to the classification technique. The data transformation activity of IDCS involves normalization, aggregation and generalization of data. The normalization concentrates to scale the data in a small specified range. The aggregation performs functions that are applied to the data for summarization. In generalization of data the raw data is replaced by higher level concepts using concept hierarchy.

Data Reduction: The data reduction is a technique that reduces volume of data set much smaller, at closely maintains the integrity of original data. The data reduction activities of IDCS uses attribute subset forward selection and reverse elimination strategy for data reduction. This strategy reduces the data set size by removing irrelevant attributes. The role of this strategy is to find a minimum set of attributes such that, the resulting probability distribution is as close as possible to the original distribution obtained using all attributes.

The whole activity of data preparation elevates the quality of the data to the required classification techniques. The resultant processed data feeds to the next stage of IDCS as an input to successfully classify the rural women based on their nutritional status.

Classification of Rural Women

The data collected and prepared from various resources is potentially high diversified and having poor nutritional rural women data as well as rich nutritional rural women data. To adapt interventions of science and technology the data of rural women having rich nutrition is not required. Thus it is essential to classify the rural women with respect to nutritional levels into two groups as Rich Nutritional Women Group (RNWG) and Poor Nutritional Women Group (PNWG). To perform this task the investigators in the present paper proposes Intelligent Data Classification System (IDCS) that acquires the knowledge from the facts of standard nutritional data. Together with this intelligence the IDCS classifies RNWG and PNWG accurately. In addition, it helps in implementing the interventions of science and technology for the successful implementation of overall proposed work.

The IDCS takes the prepared data and classifies RNWG and PNWG intelligently with less time. The prepared data collected from various resources includes dietary habits data, health data and family history data. Towards this, the IDCS algorithm is designed in such a way that it acquires intelligence from the related and standard nutritional facts. For the proposed IDCS derived attributes can be taken as set of related facts, which are specific in nature and classify the set of classes as RNWG and PNWG. These derived attributes from the original collected data used as training data to the IDCS. The complete collected and prepared data becomes testing data for IDCS.

In order to classify any new data in the testing data either RNWG or PNWG, IDCS using learning intelligence approach by posing a series of questions about the characteristics of newly entered data. In each time IDCS receives the answer and follow up question is asked until IDCS reaches a conclusion about the class label of new entry. The series of questions and their respective answers can be organized in form of a tree which is hierarchical structure consisting of nodes and edges. The IDCS tree consist root node, internal node and leaf node. A root node is a node that has no incoming edges and two or more outgoing edges. Any internal node has exactly one incoming edge and two or more outgoing edges. The leaf node is node which has exactly one incoming edge and no outgoing edges. In the process of IDCS tree development, the root node and other internal nodes are assigned with the nutritional knowledge, attitude & practices of rural women and each leaf node is generates a class labels. The class labels are either RNWG or PNWG.

To classify the rural women data accurately, one can use the intelligent systems to reduce the time for overall processing of proposed work. In principal, there are exponentially many tree algorithms that can be constructed from a given set of attributes. But finding the learning tree algorithm, which is computationally feasible and capable of acquiring intelligence for exponential size of rural women data is highly complicated. These algorithms specifically utilize a sensible line of attack to induce the intelligence. One such algorithm is IDCS, which is customized and is suitable for typical nature of rural women data. As stated earlier, due to the complex nature of data one has to focus on creation of the related facts-base of derived attributes. The input for the IDCS consists of training data and testing data. Finally IDCS generates the learning tree structure based on a set of rules and which consists of nodes and class labels.

The input for IDCS is obtained from the data prepared from the survey forms of rural women, which is having highly diverse and voluminous entries. Each entry corresponds to a details of individual rural women consists of numerous attributes of nutritional knowledge, attitude, dietary practices and details of health status. For the proposed IDCS, all these numerous attributes treated as testing data (TA), which consists of entries with unknown labels.

Among all the attributes, some of the attributes are useful to classify the rural women. These attributes are considered as training data or derived attributes. Based on the notable features of tree these attributes are identified, which are discrete, continues and specific in nature. The entries in training data have known labels. The set of derived attributes DA = {DA1, DA2, ... DAn}, which forms knowledge base for IDCS. This knowledge base makes the IDCS to acquire enough learning to classify the women automatically and reduces the processing time significantly. The proposed IDCS derives many attributes from details of nutritional knowledge, nutritional attitudes, dietary practices and health status of rural women collected and prepared from various resources to classify as RNWG and PNWG. A sample of derived attributes is shown in Table 1 and a sample characteristics of rich and poor nutritional rural women are listed in Table 2 respectively.

