IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, July 2024

Flask App for Diabetic Prediction by Improvising SVM

Sankalp Dwivedi
MTech, Department of Computer Science
Sat Kabir Institute of Technology and Management, Bahadurgarh, India

Abstract*: Chronic illnesses, due to their long-lasting and persistent nature, are among the leading causes of adult deaths worldwide. This research aims to improve diabetes prediction through an enhanced support vector machine (SVM) model. The system predicts diabetes using user-supplied data, providing precise outcomes. Earlier studies were constrained by limited datasets, focusing mainly on gestational diabetes, which affects pregnant women. This project advances the prediction system by utilizing a larger dataset and refining the SVM model, thereby increasing its accuracy in predicting diabetes in both men and women*

Keywords: Gestational, Diabetes Disease Prediction, Machine Learning, Support Vector Machine

I. INTRODUCTION

Diabetes is a widespread and serious chronic condition that impacts the entire body and is characterized by elevated blood sugar levels (Lyngdoh et al., 2021). Chronic illnesses, including diabetes, have long-term effects and significantly diminish the quality of life, representing a global health threat and economic burden (Ahmed et al., 2021; Lai et al., 2019). Type 1 diabetes, which affects 5–10% of diabetics, is due to insufficient insulin production, whereas Type 2 diabetes, which is more common, results from the body's cells not responding effectively to insulin (Deberneh& Kim, 2021).

In the United States, diabetes mellitus is a major cause of death, highlighting the necessity for early detection and diagnosis (Deberneh & Kim, 2021; Saeedi et al., 2019). As of 2019, about 463 million people aged 20 to 79 worldwide had diabetes, with projections suggesting an increase to 700 million by 2045 (International Diabetes Federation, Gulshan et al., 2016; Soni & Varma, n.d.). The disease is particularly prevalent in low- and middle-income countries, where 79% of the adult diabetic population lives. This rising number of cases underscores the urgent need for effective monitoring and management strategies (Nayak & Pandi, 2021; Perveen et al., 2016).

In India, diabetes affects around 70% of the population, with 25% of deaths attributed to the disease due to early neglect (Vizhi & Dash, 2020; Zhou et al., 2020). The primary objective of this project is to offer a convenient tool for individuals to assess their health and detect diabetes early. Many people neglect regular health check-ups due to busy lifestyles, leading to severe health outcomes. Our diabetes prediction platform, enhanced by a support vector machine (SVM), aims to address this issue by enabling users to easily check their diabetes risk and seek medical advice if necessary.

Machine learning (ML), a branch of artificial intelligence (AI), improves the accuracy and efficiency of predictions by utilizing historical data (Kaur, 2019; Kumar et al., 2022). The SVM algorithm, a supervised learning model, is particularly effective for classification tasks. It identifies a hyperplane in an n-dimensional space to separate data points, maximizing the margin between different classes (Pranto et al., 2020; Rani, 2020). For instance, in classifying two variables, the SVM selects the hyperplane that maximizes the distance to the nearest data points on either side, known as the maximum-margin hyperplane or hard margin (Shafi & Ansari, 2021)."

II. METHODOLOGY

Data Collection

The dataset utilized in this study was obtained from the UCI Machine Learning Repository. It comprises 16 attributes related to 520 patients, including both male and female participants. For analysis purposes, categorical string values like

DOI: 10.48175/568

"YES" or "NO" were converted into binary values, where 0 denotes "NO" or negative, and 1 denotes "YES" or positive. Additionally, gender was encoded with 0 for males and 1 for females.

Data Pre-processing

Data pre-processing was conducted to prepare the dataset for analysis and to obtain a thorough understanding of the data. This involved identifying and manually correcting any anomalies within the dataset. The Pandas and NumPy libraries were employed to efficiently manage the data (Bano et al., 2021). During this process, categorical values such as "YES" and "NO" were converted to binary values, with 0 representing "NO" and 1 representing "YES," while gender was encoded as 0 for males and 1 for females.

Setting Classification Metrics

To categorize the disease and obtain the final results, specific metrics are essential for predicting diabetes. In our experiment, we used the Scikit-learn machine learning library (Jakka & Vakula Rani, 2019) and employed the confusion matrix as our classification metric. The primary metric used in our analysis is accuracy, which is defined as follows (Sahoo et al., 2020):

Accuracy (A) is defined as follows.

$$\mathbf{A} = \frac{\mathbf{Tp + Tn}}{\mathbf{Tp + Tn + Fp + Fn}}$$

Precision quantifies the ratio of correctly identified true positive diabetic patients to the total number of positive predictions. Precision (P) is expressed as follows:

$$P = \frac{Tp}{Tp + Tn}$$

III. MODELING AND ANALYSIS

The system utilizes the SVM algorithm to predict the presence of diabetes. Users input their details and respond to several yes/no questions. The SVM algorithm processes this information and produces a prediction, indicating either YES or NO.

