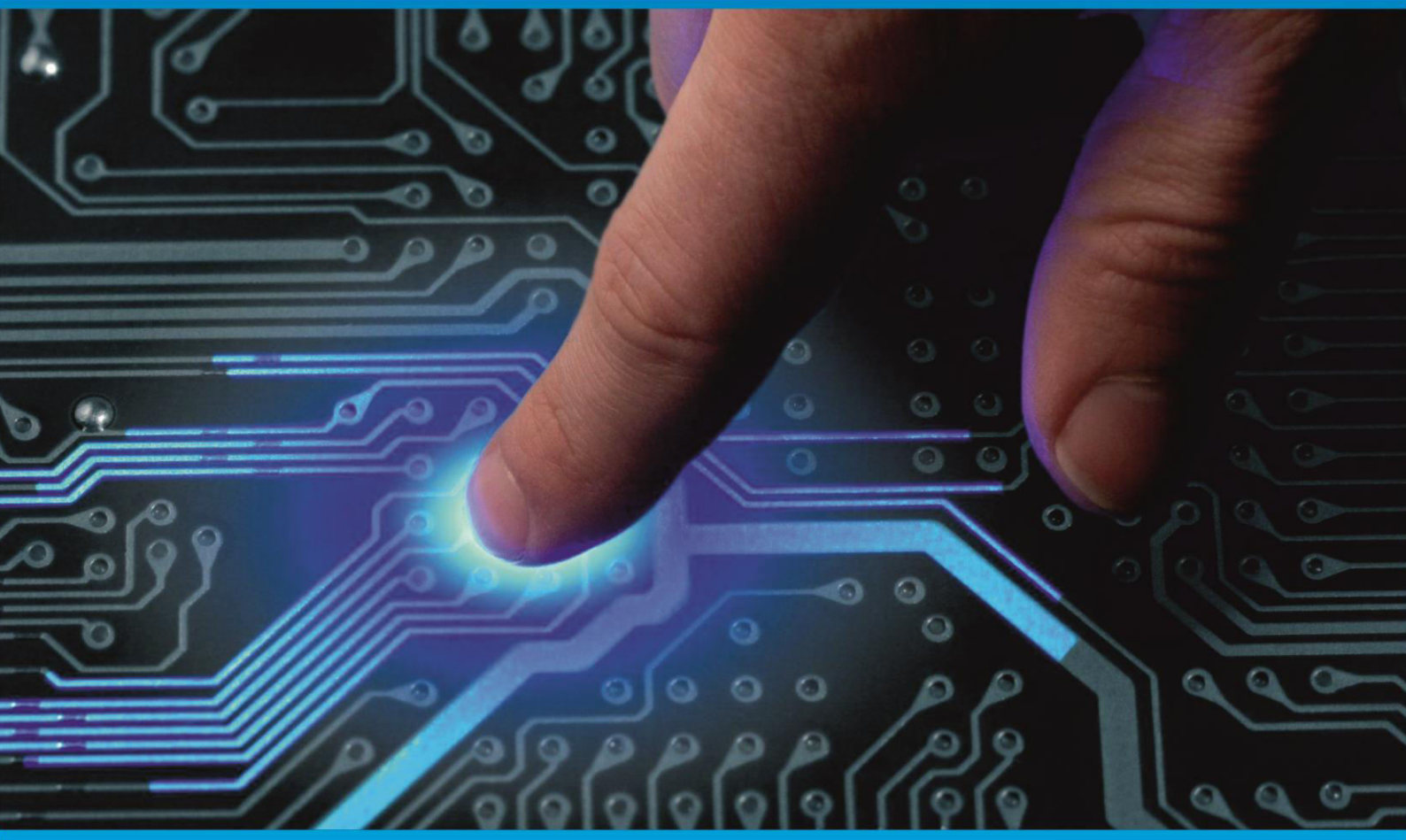
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 4, April 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904144 |

Prediction and Analysis of Diabetes Complication Disease Utilizing Data Mining Algorithm

Ankita Rajendra Patil , Roshani Rajendra Patil, Zeba Khalil Sayyed, Ashwini Gorakh Wagh

