Minor Research Project - Final Report
(July 2017 – June 2019)

DEVELOPING PREDICTIVE MODELS TO FORECAST AND PRE - TREAT DIABETES MELLITUS IN TAMIL NADU USING CLINICAL BIG DATA

(UGC Ref. No.: F. MRP-6517/16 (SERO/UGC) June - 2017)

(Link No: 6517 Comcode: TNBD007 UniqueID: RHTSJC)

Submitted to

UNIVERSITY GRANTS COMMISSION

South Eastern Regional Office (SERO), Hyderabad – 500 001

Principal Investigator

Dr. L. AROCKIAM, MCA., M.Tech., MBA., M.Phil., Ph.D.,

Associate Professor in Computer Science

PG & RESEARCH DEPARTMENT OF COMPUTER SCIENCE

ST. JOSEPH'S COLLEGE (AUTONOMOUS)

Special Heritage Status Awarded by UGC

Nationally Accredited at 'A++' Grade (4th Cycle) by NAAC

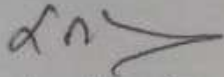College with Potential for Excellence by UGC

DBT-STAR & DST-FIST Sponsored College

Tiruchirappalli – 620 002

UNIVERSITY GRANTS COMMISSION
SOUTH EASTERN REGIONAL OFFICE (SERO)
A.P.S.F.C. Building (4ᵗʰ Floor), 5-9-194, P.B. No. 152, Chirag-Ali-Lane
HYDERABAD– 500 001. ,

Utilization Certificate

Certified that the grant of **Rs. 1,90,000 (Rupees One lakh ninety thousand** only) received from the University Grants Commission under the scheme of support for Minor Research Project entitled **"Developing Predictive Models to Forecast and Pre - Treat Diabetes Mellitus in Tamil Nadu using Clinical Big Data"** vide UGC letter no. **F.MRP-6517/16 (SERO-UGC)** dated **June 2017** has been fully utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grants Commission.

(Dr. L. Arockiam)

SIGNATURE OF THE
PRINCIPAL INVESTIGATOR

Dr. L. AROCKIAM
MCA.,M.Tech.,MBA.,CSM.,BLIS.,M.Phil.,Ph.D.,
Associate Professor in Computer Science
St.Joseph's College (Autonomous)
Tiruchirappalli- 620 002,Tamilnadu,India

(Rev. Dr. M. Arockiasamy Xavier SJ)

PRINCIPAL

PRINCIPAL
St. JOSEPH'S COLLEGE
(AUTONOMOUS)
TIRUCHIRAPPALLI 620 002

STATUTORY AUDITOR

ROY JOHN THOMAS, B.COM.,F.C.A.,
CHARTERED ACCOUNTANT
M. No: 200 / 25188

D-4, BISHOP COMPLEX
13A, CONVENT ROAD
MELAPUDUR,
TRICHY - 620 001.

Appendix - A1

FINAL REPORT OF THE WORK DONE

Title of the Project: Developing Predictive Models to Forecast and Pre-Treat Diabetes Mellitus in Tamil Nadu using Clinical Big Data

(UGC Reference No. : **F. MRP-6517/16 (SERO/UGC) June-2017**)

1. Introduction:

Diabetes mellitus is one of the non-communicable diseases that pose a threat to human health. It has become a major global health problem. It is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin which it produces. It is found that diabetes causes blindness, amputation and kidney failure. Lack of awareness about diabetes, insufficient access to health services and essential medicines can lead to the above mentioned complications. According to a study by the World Health Organization (WHO), number of diabetic patients will raise to 552 million by 2030, which means that one in 10 adults will have diabetes by 2030. In 2014, the global prevalence of diabetes was estimated to be 9 % among adults aged 18+ years [1]. WHO insisted with an alarm that Diabetes is the 7th leading cause of death in the world. In 2012, an estimated 1.5 million deaths were directly caused by diabetes. Total deaths due to diabetes are projected to rise by more than 50 % in the next 10 years.

Moreover, the International Diabetes Federation said that nearly 52 % of Indians are not aware that they are suffering from high blood sugar. More than 62 million diabetic individuals are currently diagnosed with the disease. It is predicted that, by 2030 diabetes mellitus may affect up to 79.4 million individuals in India. A nation-wide study, conducted by the Indian Council of Medical Research`s INDIAB (India Diabetes) has confirmed that one out of 10 people in Tamil Nadu is affected by diabetes, and every two persons with age group of 25 are in the pre-diabetic stage. It is stated that 14.8 per cent of urban population and 11 per cent of rural population of Tamil Nadu are affected by diabetes. Madras Diabetes Research Foundation suggested that about 42 lakh individuals have diabetes and 30 lakh people are in pre-diabetes stage. At least, 1,000 people avail treatment for diabetes out of the 12,000 outpatients who visit Rajiv Gandhi Government General Hospital (RGGGH), a leading Government hospital in Chennai.

2. Review of the Literature

Raghupathi et al. [2] had done a review on big data analytics in healthcare. The promises and potentials of big data in healthcare were pointed out and an architectural framework and methodology for applying big data in healthcare were outlined. The advantages of using big data analytics in healthcare were elaborated and the available platforms and tools for applying big data analytics in healthcare were presented. Aiswarya Iyer et al. [3] used Decision Tree and Naïve Bayes algorithms for the prediction of diabetes in pregnant women. Tenfold cross validation was done to prepare training and test data and the J48 algorithm was employed on the dataset using WEKA on the Pima Indians Diabetes Database of National Institute of Diabetes and Digestive and Kidney Diseases. The authors concluded that both algorithms were efficient for the diagnosis of diabetic and Naïve Bayes technique resulted in least error rate.

A.A. Aljumah et al. [4] suggested a predictive analysis of diabetic treatment using a regression based data mining technique. Oracle Data Miner (ODM) tool was deployed for predicting diabetics and support vector machine algorithm was employed for experimental analysis on Datasets of Non Communicable Diseases (NCD) risk factors in Saudi Arabia. Mohammed et al. [5] presented a review of existing applications of the Map Reduce programming framework and its implementation platform Hadoop in clinical big data and related medical health informatics.

