

www.ijprems.com

editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

5.725

e-ISSN:

Vol. 03, Issue 05, May 2023, pp : 1313-1316

PREDICTING EARLY ONSET DIABETES USING J48 ALGORITHM: A MACHINE LEARNING PERSPECTIVE

Shobhit Mohadikar¹, Gopal Malani², Prasanna Kedare³, Saurav Pardeshi⁴

^{1,2,3,4}Student, B.E Information Technology, NDMVPS ADV. Baburao Ganpatrao Thakre College of

engineering, Nashik, Maharashtra, India

DOI: https://www.doi.org/10.58257/IJPREMS31512

ABSTRACT

Diabetes, also known as Diabetic Mellitus, is a collection of metabolic disorders that impact various organs in the human body. Early prediction and control of this disease are crucial in preventing complications resulting from prolonged untreated diabetes, such as high blood sugar levels. With the exponential growth of data in the healthcare industry, Machine Learning algorithms and statistical techniques have been employed to analyze past and present data, aiding in the prediction of diabetes. These machine learning techniques offer valuable tools for doctors to detect diabetes at an early stage. In this study, patient medical records and diverse algorithms are integrated into a dataset for experimental analysis. Logistic regression, random forest, decision tree classifier, and gradient boosting algorithms are utilized to predict whether a patient, based on diagnostic measurements, has diabetes or not. Consequently, the project compares and discusses the performance and accuracy of these applied algorithms.

Keywords: Diabetes, prediction, healthcare, data, Machine Learning algorithms, J48, SVM

1. INTRODUCTION

The early stage prediction of diabetes using Machine Learning is a critical research area in healthcare, leveraging advanced algorithms and data analysis techniques to identify individuals at risk at an early stage. This topic holds immense importance due to the increasing prevalence of diabetes worldwide and its significant impact on public health. Accurate prediction at an early stage enables prompt interventions and preventive measures, mitigating the disease's progression and associated complications. Diabetes, characterized by elevated blood sugar levels and metabolic disorders, poses a substantial burden on individuals and healthcare systems globally. Machine Learning algorithms, by analyzing extensive datasets including patient medical records, can identify patterns and predictive factors contributing to diabetes development. Current research focuses on developing and evaluating algorithms that effectively analyze diagnostic measurements and clinical parameters to assess diabetes risk. Integration of Machine Learning models with electronic health records and other medical databases enhances predictive accuracy of various algorithms, such as logistic regression, random forest, decision tree classifiers, and gradient boosting, guides refinement and optimization of predictive models for early stage diabetes detection. Early stage diabetes prediction using Machine Learning plays a crucial role in proactive healthcare, enabling timely interventions and preventive strategies to combat the rising prevalence of diabetes and improve patient outcomes while reducing the burden on healthcare systems.

2. METHODOLOGY

- Further developed J48 Grouping Calculation for the Prediction of Diabetes.
- This work manages productive information digging methodology for anticipating the diabetes from clinical records of patients. Diabetes is an extremely normal illness nowadays in all populaces and in all age gatherings. Diabetes adds to coronary illness, builds the dangers of creating kidney infection, nerve harm, vein harm and visual impairment.
- In this way, mining the diabetes information in proficient way is a basic issue. The Pima Indians Diabetes Informational collection is utilized in this task; which gathers the data of patients with and without having diabetes. The altered J48 classifier is utilized to expand the precision pace of the information mining technique.

2.1 J48 ALGORITHM

J48 algorithm is a kind of decision tree which belongs to the supervised learning algorithm. It is one of the most important classifiers as it is easy and simple to implement. Using the decision tree, a dataset is broken down into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The algorithm uses equation (1) to (2) to find information gain for our dataset to predict the outcome.



www.ijprems.com

editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062 Impact Factor :

Vol. 03, Issue 05, May 2023, pp : 1313-1316

 $E(P) = -\sum_{j=1}^{n} \frac{|P_j|}{|P|} \log \frac{|P_j|}{|P|}$ $|P_j| = |P_j|$

$$E(j|P) = \frac{|P_j|}{|P|} \log \frac{|P_j|}{|P|}$$

$$Gain(P, j) = E(P - E(j|P))$$

Equation (1) and (2)

P represents total instance, n represents total number of classes, and j represents total number of attributes in the dataset.



