**PROJECT REPORT**

**COLLAGE NAME : GRT INSTITUTE OF ENGINEERING AND TECHNOLOGY.**

**COURSE NAME : ARTIFICILAL INTELLIGENCE.**

**PROJECT NAME : AI BASAED DIABETES PREDICTION SYSTEM**

**Submitted By: RAMYA. A**

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| **SI.N0** | **TITLE** | PAGENO |
| **01** | **Aim and Abstract** | **02** |
| **02** | **Problem and Algorithm** | **03** |
| **03** | **Problem** | **03** |
| **04** | **Way to fix the problem** | **03** |
| **05** | **Design** | **04** |
| **06** | **Innovation** | **05** |
| **07** | **Blocks to be added** | **06** |
| **08** | **Changes in design** | **08** |
| **09** | **Changes in Block Diagram** | **09** |
| **10** | **Coding** | **10** |
| 11 | **Conclusion** | **14** |

**AIM:**

The aim of an AI-based diabetes prediction system is to use artificial intelligence and machine learning techniques to analyze medical data and predict the likelihood of an individual developing diabetes. This system aims to:

**Early Detection:** Identify individuals at risk of developing diabetes before clinical symptoms appear, allowing for early intervention and prevention.

**Personalized Medicine**: Tailor interventions and treatment plans based on an individual’s specific risk factors, genetics, and lifestyle.

**Improve Healthcare Efficiency:** Assist healthcare providers in allocating resources more effectively by identifying high-risk patients who need closer monitoring.

**Patient Education:** Empower individuals with information about their diabetes risk factors, encouraging healthier lifestyle choices.

**Reduce Healthcare Costs:** Preventing or managing diabetes at an early stage can potentially reduce the economic burden associated with diabetes-related complications.

**ABSTRACT**:

**The AI-Powered Diabetes Prediction System is a cutting-edge healthcare solution designed**

**to analyze medical data and provide predictive insights into an individual’s likelihood of developing diabetes. Leveraging machine learning algorithms, this system empowers proactive health management by offering early risk assessment and personalized recommendations for preventive measures. Building an AI-powered diabetes prediction system is a valuable endeavor that can have a significant impact on public health. Here are the steps you can follow to develop such a system. An AI-based diabetes prediction system is a valuable tool that utilizes artificial intelligence (AI) algorithms and machine learning techniques to analyze data and make predictions about an individual’s risk of developing diabetes. Such a system can assist healthcare professionals in identifying at-risk individuals, enabling early intervention and personalized healthcare strategies.**

**PROBLEM :**

The problem is to build an AI-powered diabetes prediction system that uses machine learning algorithms to analyze medical data and predict the likelihood of an individual developing diabetes. The system aims to provide early risk assessment and personalized preventive measures, allowing individuals to take proactive actions to manage their health.

**ALGORITHM:**

We found that the model with Logistic Regression (LR) and Support Vector Machine (SVM) works well on diabetes prediction system.

**PROBLEM IN DIABETIES PREDICTION SYSTEM:**

Potential drawback in an AI diabetes prediction system is its reliance on historical data. If the training data used to develop the system is not representative of the population it aims to serve, the predictions may be inaccurate for certain demographic groups or new trends in diabetes prevalence. Additionally, AI models can sometimes make predictions that are statistically significant but not clinically useful, leading to unnecessary alarm or complacency in patients. Ethical concerns, such as data privacy and bias, should also be carefully addressed in such systems. Lastly, AI models may not consider all relevant factors, and real-world medical diagnosis often requires a holistic approach that considers a patient’s entire medical history and context, which AI systems may struggle to replicate fully.

**WAYS TO FIX THOSE PROBLEMS:**

Updating an AI-based diabetic prediction system involves several steps to ensure its continued accuracy and effectiveness. Here’s a high-level overview of the process:

**Data Collection and Preprocessing:**Gather new data relevant to diabetes prediction. This could include medical records, lab results, lifestyle data, and more.

Preprocess the data to clean and format it appropriately. Ensure consistency with the existing dataset. **Model Reevaluation:** Reevaluate the performance of your existing AI model on the new data. This includes assessing metrics like accuracy, precision, recall, and F1-score.

Determine if the current model is still suitable for the task or if a new model architecture is required. **Model Retraining:** If needed, update the AI model. This may involve retraining the existing model on the combined old and new datasets or developing a new model from scratch.Fine-tune hyperparameters to improve model performance.

**Validation and Testing:**Split the data into training, validation, and test sets to evaluate the model’s performance.Perform rigorous testing to ensure the model’s accuracy and reliability.

