Impact Factor Value 4.046 **e-ISSN: 2456-3463**
International Journal of Innovations in Engineering and Science, Vol 4, No.10 ,2019
National Level Technical Paper Presentation- PHOENIX-19
Organized by Godavari College of Engineering, Jalgaon - 425003
www.ijies.net

Prediction and Severity Estimation of Diabetes Using Data Mining

Prof. Pramod B Gosavi[1], Gopal Yelve[2], Khushal Patil[3], Chetan Sonawane[4], Sandip Pawar[5]
[1]Asso. Prof. [2,3,4,5] B.E. students
[12345] Department Of Computer Engineering, GF's Godavari College Of Engineering, Jalgaon-425003, India,

Abstract: - *This paper proposes a novel temporal knowledge representation and learning framework to perform large-scale temporal signature mining of longitudinal heterogeneous event data. The framework enables the representation, extraction, and mining of high order latent event structure and relationships within single and multiple event sequences. The proposed knowledge representation maps the heterogeneous event sequences to a geometric image by encoding events as a structured spatial-temporal shape process. We present a doubly constrained convolutional sparse coding framework that learns interpretable and shift-invariant latent temporal event signatures. We show how to cope with the sparsity in the data as well as in the latent factor model by inducing a double sparsity constraint on the β-divergence to learn an over complete sparse latent factor model. A novel stochastic optimization scheme performs large-scale incremental learning of group-specific temporal event signatures. We validate the framework on synthetic data and on an electronic health record dataset.*

We have proposed clinical assessment for visual interactive knowledge discovery in large electronic health record databases.

.***Keywords-Prediction and severity estimation of diabetes.***

1. INTRODUCTION

Data mining can be defined as an activity that extracts some new nontrivial information contained in large databases. The goal is to discover hidden patterns, unexpected trends or other subtle relationships in the data using a combination of techniques from machine learning, statistics and database technologies. This new discipline today finds application in a wide and diverse range of business, scientific and engineering scenarios. For example, large databases of loan applications are available which record different kinds of personal and financial information about the applicants (along with their repayment histories). These databases can be mined for typical patterns leading to defaults which can help determine whether a future loan application must be accepted or rejected. Several terabytes of remote-sensing image data are gathered from satellites around the globe. Data mining can help reveal potential locations of some (as yet undetected) natural resources or assist in building early warning systems for ecological disasters like oil slicks etc. Other situations where data mining can be of use include analysis of medical records of hospitals in a town to predict, for example, potential outbreaks of infectious diseases, analysis of customer transactions for market research applications etc.

2. LITERATURE SURVEY

Early detection of diabetes is important for the prevention of diabetic complications. The best adiposity index for indicating Type 2 diabetes mellitus remains unclear. We aimed to identify the optimal adiposity measure among BMI, waist circumference, waist-hip ratio and waist-to-height ratio to indicate undiagnosed Type 2 diabetes and impaired fasting glucose in Chinese adults. [2].

2.1 Problem Statement

Finding latent temporal signatures is important in many domains as they encode temporal concepts such as event trends, episodes, cycles, and abnormalities. For example, in the medical domain latent event signatures facilitate decision support for patient diagnosis, prognosis, and management. In the surveillance domain temporal event signatures aid in detection of suspicious events at specific locations. Of particular interest is the temporal aspect of information hidden in event data that may be used to perform intelligent reasoning and inference about the latent relationships between event entities over time. An event entity can be a person, an object, or a location in time. For instance, in the medical domain a patient would be considered as an event entity, where visits to the doctor's office would be considered as events.

2.2 Proposed System: -

This paper proposes a novel Temporal Event Matrix Representation (TEMR) and learning framework to perform temporal signature mining for large-scale longitudinal and heterogeneous event data. Basically, our TEMR framework represents the event data as a spatial-temporal matrix, where one dimension of the matrix corresponds to the type of the events and the other dimension represents the time information. In this case, if event i happened at time j with value k, then the $(i,j)^{th}$ element of the matrix is k. This is a very flexible and intuitive framework for encoding the temporal knowledge

information contained in the event sequences. To improve the scalability of the proposed approach, we further developed an online updating technology. Finally, the effectiveness of the proposed algorithm is validated on a real-world healthcare dataset.