Table 1. Sample Derived Attributes

Derived Attribute	Description
Adequacy of food	Foods provide enough of each essential nutrient, fiber and energy
Balanced diet	Food choices do not overemphasize on nutrient or food at the expense of fiber
Calorie control	Foods provide amount of energy you need to maintain ideal weight
Moderation of food	Foods do not provide excess fat, salt, sugar, or other unwanted constituents
Variety of food	Foods differ from day to day
Health status	Details of present health condition
Hereditary complaints	Details of a diseases or disorders that is inherited genetically
Life Style Habits	Details of Common Lifestyle Habits that Cause Diseases
Vitamin Deficiency Symptoms	Details of signs & symptoms of Vitamin deficiencies

Table 2 Sample Characteristics of Rich and Poor Nutritional Rural Women

Rich Nutritional Women Group (RNWG):	Poor Nutritional Women Group (PNWG):
➢ Take adequate food that is essential nutrient, fiber and energy ➢ Maintain balanced diet choices. ➢ Have the knowledge of calorie control on food. ➢ Have the conscious on the foods do not provide excess fat, salt, sugar, or other unwanted constituents. ➢ Eat many varieties Foods from day to day ➢ Good condition of present health. ➢ No diseases or disorders that is inherited genetically ➢ Good lifestyle habits ➢ No signs and symptoms of Vitamin deficiencies	➢ Take less food that is essential nutrient, fiber and energy ➢ Unable to maintain balanced diet choices. ➢ No knowledge of calorie control on food. ➢ Less conscious on the foods do not provide excess fat, salt, sugar, or other unwanted constituents. ➢ Eat same kind of foods daily ➢ Suffer from health conditions. ➢ Have diseases or disorders that are inherited genetically. ➢ Bad lifestyle habits and addictions ➢ Observe signs and symptoms of Vitamin deficiencies.

The input to the IDCS algorithm consists of training data (DA) and testing data (TA). The learning algorithm works by recursively selecting DeriveBestAttribute() (step 07) and expanding the leaf nodes of the tree (step 11 & 12) until ConditionEnd() is met (step 01).

Algorithm of IDCS

Algorithm : Intelligent Data Classification System - IDCS	
TreeCreate(DA, TA)	
Step 01:	If (ConditionEnd(DA, TA) = True) then
Step 02:	LastNode = CreateNewNode( )
Step 03:	LastNode.Label = GiveLabel(DA)
Step 04:	Return LastNode
Step 05:	Else
Step 06:	Root = CreateNewNode( )
Step 07:	Root.ConditionCheck = DeriveBestAttribute(DA, TA)
Step 08:	Let V ={v /v is a possible outcome of ConditionCheck()}
Step 09:	For each v Є V do
Step 10:	DAv ={da / Root.ConditionCheck(da) = v and d Є DA}
Step 11:	Child = TreeCreate(DAv, TA)
Step 12:	Add Child as descendant of root and label the edge as v
Step 13:	End for
Step 14:	End if
Step 15:	Return root

The details of methods used in the learning algorithm are follows,

CreateNewNode() : The IDCS make use of this function to create the tree by creating a new node. A new node in IDCS tree is given either a test condition or a class label.

ConditionCheck() : Each recursive step of TreeCreate() of IDCS must select an attribute check condition to divide into two subsets namely RNWG and PNWG. To implement this step, IDCS learning algorithm uses a method ConditionCheck() for measuring goodness of each attribute condition.

ConditionEnd(DA, TA) : The IDCS adapts this function to stop the IDCS tree extension process by testing whether all the records have either the same class label or the same attribute values. Another way of ending the function is to test whether the number of records has fallen below minimum value.

GiveLabel() : The IDCS make use of this function to determine the class label to be assigned to a last node. For each last node t, Let p(i/t) denotes the rate of training records from class i associated with the node t. In most of the cases the last node is assigned to the class that has more number of training records.

DeriveBestAttribute() : The IDCS process this function to find out which attribute should be selected as a test condition for deriving the training records. To ensure the goodness of split, the Entropy and Gini index are used.

