IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Discovering Hidden Patterns in Diabetes Data Using K-Means Clustering Algorithm and Association Rules

O.Turoti[1], O. O. Obe[2]

[1,2]Department of Computer Science, Federal University of Technology Akure, Ondo, Nigeria.

*Corresponding Author: mturoti@gmail.com

Available online at: www.ijcseonline.org

Accepted: 13/Jun/2018, Published: 30/Jun/2018

Abstract- Diabetes is considered as one of the deadliest diseases in the world, therefore medical professionals need a reliable prediction and decision making methodology. The main aim of this paper is to implement data mining in diabetes diagnosis to discover new patterns and to interpret the data patterns to provide meaningful and useful information for medical practitioners. The analytical technique for this project consists of five stages data collection, preprocessing, feature extraction, implementation and then interpretation and evaluation.

In this research, the analytical technique implores the use of k-means clustering algorithm and association rules (A-priori algorithm) was used to analyse diabetes dataset collected from two hospitals in Ondo State, Nigeria. The analytical technique proposed was implemented in PyCharm Community Edition. Three clusters were generated using K-means clustering algorithm and A-priori algorithm was used to generate patient's profile for each cluster. Performance evaluation on the technique was carried out using accuracy and showed result of 85% which indicates that the technique is efficient.

Keywords: Data Mining, Diabetes, K-Means Clustering, Association Rule Mining.

I. INTRODUCTION

The development of information technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. The amount of data continues to grow at an enormous rate even though the data stores are already vast. The major challenge is how to make the database a competitive business advantage by converting seemingly meaningless data into useful information. How this challenge is met is through Data Mining.

Data mining procedure is concerned with the investigation of data using some product methods or software techniques for finding covered up and unforeseen patterns and connection in sets of information [1].

Data mining is a process of extraction of useful information and patterns from huge data. [2].

Data mining is a core of the knowledge discovery process [3] and it is also known as knowledge mining from data, knowledge extraction or data /pattern analysis.

There are two types of data mining; Descriptive and Predictive data mining. Descriptive data mining can be defined as to discover regularities in the data and to uncover patterns can also be used to find interesting subgroups in the bulk of data. Predictive data mining is a process that uses data mining and probability to forecast outcomes.

The rest of the paper is organized as follows; Section I contains the introduction of data mining, Section II contains literature review of related works, Section III explains the methodology with flowchart, Section IV describes results of the methodology's implementation and discussion and Section V concludes the research work with future work.

I.I DATA MINING METHODS

Data mining methods refer to the function types that data mining tools provide. Data mining methods are either predicting the future or explaining the past. These methods are classified into four major categories: Association, Classification, Clustering and Regression.

Association: This discovers relationship or dependencies between multiple things, such as link analysis, market basket analysis and variable dependencies. Association exists in two levels: structured and quantitative. The structural analysis technique and are used to discover relationships between attributes and items. In this technique level specifies which things are related, the quantitative level specifies the strengths of the relationship using some numerical scale.

Classification: This maps a data item into one of several predefined categorical classes. There are two steps to implement classification function. Firstly, the model is built describing a predetermined set of classes or concept.

Secondly, the model is used for classification. Classification problems aim to identify the characteristics that indicate the group to which each data belongs.

Clustering: The goal of clustering is to find groups that are very different from each other, and whose members are very similar to each other. Unlike classification, the clusters are not known at the beginning as well as the attributes the data will be clustered.

Regression: This uses existing values to forecast what other values will be. In the simplest case, regression uses standard statistical techniques such as linear regression.

This paper therefore aims at analyzing diabetes data using data mining techniques; k-means clustering and Association Rule Mining to uncover hidden patterns.

I.II DATA MINING PROCESS

According to [4], Data mining combines techniques from machine learning, pattern recognition, statistics, database theory, and visualization to extract concepts, concept interrelations, and interesting patterns automatically from large corporate databases.

Data acquisition: The first step is to select the types of data to be used. Although a target data set has been created for discovery in some applications, Data mining can be performed on a subset of variables or data samples in a larger database.

Pre-processing data: Once the target data is selected, the data is then pre-processed for cleaning, scrubbing, and transforming to improve the effectiveness of discovery. During this pre-processing step, developers remove the noise or outliers if necessary and decide on strategies for dealing with missing data fields and accounting for time sequence information or known changes.