UG Student, Dept. of Information Technology, Sandip Institute of Technology and Research Centre, Maharashtra, India

ABSTRACT: Diabetes is perhaps the most perilous persistent sickness that could prompt others genuine muddling illnesses. In Indonesia, the most widely recognized diabetes microvascular confusions illnesses are retinopathy, nephropathy and neuropathy. To forestall these complexities to show, information mining strategy to remove information on hazard factor for every inconvenience gets pivotal. The objective of this examination is to build an expectation model for three significant diabetes difficulty illnesses in Indonesia and discover the critical highlights corresponded with it. In this exploration, the diabetes hazard calculate limited seven highlights, which are Age, Gender, BMI, Family history of diabetes, Blood pressure, term of diabetes endures and Blood glucose level. Subsequently, Naive Bayes Tree and C4.5 choice tree-based arrangement strategies and k-implies grouping procedures were utilized to investigate this dataset. After this examination, we assessed the presentation of every method and tracked down the related element and sub element as a sickness hazard factor for them. Coming about the most compelling danger factor for Retinopathy is a female patient that having a hypertension emergency. With respect to Nephropathy, the most unmistakable danger factor is the span of diabetes over 4 years. However, for Neuropathy, it ruled for female patients, with BMI more than 25. Concerning family background of diabetes, there is no unmistakable huge connection with these complexity infections. The general exactness of the proposed model is 68% so it, could be utilized to as an elective strategy to help anticipate diabetes entanglement sicknesses at a beginning phase.

KEYWORDS: decision tree, Data mining, Diabetes complication disease,k-means,NaiveBayes, Prediction model.

I. INTRODUCTION

Diabetes Mellitus (DM) characterized as gathering of metabolic problems predominantly cause by overabundance glucose inside the circulation system. The World Health Organization expresses that roughly in excess of 700 million individuals were extended experiencing diabetes by 2030.

Diabetes patients happen all through the world, yet is more normal in created nations [1]. In Indonesia, the pervasiveness of diabetes was 10,9%, and the pattern is slowly expanding [2]. Diabetes as metabolic problems could harm the veins,

.

which increment the danger of genuine unexpected issues that harming the heart, eyes, kidneys and nerves. The most widely recognized diabetes confusions infections are isolated into two assembled dependent on its harm to little veins (microvascular) and harm to the corridors (macrovascular). Microvascular illness bunch into which organ the infection assault, which are eye (retinopathy), kidney (nephropathy) and neural harm (neuropathy). The major macrovascular entanglements incorporate sped up cardiovascular illness showing as strokes among other genuine infections. As indicated by Indonesian Ministry of Health, the main three of the diabetes microvascular inconvenience illnesses are retinopathy, neuropathy and nephropathy [3]. Besides, to forestall the confusions deteriorating, one of the manners in which that should be possible is by acquiring data in regards to its danger factor. Because of high mortality and bleakness of diabetes confusion infections, counteraction and hazard factor forecast become significant and arising pattern of examination subject and studies.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904144 |

Numerous examinations have been led to acquire information related of hazard elements and finding of diabetes and pre-diabetes. Notwithstanding, barely any investigations have been led to assess the diabetes complexity sicknesses, particularly its danger factors. Thus, diabetes intricacy infections keep on being underutilized in sickness counteraction and not regularly possibly discovered when the illness previously showed in the hurting condition. Diabetes is otherwise called the quiet executioner in light of this explanation. Diabetes hazard figure partitioned two gatherings, which are changed and unmodified. Changed identified with trait like racial, ethnic, age, sex, and so forth Unmodified identified with undesirable and stationary way of life. In this examination, we utilized credits from clinical records of the Indonesian diabetes patient to establish hazard factor for every one of three significant diabetes intricacy illnesses in Indonesia.

