ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

PREDICTING DIABETES WITH MACHINE LEARNING ON AWS CLOUD

S. Sundeep Kumar[1], K.Anitha[2], M.Navya[3], M.Prananya[4], K.Yashwanth Chary[5]

Assistant Professor, Department of Computer Science and Engineering[1]

Student, Department of Computer Science and Engineering[2,3,4,5]

[1,2,3,4,5]Sree Dattha Institute of Engineering and Science, Sheriguda, Telangana.

ABSTRACT

Machine learning (ML) has revolutionized various sectors, including healthcare and medical facilities, by enabling predictive analysis on large datasets obtained from diverse devices. This study explores the application of ML algorithms in the healthcare domain for predictive analysis. The research focuses on evaluating six ML algorithms—Artificial Neural Networks (ANN), XGBoost, AdaBoost, K Nearest Neighbours (KNN), Support Vector Machine (SVM), and Decision Tree (DT)—using a training dataset comprising patient health records.

The study analyzes the performance and effectiveness of these algorithms in predicting diabetes and associated symptoms. By comparing and validating these ML techniques, the research aims to identify reliable algorithms suitable for early diabetes detection. The ultimate goal is to assist medical practitioners and healthcare providers in making strategic assessments during patient prognosis and treatment.

Index Terms: Machine learning, healthcare, predictive analysis, diabetes prediction, artificial neural networks (ANN), XGBoost, AdaBoost, KNN, SVM, Decision Tree (DT).

I.INTRODUCTION

The medical sector manages vast amounts of sensitive and private data that require secure handling to prevent unauthorized access and modification. Diabetes Mellitus is increasingly prevalent due to sedentary lifestyles, making it one of the most widespread and serious global diseases. Medical practitioners urgently need robust and secure prediction systems to accurately detect and manage diabetes, enabling customized healthcare services.

Various machine learning techniques are crucial for analyzing diverse perspectives of medical data to extract valuable insights and validate their relevance. Data mining methodologies play a pivotal role in understanding the significance of information from large, accessible datasets. The primary goal is to identify patterns within datasets and derive meaningful relationships that provide actionable insights for stakeholders.

Undetected and uncontrolled diabetes can lead to severe health complications such as heart disease, kidney infections, nerve damage, and vision loss. Effectively mining diabetes-related information is critical but

complex, as subsequent analyses hinge on initial data quality.

This research employs machine learning for precise diabetes prediction, using the WEKA mining tool for effective diagnosis. The study utilizes the Pima Indian diabetes dataset from the UCI repository, analyzing it to develop a robust framework for predicting and diagnosing diabetes. The methodology includes employing bootstrapping techniques to enhance prediction accuracy and performance, followed by applying ML algorithms such as Naive Bayes (NB), Decision Tree (DT), and K-Nearest Neighbors (KNN) to validate and analyze their efficacy.

Overall, this study aims to enhance diabetes prediction capabilities in healthcare settings through advanced data mining and machine learning approaches, facilitating proactive and personalized patient care.

II.LITERATURE SURVEY

- Vishwanath S Mahalle,Aniket K Shahade proposed this paper presents hybrid (RSA & AES) encryption algorithm to safeguard data security in Cloud. Security being the most important factor in cloud computing has to be dealt with great precautions. This paper mainly focuses on the following key tasks: 1. Secure Upload of data on cloud such that even the administrator is unaware of the contents. 2. Secure Download of data in such a way that the integrity of data is maintained. 3. Proper usage and sharing of the public, private and secret keys involved for encryption and decryption. The use of a single key for both encryption and decryption is very prone to malicious attacks. But in hybrid algorithm, this problem is solved by the use of three separate keys each for encryption as well as decryption. Out of the three keys one is the public key, which is made available to all, the second one is the private key which lies only with the user. In this way, both the secure upload as well as secure download of the data is facilitated using the two respective keys. Also, the key generation technique used in this paper is unique in its own way. This has helped in avoiding any chances of repeated or redundant key.

- Yussuf Ahmed,Syed Naqvi and Mark Josephs proposed cybersecurity incidents are on the rise in the health-care sector and it is becoming a growing concern for the senior executives. The attack surface is expanding due to the large number of connected medical devices and the proliferation of portable devices such as smart phones, tablets and USB devices. In this paper, we will discuss some of the security challenges facing this sector and propose a set of cybersecurity metrics that could be used to enhance the protection of the IT systems.