Data exploration and model building: The third step of data mining refers to a series of activities such as deciding on the type of data mining operation; selecting the technique; the algorithm; and mining the data.

Interpretation and evaluation: The fourth step of the data mining process is the interpretation and evaluation of discovered patterns. This task includes filtering the information to be presented by removing redundant or irrelevant patterns, visualizing graphically or logically the useful ones, and translating them into understandable terms by users. In the interpretation of results, we determine and resolve potential conflicts with previously found knowledge or decide to redo any of the previous steps.

II. LITERATURE REVIEW

Diabetes mellitus is one of the most common and serious chronic diseases in the in the world. Diabetes is a metabolic disease in which the body does not produce or properly use insulin, a hormone that is needed to convert sugar, starches, and other food into energy needed for daily life. Diabetes is characterized by high levels of blood glucose (sugar). There are three main types of diabetes: Type 1 diabetes, Type 2 diabetes and Gestational diabetes.

Presently, regarding the application of data mining for purpose of diabetes diagnosis, the following are the researches that have been carried out to mention a few;

Aljumah et al [5] presents application of data mining in diabetes health care of both young and old patients. The data set used was the 2005 dataset is a standard NCD risk factor report from the Ministry of Health, Saudi Arabia. Oracle Data Miner was used to implement the algorithm. A regression technique that employed a support vector machine algorithm was used.

Nagarajan and Chandrasekaran [6] proposed a design and implementation of expert clinical system for diagnosing diabetes using data mining techniques. The research design made use of primary and secondary data. The clustering algorithm Simple K-Means was used to develop a model for this paper.

This model groups the dataset into type-1, type-2 or gestational diabetes. After clustering the type of diabetes, the model also applied classification algorithms like RandomTree, NaiveBayes, SimpleCart and Simple Logistics for predicting the risk levels of diabetes as mild, moderate and severe.

Padmaja et al [7] presents their research aimed at finding out the characteristics that determine the presence of diabetes and to track the maximum number of women suffering from diabetes. They used Data mining functionalities like clustering and attribute oriented induction techniques to track the characteristics of the women suffering from diabetes.

Pramanand and Sankaranarayanan [8] proposed an idea about diabetes mellitus its diagnosis using data mining with minimum number of attributes applied to classification algorithms. They worked on Apriori and FP-growth techniques and realized that with the help of data Apriori and FP-growth algorithms, the computation cost decreases and also the classification performance increases

Pala and Yucedag [9] presents a data mining approach for diagnosis of diabetes using association rules and clustering. The data analysed was the Pima Indian database. After data collection, data preprocessing was carried out on the dataset. A hybrid of techniques namely; k-means clustering algorithm and A-priori algorithm was proposed for effectively diagnosing type 2 diabetes.

Kumar and Umatejaswi [10] presents diagnosing diabetes using data mining technique. The authors tried to diagnose diabetes based on the 650 patient's data. Simple k-

means algorithm was used for clustering the entire dataset into 3. This clustered dataset was given as input to the classification model using different classification algorithm such as Random Tree, Naive Bayes, C4.5 and simple Logistics which further classifies each patient's risk levels of diabetes as mild, moderate and severe.

Iyer et al [11] presents diagnosis of diabetes using classification mining techniques. In this paper two algorithms namely, J48 (decision tree algorithm) and Naïve Bayes were used to create the model for diagnosis. After data pre-processing (CSV format), the J48 algorithm was employed on the dataset using WEKA (Java Toolkit for various data mining technique).

Vani and Priyadharshni [12] presents discovering the diagnosis of diabetes mellitus by using association rule mining. Hierarchal Clustering Explorer (HCE) and RPC (Recursive Partition Clustering) Tree algorithm were used for clustering the dataset and then association rules were applied.

Al-Rofiyee et al [13] presents using prediction methods in data mining for diabetes diagnosis. The PIMA India diabetes dataset is converted to the ARFF format and was inserted in WEKA software. The WEKA (Classifier) is used to train and test the model that was build based on the training results, under the Multi-layer perception function.

Tugba [14] presents defining characteristics of diabetic patients by using data mining tools. The dataset contained 21 variables and 148 records. K-means clustering algorithm was used to identify the profile of the patients and a-priori algorithm were used to find which illness occurred together. IBM Modeler was used to implement the algorithms.