In this data time, Data mining has effectively become a significant method to help analysts to separate information from huge and complex information, for example, from patient clinical records. In this exploration, Indonesian diabetic's patient dataset was prepared with information mining strategies to discover decides that can assist with deciding the danger factor quality and its worth of potential diabetes complexity infection. We utilized information mining to acquire this information from patient clinical record ascribes, from changed and unmodified danger factor. Subsequently, in the engineering of this exploration approach, an exertion was made to track down the most appropriate information mining procedure to create the standard and the most impact quality and its worth from altered and unmodified danger factors.

This examination is coordinated as follows: segment 2 gives the vital foundation information on information mining and related exploration in this subject and the distinction with the proposed of this examination. Segment 3 presents the technique approach and area 4 give the outcome and conversation of this examination, with segment 5 giving ends.

.

II. RELATED WORK

1. **Global estimates of the prevalence of diabetes for 2010 and 2030: Design of Knowledge Management System for Diabetic Complication Diseases.**

System, , Institut Teknologi Harapan Bangsa (ITHB).

Diabetes is a complex, chronic illness that requires continuous medical care with multifactors risk reduction strategies to control the blood sugar level. Diabetes could increase the risk of developing a series of serious health problems. This happens because the high level of glucosecould lead to serious diseases affecting the heart and blood vessels, eyes, kidneys, nerves and teeth. In addition, people with diabetes also have a higher risk of developing infections. People with diabetes need self-management education and they also need support for preventing and reducing the risk of other complications diseases. People with diabetes should receive medical care from a collaborative, integrated team with expertise in diabetes and also a support group from family, friends and other people that diagnose with diabetes. On the other hand, it is well known that every medical treatment has risks as well as benefits. The medical ethical principle state that a patient has the right to decide what's appropriate treatments for them, taking into account their personal circumstances, lifestyle, beliefs, and priorities. Moreover, to choose the right medical treatment is a long and confusing process.Especially if one of the options is whether to take surgery or other high risk and expensive treatments. Informed decision-making is part of process to give a relevant information regarding to the medical treatments, by making collaborative communication between patient, family and one or more medical practitioners. But sometime, this information still not enough, patient also has the need for finding and get information from another patient with similar case to make the most suitable decision. The better information and knowledge to other people with a similar case, lifestyle and treatments history, personal circumstances and lifestyle are crucial in making the decision wait for time δt and collects all the packets. After time δt it calls the optimization function to select the path and send RREP. Optimization function uses the individual node's battery energy; if node is having low energy level then optimization function will not use that node.

2. Journal of Diabetes Research

Type 2 diabetes (T2D) is a public health problem worldwide, and the main risk factor for its development is obesity. The Yaqui ethnic group of Sonora has serious obesity problems,resulting in an increased risk of T2D in its inhabitants. The objective of this study was to evaluate the effectiveness of a health promotion program on obesity parameters and cardiovascular risk factors in short- (6 months) and medium-term periods (12 months) in indigenous Yaquis of Sonora. The design is a translational clinical study of a single cohort with prepost intervention measurements in a sample of 93 subjects. The effectiveness of the program was evaluated by comparing obesity parameters, metabolic markers, and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904144 |

physical activity 6 and 12 months with those measured under basal conditions using a paired -test or Wilcoxon rank-sum test.
a) Seasonal crops— crops can be planted during a season. eg. wheat, cotton.
b) Whole year crops— crops can be planted during entire year. eg. vegetable, paddy, Toor.
These research paper tells about various machine learning techniques and algorithms.

3. Review on Determining Number of Cluster in K-Means Clustering

Clustering is widely used in different field such as biology, psychology, and economics. The result of clustering varies as number of cluster parameter changes hence main challenge of cluster analysis is that the number of clusters or the number of model parameters is seldom known, and it must be determined before clustering. The several clustering algorithm has been pro-posed. Among them k-means method is a simple and fast clustering technique. We address the problem of cluster number selection by using a k-means approach We can ask end users to provide a number of clusters in advance, but it is not feasible end user requires domain knowledge of each data set. There are many methods available to estimate the number of clusters such as statistical indices, variance based method, Information Theoretic, goodness of fit method etc...The paper explores six different approaches to determine the right number of clusters in a dataset.