- Basel Kayyali , David Knott , and Steve Van Kuiken proposed over the last decade, pharmaceutical companies have been aggregating years of research and development data into medical databases, while payors and providers have digitized their patient records. Meanwhile, the US federal government and other public stakeholders have been opening their vast stores of health-care knowledge, including data from clinical trials and information on patients covered under public

insurance programs. In parallel, recent technical advances have made it easier to collect and analyze information from multiple sources—a major benefit in health care, since data for a single patient may come from various payors, hospitals, laboratories, and physician offices.

- Mohammad Nuruzzaman Bhuiyan and Md Mahbubur Rahman proposed the Internet of Things (IoT) is a methodology or a system that encompasses real-world things to interact and communicate with each other with the assistance of networking technologies. This article describes surveys on advances in IoT-based healthcare methods and reviews the state-of-the-art technologies in detail. Moreover, this review classifies an existing IoT-based healthcare network and represents a summary of all perspective networks. IoT healthcare protocols are analyzed in this context and provide a broad discussion on it. It also initiates a comprehensive survey on IoT healthcare applications and services. Extensive insights into IoT healthcare security, its requirements, challenges, and privacy issues are visualized in IoT surrounding healthcare. In this review, we analyze security and privacy features consisting of data protection, network architecture, Quality of Services (QoS), app development, and continuous monitoring of healthcare that are facing difficulties in many IoT-based healthcare architectures. To mitigate the security problems, an IoT-based security architectural model has been proposed in this review. Furthermore, this review discloses the market opportunity that will enhance the IoT healthcare market development. To conduct the survey, we searched through established journal and conference databases using specific keywords to find scholarly works. We applied a filtering mechanism to collect only papers that were relevant to our research works. The selected papers were then examined carefully to understand their contributions/research focus. Eventually, the paper reviews were analyzed to identify any existing research gaps and untouched areas of research and to discover possible features for sustainable IoT healthcare development.

III.EXISTING SYSTEM

- Patil (2010) used K-means clustering to validate the selected label and achieved approximately 93 percent accuracy in predicting type 2 diabetes. This model can be used to detect T2DM at early stages and can also determine if the eyes are impacted due to diabetes.
- Senthil Kumar et al. created a simple and computationally efficient algorithm for gradient-based optimization of stochastic objective functions. This study focuses on the capability of AdaGrad to manage sparse gradients and the capability of RMSProp to manage nonstationary items. This scheme is less complex to deploy and requires minimal hardware.

IV.PROPOSED SYSTEM:

In this study,most significant machine learning clas- sification algorithms have

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

been employed topredict di- abetes. K Nearest Neighbours (KNN), Ada boosting , Support Vector Machine (SVM), Decision Tree (DT),Artifical neural network and Xg boosting are the ML algorithms taken into account to perform analysis and whose subsequent comparison of these machine learn- ing techniques discovers the best algorithm that assists the medical practitioners towards prediction of diabetes. Ultimately, an application interface has been designed and developed that receives user inputs to predict the existence of diabetes from their data.