III. METHODOLOGY

System design and implementation are the most important phase of any project. The analytical technique for this project consists of five stages data collection, preprocessing, feature extraction, implementation of K-means clustering algorithm and association rules and then interpretation and evaluation.

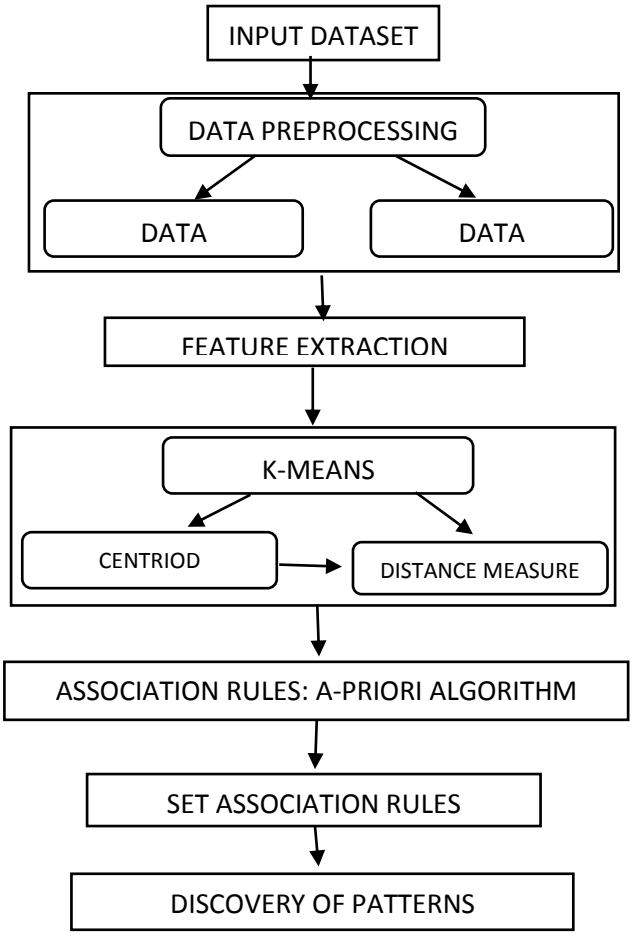


Figure 1: Framework for K-means clustering and Association rule technique

III.1 DATA COLLECTION

Data collection is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables one to answer relevant questions, test hypothesis and evaluate outcomes.

III.I.I Method of data Collection

The method of data collection employed for the data collection was documentation. Document analysis involves analysis of content from written documents in order to make certain deductions based on the study parameters. The areas of study are 2 hospitals drawn from South West Nigeria. The dataset was gathered from their diabetic patient's records. Data collected through documentation were represented in tables in excel spreadsheet.

III.I.II Collection of initial data

For the purpose of this project, data was collected from the records of two hospitals in Ondo State, Nigeria. The first hospital is First Mercy Hospital Akure, Ondo State and the second is State Specialist Hospital, Ondo town, Ondo State

with parameters such as age, sex, weight, route of drug administration (RODA), dose frequency (DF), risk of heart disease (ROHD), response to drugs (RTD), blood pressure control (BC), risk of hypoglyceria (ROH), risk of obesity (ROO) and lifestyle changes (LC).

III.II PRE-PROCESSING DATA.

Since the target data has been selected, the data is then pre-processed for cleaning, scrubbing, and transforming to improve the effectiveness of discovery. During this pre-processing step, noise or outliers were removed.

III.II.I Data Integration

Data integration is the process of integrating data from multiple sources and probably have a single view over all these sources and answering queries using the combined information. In this stage, the data gathered from the two hospitals were be combined physically into one spreadsheet.

III.II.II Data Cleaning

This involves filling in missing values, smoothening noisy data, identifying or removing outliers and noisy data, and resolve inconsistencies. Data cleaning is the number one problem in data analysis.

The following data cleaning tasks were performed on the data includes;

1. Fill in missing values: Data is not always available e.g many tuples have no recorded values for several attributes. Missing data may be due to, equipment malfunction, inconsistent with other recorded data and thus deleted, data not entered due to misunderstanding, certain data may not be considered important at the time of entry, no register history or changes of the data and expansion of data schema
2. Case Transformation: In this step, all characters of text are converted to the same form and numbers are transformed to the same format.
3.