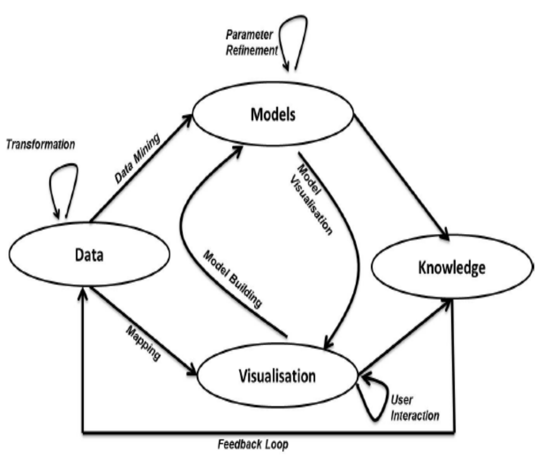


Fig 1. The process of Visual Analytics

Visual Analytics in the Context of the Knowledge Discovery from Data Process and Data Mining

Models or patterns shown in Figure 1 are generated either through a data mining process, mathematical or statistical methods and the visualization is used to evaluate and refine the model. However, data mining is often regarded as just one but very important step in the Knowledge Discovery from Data process.

Hence from a data analytics perspective it is important to discuss how Visual Analytics fits in the overall KDD process.

Figure 2 describes the KDD process augmented with the Visual Analytics process discussed in the 'The Principle of Visual Analytics' section. The dashed printed elements are taken from the Visual Analytics process, whereas the solid printed elements correspond to the KDD process. Please note that none of the KDD steps have been removed. Also note that as opposed to the KDD process, the Visual Analytics process is interactive through interaction with the user of the system and the feedback loop is feeding changes back to the data input as shown in Figure 1.

The following steps are extracted from Figure 2 and describe the main activities involved in the KDD process augmented by the process of Visual Analytics as it is depicted in Figure 1:

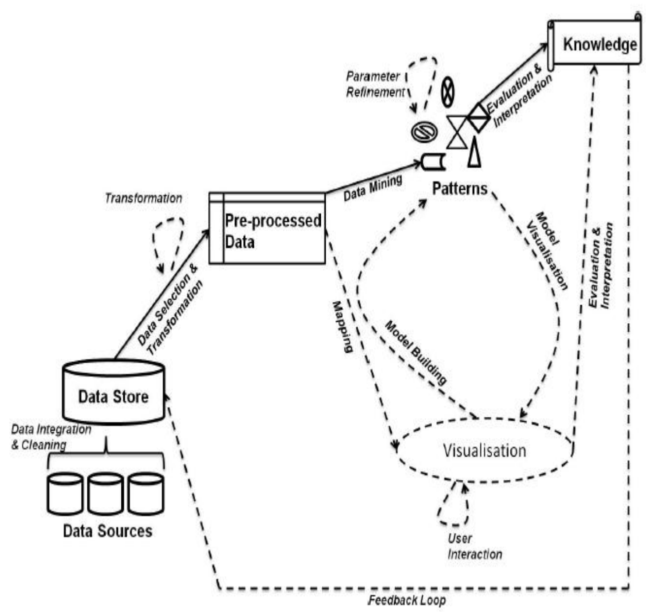


Fig 2. The Knowledge Discovery from Data process

3. METHODOLOGY

3.1 Advantages:

1.Easier to build and maintain.

2.Transition from object-oriented analysis to object oriented design will be easy.

3.OOA is more immune to change because objects are more stable than function.

4.Objects are likely to stay the same even if the exact nature of the problem changes

3.2 Basic concepts:

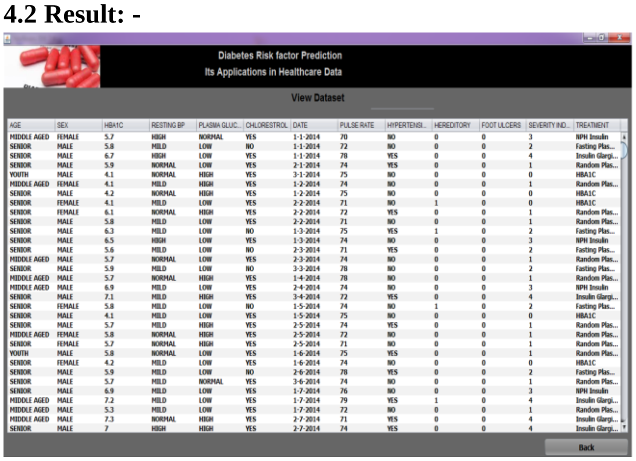
Object contain attributes that define the state of the object of similar type are grouped together to form an object class (or class). An object also provides some services or operations which are used to view or modify the state of an object from outside with the help of messages sent to that object.