4.An Interpretable Rule-Based Diagnostic Classification of Diabetic Nephropathy

Among Type 2 Diabetes Patients
The prevalence of type 2 diabetes is increasing at an alarming rate. Various complications are associated with type 2 diabetes, with diabetic nephropathy being the leading cause of renal failure among diabetics. Often, when patients are diagnosed with diabetic nephropathy, their renal functions have already been significantly damaged. Therefore, a risk prediction tool may be beneficial for the implementation of early treatment and prevention.Methodologies used in these research paper are:Data set Collection: These data sets contain various attributes and their respective values of soil samples taken from 3 regions of Pune District . Decision tree algorithm for soil fertility prediction: J48 (C4.5): J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool.NBTree: This algorithm is used for generating a decision tree with naive Bayes classifiers at
the leaves.Simple Cart :It is a non-parametric decision tree learning technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numeric, respectively.
Advantage:
The main advantage of researched study is prediction of soil fertility.
Use in our system:
Hence we analyzed and used the results of these study to classify the soil into low,medium and
high according to their Ph level.

5. Combined Methods for Diabetic Retinopathy Retinopathy Screening Using Retina

Photographs and Tear Fluid Proteomics Biomarkers
it is estimated that 347 million people suffer from diabetes mellitus (DM), and almost 5 million are blind due to diabetic retinopathy (DR). The progression of DR can be slowed down with early diagnosis and treatment. Therefore our aim was to develop a novel automated method for DR screening. Methods. 52 patients with diabetes mellitus were enrolled into the project. Of all patients, 39 had signs of DR. Digital retina images and tear fluid samples were taken from each eye. The results from the tear fluid proteomics analysis and from digital microaneurysm (MA) detection on fundus images were used as the input of a machine learning system. Results. MA detection method alone resulted in 0.84 sensitivity and 0.81 specificity. Using the proteomics data for analysis 0.87 sensitivity and 0.68 specificity values were achieved. The combined data analysis integrated the features of the proteomics data along with the number of detected Mas in the associated image and achieved sensitivity/specificity values of 0.93/0.78. Conclusions.As the two different types of data represent independent and complementary information on the outcome, the combined model resulted in a reliable screening method that is comparable to the requirements of DR screening programs applied in clinical routine.

6. Simplified Polynomial Neural Network for classification task in data mining

In solving classification task of data mining the traditional polynomial neural network (PNN) algorithm takes longer time while generating complex mathematical models. PNN algorithm takes the combinations two or three inputs to generates one partial description (PD) for the next layer. The output of the PDs becomes the input to the next layer. The number of PDs in each layer increases very fast, which consume lot of time for evaluation of the coefficients of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904144 |

the PDs, consume huge memory and increase complexity of the model. We propose simplified polynomial neural network (SPNN) for the task of classification. PDs for a single layer of the PNN model are developed. The outputs of these PDs along with the original inputs from the dataset are fed to a single perception model of artificial neural network (ANN) without any hidden layers. The ANN is trained with gradient descent method as well as with particle swarm optimization (PSO) technique. The results of both techniques for training are considered for the comparison of the performance. Simulation and result shows that the performance of SPNN is
better than PNN model.

7. Diagnosing Diabetic Dataset using Hadoop and K-means Clustering Techniques

The articles display how enormous measure of information in the field of social insurance frameworks can be dissected utilizing grouping method. Removing helpful data from this gigantic measure of information is profoundly compound, exorbitant, and tedious, in such territory information mining can assume a key part. Specifically, the standard information dig-ging calculations for the examination of colossal information volumes can be parallelized for speedier preparing. Methods/Statistical Analysis: This paper concentrate on how grouping calculation to be specific K-means can be utilized as a part of parallel handling stage in particular Apache Hadoop bunch (MapReduce paradigm huge) so as to dissect the gigantic information quicker. Findings: As an early point, we complete examination keeping in mind the end goal to evaluate the adequacy of the parallel preparing stages as far as execution. Applications/Improvements: Based on the final result, it shows that Apache Hadoop with K-means cluster is a promising example for versatile execution to anticipate and analyze the diabetic infections from huge measure of information. The proposed work will give an insight about the big data prediction of diabetic dataset through Hadoop. In future this technology has to be extended on cloud so as to connect various geographic districts around Tamil Nadu to predict diabetic related diseases.
Algorithms used in these paper are