III.III FEATURE EXTRACTION

The word feature in the technique are combination of attributes in the data that have special important characteristics of the data. Feature extraction is mostly applicable to latent semantic analysis, data compression, data decomposition and projection, and pattern recognition, etc. Using feature extraction process the speed and effectiveness of supervised learning can also be improved [15].

Feature selection is a process that selects a subset of original features by rejecting irrelevant and/ or redundant features according to certain criteria. Relevancy of features is typically measured by discriminating ability of a feature to enhance predictive accuracy of classifier and cluster goodness for clustering algorithm.

The approach that was used in this project is X^2 statistical test. This approach Feature Chi - square testing is based on Pearson's X^2 (chi square) tests. The Chi square test of independence helps to find out the variables X and Y are related to or independent of each other. In feature selection, the Chi - square test measures the independence of a feature and a category. The null-hypothesis here is that the feature and category are completely independent [16].

Chi-square statistical test can be expressed as thus;

$$X^2 = \sum \frac{(O - E)^2}{E} \qquad (1)$$

X^2 is the chi-square value

O is the observed frequency

E is the expected frequency

When calculating the relevancy of an input attribute to an output attribute, the observed frequency (O) is how frequent the attribute occurs and the expected frequency (E) is how frequent the attribute is expected to occur.

III.IV DATA EXPLORATION

III.IV.I Selection of Analytical Technique

The technique selected is the combination of a clustering technique (k- means) and association rules (A-priori algorithm).

K-means clustering is an exclusive clustering algorithm. Each object is assigned to precisely one of a set of clusters. For this method of clustering we start by deciding how many clusters we would like to generate from our data. We call this value k. The value of k is generally a small integer, such as 2, 3, 4 or 5, but may be larger.

The algorithm is composed of the following steps:

1. Place K points into the space characterized by the objects that are being clustered. These points represent initial group centroids and are defined in a canny (better placed far away from each other) way because different locations cause different results.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

This algorithm aims at minimizing an objective function known as squared error function given by:

$$J = \sum_{i=1}^{c_i} \sum_{j=1}^{c} \|x_i - v_j\|^2 \qquad (2)$$

where,

x_i is the itemset of an attribute.

v_j is the centroid of the itemset.

$\|x_i - v_j\|$ is the Euclidean distance between x_i and v_j.

C_i is the number of data points in i^{th} cluster.

In the first phase of implementation, K-means clustering was used to classify patients into clusters with shared characteristics. In the second phase, the employment of mining association rules was be used to create patients profile in each cluster. The purpose of association rule extraction is to discover significant relationships between features/attributes that occur frequently in the database.

A-priori algorithm is one of most successful algorithm has been proposed for mining association rules in a database. Rule candidates are considered useful and become association rules only if their support is larger than a minimum support (min supp) threshold and whose confidence is larger than a minimum confidence (min conf) threshold. There are two major ways to measure association within clusters; Support and Confidence.

Support says how popular an item is as measured by the number of times in which an itemset appears. It is thus expressed;

$$s(x \rightarrow y) = \frac{\sigma(x \cup y)}{N} \qquad (3)$$

Where N is the total number of itemset and Ơ is the frequency of X and Y occurring together.

Confidence says how likely item Y appears when item X occurs. Its is thus expressed as;

$$c(x \rightarrow y) = \frac{\sigma(x \cup y)}{\sigma(y)} \qquad (4)$$

Where Ơ is the frequency of X and Y occurring together

III.IV.II Implementation of selected technique

The algorithms were implemented using python programming language in PyCharm Community Edition. PyCharm is an Integrated Development Environment (IDE) used in computer programming, specifically for the python language. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCS).

The three major python libraries that will be used for the analysis of the data includes: Numpy, Matplotlib and Pandas.

III.V INTERPRETATION AND EVALUATION

The patterns obtained in the data mining stage are converted into knowledge, which in turn is used to support decision making. The final step of knowledge discovery from data is to verify that the patterns produced by the data mining algorithms occur in the data set because not all patterns found by the data mining algorithms are necessarily valid. Accuracy was used as the method of evaluating the effectiveness of the analytical technique used.