4. RESULTS

4.1 Main Menu: -

Fig 3: - Main Menu

4.2 Result: -



AGE	SEX	HBA1C	RESTING BP	PLASMA GLUC...	CHLORESTROL	DATE	PULSE RATE	HYPERTENSI...	HEREDITORY	FOOT ULCERS	SEVERITY IND...	TREATMENT
MIDDLE AGED	FEMALE	5.7	HIGH	NORMAL	YES	1-1-2014	70	NO	0	0	3	NPH Insulin
SENIOR	MALE	5.8	MILD	LOW	NO	1-1-2014	72	NO	0	0	2	Fasting Plas...
SENIOR	MALE	6.7	HIGH	LOW	YES	1-1-2014	78	YES	0	0	4	Insulin Glargi...
SENIOR	MALE	5.9	NORMAL	LOW	YES	2-1-2014	74	YES	0	0	1	Random Plas...
YOUTH	MALE	4.1	NORMAL	HIGH	YES	3-1-2014	75	NO	0	0	0	HBA1C
MIDDLE AGED	FEMALE	4.1	MILD	HIGH	YES	1-2-2014	74	NO	0	0	1	Random Plas...
SENIOR	MALE	4.2	NORMAL	HIGH	YES	1-2-2014	75	NO	0	0	0	HBA1C
SENIOR	FEMALE	4.1	MILD	LOW	YES	2-2-2014	71	NO	1	0	0	HBA1C
SENIOR	FEMALE	6.1	NORMAL	HIGH	YES	2-2-2014	72	YES	0	0	1	Random Plas...
SENIOR	MALE	5.8	MILD	LOW	YES	2-2-2014	71	NO	0	0	1	Random Plas...
SENIOR	MALE	6.3	MILD	LOW	NO	1-3-2014	75	YES	1	0	2	Fasting Plas...
SENIOR	MALE	6.5	HIGH	LOW	YES	1-3-2014	74	NO	0	0	3	NPH Insulin
SENIOR	MALE	5.6	MILD	LOW	NO	2-3-2014	71	YES	0	0	2	Fasting Plas...
MIDDLE AGED	MALE	5.7	NORMAL	LOW	YES	2-3-2014	74	NO	0	0	1	Random Plas...
SENIOR	MALE	5.9	MILD	LOW	NO	3-3-2014	78	NO	0	0	2	Fasting Plas...
MIDDLE AGED	MALE	5.7	NORMAL	HIGH	YES	1-4-2014	78	NO	0	0	1	Random Plas...
MIDDLE AGED	MALE	6.9	MILD	LOW	YES	2-4-2014	74	NO	0	0	3	NPH Insulin
SENIOR	MALE	7.1	MILD	HIGH	YES	3-4-2014	72	YES	0	0	4	Insulin Glargi...
SENIOR	FEMALE	5.8	MILD	LOW	NO	1-5-2014	74	NO	1	0	2	Fasting Plas...
SENIOR	MALE	4.1	MILD	LOW	YES	1-5-2014	75	NO	0	0	0	HBA1C
SENIOR	MALE	5.7	MILD	HIGH	YES	2-5-2014	74	YES	0	0	1	Random Plas...
MIDDLE AGED	FEMALE	5.8	NORMAL	HIGH	YES	2-5-2014	72	NO	0	0	1	Random Plas...
SENIOR	FEMALE	5.7	NORMAL	HIGH	YES	2-5-2014	71	NO	0	0	1	Random Plas...
YOUTH	MALE	5.8	NORMAL	LOW	YES	1-6-2014	75	YES	0	0	1	Random Plas...
SENIOR	FEMALE	4.2	MILD	LOW	YES	1-6-2014	74	NO	0	0	0	HBA1C
SENIOR	MALE	5.9	MILD	LOW	NO	2-6-2014	78	YES	0	0	2	Fasting Plas...
SENIOR	MALE	5.7	MILD	NORMAL	YES	3-6-2014	74	NO	0	0	1	Random Plas...
SENIOR	MALE	6.9	MILD	LOW	YES	1-7-2014	76	NO	0	0	3	NPH Insulin
MIDDLE AGED	MALE	7.2	MILD	LOW	YES	1-7-2014	79	YES	1	0	4	Insulin Glargi...
MIDDLE AGED	MALE	5.3	MILD	LOW	YES	1-7-2014	72	NO	0	0	1	Random Plas...
MIDDLE AGED	MALE	7.3	NORMAL	HIGH	YES	2-7-2014	71	YES	0	0	4	Insulin Glargi...
SENIOR	MALE	7	HIGH	HIGH	YES	2-7-2014	74	YES	0	0	4	Insulin Glargi...

