

An Effective Diabetes Disease Prediction Model Using Machine Learning Algorithms

Dr. D. Suneetha, Professor & HOD, Department of CSE, NRI Institute of Technology, A.P
B N V Basaveswara Rao, M. Tech Student, CSE, NRI Institute of Technology, Aagiripalli, A.P

Abstract:

Diabetes mellitus is a serious health condition that affects many people. Diabetes mellitus can be caused by a variety of factors, including age, obesity, lack of physical activity, family history, poor diet, hypertension, etc. Diabetes can lead to a variety of health problems, including heart disease, kidney disease, stroke, impaired vision and nerve damage, and other conditions. Today, hospitals use a series of tests to diagnose a patient's diabetes type, and once a diagnosis is made, patients receive care that is customized to their individual needs. The current methods for classifying and predicting diabetes are not very accurate. In this paper, we provide a model for predicting diabetes that includes traditional parameters such as glucose, BMI, age and insulin levels, as well as several external factors that contribute to diabetes. The new model improves the classification precision when compared to the initial dataset. Additionally, we impose a pipeline model to predict diabetes with the aim of increasing the classification precision.

Keywords: - Classification, Diabetes System

1.INTRODUCTION:

Huge databases are common in the healthcare industry. These databases can contain organized, semi-structured or even completely unstructured information. The process of analyzing large data sets, often called "big data analytics", aims to uncover previously unknown facts and trends in order to gain new insights from the given data. Considering the current situation, diabetes mellitus (DM) has evolved into a very serious disease in developing countries like India. The condition known as diabetes mellitus (DM) is an example of a non-communicable disease (NCD), and it affects a significant number of individuals. According to the estimates from 2017, there are over 425 million

individuals living with diabetes. Diabetes is the leading cause of death worldwide, taking the lives of around 2-5 million people per year. It is projected that by the year 2045, this number would have increased to 629 million. [1] Type-1 diabetes mellitus, often known as insulin-dependent diabetes mellitus, is a subtype of diabetes mellitus (DM) (IDDM). Diabetes mellitus type 1 is characterized by the body's inability to produce an adequate amount of insulin, which is why patients with this form of the disease must get insulin injections. Diabetes mellitus type 2, sometimes referred to as non-insulin-dependent diabetes (NIDDM). Diabetes mellitus type 1 is characterized by the inability of body cells to make appropriate utilization of insulin. Diabetes type 3, often known as gestational diabetes, is caused by an increase in the amount of sugar in a pregnant woman's blood when the condition is not recognized as diabetes at an earlier stage. Diabetes Mellitus is accompanied by a number of long-term consequences. A person with diabetes is also at a significantly increased risk for a variety of health complications. A method known as Predictive Analysis makes use of data from the present as well as the past in order to gain information and make forecasts about the future. This approach is comprised of a number of different machine learning algorithms, data mining techniques, and statistical methodologies. The use of predictive analysis to data pertaining to healthcare enables crucial judgments to be made as well as predictions to be made. Machine learning and regression approach are two methods that may be used to perform predictive analytics. The goal of predictive analytics is to improve clinical outcomes while simultaneously maximizing resources, boosting patient care, and detecting diseases with the highest possible degree of precision. [1] Machine learning is widely regarded as one of the most essential characteristics of artificial intelligence. It enables the construction of computer systems that can learn from their own experiences without requiring explicit programming for each scenario. It is generally agreed that machine learning is an absolute need in the current climate in order to eradicate the need for human labor and to facilitate automation with a minimum of errors. Laboratory testing, such as blood glucose levels while fasting and oral glucose tolerance, are now the standard way for diagnosing diabetes. On the other hand, this approach requires a lot of time. Building a predictive model for diabetes using machine learning algorithms and data mining approaches is the primary emphasis of this article. The paper is structured in the following manner: In Section II, a literature overview of the prior work done on diabetes prediction is presented, as well as a classification system for machine learning algorithms. The rationale for working on this issue is presented in Section III.

In Section IV, a potential model for diabetes prediction is presented and addressed. The findings of the experiment are presented in Section V, which is then followed by a Conclusion and References.