N.M. Saravana Kumar et al. [6] presented Predictive Analysis System Architecture with various stages of data mining. Prediction was carried out in Hadoop / Map Reduce environment. Predictive Pattern matching system was deployed to compare the analyzed threshold value with the obtained value. Saumya et al. [7] applied analytical techniques to reduce the hospital readmission of diabetic patients. In the proposed methodology, Hive was used as the preprocessing tool and R Hadoop as the analytics and predictive modeling tool. Classification was done using Logistic Regression, SVM, KNN and Decision Tree methods. Miss-classification error rates were also calculated.

D. Peter Augustine [8] presented a concept paper on analyzing the data flowing from health monitoring devices. The present status of healthcare in India was presented. The application of Hadoop's map reduce in healthcare data was expounded. An interface HIPI (Hadoop Image Processing Interface) in Hadoop environment was also explicated. Sadhana et al. [9] analyzed the Pima Indians Diabetes Database of National Institute of Diabetes and Digestive and Kidney Diseases data set using a proposed architecture which comprised of Hive and R. The raw data (CSV file) was given as input to

the Hive for analyzing and partitioning. The output file was passed to the through R system for statistical computing.

MuniKumar N et al. [10] has pointed out the massive shortage of proper healthcare facilities and addressed the concern to provide greater access to primary healthcare services in rural India. The ability of Big Data analytics in processing huge volumes of data in real-time situations to turn the dream of Swachh Bharath (Clean &Healthy India) into reality was explicated. AditiBansal et al. [11] proposed an architecture consisting of Dynamic Hadoop Slot Allocation (DHSA) which used the slot based resource model. Two more alternatives for DHSA namely, Pool Independent DHSA (PIDHSA) and Pool dependent DHSA (PDDHSA) were also presented. It was found that DHSA focused on the maximum utilization of slots by allocating map (or reduce) slots to map and reduce tasks dynamically. K. Sharmila et al. [12] presented a survey paper on the advancement in the field of data mining, the latest adoption of Hadoop platform, deployment of big data algorithms and consequently the open challenges in the Indian medicinal data set.

3. Objectives

The predominant aim of this project is to propose novel predictive model to predict diabetes mellitus using the clinical and e-diabetic Big Data. The objectives of the proposed work are formulated as below:

- ➢ To create an e-diabetic portal
- ➢ To build data warehouse using cloud computing technology
- ➢ To apply Big Data analytics to derive patterns
- ➢ To predict diabetes using the generated patterns

4. Methodology:

In this modern era, human beings encounter different health issues. Most of the health issues are due to the food habits of the individuals. In this project work, a predictive approach is proposed to pre-treat Diabetic Mellitus. The proposed approach has three phases namely data collection, data storage and analytics. This approach plays an important role in predicting diabetes and pre-treating diabetic patients. The phases in the proposed approach for diabetic prediction are presented in Fig. 1.
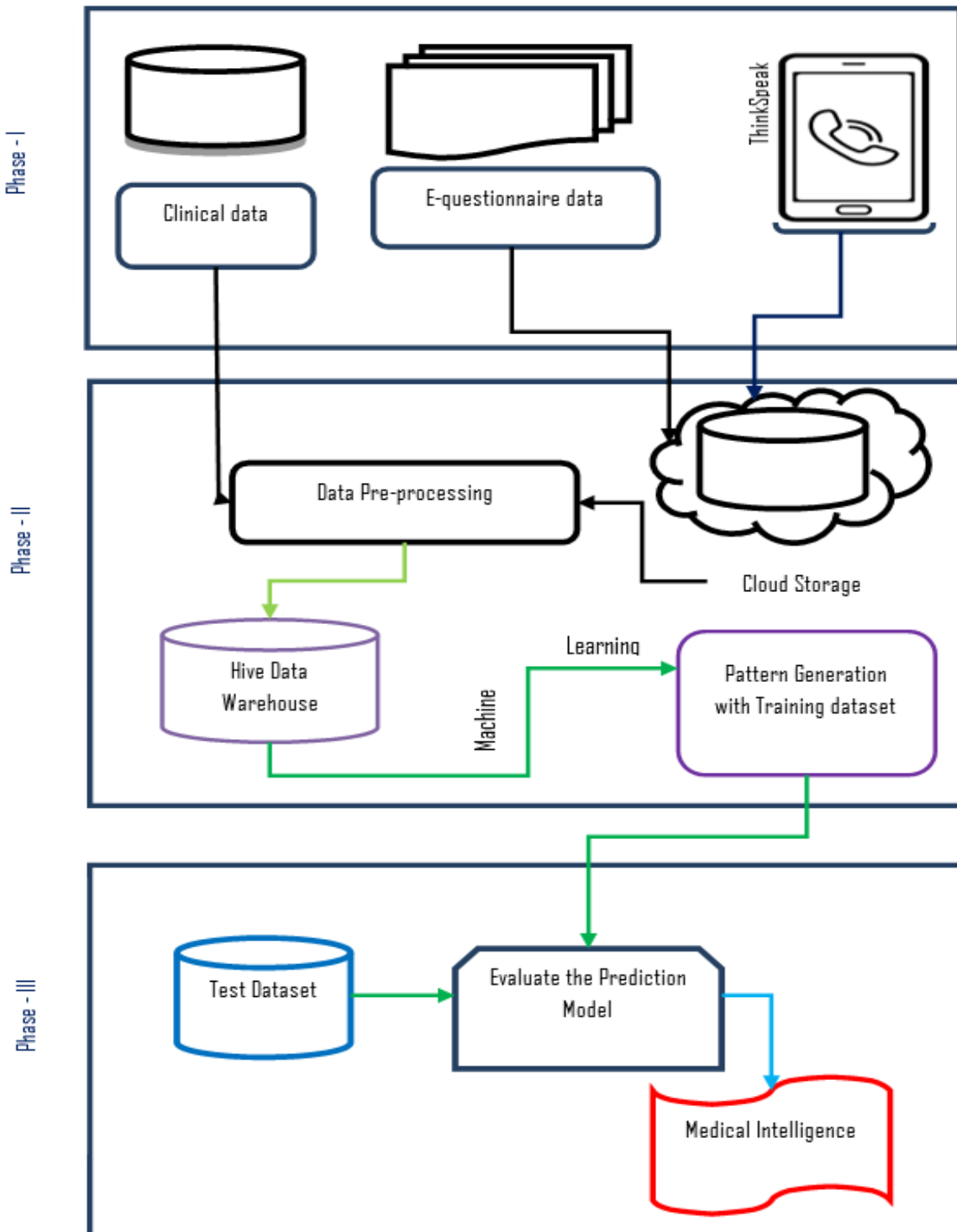
Fig 1. Methodological Diagram for Predicting Diabetes Mellitus