Mathematical Model to estimate the Learning Performance of IDCS

The learning performance of any algorithm is proportionate on the training of algorithm, which directly depends on the training data. As testing data is continuously growing the training data is also continuous. Hence to estimate the training data one can use predictive modeling technique called regression. The goal of regression is to estimate the testing data with minimum errors.

Let S denote a survey data of rural women that contains N observations,

$$S = \{(D_i, T_i) / i = 1,2,3, \ldots, N\}$$

Suppose to fit the observed data into a linear regression model, the line of regression D on T is

$$D = a + bT \qquad (1)$$

Where a and b are parameters of the linear model and are called regression coefficients. A standard approach for doing this is to apply the method of least squares, which finds the parameters (a, b) that minimize the sum of squared error say E.

$$E = \sum_{i-1}^{n} (D_i - a - bT_i)^2 \qquad (2)$$

The optimization problem can be solved by taking partial derivative of E w.r.t a and b, equating them to zero and solving the corresponding system of linear equations.

$$\frac{\partial E}{\partial a} = 0 \quad \Rightarrow \quad \sum_{i-1}^{n} D_i = na + b\sum_{i-1}^{n} t_i \qquad (3)$$

$$\frac{\partial E}{\partial a} = 0 \quad \Rightarrow \quad \sum_{i-1}^{n} D_i t_i = a\sum_{i-1}^{n} t_i + b\sum_{i-1}^{n} t_i^2 \qquad (4)$$

Equations (3) and (4) are called normal equations. By solving equations (3) and (4) for a given set of Di, Ti values, we can find the values of ‘a’ and ‘b’, which will be the best fit for the linear regression model. By dividing equation (3) by ‘N’ we get

$$\bar{D} = a + b\bar{t} \qquad (5)$$

Thus the line of regression D on T passes through the point ($\bar{D}, \bar{T}$)

$$\mu_{11} = Cov(D, T) = \frac{1}{n}\sum_{i-1}^{n} D_i T_i - \bar{D}\bar{T}$$

We can define,

$$\Rightarrow \frac{1}{n}\sum_{i-1}^{n} D_i T_i = \mu_{11} + \bar{D}\bar{T} \qquad (6)$$

Also,

$$\frac{1}{n}\sum D_i^2 = \sigma_d^2 + \bar{D}^2 \qquad (7)$$

From equations (4), (6) and (7) we get

$$\mu_{11} + \bar{D}\bar{T} = a\bar{D} + b(\sigma_d^2 + \bar{D}^2) \qquad (8)$$

And on simplifying (8), we get

$$\mu_{11} = b\sigma_d^2 \Rightarrow b = \frac{\mu_{11}}{\sigma_d^2} \qquad (9)$$

b is called the slope of regression D on T and the regression line passes through the point ($\bar{D}, \bar{T}$). The equation of the regression line is

$$D - \bar{D} = b(T - \bar{T}) = \frac{\mu_{11}}{\sigma_d^2}(T - \bar{T})$$

$$D - \bar{D} = r\frac{\sigma_d}{\sigma_t}(T - \bar{T})$$

$$\Longrightarrow D = \bar{D} + r\frac{\sigma_d}{\sigma_t}(T - \bar{T}) \qquad (10)$$

The linear regression coefficient 'r' is used to predict the error between testing data and training data. It also used to study the nature of the relationship between training data and testing data. The learning performance can also be expressed in terms of training error rate of the learning algorithm. The training error rate is given by the following equation,

$$\text{Training Error Rate} = \frac{\text{Number of wrong characteristic definitions}}{\text{Total number of characteristic definitions}}$$

Experimental Analysis

Learning performance of IDCS: The sample rural women data collected across 30 villages in 3 mandals is experimented over a period of twelve months under standard execution environment. The experiments prove that intelligent data classification system reduces the human intervention to classify the data. The error rate between the testing data and training data is almost minimized in IDCS and is found to be 0.2 on the average. Hence the experimental study is in line with the mathematical modelling of, goal of regression. The nature of relation between testing data and training data is studied and both are proven as continuous as shown in Fig 2.