2. LITERATURE SURVEY

2.1 A Predictive Analysis of Data Collected from Diabetic Patients Employing Machine Learning and Hadoop

Authors: Gauri D. Kalyankar, Shivananda R. Poojara and Nagaraj V. Dharwadkar

Abstract: A significant amount of data is being produced these days by many health care companies. It is required to gather, store, and interpret this data in order to draw conclusions and make important choices based on what can be learned from it. A significant number of individuals are afflicted with the non-communicable disease known as diabetes mellitus (DM), which belongs to the NCD category. Diabetes mellitus has evolved into a significant public health problem in recent years for emerging nations like India. Diabetes mellitus is one of the most dangerous illnesses since it may cause difficulties over a lengthy period of time and it can also lead to a variety of other health issues. It is essential to create a system that, with the assistance of modern technology, can save diabetes data, do analysis on that data, and identify probable dangers in accordance with those findings. The term "predictive analysis" refers to a strategy that combines a number of data mining techniques, machine learning algorithms, and statistical methods. Predictive analysis makes use of both present and historical data sets in order to acquire insight and anticipate potential dangers. In this work, a machine learning method is constructed for the Pima Indian diabetes data set using the Hadoop MapReduce environment. The goals of this work are to identify missing values within the data set and to detect patterns from the data. Because of this study, it will be possible to forecast which varieties of diabetes are most common, the associated dangers for the future, and the sorts of treatments that may be administered to patients based on their individual risk levels.

2.2 Diabetes Risk Assessment Based on Individual Lifestyle Factors

Ayush Anand and Divya Shakti are the authors of this piece.

Abstract:

Diabetes Mellitus, which is more often referred to as diabetes, has been presented as being more severe than both cancer and HIV (Human Immunodeficiency Virus). It manifests as when

elevated quantities of sugar are present in the blood for an extended length of time. In recent times, it has been suggested that it may be a role in the development of Alzheimer's disease, in addition to being a primary cause of blindness and renal failure. Within the medical research community, the issue of most interest right now is the disease's avoidance. There are a variety of approaches that have been developed to investigate the reasons for and treat diabetes. This research paper is a discussion on establishing a relationship between diabetes risk likely to be developed from a person's daily lifestyle activities such as his or her eating habits, sleeping habits, physical activity along with other indicators like BMI (Body Mass Index), waist circumference, etc. This paper is a discussion on establishing a relationship between diabetes risk likely to be developed from a person's daily lifestyle activities such as his or her eating habits, sleeping habits, physical activity along with First, a Chi-Squared Test of Independence was carried out, then the data were subjected to the machine learning method known as CART (Classification and Regression Trees), and lastly, the bias in the findings was eradicated by means of cross-validation.

2.3 Applications of Predictive Analytics in the Health Care Industry Utilizing the Resources and Methods of Machine Learning

B. Nithya and Dr. V. Ilango are the authors of this piece.

Machine learning is the way to go when we have a massive data collection on which we would want to do predictive analysis or pattern identification. This is because machine learning can learn from its own mistakes. The field of computer science that is seeing the most rapid growth right now is machine learning (ML), and the field of health informatics is very difficult. The purpose of machine learning is to create algorithms that are capable of learning and improving themselves over the course of time, as well as being able to make accurate predictions. The principles of machine learning are extensively applied in a variety of industries, but the health care business in particular has benefited greatly from the use of machine learning prediction techniques. It provides a wide range of alerting and risk management decision support tools, all of which are geared at enhancing the safety of patients and the quality of healthcare. The healthcare business is facing problems in important areas such as electronic record administration, data integration, computer-aided diagnosis, and illness prediction as a result of the desire to lower healthcare costs and the push towards individualized treatment. The field of machine learning provides a broad variety of tools, strategies, and frameworks that may be used

to overcome these issues. This article presents the results of a research on several prediction approaches and tools that may be used in the context of machine learning. This article also provides a brief overview of the applications of machine learning in a variety of fields while putting an emphasis on the prominent role it plays in the health care business.