In the first phase, data collection is done through IoT devices and other sources. The collected data are cleansed using pre-processing techniques.

Phase two deals with data storage. The pre-processed data are stored in warehouses. To store massive amount of data, cloud storage is used. The data stored in the cloud are analyzed to establish association between the various parameters such as BP, BMI, Air Pollution level etc., with Diabetic Mellitus.

The third phase of the proposed approach deals with Predictive Analytics where the decisions are taken based on association rules with respect to diet pattern, physical fitness, current medicine intake etc.

5. Work Completed:

The project Titled "Developing Predictive Models to Forecast and Pre-Treat Diabetes Mellitus in Tamil Nadu using Clinical Big Data" aims to propose novel predictive model to predict diabetes mellitus. It is achieved through three phases viz. phase-1, phase-2 and phase-3. In the first phase, data collection is carried out in the state of Tamil Nadu. For the purpose of data collection, the whole region was divided into four regions viz. Chennai City Region, Central Region, Western Region, and Southern Region. The study was carried out randomly in urban and rural places in the districts. A standardized questionnaire, including items of common risk factors of diabetes was sent to the participants to obtain information on demographic characteristics, family diabetes history, and lifestyle risk factors. It contains a set of questions under three categories namely Category A, B and C. Category A (Demographic) contains personal information including Name, Age, Sex, and Address. Category B contains details about family that include Diabetic history of parents, financial status, and educational status. Category C includes the physical data such as Height, weight, smoking habit, consumption of Alcohol, work type and physical activity. The study has included 15344 participants of both male and female of age above 18 years.

In the second phase, the collected data have been cleansed and stored in the cloud storage. DataPreparator tool is used for data Preparation. It is a software tool designed to perform common tasks of data preprocessing in data analysis. The cleaned data set named tnset926 has been obtained after performing removal, reduction and integration. Diabetic Data Warehouse has been created using SQL Server Management Studio (SSMS) which holds 14745 item set after preprocessing process. A model has been created to form OLAP cube with the dimension name Diabetics. It is a master table consisting of non-overlapping data elements. Transact SQL (T-SQL) is used in SSMS for creating a

star schema relational dimension with study id as a primary key in the fact table. Thus the non-overlapped diabetic data warehouse has been created.

Model building for type-2 diabetic mellitus has been carried out in the third phase. An account has been created in BigML web interface for model building and model evaluation. It provides various Machine Learning algorithms as Software as a Service (SaaS). Discretized tnset926 data base with 14745 has been uploaded as input source. The Filteredtop-K Technique is used for association discovery between the variables. Based on the result model with maximum confidence level is chosen for creating the E- Diabetic portal.

5.1 Sample Selection for Data Collection

State of Tamil Nadu has been chosen for the purpose of data collection. It is divided into four regions viz Chennai City Region, Central Region, Western Region, and Southern Region for data collection. The Central Region includes 10 districts (Cuddalore, Thanjavur, Perambalur, Tiruchirappalli, Ariyalur, Karur, Nagapattinam, Thiruvarur, Pudukottai, and Karaikal). The western region consists of 6 districts namely Coimbatore, Erode, Nammakal, Salem, Dharmapuri and Nilgiris. The Districts Dindigul, Madurai, Theni, Sivaganga, Virudunagar, Ramanathapuram, Tirunelveli, Thoothukudi and Kanyakumari are included in the southern region. Finally, Chennai, Thiruvalluvar, Kancheepuram, Vellore and Tiruvannamalai districts are grouped under the Chennai city region.

5.2 Phase I: Data Collection (1 – 6 Months)

(a) Collection of clinical data from Medical Laboratories and storing it in raw Data Store (1–3 Months)

The data set selection is one of the important processes in data mining. For this, most relevant data should be collected from a particular domain to attain values. The derived values should be more flexible and informative in that domain. In this study, the data collection is done through questionnaire, sensors and from the clinical experiments.

The data are collected using pre-tested questionnaire, medical devices and sensors. The survey included participants above 18 years of age who belong to Tamilnadu.

(b) Collection of e-questionnaire data from users through online survey (1–3 Months)

A standardized questionnaire, including items of common risk factors of diabetes, has been sent to the participants to obtain information on demographic characteristics, family diabetes history, and lifestyle risk factors. It contains a set of questions under three categories.

Category A (Demographic): It contains **personal information** including viz. Name, Age, Sex, and Address.

Category B: It contains details about **family such as** Diabetic history of parents, financial status, and educational status

Category C: It includes the **physical data** such as Height, weight, smoking habit, consumption of Alcohol, work type and physical activity.