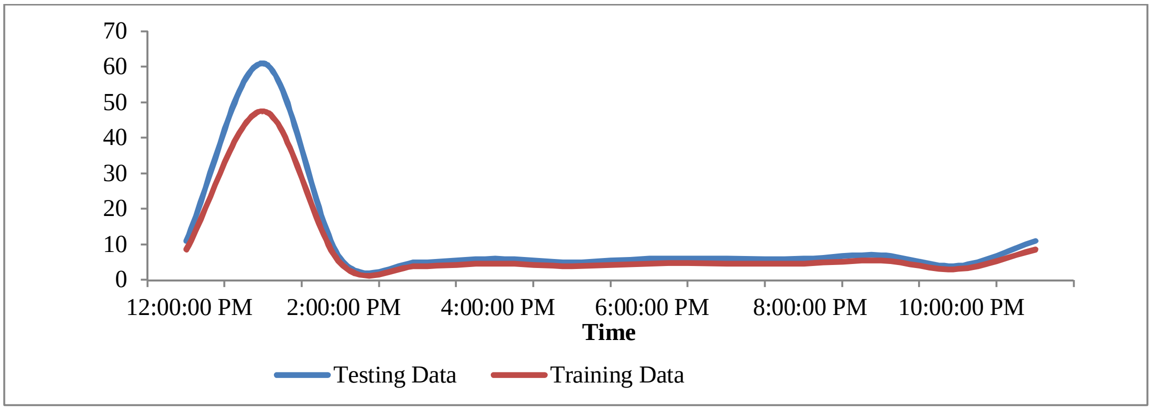

Figure 2. Error rate between training and testing data

Processing Performance of IDCS: Several experiments are conducted in a standard environment with respect to the processing time of IDCS. The results clearly indicate that IDCS is consistently taking less processing time as shown in Fig.3.

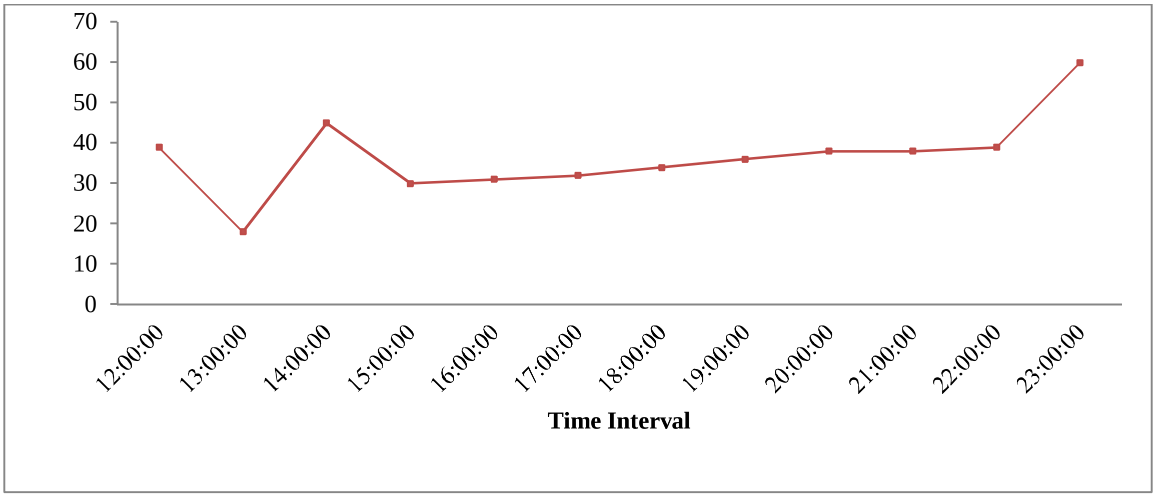

Figure 3. Processing performance of IDCS

Conclusion

Classification is a promising research area in the era of data mining to study the insights of the rural women data and segment the rural women with respect to nutritional levels. The investigators in the present paper outlined the importance of all the stages namely selection of suitable sources of data, data preparation, classification and knowledge interpretation. The investigators introduced an intelligent learning algorithm IDCS to separates the Rich Nutritional Women Group and Poor Nutritional Women Group efficiently. The IDCS reduces the human intervention and classify the said groups automatically in less time. Also, mathematical model is presented to ensure the goodness of error rate. Finally the proposed system helped to adapt right interventions of science and technology like creating awareness about nutritional knowledge, arranging training workshops to prepare the food without losing nutrients & ready to cook nutritious complementary food, promoting homestead gardens and backyard poultry etc. among the less nutritional rural women.

