

A Survey about Role of Data Mining Techniques and its Applications in Healthcare Sector

Ramya. A
Research Scholar
Department of Computer Science
Vels Institute of Science Technology and Advanced Studies
 Chennai, India
 ramyaimayaa7@gmail.com

Dr. K. Rohini
Associate Professor
Department of Computer Science
Vels Institute of Science Technology and Advanced Studies
 Chennai, India
 rrohini16@gmail.com

Abstract - The Data mining is a goal part of different areas looking like financing, quantifiable exploration, instructing, retail, e-business, advertising, human services and so forth. Numerous analysts have been deliberately inspected and reviewed in social insurance, which is a functioning interdisciplinary region that is the degree of information mining. The assignment of understanding evacuation in the social insurance records is a requesting undertaking and complicated as well. The interoperability of heterogeneous clinical gadgets, security of the patient data, and customized representation of the handled information is by all accounts exceptionally fundamental. In this paper analyzed Data mining techniques with machine learning algorithms and tools.

Keywords: Data Mining, Healthcare System, DNA Analysis, KDD

I. INTRODUCTION

Clinical medicinal services has been as of late buy expanding fixation and notoriety. In apparatus taking over nuclear, bio clinical system, remedial imaging, and restorative records of quiet, tremendous amount of wellbeing records all are produced every day due close to propels. Associations of social insurance in enormous volumes of data are created and gathered to a consistent schedule. Information mining is created step by step in current presence as new data hardware. The networking of certain various sensor gadgets has been prompting mechanization approximately in the entire areas and together authorizing applications that pertain innovative advancements. This is leading into prompted and improved precision, intensity, and financial benefit along with diminished man-made intercession. Semantic explanation of physical welfare information helps in handling the catchphrases connected to the semantic highlights and determining the association of every component into another and the way these highlights are utilized in the characterization of sicknesses. In the clinical segment, immense measure of data, viz. the complete clear record of the sick person, investigative strategies, schemes to treat the patient, assurance of insurance plans, also strict methods can be seen available and here emerges the necessity to comprehend the correlation among all the concerned information. The issue of semantic information mix is to distinguish if two substances in the assortment of datasets coordinate or don't coordinate the equivalent real world element. A substance in a dataset comparison to the genuine

world idea and its properties can be portrayed in an organized or unstructured way. When comparable substances have been coordinated, distinctive combination approaches should be performed for consolidating them into a solitary substance. Considering the wide idea of elements, the best in class has concentrated on techniques that decrease manual work and boost exactness and accuracy.

II. LITERATURE REVIEW

The paper proposed about Modern Industrial Automation System in machine learning model [1]. The paper discusses about the IOT based Data mining techniques [15] to processing data [2]. The paper proposed Health care domain with Heterogeneous IOT devices [3]. The paper focuses on efficient techniques based on Ada Boost [4]. The paper Proposed about the Artificial Intelligence Techniques as the Predictive Analytics tool [5]. This paper discussed about Prediction using Classification techniques [6]. The paper discussed findings in chronic disease treatments [7]. This paper discussed about the uncover risk factors of health examinations [8]. This paper discusses about breast cancer diagnosis in data mining approach [9]. The paper proposed the classification based on the different soft computing tool [10].

III. THE PROCESS OF DATA MINING

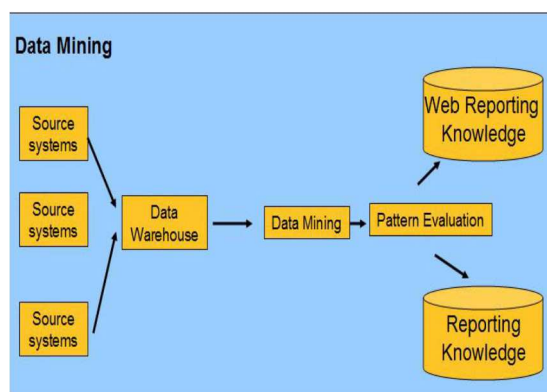


Fig. 1. Data mining process

Knowledge Discovery in Data (KDD) everyone had the option to find a few examples gathering and examining all of these data. Indeed, even the entire data set is a garbage, it can be seen that there certain shrouded designs which would be a

2021 2nd International Conference on Intelligent Engineering and Management (ICIEM) | 978-1-6654-1450-0/21/\$31.00 ©2021 IEEE | DOI: 10.1109/ICIEM51511.2021.9445300

possibility to be pulled out through consolidating various data resources to give important bits of knowledge. The Data mining is called like this.