Anthropometric measurements have been taken from participants. Body Mass Index (BMI) has been calculated (weight in kilograms divided by the square of height in meters (kg/m2)). BMI ≥25 is defined as overweight. In category C, the data related to the following physical activities have been collected: cigarette smoking at least 500 cigarettes in one's life, Alcohol consumption of at least 100 g per week for 1 year or longer and participation in moderate or vigorous physical activity for 30 minutes or more per day for at least 3 days in a week.

c) Collection of IoT data from users through ThingSpeak IoT cloud (4–6 Months)

IoT data are accumulated from the devices that are connected to the Internet. MQ-3, MQ-7 and Honeywell HPm Particle Sensor (represented in fig 3.) are used to collect air quality data. These sensors are connected to ThingSpeak cloud storage through NodeMCU board. ThingSpeak enables sensors, instruments, and websites to send data to the cloud and it is stored in a private channel. The platform is based on a MSP430 16-bit CPU running at 3.9 MHz. It provides a CC2420 radio chip, 48 kB of program flash and 10 kB of RAM. One Tmote Sky implements the 6LoWPAN border router connected to a computer running on Windows OS. The 802.15.4 radio is configured to be channel 15, which underlies WiFi interference. NodeMCU ESP8266 is a microcontroller with WiFi functionality that is used as a bridge between sensors and cloud. The readings of the sensors are stored in this microcontroller temporarily and sent to the cloud. In this data collection process, ThingSpeak web API is used for storing data into the cloud. The ThingSpeakIoT analytics platform service allows to aggregating, visualizing and analyzing live data streams in the cloud.

The data related to blood pressure (diastolic and systolic) and blood sugar along with their screening data, the Body Mass Index (BMI), dietary history, physical activity, Air pollution level(Pm2.5& Pm10) are collected using this system. The selected populations are also included in the blood tests for collection of Random Blood Sugar level. The blood pressure is screened using Arm Bp digital monitor. The data relating dietary history and physical activity are collected through the questionnaire.

Phase II: Data Storage and Analytics (7–18 Months)

Pre-Processing and Diabetic Data Warehouse creation (10 – 12 Months)

Pre-processing of large volume of clinical data becomes essential. Data pre-processing is an important step in the data mining process. Observing the data which has not been carefully examined for such issues can produce misleading outcomes. If there is an ample amount of incorrect and redundant information or noisy and unreliable data, then knowledge discovery becomes challenging. Data pre-processing includes various steps such as cleaning, normalization, transformation, feature extraction and selection. The outcome of data pre-processing is the complete data set with reduced attributes. The major drawback with clinical data set is the existence of redundant records. These redundant records cause the learning algorithm to be biased. So, eliminating redundant records is essential to enhance the detection accuracy. During pre-processing, the raw data is supplied as input and several suitable data pre-processing methods are applied thereby decreasing the invalid instances in the dataset. In our diabetes dataset, data transformation and data validation are considered as the important pre-processing techniques for eliminating impure and invalid data. The preprocessed data with zero missing value and zero error is shown in **fig 2.**
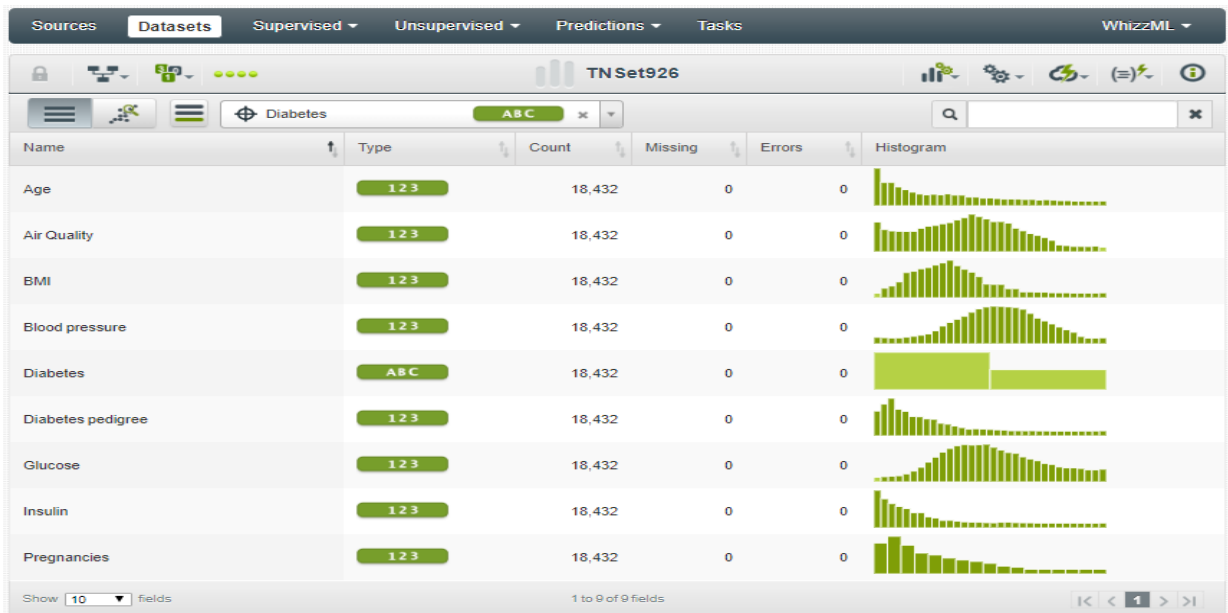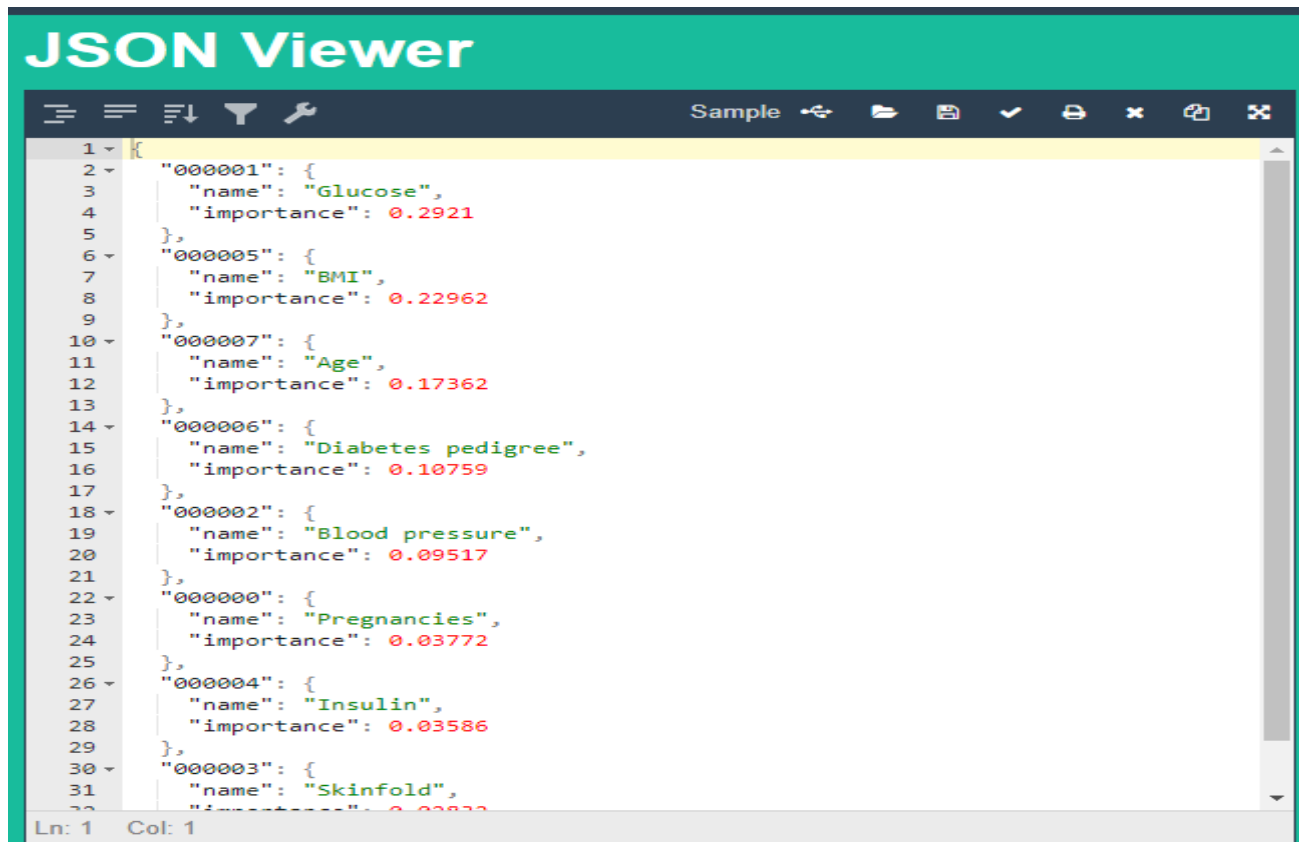