2.4 Diagnosis of Diabetes Through the Utilization of Different Classification Mining Methods

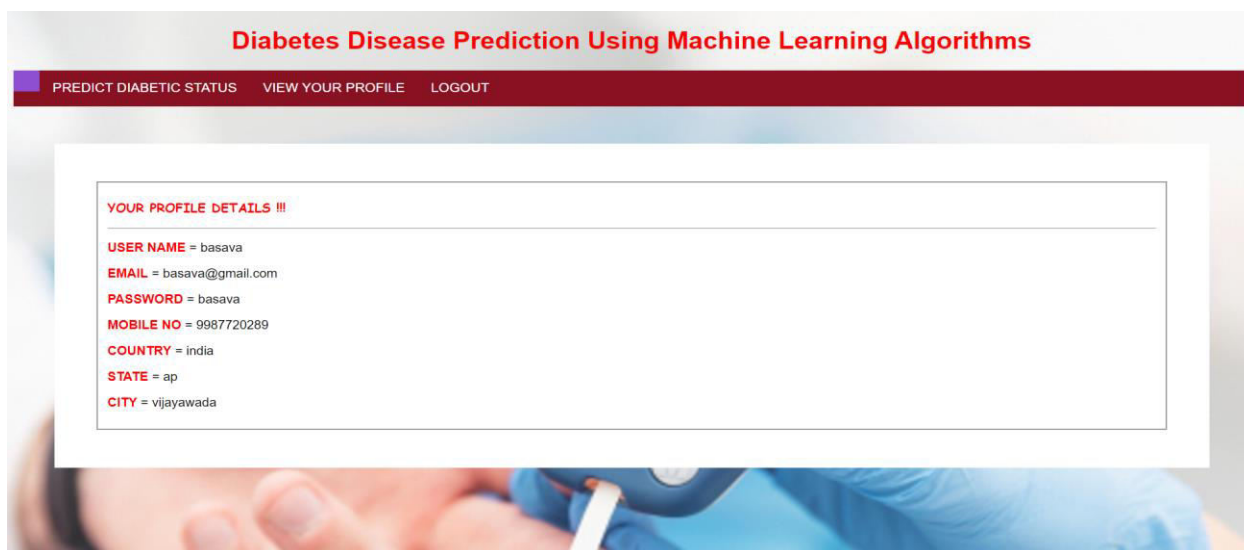
Aiswarya Iyer, S. Jeyalatha, and Ronak Sumbaly are the authors of this piece.
Abstract: Over 246 million people throughout the globe are living with diabetes, with women making up the vast majority of those afflicted. According to the data from the WHO, it is anticipated that by the year 2025, this number would have increased to more than 380 million. The illness has been ranked as the fifth most lethal disease in the United States, and there is currently no sign of an impending cure. Cases of diabetes, in addition to the symptoms that accompany it, have been meticulously recorded thanks to the emergence of information technology and its ongoing introduction into the medical and healthcare industries. Through the use of Decision Tree and Naive Bayes algorithms, the purpose of this study is to identify solutions that can be used to diagnose the illness. These solutions will be identified by doing an analysis of the patterns that can be found in the data. The purpose of the study is to investigate the possibility of developing a method that is both speedier and more effective in detecting the illness, which might ultimately result in patients receiving treatment at an earlier stage.

3. PROPOSED SYSTEM

Diabetes is more easily managed if it is diagnosed at an earlier stage. In order to accomplish this mission, the work that we are doing for this project entails making an early prediction of diabetes in a human body or in a patient by using a number of different machine learning techniques. This will ensure that the aim is met. Machine learning approaches To get more accurate results when making predictions, construct models using patient data that has been gathered. In this study, we will attempt to forecast cases of diabetes by using Machine Learning Classification and ensemble methods to a dataset. Which of the following algorithms are: K-Nearest Neighbor (KNN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Gradient Boosting (GB), and Random Forest (RF). When compared to the accuracy of other models, each model's own accuracy varies. The work done on the project provides an accurate or better accuracy model, which demonstrates that the model is capable of making an accurate prediction

of diabetes. According to the results of our analysis, Random Forest attained a greater level of accuracy when compared to other machine learning strategies. This model also predicts the diabetes in type-1 and type-2 with good accuracy and classification based on aforesaid algorithms.