V.SYSTEM ARCHITECTURE

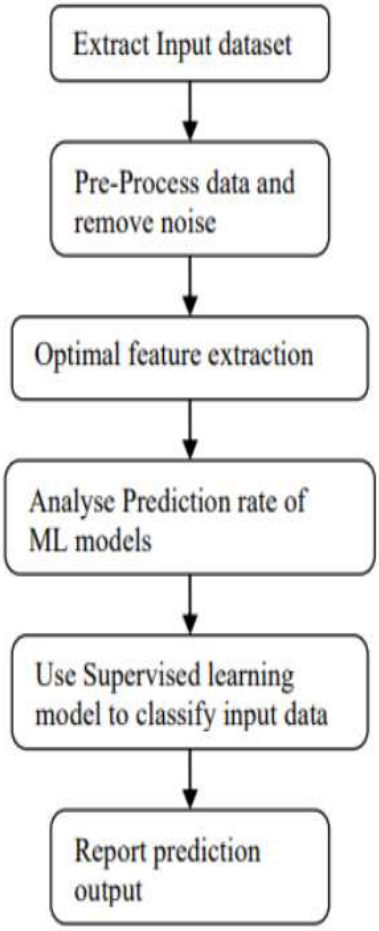


Figure.1 System Architecture

VI.IMPLEMANTATION

MODULES

- **Data Collection and Preprocessing Module:**
  This module focuses on acquiring a diverse and representative dataset of patient health records. It includes processes for cleaning, handling missing values, and preprocessing the data to ensure its quality and suitability for machine learning model training.
- **Feature Selection and Optimization Module:**
  In this module, advanced feature selection techniques are implemented to identify the most relevant factors for diabetes prediction. The goal is to enhance the efficiency of the machine learning algorithms by focusing on key features and mitigating potential biases in the dataset. Optimization techniques are also applied to fine-tune the model parameters.
- **Machine Learning Algorithm Implementation Module:**
  This module incorporates the implementation of various machine learning algorithms, including Artificial Neural Networks (ANN), XG Boosting, Ada Boosting, K Nearest Neighbours (KNN), Support Vector Machine (SVM), and Decision Tree (DT). Each algorithm is trained on the preprocessed data to create prediction models for diabetes.
- **AWS Integration and Cloud Deployment Module:**
  The integration with AWS cloud services is a crucial aspect of the system. This module involves deploying the machine learning models on the AWS platform,

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

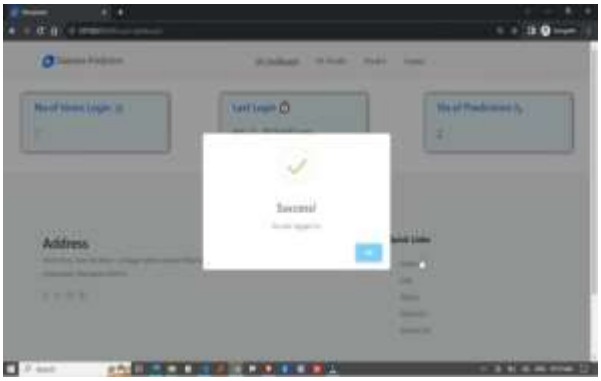
www.ijasem.org

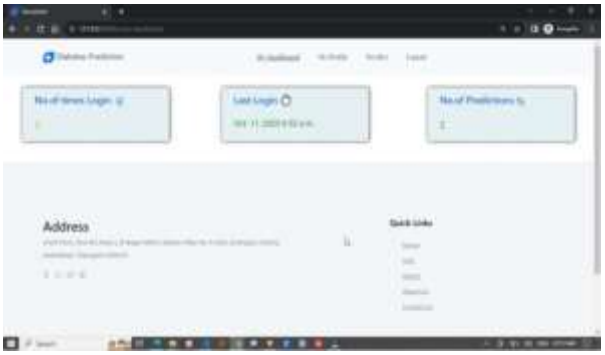
Vol 18, Issue 2, 2024

leveraging services such as Amazon SageMaker for model training and hosting. This ensures scalability, resource optimization, and efficient utilization of cloud infrastructure.

- **User Interface and Interaction Module:**
  The user interface module is designed to create a user-friendly platform for medical professionals. It allows users to input relevant patient details and obtain diabetes predictions. This module ensures seamless interaction between the machine learning models and end-users, promoting easy adoption and integration into the clinical workflow.

VII.RESULTS:



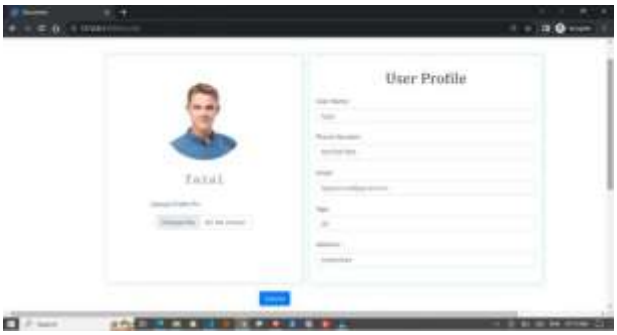


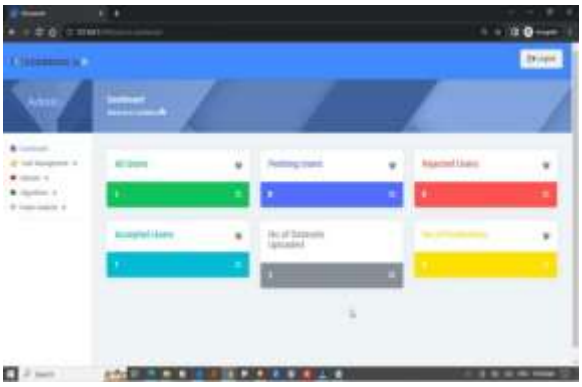
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