IV. RESULTS

After preprocessing, feature extraction was carried out. The approach used in this project is the Chi-square (X^2) statistical test. In this stage, the chi square test algorithm was implemented in PyCharm Community Edition.

The result of the implementation is shown below:

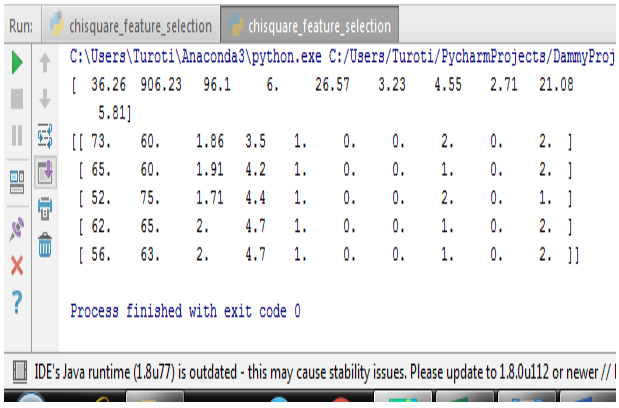


Figure 2: Result of X^2 statistical test

The result of the test shows the attributes in the dataset in order of their relevancy to the output variable with Age being the attribute with the highest relevancy, weight, blood pressure control, risk of hypoglyceria, sex, route of drug administration, dose frequency, risk of heart disease, response to drugs and risk of obesity.

IV.I IMPLEMENTATION OF TECHNIQUE

The implementation was done in PyCharm Community Edition 2017 (2.1 version) environment and compiled with Python Anaconda Interpreter.

IV.I.I Implementation of algorithms

K-means implementation:

K-means was employed to segment patients into three clusters (high, medium or low) based on their attributes. The number of clusters which is K was predetermined as three. Figure 3 shows the graphical representation of the data after k-means clustering. The figure shows cluster 0 as high level of diabetes patient's group, cluster 1 as medium level of diabetes patient's group and cluster 2 as low level of diabetes patient's group. The model test shows that the total percentage of patients that belong to cluster 0 is 13.1%, cluster 1 is 37.4% and cluster 2 is 49.5%. The results shows that most patient have low level of diabetes.

Figure 3: Final results of K-means clustering

A-priori rules implementation:

Once the clusters and the associated statistical summarized data are made by K-means algorithm. The patients fall into three groups. The a-priori algorithm is used to create and verify the patient's profiles. The variables chosen as predicate variables for association rule analysis were the four major variables age, weight, risk of hypoglyceria and blood pressure control.

Firstly, the four major input variables were set in ranges manually for patients clustering in the table below;

Table 2: Input variables for patients clustering

VARIABLE NAME	SEGMENT 1	SEGMENT 2	SEGMENT 3
Risk of hypoglyceria	>2 to <=5	>6 to <=9	Greater than 9
Age	>=17 to <35	>=35 to <=65	Greater than 65
Blood pressure	>=1 to <=1.5	>=1.6 to <=2	Greater than 2
Weight	>=0 to <=30	>=31 to <=80	Greater than 80

The next phase includes applying the a-priori algorithm to this input variable ranges to derive the support and confidence which was used to generate rules for each cluster where support is how often an attribute occurs in a database and confidence is how likely it is for an attribute to occur when another does.

The a-priori algorithm was applied to cluster 0 and the total number of rules generated for cluster 0 is 129, then the algorithm was set to filter out the best rules based on the minimum support which was selected as 30% and the minimum confidence was set as 50%.

Table 3: Best rules generated for cluster 0

antecedants	consequents	support	confidence
BP: 1 - 1.5	AGE: 35 - 65	0.393638171	0.545454545
AGE: 35 - 65	RISK: 2 - 5	0.491053678	0.615384615
AGE: 35 - 65	WEIGHT: 31 - 80	0.491053678	1
BP: 1.6 - 2	AGE: greater than 65	0.4055666	0.529411765
BP: 1 - 1.5	RISK: 2 - 5	0.393638171	0.585858586
BP: 1.6 - 2	WEIGHT: 31 - 80	0.4055666	0.862745098
RISK: 2 - 5	WEIGHT: 31 - 80	0.554671968	0.863799283
BP: 1 - 1.5, WEIGHT: 31 - 80	AGE: 35 - 65	0.330019881	0.65060241
BP: 1.6 - 2, WEIGHT: 31 - 80	AGE: 35 - 65	0.349900596	0.545454545
AGE: 35 - 65, RISK: 2 - 5	WEIGHT: 31 - 80	0.302186879	1
BP: 1 - 1.5, WEIGHT: 31 - 80	RISK: 2 - 5	0.330019881	0.602409639
BP: 1.6 - 2, WEIGHT: 31 - 80	RISK: 2 - 5	0.349900596	0.5625