C4.5 Decision tree:
C4.5 performs top down induction of Decision trees from a set of examples which have each been given a classification. Typically, a training set will be specified by the user. The root of the tree specifies an attribute to be selected and tested first, and the subordinate nodes dictate tests on further attributes. The leaves are marked to show the classification of the object they represent.
The ID3 Algorithm:

8. Predicting realizations of daily weather data for climate forecasts using the non-
parametric nearest-neighbor re-sampling technique The goal of this study was to verify the k-nearest neighbors (k-NN) approach for the prediction of daily weather sequences. This method can be employed on the assumption that weather during the target year is analogous to the weather recorded in the past.
Algorithm used during research are given below:
1.Multi-site verification of k-NN approach.
2. k-NN method.

9. Use of Data Mining Technique for Prediction of tea yield in the face of climatic change of Assam India

A research has been conducted to focus on the application of data mining techniques in tea plantation in the face of climatic change to help the farmer in taking decision for farming and Achieving the expected economic returns.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904144 |

III. PROPOSED ALGORITHM

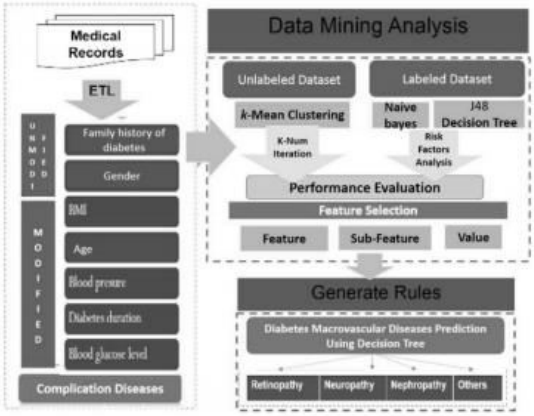


Figure 3.1: System Architecture

The system architecture for the Diabetes Management System presented below in Figure 1 is the conceptual model that defines the structure behavioural interactions, and multiple system views that underpins the system development. It presents the formal descriptions of the systems captured graphically that supports reasoning, and the submodules developed as well as the dataflows between the developed modules.

The main objective of this research is to construct models by feature selection process from modi fied and unmodified diabeties risk factor, adopt the data mining algortihm with the best accuracy per formance and also compare data mining learning technique for diabetes complication diseases There are three main phases of this research approach, which are data attribute selection and data mining pre-processing, data mining algorithm and its evaluation criteria analysis, then generated rules for the three major microvascular diabetes complication diseases to construct a general model. In the data mining analysis stage, data would undergo clustering and classification technique of data mining. In this research, data mining analysis created with the data mining programs Waikato Environment for Knowledge Analysis (WEKA) using the diabetes patient medical record.

IV. MATHMATICAL MODEL

When solving problems we have to decide the difficulty level of our problem. There are three types of classes provided for that. These are as follows

1. P Class
2. NP-hard Class
3. NP-Complete Class

• P Class:

Informally the class P is the class of decision solvable by some algorithm within a number of steps bounded by some fixed polynomial in the length of the input. Turing was not concerned with the efficiency of his machines, but rather his concern was whether they can simulate arbi trary algorithms given sufficient time. However it turns out Turing machines can generally sim ulate more efficient computer models by at most squaring or cubing the computation time.Thus P is a robust class and has equivalent definitions over a large class of computer models.

• NP-Class:

A problem is NP-hard if solving it in polynomial time would make it possible to solve all problems in class NP in polynomial time. Some NP-hard problems are also in NP some are not.If you could reduce an NP problem to an NP-hard problem and then solve it in polynomial time,you could solve all NP problem. Also, there are decision problem s in NP-hard but are not NP complete,such as the halting problems

• NP-complete:

The complexity class NP-complete is the set of problems that are the hardest problems in NP, in the sense that they are the ones most likely not to be in P. If you can find a way to solve an NP complete problem quick,then you can use that algorithm to solve all NP problems quickly.