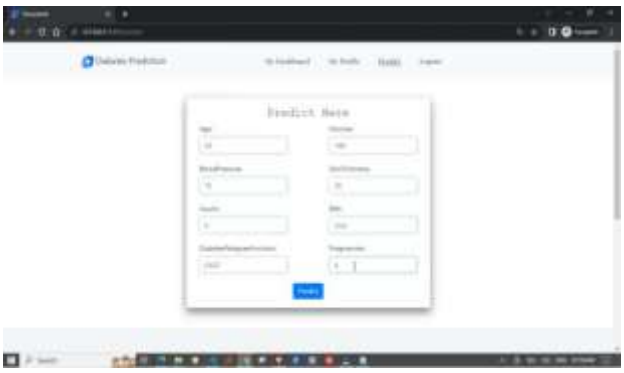
ISSN2454-9940

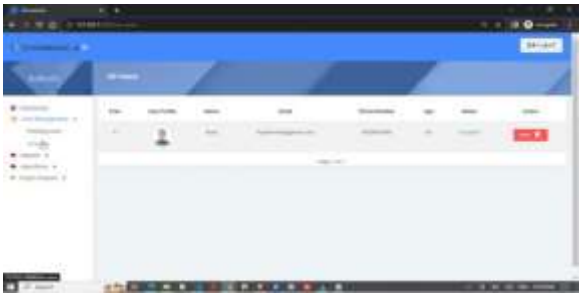
www.ijasem.org

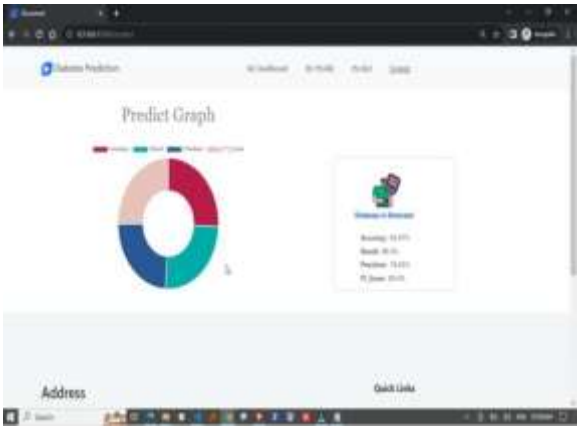
Vol 18, Issue 2, 2024

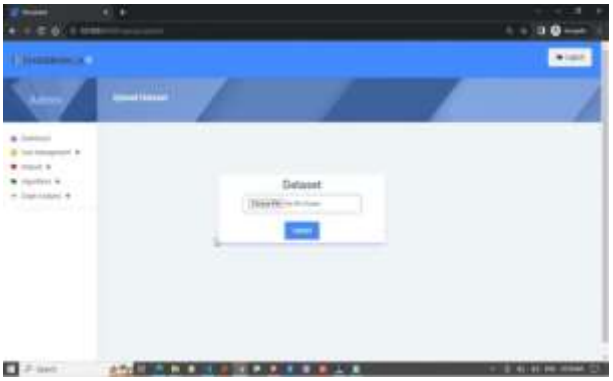


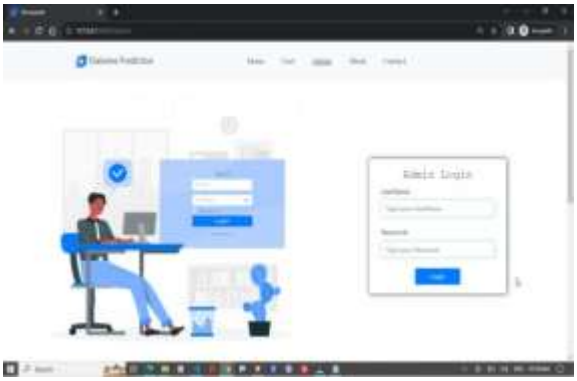










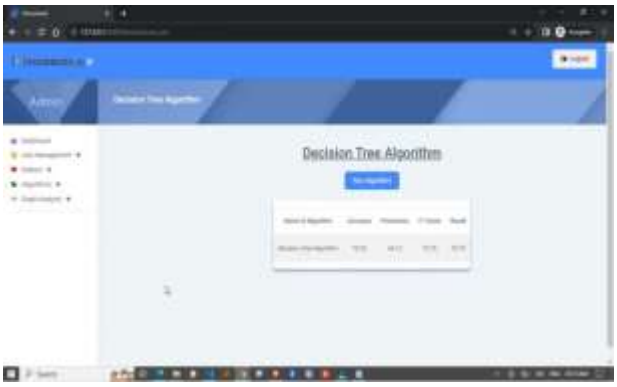


INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024



VIII.CONCLUSION

Multiple benchmark performance metrics like accuracy, precision and error in classification it will take into account to estimate the performance and efficiency of the proposed model