IV. KNOWLEDGE DRIVEN FRAME WORK

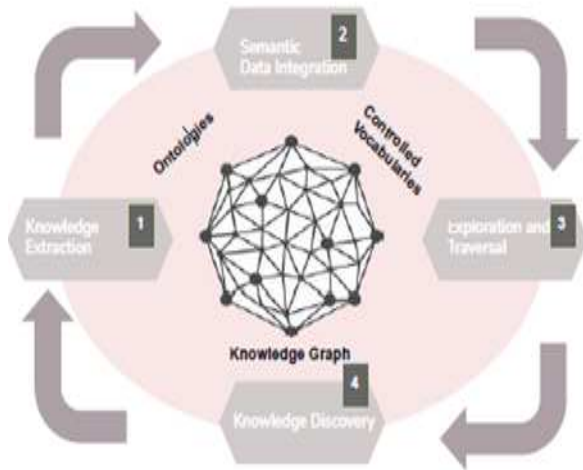


Fig. 2: Knowledge-driven Framework

The necessity to be integrated different assorted gadgets which replace the data got from the sensors to the cyberspace medical management cloud. Because of the absence of overall palatable principles, the issue of compatibility in the midst of the different gadgets proceeds in the equipment level. However, semantic representations can possibly be worked to beat the subject of compatibility in assorted framework through clarifying the data from the technological medical documents. In the medical management area, the sick people and doctors require to speak to the concerned persons in the event of far off medical checking framework.

V. DATA MINING TECHNIQUES

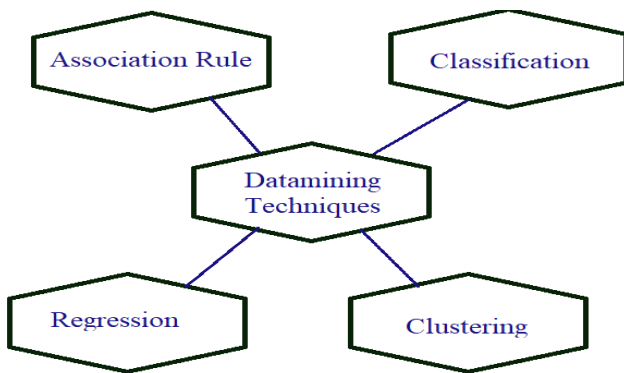


Fig.3 Data mining Techniques

These Techniques are used to understand the process and derive the conclusion by using statistics and Machine learning algorithm from extreme amount of information.

VII. TOOLS IN DATA MINING

Classification Analysis: This data mining method helps to grouping data in different classes. This technique is more complex that forces to collect various attributes together in to observed class. Association Rule mining, Clustering Analysis, Regression analysis are the techniques of data mining.

VI. DATA MINING ALGORITHM

The Adaboost is one of the boosting algorithm in machine learning to boost or improve their performances. K-Means [11]- K-means algorithm cluster variation is less. K-Nearest- K-nearest is used for the classification [12][13][14][16] and regression. Naive Bayes Algorithm- The algorithm solve problems in classification. Apriori Algorithm- It proceeds by showing the frequent individual items in the database and it is used to determine association rules. Decision based Algorithm-It has the efficiency to eliminate, introduce or to expand biases or discrimination. CART Algorithm- Classification and Regression algorithm can be represented as Binary Tree. Predictions are made with CART by travelling the Binary Tree given a new input record. C4.5 Algorithm- It is a data mining classification Algorithm.