• Summary: As we have seen all the classes of problems. Our topic is Recommendation system for career opportunities using data mining approaches is of P class because: Problem can be solve in polynomial time

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904144 |

What is p ?

• P is set of all decision problems which can be solved in polynomial time by a deterministic. • Since it can be solved in polynomial time, it can be verified in polynomial time. • Therefore P is a subset of NP.

p :

The present study focuses on the applications of Machine learning and data mining techniques in yield prediction in the face of climatic change to help the farmer in taking decision for farming and achieving the expected economic return. The problem of yield as well as disease prediction is a major problem that can be solved based on available data. Hence we proposed an system Prediction of "An efficient approach for classification prediction and Recommendation for farmers using Machine Learning".

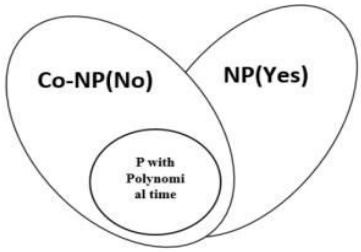


Figure:4.1.p class

What is NP?

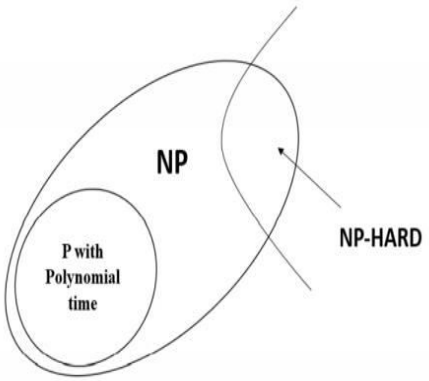
• "NP" means "we can solve it in polynomial time if we can break the normal rules of step-by step computing".

What is NP Hard?

A problem is NP-hard if an algorithm for solving it can be translated into one for solving any NP problem (non-deterministic polynomial time) problem. NP-hard therefore means "at least as hard as any NP-problem," although it might, in fact, be harder.

NP-Hard:

NP-hardness stands for Non-deterministic polynomial-time hard. Informally, "at least as hard as the hardest problems in NP" are called as NP hard class problem



4.2: NP-hard class

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904144 |

What is NP-Complete?

• Since this amazing ”N” computer can also do anything a normal computer can, we know that ”P” problems are also in ”NP”.

• So, the easy problems are in ”P” (and ”NP”), but the really hard ones are *only* in ”NP”, and they are called ”NP-complete”.

• is like saying there are things that People can do (”P”), there are things that Super People can do (”SP”), and there are things *only* Super People can do (”SP-complete”).

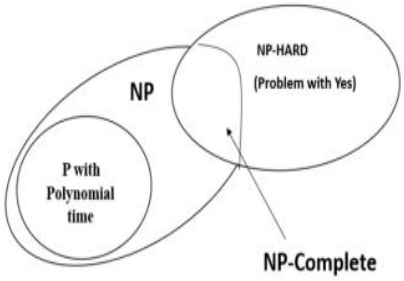


Figure 4.3: NP-complete class

NP-Complete:

In computational complexity theory Equivalently, the formal definition of NP is the set of deci sion problems solvable in polynomial time by a theoretical non-deterministic Turing machine. This second definition is the basis for the abbreviation NP, which stands for non deterministic, polyno mial time. However, the verifier-based definition tends to be more intuitive and practical in common applications compared to the formal machine definition. The two definitions are equivalent because the algorithm for the machine definition consists of two phases, the first of which consists of a guess about the solution, which is generated in a non-deterministic way, while the second phase consists of a deterministic algorithm that verifies or rejects the guess as a valid solution to the problem.