Fig 2: Pre-processed data sorted in alphabetical order.

The Cleaned dataset is named as TNSet926 to form a Data warehouse in .JSON format as shown in figure 3.

Fig 3: TNSet926 Data warehouse

Model Construction and Validation

The preprocessed data is now deployed in the analysis to build prediction model. In this project work, three type –II Diabetes predicative models are developed using BigML tool. BigML provides machine learning algorithms as Software as a Services (SaaS). It can be accessed through three models namely web interface, command line and Restful API. The Model is developed using the available web interface. The model creation process using BigML is shown in fig 4.

Fig 4: Model Creation Process in BigML

In standard statistical analysis, log_lenier is complex due to exponential number of variables. Association discovery approaches are very much suitable for high dimensional data. The relationship between the values is discovered using association discovery rather than the relation between the variables. FilteredTop_K Technique is used for association discovery. The created model is named as Tamil Nadu Diabetes Diagnosis Model I which is shown in **fig 5**. The association discovered using the developed model is shown in the fig 6 a, b and c.
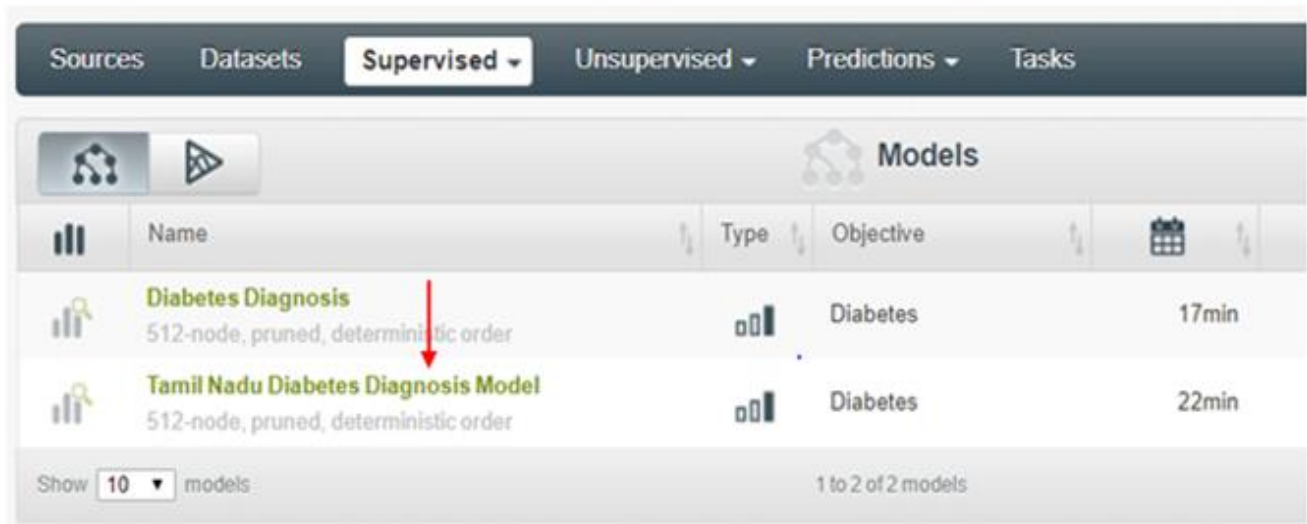