Fig 4: - Result

4.3 Classification: -

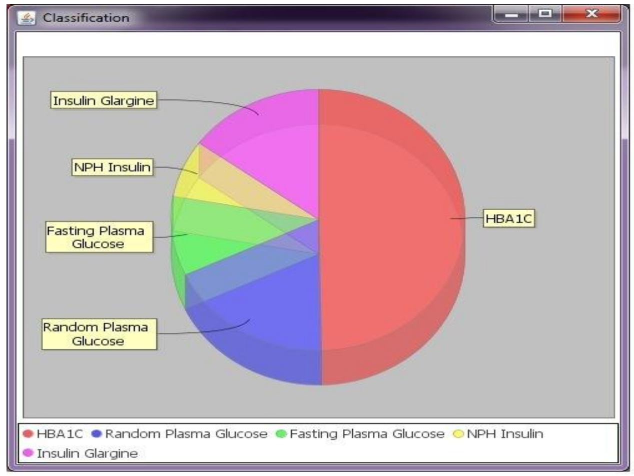


Fig 5: - Classification

4.4 Graph Show: -

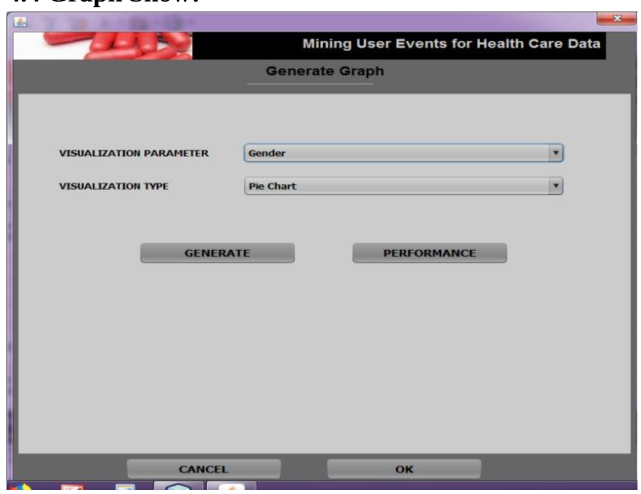


Fig 6: - Graph Show

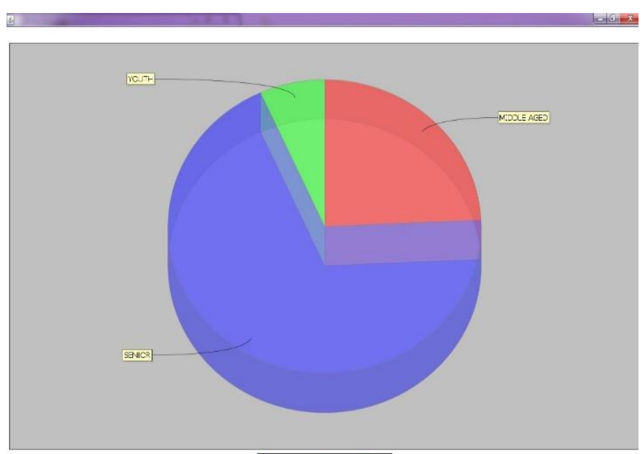


Fig 7: - Graph

5. CONCLUSION

The Amount of Research work has been done for Prediction of diabetes using data mining technique. The bottom up summarization technique uses when patient has high risk of diabetes. The K-Nearest Neighbor Algorithm, Bayesian Classifier, Naïve Bayesian Classifier, Artificial Neural Network, Bayesian Network, Association Rule Mining all methods used for prediction of diabetes which gives patient's condition of Normal, Pre-diabetes, Diabetes. In K-Nearest neighbor algorithm always need to determine the value of K. All above methods used to predict diabetes. But if Patient is detected as diabetes firstly there is a need of finding Control and Un control condition of diabetes. Because if Patient has diabetes in Un- control condition, may be the patient has severe effect on Patient's Organ like Heart, Eye, Kidney etc. So, there is need of finding early Severity which may be help patient for reducing the Severity on Organ or Halting the Severe Effect on Organ.