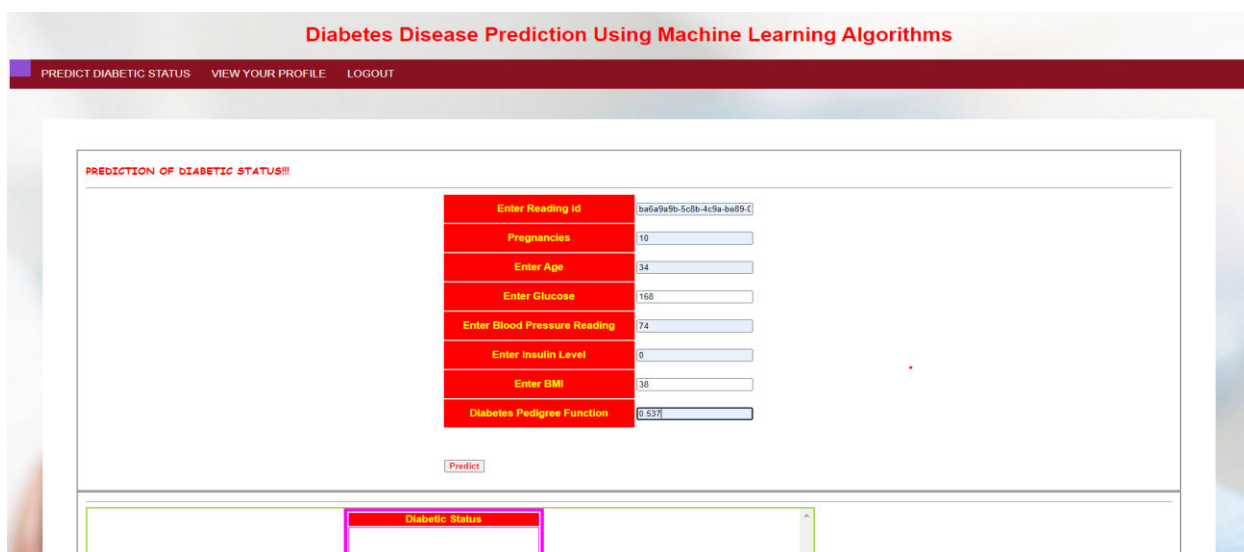
4. RESULTS



The screenshot shows the 'Diabetes Disease Prediction Using Machine Learning Algorithms' web application. The navigation bar includes 'PREDICT DIABETIC STATUS', 'VIEW YOUR PROFILE', and 'LOGOUT'. The main content area displays 'YOUR PROFILE DETAILS !!!' with the following information:

USER NAME	= basava
EMAIL	= basava@gmail.com
PASSWORD	= basava
MOBILE NO	= 9987720289
COUNTRY	= india
STATE	= ap
CITY	= vijayawada

Figure 1: Registration of user



The screenshot shows the 'Diabetes Disease Prediction Using Machine Learning Algorithms' web application. The navigation bar includes 'PREDICT DIABETIC STATUS', 'VIEW YOUR PROFILE', and 'LOGOUT'. The main content area displays 'PREDICTION OF DIABETIC STATUS!!!' with a form containing the following input fields and values:

Enter Reading Id	ba6a9a9b-5c8b-4c9a-ba89-c
Pregnancies	10
Enter Age	34
Enter Glucose	168
Enter Blood Pressure Reading	74
Enter Insulin Level	0
Enter BMI	36
Diabetes Pedigree Function	0.5371

Below the form is a 'Predict' button. At the bottom, there is a section labeled 'Diabetic Status'.

Figure 2: Results of the proposed algorithm with type-2 prediction

Diabetes Disease Prediction Using Machine Learning Algorithms

PREDICT DIABETIC STATUS VIEW YOUR PROFILE LOGOUT

PREDICTION OF DIABETIC STATUS!!!

Enter Reading Id	<input type="text"/>
Pregnancies	<input type="text"/>
Enter Age	<input type="text"/>
Enter Glucose	<input type="text"/>
Enter Blood Pressure Reading	<input type="text"/>
Enter Insulin Level	<input type="text"/>
Enter BMI	<input type="text"/>
Diabetes Pedigree Function	<input type="text"/>

Diabetic Status

Type2

Figure 3: Type-2 diabetes prediction

Diabetes Disease Prediction Using Machine Learning Algorithms

PREDICT DIABETIC STATUS VIEW YOUR PROFILE LOGOUT

PREDICTION OF DIABETIC STATUS!!!