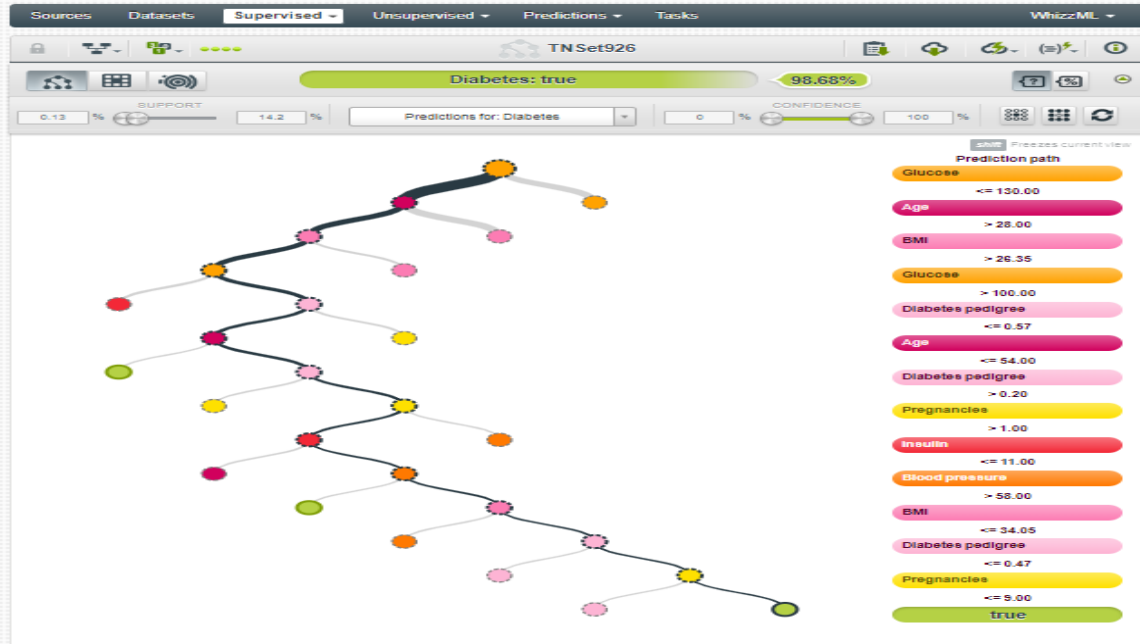
Fig 5: Tamil Nadu Diabetes Diagnosis Model

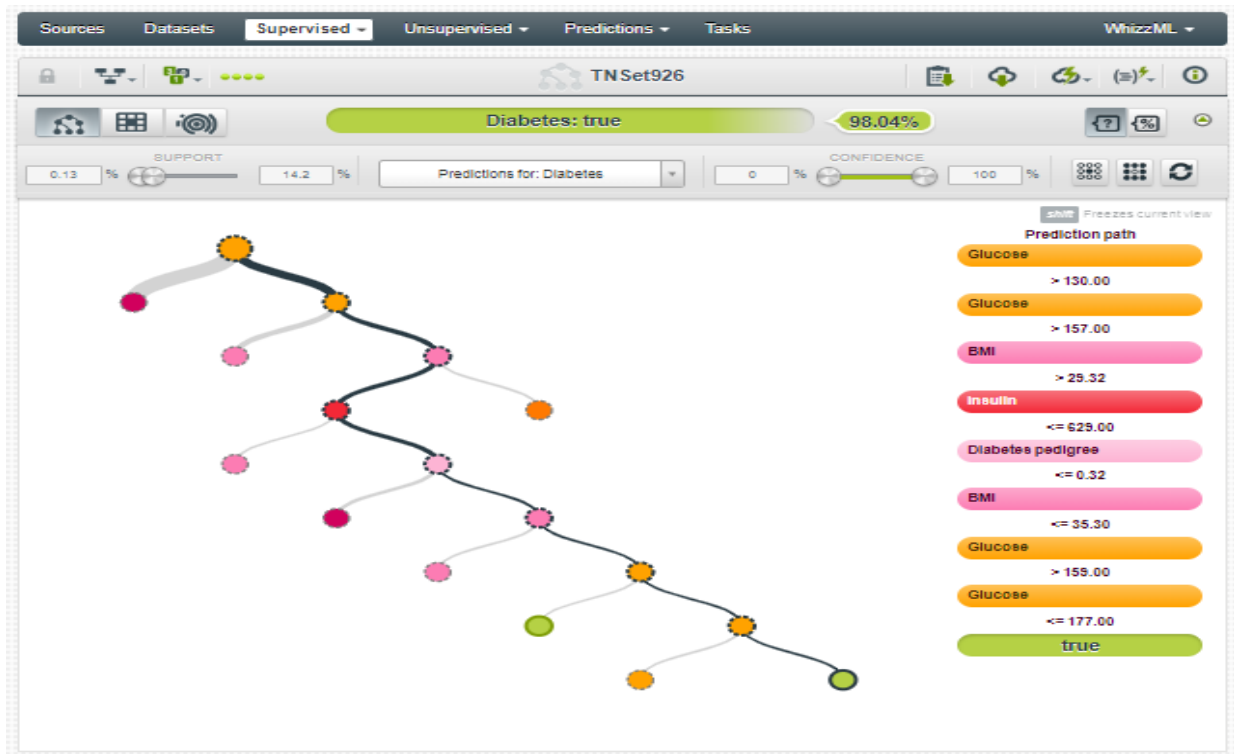**Fig 6a: Association between variables with confidence of 98.68%**