V. CONCLUSION AND FUTURE WORK

Grouping and characterization of information mining strategy and its calculation were concentrated to fabricate the expectation model of diabetes difficulty illness. The model produces rule from diabetic clinical information into four gatherings, which are nephropathy, retinopathy, neuropathy and blended entanglements (other). To assemble the most reasonable standard based model for the forecast reason, we assess the exhibition from grouping and arrangement method. It tends to be seen that contrast with grouping method, arrangement procedure gives better data, execution and could order highlights and sub element into three significant microvascular diabetes complexity infection. From the information mining investigation, we can close the most powerful danger factor for every diabetes confusion infection. Turn out that despite the fact that the blood glucose level and the length of diabetes endure lead to complexity illness, yet it's generally unmistakable on Neprophaty. It additionally presumes that glucose level and quality (family background of diabetes) turn out don't impact to explicit diabetes inconvenience. Likewise, we acquire information that the most well-known danger factor for Retinopathy are the circulatory strain in a reach hypertension emergency. With respect to Nephrophaty the most unmistakable danger factor is the span of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 4, April 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0904144 |

diabetes endure, particularly that over 10 years. Diabetes patients that overweight and fat, having more danger to Neurophaty. Given the exactness of the proposed model is 68%, with the higest precision on.

REFERENCES

1. Shaw, J. E., R. A. Sicree, and P. Z. Zimmet. (2010) "Global Estimates of the Prevalence of Diabetes For 2010 And 2030." Diabetes Res Clin Pract 87: 4–14.
2. Health Research and Development Division of Ministry of Health Republic of Indonesia. (2018) Basic Health Research Survey.
3. Data and Information Center of Ministry of Health Republic of Indonesia. (2014) Analysis and Situation of Diabetes.
4. Tarigan, Tri J.E., E. Yunir, I. Subekti, A. Laurentius, A. Pramono, and Diah Martina. (2015) "Profile and Analysis of Diabetes Chronic Complications in Outpatient Diabetes Clinic of Cipto Mangunkusumo Hospital, Jakarta." Medical Journal of Indonesia.
5. Harleen, Bhambri (2016) "A Prediction Technique in Data Mining for Diabetes Mellitus." Journal of Management Sciences And Technology
6. 4 (1).
7. Misra, (2007) "Simplified Polynomial Neural Network for Classification Task in Data Mining", in International Conf. on Evolutionary Computation, 721 – 728.
8. Sharmila, K., and S. A. Vetha Manickam. (2016) "Diagnosing Diabetic Dataset using Hadoop and K-means Clustering Techniques." Indian Journal of Science and Technolog, 9 (40). [8] Fiarni, C. (2016) "Design of Knowledge Management System for Diabetic Complication Diseases" in International Conference on Computing and Applied Informatics, IOP Publishing.
9. Huang, G-M., K-Y. Huang, T-Y. Lee, and J. Weng. (2015) "An Interpretable Rule-Based Diagnostic Classification of Diabetic
10. Nephropathy Among Type 2 Diabetes Patients." BMC Bioinforma, 16 (S-1): S5.
11. DuBrava S, J. Mardekian, A. Sadosky, E. J. Bienen, B. Parsons,
12. MD, M. Hopps, and J. Markman. (2017) "Using Random Forest
13. Models to Identify Correlates of a Diabetic Peripheral Neuropathy
14. Diagnosis from Electronic Health Record Data." Pain Medicine,18: 107-115.
15. Torok, Zsolt, Tunde Peto, Eva Csosz, Edit Tukacs, Agnes M. Molnar, Andras Berta, Jozsef Tozser, Andras Hajdu,Valeria Nagy, Balint Domokos, and Adrienne Csutak. (2015) "Combined Methods for Diabetic Retinopathy Retinopathy Screening Using Retina Photographs and Tear Fluid Proteomics Biomarkers." Journal of Diabetes Research. pp. 1-9.
16. Zhang, Bob, B.V.K. Vijay Kumar, David Zhang. (2014)
17. "Detecting Diabetes Mellitus and Nonproliferative Diabetic
18. Retinopathy Using Tongue Color, Texture and Geometry Features." IEEE Transaction on Biomedical Engineering 61 (2): 491-501.
19. Kavakiotis, I., O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and Chouvadra. (2017) "Machine Learning and Data Mining Methods in Diabetes Research." Computational and structural biotechnology Journal 15: 104-116.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details