Enter Reading Id	<input type="text"/>
Pregnancies	<input type="text"/>
Enter Age	<input type="text"/>
Enter Glucose	<input type="text"/>
Enter Blood Pressure Reading	<input type="text"/>
Enter Insulin Level	<input type="text"/>
Enter BMI	<input type="text"/>
Diabetes Pedigree Function	<input type="text"/>

Diabetic Status

Type1

Figure 4: Type-1 diabetes prediction

5.CONCLUSIONS:

The dataset in this study was subjected to several different machine learning algorithms and several different techniques were also used for classification, with the logistic regression algorithm providing the highest accuracy. After applying the sliding line, the AdaBoost classifier was determined to be the most accurate model. The accuracy of the machine learning algorithms was evaluated on two separate datasets and the results were compared. Compared to the previous data set, it is quite clear that the model improves both accuracy and precision of diabetes prediction using this data set. This approach can also be extended to determine the probability that people who do not currently have diabetes will develop diabetes in the following years.

References

- [1] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big Data Analytics in Healthcare," Hindawi Publ. Corp., vol. 2015, pp. 1–16, 2015.
- [2] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G.-Z. Yang, "Big Data for Health," *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 4, pp. 1193–1208, 2015
- [3] E. Ahmed et al., "The role of big data analytics in Internet of Things," *Comput. Networks*, vol. 129, no. December, pp. 459–471, 2017
- [4] "The big-data revolution in US health care: Accelerating value and innovation | McKinsey & Company." [Online]. Available: <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care>. [Accessed: 12-May-2018].
- [5] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities," *IEEE Access*, vol. 5, no. c, pp. 8869–8879, 2017.
- [6] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, May 2017.
- [7] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," *Appl. Stoch. Model. Bus. Ind.*, vol. 33, no. 1, pp. 3–12, Jan. 2017.
- [8] K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, "Enhanced Fingerprinting and Trajectory Prediction for IoT Localization in Smart Buildings," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 3, pp. 1294–1307, Jul. 2016.

- [9] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, "Localization Based on Social Big Data Analysis in the Vehicular Networks," *IEEE Trans. Ind. Informatics*, vol. 13, no. 4, pp. 1932–1940, Aug. 2017.
- [10] P. A. Chiarelli, J. S. Hauptman, and S. R. Browd, "Machine Learning and the Prediction of Hydrocephalus," *JAMA Pediatr.*, vol. 172, no. 2, p. 116, Feb. 2018.
- [11] A. Jindal, A. Dua, N. Kumar, A. K. Das, A. V. Vasilakos, and J. J. P. C. Rodrigues, "Providing Healthcare-as-a-Service Using Fuzzy Rule-Based Big Data Analytics in Cloud Computing," *IEEE J. Biomed. Heal. Informatics*, pp. 1–1, 2018.
- [12] N. M. S. kumar, T. Eswari, P. Sampath, and S. Lavanya, "Predictive Methodology for Diabetic Data Analysis in Big Data," *Procedia Comput. Sci.*, vol. 50, pp. 203–208, Jan. 2015.
- [13] J. Zheng and A. Dagnino, "An initial study of predictive machine learning analytics on large volumes of historical data for power system applications," in *2014 IEEE International Conference on Big Data (Big Data)*, 2014, pp. 952–959.
- [14] *International Journal of Advanced Computer and Mathematical Sciences*. Bi Publication-BioIT Journals, 2010.
- [15] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
- [16] R. A. Taylor et al., "Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big DataDriven, Machine Learning Approach," *Acad. Emerg. Med.*, vol. 23, no. 3, pp. 269–278, Mar. 2016
- [17] S. Das and A. Thakral, "Predictive analysis of dengue and malaria," in *2016 International Conference on Computing, Communication and Automation (ICCCA)*, 2016, pp. 172–176.
- [18] M. S. Simi, K. S. Nayaki, M. Parameswaran, and S. Sivadasan, "Exploring female infertility using predictive analytic," in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, 2017, pp. 1–6.
- [19] R. Lafta, J. Zhang, X. Tao, Y. Li, and V. S. Tseng, "An Intelligent Recommender System Based on Short-Term Risk Prediction for Heart Disease Patients," in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2015, pp. 102–105.

[20] S. T. Prasad, S. Sangavi, A. Deepa, F. Sairabanu, and R. Ragasudha, “Diabetic data analysis in big data with predictive method,” in 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 2017, pp. 1–4.