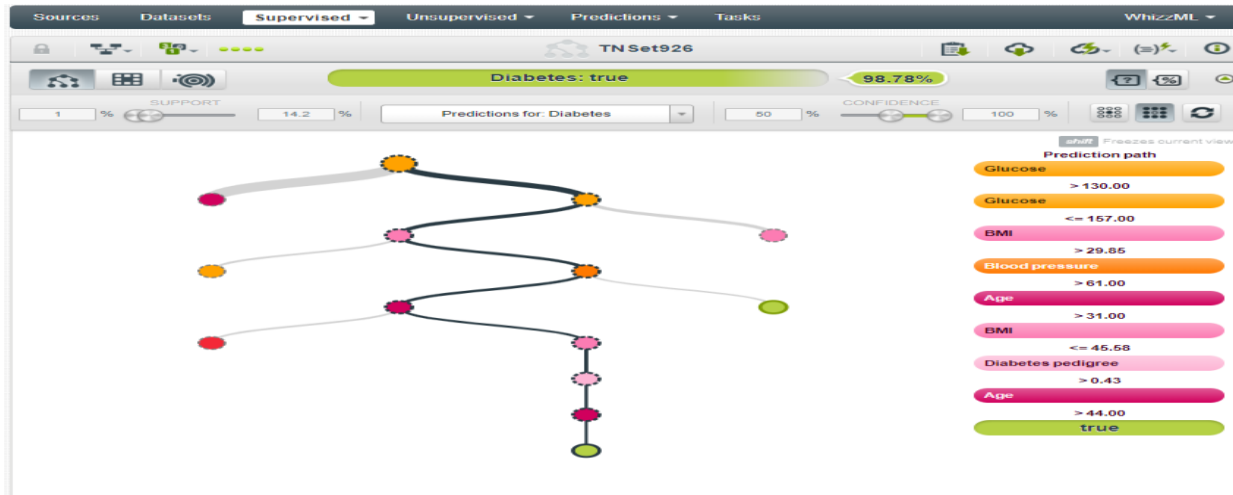**Fig 6b: Association between variables with confidence of 98.04%**

Fig 6c: Association between variables with confidence of 98.78%

The maximum number of associations that can be obtained from variables in BigML is 500. It ranges between 0- 500. The obtained associations are sorted from maximum to minimum confidence value. It is observed that the potential association increases exponentially when there is large number of fields, values and instances. The associations are discovered for the item in the antecedent ranges from 0 to 10. Leverage and lift are the two commonly preferred search strategies. Leverage produces the result with most relevant case whereas Lift produces least relevant case. The Association graph shown in figure 7 is for TNSet926 data. The number of instances is 14745, the number of item fields is 27, the number of associations discovered is 100, the search strategy chosen is leverage, the percentage set for the minimum support is 4.2970%, and the percentage set for the minimum confidence is 23.8000%. Actionable Association for Dataset TNSet926 and JavaScript Object Notation Predictive Markup Language (JSON- PML) is attached in Appendix-B1.
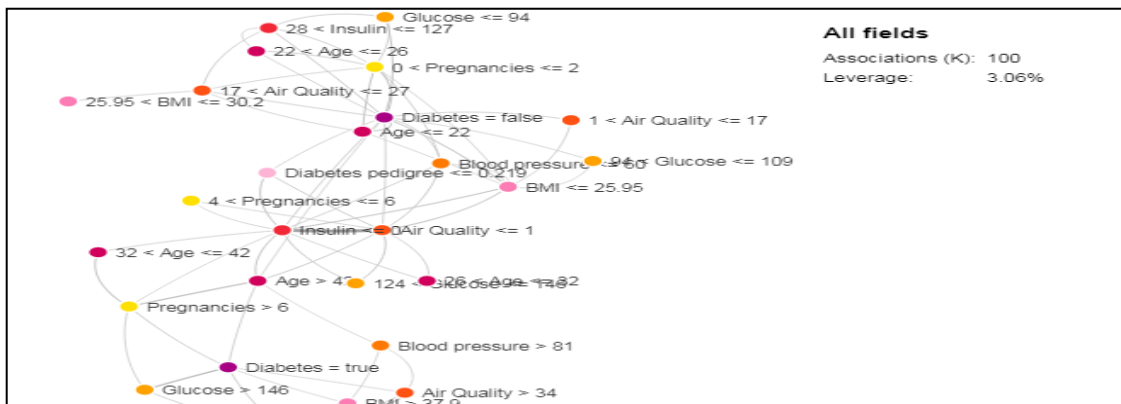

Fig 7: Association graph for Dataset TNSet926

6. Conclusion and Further Work to be done:

Diabetes Mellitus is the most severe chronic ailment that can seriously impact the quality of living of the affected persons. In this research work, a novel approach is proposed to predict the diabetes mellitus at an early stage using the generated patterns obtained through the deployment of big data analytics techniques. The proposed model comprises of three well defined layers with defined functionalities. Data collection has been done from the selected cities in Tamil Nadu through various methods viz. questionnaire, interviews and surveys ect. Diabetic Data Warehouse has been created using cloud computing technology after the rigorous preprocessing process. A dimensional model has been created to form OLAP cube with the dimension name Diabetics. The model with maximum confidence level is chosen for creating the E- Diabetic portal. An application has also been filed to get patent approval for the proposed work. In future, an IoT-based diabetes mellitus recommender model will also be built and deployed to predict disease risk factors and recommend lifestyle changes, diagnosis and alternative medicines.

7. IPR and Publication:

IPR Applied

1. L. Arockiam, A. Dalvin Vinoth kumar, "System to Predict Risk factor of Diabetes Mellitus using Internet of Things", Indian patents. (**Application Number : 201941046379**)

2. L. Arockiam, A. Dalvin Vinoth kumar, "E- Portal For Type -2 Diabetics Predication Using Internet of Things" Indian Copy Rights Under Computer Software.
   (Applied: **Diary No: 378/2020-CO/SW**)

Paper Published

1. L. Arockiam, A. Dalvin Vinoth kumar, "A Model to Predict and Pre-Treat Diabetes Mellitus", International Journal of Engineering and Techniques, ISSN: 2395-1303, pp. 1 – 6 **(Impact Factor: 3.546).**

2. L. Arockiam, A. Dalvin vinoth kumar and Sathyapriya, "Performance Analysis of Classification Algorithms For Diabetic Prediction Using Pima- Indian Dataset", International Journal of Emerging Technologies and Innovative Research, vol. 5, No. 12, pp. 563-569, ISSN: 2349-5162 **(Impact Factor: 5.87) (UGC Listed)**