The acquired results are validated by comparing them with the outcomes of traditional approaches employedin health sector domain and is observed to have shown promising performance. The inputs of several diabetic patients has been obtained from UCI laboratory which is further utilised to understand and locate patterns using ML algorithms like K Nearest Neighbours (KNN), ANN, XG Boosting and ada boosting, SVM . The simulated performance are compared and evaluated for performance and accuracy aspects. The proposed model produces a staggering outcome of around 98 percent, in comparison with other conventional approaches.

IX.FUTURE ENHANCEMENT

Future research work has been on to enhance the security of the proposed system through adding intrusion detection based techniques.

X.REFERENCES

[1] A. Belle, R.. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, ―medicalcare using with Big DataAnalytics ,‖ Hindawii Publ. Corp., vol. 2015, pp. 1–16, 2015..

[2] 2.. D. Yadav, A. Shinde, A. Nair, Y. Patil and S. Kanchan, ‖Enhancing the Data Security in Cloud ,‖ 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 753-757,, doi: 10.1109/ICICCS48265..2020.9121109.

[3] 3. Y. Ahmed, S. Naqvi and M. Josephs, ‖Cybersecu- rity Metrics for Protection of medicalcare IT Systems,..‖ 2019 13th International Symposium on Medical Informa- tion and Communication Technology (ISMICT), 2019, pp. 1-9, doi: 10.1109/ISMICT.2019.8744003...

[4] ―The big-data revolution in US health care: Accelerating value and innovation — McKinsey amp; Company.‖ [Online]. Available: https://www.mckinsey.com/industries/health caresystemsandservices/ our-insights/the-big-datarevolution-in-us- healthcare. [Accessed: 12-May2018]..

[5] 5. M. N. Bhuiyan, M. M. Rahman, M. M. Billah and D. Saha, ‖Internet of Things (IoT): A Review of Its En- abling

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

Technologies in Healthcare Applications, Standard Protocols, Security, and Market Opportunities,‖ in IEEE Internet of Things Journal, vol. 8, no. 13, pp. 10474- 10498, 1 July1, 2021, doi: 10.1109/JIOT.2021.3062630. 2017.

[6] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, ―big data in machine learning : Opportunitie and challenges,‖ Neurocomputing, vol. 237, pp. 350–361, May 2017.

[7] J. B. Heatonn.., N. G. Polson, and J. H. Witte, ―Deep learning techniques for finance: deep portfolios,‖ Appl. Stoch.. Model. Bus. Ind., vol. 33, no. 1, pp. 3–12, Jan. 2017.

[8] 8. Z. Ying,, W. Jiangg, X. Liu and S. Xu, ‖Implement- ing Security-Enhanced PHR System in the Cloud Using FAME,‖ 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9014230.

[9]. L. -e. Wang, Y. Bai, Q. Jiang, V. C. M. Leung, W. Cai and X. Li, ‖Beh-Raft-Chain: A Behavior-Based Fast Blockchain Protocol for Complex Networks,‖ in IEEE Transactions on Network Science and Engineering, vol. 8, no. 2, pp. 1154-1166, 1 April-June 2021, doi: 10.1109/TNSE.2020.2984490

[10] D. Moussaoui, B. Kadri, M. Feham and B. Ammar Bensaber, ‖A Distributed Blockchain Based PKI (BCPKI) architecture to enhance privacy in VANET,‖ 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), 2021, pp. 75-79, doi: 10.1109/IHSH51661.2021.9378727

[11] 11. W. Zeng and J. Zhang, ‖Leakage-Resilient and Lightweight Authenticated Key Exchange for E- Health,‖ 2020 6th International Conference on Infor- mation Management (ICIM), 2020, pp. 162-166, doi: 10.1109/ICIM49319.2020.244691.

[12] S. K and C. Priya, ‖Enabling the efficiency of Blockchain Technology in Tele-Healthcare with En-

[13] hanced EMR,‖ 2020 International Conference on Com- puter Science, Engineering and Applications (ICCSEA), 2020, pp. 1- 6, doi: 10.1109/ICCSEA49143.2020.