Table 3 shows the best rules derived from cluster 0. Rule 1 states that if a patient's blood pressure is between 1 and 1.5 and also between the age of 35 and 65, the patient may likely belong to cluster 0 and have high level of diabetes.

IV.II Evaluation of technique

Using accuracy as the evaluation method for this research, the analytical technique was evaluated and benchmarked against other models reviewed in related literature. Figure 4 shows the accuracy of the techniques' estimates.

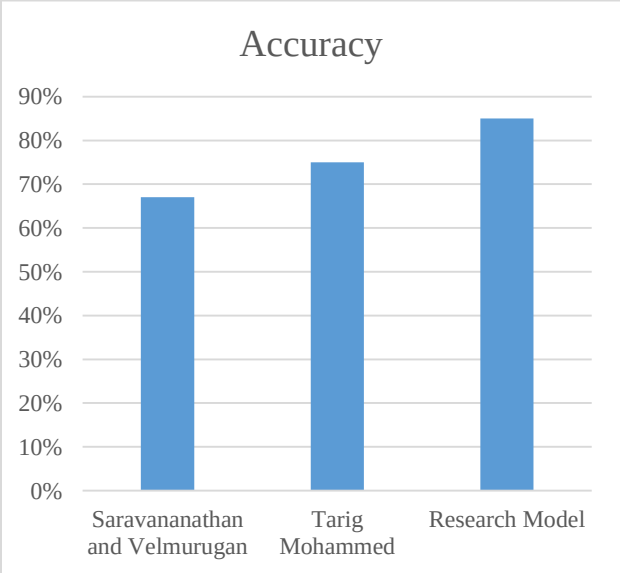


Figure 4: Results of technique's accuracy

V. CONCLUSION

This study proposes a five-stage framework of analysis using K-means clustering algorithm and an association rule inducer for analyzing a diabetes dataset. For differentiation purposes, the patients were grouped according to their shared behavior and characteristics by k-means algorithm into three clusters (high level diabetes, medium level diabetes and low level diabetes).

After the classification of patients into groups, the patient's profiles for each cluster were created using the a-priori algorithm. Medical practitioners then can infer the profiles of patients in each group and propose treatment strategies appropriate to the each group. This study can help the hospitals to create an effective and better treatment plan for new patients. For future work, more reports can be fed to improve accuracy of the algorithms as machine-learning algorithms learn more when fed with more examples.

REFERENCES

[1] M. Fernanades. '*Data Mining: A Comparative Study of Its Various Techniques and Its Process'* International Journal of Scientific Research in Computer Science and Engineering (IJSRCSE), Volume 5, Issue 1, pp. 19-23, 2017.

[2] M. Ramageri. *"Data Mining Techniques and Applications",* Indian Journal of Computer Science and Engineering (IJCSE), Volume 1, Issue 4, pp. 301-305, 2010.

[3] N. Gbuse, P. Pawar and A. Potgantwar. "*An Improved Approach for Fraud Detection in Health Insurance Using Data Mining Techniques"* International Journal of Scientific Research in Network Security and Communication (IJSRNSC), Volume 5, Issue 5, pp. 27-32. 2017.

[4] L. Guo. *"Applying Data Mining Techniques in Property Casualty Insurance",* Casualty Actuarial Society Forum, Casualty Actuarial Society, pp. 1-25, 2003.

[5] A. Aljumah, M. G. Ahamad and M. K. Siddiqui "*Application of data mining: Diabetes health care in young and old patients*", Journal of King Saud University – Computer and Information Sciences, Volume 2, Issue 5, pp. 127–136, 2013.