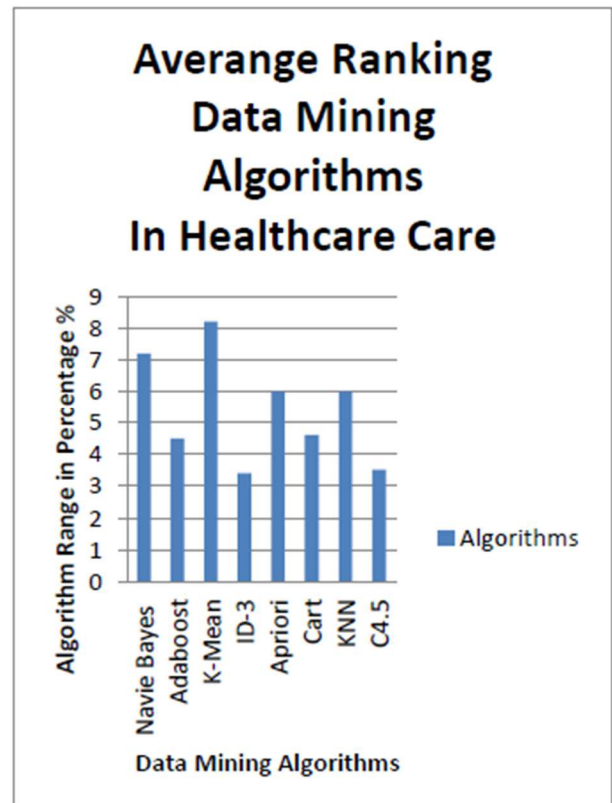


Fig.4 Data Mining Algorithms

Tools	Description
WEKA	Group of machine learning algorithms data mining tasks in weka tool. This tool is open source and free, platform independent, GUI interfaces combines into one. Primarily the tool designed to implement in agricultural sector. Now a days many the industries using weka tool. Format of data is CSV and ARFF. ODBC is used in database. Weka tool explores preprocessing.
. Rapid Miner	Predictive and powerful data mining tool is Rapid miner tool. Text mining, machine learning and deep learning predictive algorithms implemented in this algorithm. The package of this software allows to mine the data, which includes raw data. Rapid miner is a Artificial intelligence framework and positively extracted solutions in AI. Rapid miner is user friendly.
Oracle Data mining	This is graphical tool for database development. R scripts takes place in oracle data mining.
KNIME	CSV file format used in this tool. Filter node filters the data. The output visualization part will be in the form of chart.
Python	Scikit-learn, keras, SciPy, Selenium are the python data science and automation tool. It is a open source tool.

Table. 1 Tools in Data mining

Above Table explores the details about Data mining Tools and its description. In this paper Weka, Orange, python, knime, Oracle Data mining, Rapid Miner tools are discussed. WEKA-Group of machine learning algorithms data mining tasks in weka tool. This tool is open source and free, platform independent, GUI interfaces combines into one. Primarily the tool designed to implement in agricultural sector. Now a days many the industries using weka tool. Format of data is CSV and ARFF. ODBC is used in database. Weka tool explores preprocessing, classify, cluster, association and attributes selected visualized. Rapid Miner - Predictive and powerful data mining tool is Rapid miner tool. Text mining, machine learning [17][18] and deep learning predictive algorithms implemented in this algorithm. The package of this software allows to mine the data, which includes raw data. Rapid miner is a Artificial intelligence framework and positively extracted solutions in AI. Rapid miner is user friendly. Oracle Data mining – This is graphical tool for database development. R scripts takes place in oracle data mining. KNIME – CSV file format used in this tool. Filter node filters the data. The output visualization part will be in the form of chart. Python – Scikit-learn, keras, SciPy, Selenium are the python data science and automation tool. It is a open source tool.

VIII. DATA IN ML

In this paper applied Obesity data from UCI Repository. Height and weight taken from dataset based upon BMI its confirmed obesity level shows in fig 6.

Gender	Age	Height	Weight	family_history_with_overweight	NObesyedad
0	1	21	162	64.0	yes Normal_Weight
1	1	21	152	56.0	yes Normal_Weight
2	0	23	180	77.0	yes Normal_Weight
3	0	27	180	87.0	no Overweight_Level_I
4	0	22	178	89.8	no Overweight_Level_II
5	0	29	162	53.0	no Normal_Weight
6	1	23	150	55.0	yes Normal_Weight
7	0	22	164	53.0	no Normal_Weight
8	0	24	178	64.0	yes Normal_Weight
9	0	22	172	68.0	yes Normal_Weight
10	0	26	185	105.0	yes Obesity_Type_I
11	1	21	172	80.0	yes Overweight_Level_II
12	0	22	165	56.0	no Normal_Weight
13	0	41	180	99.0	no Obesity_Type_I
14	0	23	177	60.0	yes Normal_Weight
15	1	22	170	66.0	yes Normal_Weight
16	0	27	193	102.0	yes Overweight_Level_II
17	1	29	153	78.0	no Obesity_Type_I
18	1	30	171	82.0	yes Overweight_Level_II

Fig 5. Display of data in Jupiter Notebook.