Figure 1: J48 Algorithm

3. MODELING AND ANALYSIS

3.1 DATA FLOW



Figure 2: Dataflow Diagram



www.ijprems.com

editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062 Impact Factor : 5.725

Vol. 03, Issue 05, May 2023, pp : 1313-1316

3.2 SYSTEM ARCHITECTURE



Figure 3: System Architecture

3.3 FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS

Functional Requirements: It might include estimations, specialized subtleties, information control and handling and other explicit usefulness that characterize what a framework should achieve. Social prerequisites depict every one of the situations where the framework utilizes the practical necessities; these are caught being used cases.

Nonfunctional Requirements: (NFRs) characterize framework credits like security, dependability, execution, practicality, versatility, and ease of use. They act as imperatives or limitations on the plan of the framework across the various overabundances.

4. RESULTS AND DISCUSSION



Figure 4: Home Page

Welcome		Patient Report
	Petertit	3
	Parient Rene	Netrol Rene
Prince Karawadiya	Cardio: Multiver	biarte(ma)
	Apr	38
	Preparatas	
	Clarge	127

Figure 5: Patient Report

www.ijprems.com editor@ijprems.com	RNAL OF PROGRESSIVE ERING MANAGEMENT CE (IJPREMS) 7 2023, pp : 1313-1316	e-ISSN : 2583-1062 Impact Factor : 5.725	
New Coveral Line Margin Starts, 19 (2001) Tage Data New Coveral Line Margin Data Strap 5, 20 (2011) Tage Margin Start Margins Start Margins Start Margins Start Margins Start Starts Received Tage year between analy. Marg	All State Designed and state of the state of	Paid type	
	endate page a	Send Send	

Figure 6: Final Report

The study achieved promising results in early-stage diabetes prediction. The enhanced J48 classifier demonstrated improved accuracy in identifying individuals at risk of developing diabetes based on their clinical records. These findings highlight the potential of the proposed methodology for early detection and intervention in diabetes management.

5. CONCLUSION

In conclusion, our research successfully developed an enhanced J48 grouping calculation for early-stage diabetes prediction. The results demonstrated the effectiveness of this methodology in accurately identifying individuals at risk of developing diabetes based on their clinical records. The findings emphasize the importance of efficient data mining techniques in improving diabetes management through early detection and intervention. Further research and implementation of this approach can contribute to better healthcare outcomes for individuals at risk of diabetes. We are trying achieved more than 85% detection accuracy.

ACKNOWLEDGEMENTS

We express our heartfelt gratitude to our guide, Dr. V.R. Sonawane, for their invaluable assistance and guidance throughout our research. Their support and valuable suggestions have been instrumental in our work. We would also like to extend our thanks to Dr. V. R. Sonawane, Head of the Information Technology Department, Maratha Vidya Prasarak Samaj's Karmaveer Adv. Baburao Thakare College of Engineering, for their unwavering support and insightful suggestions. Lastly, we extend our special thanks to Dr. S. R. Devane for providing us with essential resources, including a well-equipped laboratory, vital software platforms, and a continuous Internet connection, which were indispensable for the completion of our project.)

6. **REFERENCES**

- KM Jyoti Rani, "Diabetes Prediction Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 6 Issue 4, pp. 294-305, July-August 2020. Available at Doi: https://doi.org/10.32628/CSEIT206463 Journal URL: http://ijsrcseit.com/CSEIT206463
- [2] Aljumah, A.A., Ahamad, M.G., Siddiqui, M.K., 2013. Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University - Computer and Information Sciences 25, 127–136. Doi: 10.1016/j.jksuci.2012.10.003.
- [3] Choubey, D.K., Paul, S., Kumar, S., Kumar, S., 2017. Classification of Pima Indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection, in: Communication and Computing Systems: Proceedings of the International Conference on Communication and Computing System (ICCCS 2016), pp. 451–455
- Bamnote, M.P., G.R., 2014. Design of Classifier for Detection of Diabetes Mellitus Using Genetic Programming. Advances in Intelligent Systems and Computing 1, 763–770. doi:10.1007/978-3-319-11933-5.
- [5] https://www.kaggle.com/johndasilva/diabetes
- [6] Islam, M. M. Faniqul, Ferdousi, R., Rahman, S., & Bushra, H. (2020). Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques. In Proceedings of the Conference Name (Vol. Volume Number, Issue Number, pp. 113-125). ISBN: 978-981-13-8797-5. DOI: 10.1007/978-981-13-8798-2_12.