3. Dr. L. Arockiam, S. Sathyapriya, V.A. Jane, A. Dalvin Vinoth Kumar, "Prevalence of Type-II Diabetics Association with PM 2.5 and PM 10 in Central Region of Tamil Nadu, India", International Journal of Recent Technology and Engineering, vol.8, No. 2, pp. 6440-6444, ISSN: 2277-3878. **(Scopus Indexed) (UGC CARE Listed) IF: 5.92.**

4. Dr. L. Arockiam, S. Sathyapriya, V.A. Jane, A. Dalvin Vinoth Kumar, "Prevalence of Diabetes Mellitus in Tiruchirappalli District using Machine Learning", IJRTE Blue Eyes Intelligence Engineering & Sciences Publication, vol.8, No. 2, pp. 6400-6403, ISSN: 2277-3878. **(Scopus Indexed) (UGC CARE Listed) IF: 5.92.**

- **Papers & Posters Presented**
  1. L. Arockiam, A. Dalvin Vinoth Kumar, "Predictive Model for Diabetes Mellitus in Tamil Nadu Using IoT Data" in Two Day International Conference on Transforming Technology held at Srimath Andavar Arts and Science College, Trichy, January 2018.

  2. L. Arockiam, A. Dalvin Vinoth Kumar, "A Model to predict and pre-treat Diabetes Mellitus" in International Conference on Advances in Computers Science and Technology held at St. Francis de Sales College, Bengaluru, January 2018.

  3. L. Arockiam, A. Dalvin Vinoth Kumar, "Swatchh Bharat - IoT Based Diabetic Predication", **Young Scientist conclave at Indian International Science Festival (IISF-2017) conducted by Ministry of Earth Sciences, Govt. of India**, October 13-16, 2018.

Principal Investigator

/Dr. L. AROCKIAM
MCA..M.Tech..MBA..CSM..BLIS..M.Phil.,Ph.D.
Associate Professor in Computer Science
St.Joseph's College (Autonomous)
Tiruchirappalli- 620 002, Tamilnadu, India

Principal
PRINCIPAL
St. JOSEPH'S COLLEGE
(AUTONOMOUS)
TIRUCHIRAPPALLI 620 002

UNIVERSITY GRANTS COMMISSION
SOUTH EASTERN REGIONAL OFFICE (SERO)
A.P.S.F.C. Building (4ᵗʰ Floor), 5-9-194, P.B. No. 152, Chirag-Ali-Lane
HYDERABAD– 500 001.

STATEMENT OF EXPENDITURE IN RESPECT OF MINOR

RESEARCH PROJECT

1. Name of Principal Investigator　: **Dr. L. Arockiam**
Associate Professor, Dean

2. Dept. of College　　　　　: Research and PG Dept. of Computer Science,

　St. Joseph's College (Autonomous),

　Tiruchirappalli – 620002

　TamilNadu, India.

3. UGC approval No. and Date　: **F.MRP-6517/16 (SERO/UGC)** dated **June 2017**

4. Title of the Research Project　: **Developing Predictive Models to Forecast and Pre-Treat Diabetes Mellitus in Tamil Nadu using Clinical Big Data.**

5. Effective date of starting the project: **1ˢᵗ July 2017**

6. a. Period of Expenditure　　: From **July 2017** to **June 2019**

　b. Details of Expenditure　　:

S.No.	Item	Amount Approved Rs.	Expenditure Incurred Rs.
i.	Books & Journals	50,000.00	50,528.00
ii.	Equipment *(please enclose the quotation)*	50,000.00	50,505.00
iii.	Contingency	45000.00	47,800.8
iv.	Field Work/Travel (Give details in the proforma at Annexure- VI)	45,000.00	47, 715.91
v.	Hiring Services	-NIL-	-NIL-
vi.	Chemicals & Glassware	-NIL-	-NIL-
vii.	Overhead	-NIL-	-NIL-
viii.	Any other Items (Please specify)	-NIL-	-NIL-
TOTAL		**1,90,000.00**	**1,96,549.71**

e. Staff Date of Appointment: -NIL-

S.No.	Expenditure Incurred	From	To	Amount Approved (Rs.)	Expenditure Incurred (Rs.)
1	Honorarium to PI (Retired Teachers) Rs.12,000/- p.m.	-	-	-	-
2	Post-Doctoral Fellow Fellowship @ Rs. 12,000/- p.m.	-	-	-	-
3	Project Associate salary @ Rs.10,000/- p.m.	-	-	-	-
4	Project Fellow salary @ Rs.8000/- p.m.	-	-	-	-

1. It is certified that the appointment(s) have been made in accordance with the terms and conditions laid down by the Commission.

2. It as a result of checks or audit objective, some irregularity is noticed, later date, action will be taken to refund, adjust or regularize the objected amounts.

3. Payment @ revised rates shall be made with arrears on the availability of additional funds.

4. It is certified that the grant of **Rs. 1,90,000 (Rupees One Lakh Ninety Thousand only)** received from the University Grants Commission under the scheme of support for Minor Research Project entitled "**Developing Predictive Models to Forecast and Pre-Treat Diabetes Mellitus in Tamil Nadu using Clinical Big Data**" vide UGC letter No. **F.MRP-6517/16 (SERO/UGC)** dated **June 2017** has been fully utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grants Commission.

(Dr. L. Arockiam)

SIGNATURE OF THE

PRINCIPAL INVESTIGATOR

Dr. L. AROCKIAM

MCA.,M.Tech.,MBA.,CSM.,BLIS.,M.Phil.,Ph.D.
Associate Professor in Computer Science
St.Joseph's College (Autonomous)
Tiruchirappalli- 620 002.Tamilnadu.India

(Rev. Dr. M. Arockiasamy Xavier SJ)

PRINCIPAL

PRINCIPAL
St. JOSEPH'S COLLEGE
(AUTONOMOUS)
TIRUCHIRAPPALLI 620 002