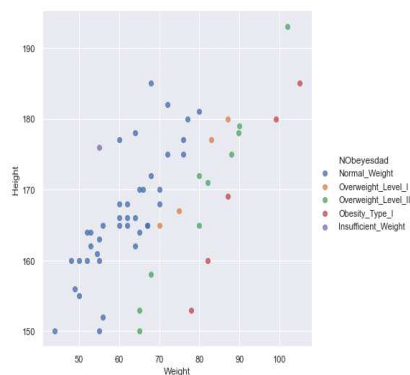


Fig 6. Type of Obesity

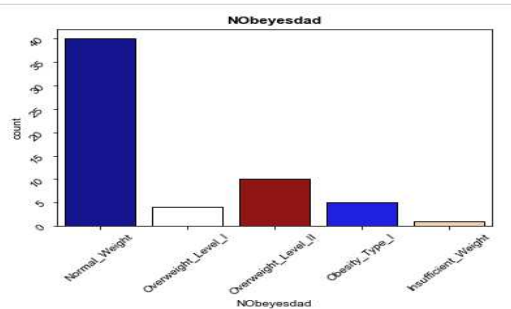


Fig 7. Obesity levels in bar plot

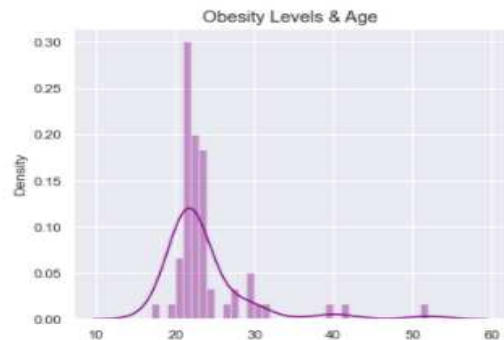


Fig 8. Obesity level and Age



Fig 9. Obesity level and height

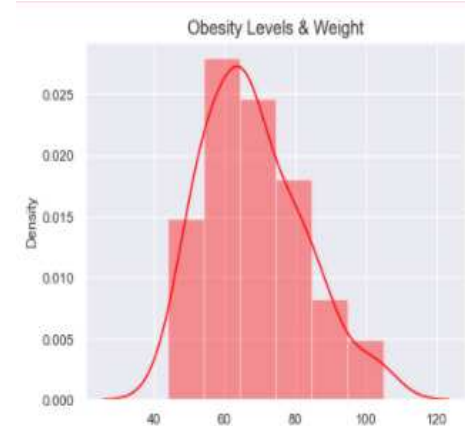


Fig 10. Obesity level and weight

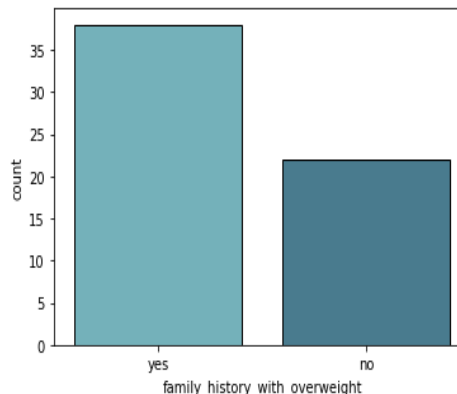


Fig 11 Family History with overweight

IX. CONCLUSION

In this paper Obesity dataset taken from UCI Repository. The responsibility of complete the necessities of the medical-care associations possibly has immense effect for region of medication, medical health welfare clinical and that stands for an extensive procedure. In this paper Data mining techniques, tools and algorithms analyzed. Using Data mining Algorithms and Tools in future to design a prediction model for medical Data.