Dataset Collection: The data is obtained from the UCI dataset, which includes 16 attributes for 520 patients.

Data Pre-processing: This essential step enhances the data's efficiency and quality. It involves:

1. Missing Value Removal: This step is skipped as there are no missing values in our dataset.
2. Splitting the Data: After cleaning, the data is normalized and divided into training and testing sets.

Applying the SVM Algorithm: Once pre-processing is complete, the SVM algorithm is applied to the UCI dataset to evaluate its accuracy.

User Activity Flow:

1. **Enter the Details**: Users provide details such as age, gender, and symptoms in a yes/no format.
2. **Match Values:** The entered values are compared with the database using the SVM algorithm.
3. **Output Generation:** The system processes the inputs with the SVM algorithm and displays the result to the user as either YES or NO

IV. IMPLEMENTATION

1. Import Dependencies: Begin by importing the necessary libraries and dependencies.
2. Load Dataset: Load the diabetes dataset into a pandas DataFrame.
3. Standardize Data: Use the scaler.transform() function to standardize the data.
4. Split Dataset: Split the dataset into training and testing sets using the train_test_split() function.
5. Set Up SVM: Initialize the SVM classifier with a linear kernel using SVC(kernel='linear').
6. Train Model: Train the model with the training data using the fit() function of the classifier.
7. Predict Outcomes: Perform predictions on the test set using the predict() function of the classifier.

1. Environment used:
a. Web App: Visual Studio Code
b. Model Training: Jupyter Notebook
c. Languages & Libraries used:
2. Web App:
a. Front-end: HTML5, CSS3, Bootstrap v4.5
b. Back-end: Flask v1.1.2
3. Model Training:
a. Language: Python v3.8
b. Libraries: pandas v1.3.2, numpy v1.19.0, seaborn v0.11.2, matplotlib v3.4.3, scikit_learn v0.24.2

V. CONCLUSION

Diabetes is a serious chronic disease that impacts the entire body. Various machine learning techniques, such as SVM, logistic regression, KNN, and XGBoost, can be used to predict the disease. In our research, we developed a diabetes prediction system using SVM. Earlier versions of the dataset had a limited number of features and focused primarily on gestational diabetes, providing medical details to assess the patient's health. In our current implementation, we expanded the dataset with additional features, improving the SVM accuracy to 93.26%. This dataset includes both male and female patients and is designed to help assess the risk of developing diabetes.

VI. RESULTS AND DISCUSSION

Our Diabetes Disease Prediction system, utilizing an enhanced SVM algorithm, achieved an accuracy of 93.26%. This was achieved by incorporating 16 attributes from both male and female patients to improve the SVM's performance.

REFERENCES

[1]. Zhou, H., Myrzashova, R., & Zheng, R. (2020). Diabetes prediction model based on an enhanced deep neural network. Eurasip Journal on Wireless Communications and Networking, 2020(1). 10.1186/s13638-020-01765-7Chan, Syed Hasnain Alam Kazmi, Hao Fu, “Financial Fraud Detection Model: Based on Random Forest,” International Journal of Economics and Finance, Vol. 7, Issue. 7, pp. 178-188, 2015.

[2]. Lai, H., Huang, H., Keshavjee, K., Guergachi, A., & Gao, X. (2019). Predictive models for diabetes mellitus using machine learning techniques. BMC Endocrine Disorders, 19(1), 101.Advance online publication. doi:10.1186/ s12902-019-0436-6 PMID:31615566 Lyngdoh, A. C., Choudhury, N. A., & Moulik, S. (2021).

[3]. Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance Analysis of Data Mining Classification Techniques to Predict Diabetes. Procedia Computer Science, 82, 115–121. doi:10.1016/j.procs.2016.04.016

[4]. Pranto, B., Mehnaz, S. M., Mahid, E. B., Sadman, I. M., Rahman, A., & Momen, S. (2020). Evaluating machine learning methods for predicting diabetes among female patients in Bangladesh. Information (Switzerland), 11(8),374. Advance online publication. doi:10.3390/info11080374

[5]. Rani, K. J. (2020). Diabetes Prediction Using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 294–305. 10.32628/CSEIT206463Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala

[6]. Deberneh, H. M., & Kim, I. (2021). Prediction of type 2 diabetes based on machine learning algorithm. International Journal of Environmental Research and Public Health, 18(6), 3317. Advance online publication.doi:10.3390/ijerph18063317 PMID:33806973

DOI: 10.48175/568