[6] S. Nagarajan and R. M. Chandrasekaran. *"Design and Implementation of Expert Clinical System for Diagnosing Diabetes using Data Mining Techniques",* Indian Journal of Science and Technology (IJST), Volume 8, Issue 8, pp. 771–776, 2015.

[7] P. Padmaja, V. , Nilofer I. S., Praveen D., A Bikkina., R. Venkata, M. V. Shaik and Raju R. *"Characteristic evaluation of diabetes data using clustering techniques',* International Journal of Computer Science and Network Security (IJCSNS), Volume 8, Issue 11, pp. 244-251, 2008.

[8] Pramanand P. and Sankaranarayanan S. *"Diabetic prognosis through Data Mining Methods and Techniques",* International Conference on Intelligent Computing Applications, Volume 2, Issue 2, pp. 162-166, 2014.

[9] T. Pala and I. Yucedag. *"A Data Mining Approach for Diagnosis of Diabetes Using Association Rules and Clustering",* International Artificial Intelligence and Data Processing Symposium. Volume 2, Issue 2, pp. 187-199, 2016.

[10] P. S. Kumar and V. Umatejaswi *"Diagnosing Diabetes using Data Mining Techniques",* International Journal of Scientific and Research Publications, Volume 7, Issue 6, pp. 705-709, 2017.

[11] A. Iyer, S. Jeyalatha and R Sumbaly. *"Diagnosis of Diabetes using Classification Mining Techniques",* International Journal of Data Mining & Knowledge Management Process (IJDKP) Volume 5, Issue 1, pp. 1-14, 2015.

[12] B. Vani and J. Priyadharshni. *"Discovering the Diagnosis of Diabetes Mellitus by using Association Rule Mining",* International Journal of Research Instinct, Volume 3, Issue 2, pp. 87-95, 2016.

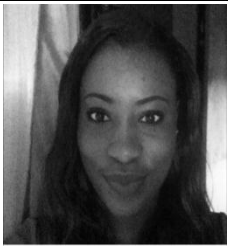
[13] A, Al-Rofiyee, M. Al-Nowiser, N. Al-Mufadi and M. A. ALHagery. *"Using Prediction Methods in Data Mining for Diabetes Diagnosis",* Third Symposium on Data Mining Applications. Volume 3, pp: 5-6, 2014.

[14] S. G. U Tugba. *"Defining Characteristics of Diabetic Patients by Using Data Mining Tools",* Journal of Hospital & Medical Management. Volume 2, Issue 2, pp. 1-8, 2016.

[15] P. P. Sondwale. *"Overview of Predictive and Descriptive Data Mining Techniques",* International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSE), Volume 5, Issue 4, pp. 262-265, 2015.

[16] M. Ramya and A. J Pinakas. *"Different Type of Feature Selection for Text Classification",* International Journal of Computer Trends and Technology (IJCTT), Volume 10, Issue 2, pp. 102-107, 2014.

Author's Profile



Miss Turoti Oluwadamilola, received the Bachelor of Science Degree in Computer Electronics from Lead City Univerisity, Ibadan, Nigeria in 2013. She is currently pursuing her Master of Technology from Federal University of Technology, Akure, Nigeria. Her main research work focuses on Data Mining.

Dr Olumide O. Obe, obtained Bachelor of Technology degree in Computer Science from Federal University of Technology, Akure, Nigeria in 2000. A Master of Science degree in Computer Science from the University of Lagos, Akoka, Nigeria in 2002 and a Ph,D degree in System Engineering from the Politehnica University of Bucharest, Romania in 2010. He was a Teaching and Research Assistant at University of Lagos between 2001 and 2003. He has been a lecturer at the Computer Science Department of Federal University of Technology, Akure since 2004. He was an adjunct lecturer to a number of Universities in Nigeria. His research interest area includes: Artificial Intelligence, hybrid adaptive intelligent control systems, Robotics and Automation, Mobile Applications, Security and Big Data. He has co-authored a book and has authored/co-aothored more than 30 articles in refereed journals and conference proceedings. He is a regular reviewer in local as well as international scientific journals. Dr. Obe is a member of the IEEE, the ACM and the IAENG, UK. He was a scholar of the Romanian Ministry of Education. He is a fellow of the Massachusetts Institutes of Technology, USA.