REFERENCES

- [1] Ringsqunadi, M., et al.: Semantic-guided feature selection for industrial automation systems. In: International Semantic Web Conference, Springer, LNCS 9367, pp. 225–240 (2015)
- [2] Mahdaveinejad, M.S., Rezvan, M., et al.: Machine learning for internet of things data analysis: a survey. Digit. Commun. Netw. 4, 161–175 (2018)
- [3] Jabbar, S. et. 1, Semantic Interoperability in Heterogeneous IoT Infrastructure for Healthcare. *Wireless Communications and Mobile Computing*, vol. 2017, 2017.
- [4] DavarGiveki et al. Automatic detection of erythematous-squamous diseases using pso-svm based on association rules. Published: Engineering Applications of Artificial Intelligence, 26(1):603–608, 2013.
- [5] A. A. Bakar, Z. Kefli, S. Abdullah and M. Sahani, "Predictive models for dengue outbreak using multiple rulebase classifiers," Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, Bandung, Indonesia, 2011, pp. 1-6, doi: 10.1109/ICEEL.2011.6021830.
- [6] Chun-Lang Chang, Chih-Hao Chen, Applying decision tree and neural network to increase quality of dermatologic diagnosis, Expert Systems with Applications, Volume 36, Issue 2, Part 2, 2009, Pages 4035-4041, ISSN 0957-4174.
- [7] Mu-Jung Huang, Mu-Yen Chen, Show-Chin Lee, Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. Expert Systems with Applications, Volume 32, Issue 3, 2007, Pages 856-867, ISSN 0957-4174.
- [8] Chih-Hung Jen, Chien-Chih Wang, Bernard C. Jiang, Yan-Hua Chu, Ming-Shu Chen, Application of classification techniques on development an early-warning system for chronic illnesses, Expert Systems with Applications, Volume 39, Issue 10, 2012, Pages 8852-8858, ISSN 0957-4174.
- [9] Kharya, Shweta. (2012). Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease. International Journal of Computer Science, Engineering and Information Technology (IJCSSEIT). 2. 10.5121/ijcseit.2012.2206.

- [10] S. Mitra, S. K. Pal and P. Mitra, "Data mining in soft computing framework: a survey," in *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 3-14, Jan. 2002, doi: 10.1109/72.977258.
- [11] K.ROHINI, .DR.G.SUSEENDRAN,"Aggregated K Means Clustering and Decision Tree Algorithm for Spirometry Data", *Indian Journal of Science and Technology*, Vol 9(44), DOI: 10.17485/ijst/2016/v9i44/103107, November 2016 ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645 (Scopus)
- [12] K.ROHINI, .DR.G.SUSEENDRAN,"Aggregated K Means Clustering and Decision Tree Algorithm for Spirometry Data", *Indian Journal of Science and Technology*, Vol 9(44), DOI: 10.17485/ijst/2016/v9i44/103107, November 2016 ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645 (Scopus).
- [13] K. KAVITHA, K. ROHINI, G. SUSEENDRAN, "A Comparative Evaluation of Meta Classification Algorithms with Smokers Lung Data", *International Journal of Engineering & Technology*, vol.7 (4.36) (2018) 845-849 Website:www.sciencepubco.com/index.php/IJET (Scopus) .
- [14] R. VIGNESH, K. ROHINI, "Analysis to Determine the Scope and Challenging Responsibilities of Ethical Hacking Employed in Cyber Security", *International Journal of Engineering & Technology*, 7 (3.27) (2018) 196-199 (Scopus)
- [15] K.ROHINI, .DR.G.SUSEENDRAN "Analysis Of Datamining Techniques Used In Spiro metric Data" , *International Journal of Control Theory and Applications*. ISSN: 0974-5572, Vol No-10, 2017.
- [16] K.ROHINI, DR.G.SUSEENDRAN, "Predicting lung disease severity evaluation and comparison of hybrid decision tree algorithm", *Indian Journal of Innovations and Developments* Vol 6(1), January, 2017 ISSN (online) : 2277-5390 ISSN (Print): 2277- 5382 .
- [17] D. R. Krithika and K. Rohini, "Blockchain with Bigdata Analytics" *ICTIDS 2019*, Vol 118, and page: 403, published Journal: springer.
- [18] Krithika, D.R. & K. Rohini. (2020). A Survey on Challenging Capabilities of Big Data Analytics in Healthcare. *International Journal of Innovative Research in Applied Sciences and Engineering*. 4. 593-597. 10.29027/IJIRASE.v4.i1.2020.593-597.