Acknowledgement

The investigators would like to thank the Department of Science & Technology (DST), Ministry of Science &Technology, and Government of India under Science for Equity, Empowerment and Development (SEED) Division for providing the fund to implement this project. The investigators also recorded their acknowledgements to the authorities of Shri Vishnu Engineering College for Women (Autonomous), Bhimavaram, Andhra Pradesh, India, for their constant support and cooperation.

References

[1] A. Anguera, J.M. Barreiro, et. al., “Applying data mining techniques to medical time series: an empirical case study in electroencephalography and stabilometry”, Computational and Structural Biotechnology Journal 14, pp: 185–199, 2016.
[2] H. Hamidi, and A. Daraei, “Analysis of Pre-processing and Post-processing Methods and Using Data Mining to Diagnose Heart Diseases”, IJE Transactions, Vol. 29, No. 7, pp: 921-930, 2016.
[3] Sheenal Patel and Hardik Patel,” Survey of Data Mining Techniques Used in Healthcare Domain”, International Journal of Information Sciences and Techniques, Vol.6, No.1/2, pp: 53-60, 2016.
[4] Olegas Niaksu, “Development and Application of Data Mining Methods in Medical Diagnostics and Healthcare Management”, Doctoral Dissertation, Vilnius University, 2015.
[5] Shubpreet Kaur and Dr. R.K.Bawa, “Future Trends of Data Mining in Predicting the Various Diseases in Medical Healthcare System” IJEIC, Vol.6, Issue 4, pp: 17-34, 2015.
[6] Health Catalyst, “It All Starts with a Data Warehouse”, 2014.
[7] Juan A. Lara, David Lizcano, “A general framework for time series data mining based on event analysis: Application to the medical domains of electroencephalography and stabilometry”, Journal of Biomedical Informatics 51, pp: 219–241, 2014.
[8] Salha M. Alzahani, Afnan Althopity, et al., “An Overview of Data Mining Techniques Applied for Heart Disease Diagnosis and Prediction”, Lecture Notes on Information Theory, Vol. 2, No. 4, pp: 310-315, 2014.
[9] Abdullah A. Aljumah and Mohammed Gulam Ahamad, “Application of data mining: Diabetes health care in young and old patients”, Journal of King Saud University – Computer and Information Sciences 25, pp: 127–136, 2013.
[10] Divya Tomar and Sonali Agarwal, “A survey on Data Mining approaches for Healthcare”, IJBSBT, Vol.5, No.5, pp. 241-266, 2013.
[11] R.S. Santos, S.M.F. Malheiros, et. al., “A data mining system for providing analytical information on brain tumors to public health decision makers” computer methods and programs in bio medicine 109, pp: 269–282, 2013.
[12] Hedwig Acham, Wilna H. Oldewage-Theron, et al., “Dietary diversity, micro nutrient intake and their variation among black women in informal settlements in South Africa: A cross-sectional study”, IJNAM, Vol. 4(2), pp. 24-39, 2012.
[13] Hian Chye Koh and Gerald Tan “Data Mining Applications in Healthcare”, Journal of Healthcare Information Management, Vol. 19, No. 2, pp: 64-72, 2012.
[14] Tipawan Silwattananusarn and Dr. Kulthida Tuamsuk, “Data Mining and its Applications for Knowledge Management: A Literature Review from 2007 to 2012”, IJDKP, Vol.2, No.5, pp; 13-24, 2012.
[15] D.Senthil Kumar, G.Sathyadevi and S.Sivanesh, “Decision Support System for Medical Diagnosis Using Data Mining”, IJCSI, Vol. 8, Issue 3, No. 1, pp: 147-153, 2011.
[16] Christos A. Frantzidis, Charalampos Bratsas, et al., “On the Classification of Emotional Bio signals Evoked While Viewing Affective Pictures: An Integrated Data-Mining-Based Approach for Healthcare Applications”, IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 2, pp: 309-318, 2010.
[17] Mohit Kumar, Rayid Ghani, and Zhu-Song Mei, “Data Mining to Predict and Prevent Errors in Health Insurance Claims Processing”, KDD’10, Washington, DC, USA, pp:65-73, 2010.
[18] Olegas Niaksu and Olga Kurasova, “Data Mining Applications in Healthcare: Research vs Practice”, Institute of Mathematics and Informatics, Vilnius University, pp: 58-70, 2010.