6. FUTURE SCOPE

The future scope of the project is day by day our population is increase and because of that there are so many diseases are found on earth like diabetics. Worldwide are suffering from diabetes. Diabetes is a metabolic disease where the improper management of blood glucose levels led to risk of many diseases like heart attack, kidney disease, eye etc. for that purpose the data of the patient will increase that time this software is use for big data mining for sort the people who are suffering from diabetes. This show the prediction and severity of that patient for the diabetes diseases and we can easily give the treatment for that.

REFERENCES

[1] *P. T. Katzmarzyk, C. L. Craig, and L. Gauvin, "Adiposity, physical fitness and incident diabetes: The physical activity longitudinal study," Diabetologia, vol. 50, no. 3, pp. 538–544, Mar. 2007.*

[2] *Z. Xu, X. Qi, A. K. Dahl, and W. Xu, "Waist-to-height ratio is the best indicator for undiagnosed type 2 diabetes," Diabetic Med., vol. 30, no. 6, pp. e201–e207, Jun. 2013.*

[3] *R. N. Feng, C. Zhao, C. Wang, Y. C. Niu, K. Li, F. C. Guo, S. T. Li, C. H. Sun, and Y. Li, "BMI is strongly associated with hypertension, and waist circumference is strongly associated with type 2 diabetes and dyslipidemia, in northern Chinese adults," J. Epidemiol., vol. 22, no. 4, pp. 317–323, May 2012.*

[4] *A. Berber, R. G´omez-Santos, G. Fangh¨anel, and L. S´anchez-Reyes, "Anthropometric indexes in the prediction of type 2 diabetes mellitus, hypertension and dyslipidaemia in a Mexican population," Int. J. Obes. Relat Metab. Disorders, vol. 25, no. 12, pp. 1794–1799, Dec. 2001.*

[5] *B. Balkau, D. Sapinho, A. Petrella, L. Mhamdi, M. Cailleau, D. Arondel, and M. A. Charles, D. E. S. I. R. Study Group, "Prescreening tools for diabetes and obesity-associated dyslipidaemia: Comparing BMI, waist and waist hip ratio. The D.E.S.I.R. Study," Eur. J. Clin. Nutr., vol. 60, no. 3, pp. 295–304, Mar. 2006.*

[6] *I. S. Okosun, K. M. Chandra, S. Choi, J. Christman, G. E. Dever, and T. E. Prewitt, "Hypertension and type 2 diabetes comorbidity in adults in the United States: risk of overall and regional adiposity," Obes. Res., vol. 9, no. 1, pp. 1–9, Jan. 2001.*

[7] *L. A. Sargeant, F. I. Bennett, T. E. Forrester, R. S. Cooper, and R. J.Wilks, "Predicting incident diabetes in Jamaica: the role of anthropometry," Obes. Res., vol. 10, no. 8, pp. 792–798, Aug. 2002.*

[8] *N. T. Duc Son le, T. T. Hanh, K. Kusama, D. Kunii, T. Sakai, N. T. Hung, and S. Yamamoto, "Anthropometric characteristics, dietary patterns and risk of type 2 diabetes mellitus in Vietnam," J. Amer. Coll. Nutr., vol. 24, no. 4, pp. 229–234, Aug. 2005.*

[9] *G. T. Ko, J. C. Chan, C. S. Cockram, and J.Woo, "Prediction of hypertension, diabetes, dyslipidaemia or albuminuria using simple anthropometric indexes in Hong Kong Chinese," Int. J. Obes. Relat. Metab. Disorders, vol. 23, no. 11, pp. 1136–1142, Nov. 1999.*

[10] *M. B. Snijder, P. Z. Zimmet, M. Visser, J. M. Dekker, J. C. Seidell, and J. E. Shaw, "Independent and opposite associations of waist and hip circumferences with diabetes, hypertension and dyslipidemia: The AusDiab study," Int. J. Obes. Relat. Metab. Disorders, vol. 28, no. 3, pp. 402–409, Mar. 2004.*