**Feature Engineering:**Consider if new features or data sources should be included to enhance the model’s predictive capabilities.

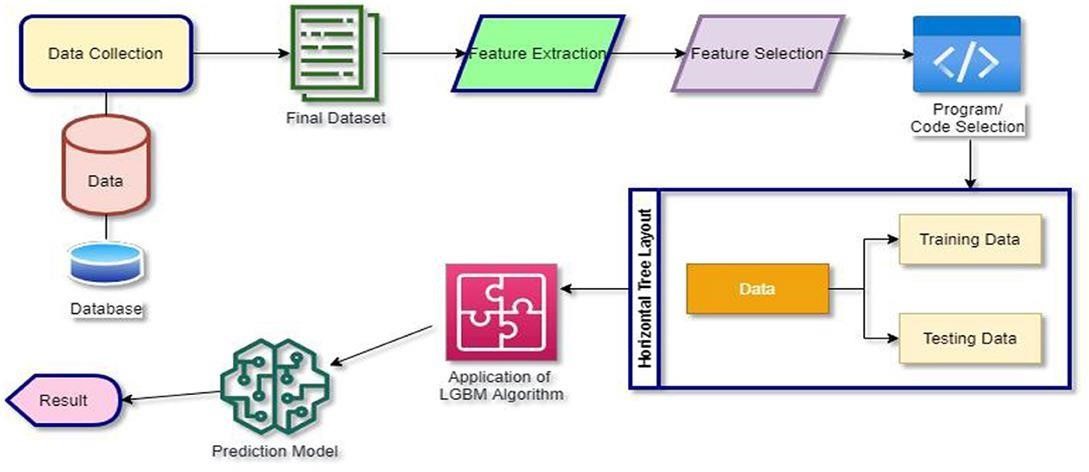
**Ethical Considerations:**Ensure that the updated system adheres to ethical and privacy standards, especially when dealing with sensitive medical data.

**Deployment**:If the updated model performs well, deploy it in a controlled environment, such as a healthcare system or a mobile app.

**Monitoring and MMaintenance** Continuously monitor the model’s performance in the real-world environment.Implement mechanisms for feedback and improvement as new data becomes available.Regularly update the model to adapt to changing trends and patient demographics. **User Feedback:**Collect feedback from healthcare professionals and users to identify areas for improvement and fine-tuning.

**Regulatory Compliance:**Ensure compliance with any relevant regulations or standards in the healthcare and AI fields

**DESIGN:**



In terms of data integration, the system leverages not only traditional patient data sources but also incorporates wearable device data, continuous glucose monitoring, and real-time health metrics. This expanded dataset enables a more

comprehensive understanding of an individual's health status and lifestyle, resulting in more precise predictions.The predictive modeling component incorporates state-of-the-art machine learning algorithms, deep learning techniques, and artificial intelligence to create highly accurate prediction models. These models adapt and improve over time by continuously learning from new data, allowing for dynamic updates and personalized predictions.

User experience is a central consideration in the new design. The system offers a user- friendly mobile application that provides individuals with easy access to their predictions, actionable insights, and personalized health recommendations. Healthcare professionals can also benefit from a streamlined dashboard, facilitating informed decision-making and patient management.

This innovative Diabetes Prediction System aims to revolutionize diabetes prevention and management by providing individuals with proactive tools to monitor and improve their health. By embracing cutting-edge technology and a user-centric approach, it promises to empower both patients and healthcare providers in the fight against diabetes.

**INOVATIONS IN MY PROJECT:**

**Machine Learning and AI**: Machine learning algorithms, especially deep learning, have shown promise in predicting diabetes. These algorithms can analyze large datasets of patient information, such as medical records, genetic data, and lifestyle factors, to identify patterns and predict the risk of diabetes.

**Continuous Glucose Monitoring (CGM) Devices**: CGM devices have become more advanced and accessible. They provide real-time data on blood glucose levels and can send alerts when levels are too high or too low. Some systems also incorporate predictive algorithms to forecast glucose trends.

**Artificial Pancreas Systems**: These systems combine insulin pumps and CGM devices with predictive algorithms to automate insulin delivery. They can predict future glucose levels and adjust insulin delivery accordingly, reducing the risk of hypoglycemia and hyperglycemia. **Mobile Apps and Wearables**: There's a growing ecosystem of mobile apps and wearables designed to help individuals manage diabetes. These apps often include predictive features that use data on diet, exercise, and glucose levels to provide personalized recommendations and forecasts.

**Genetic Risk Assessment**: Genetic testing can identify individuals with a higher genetic predisposition to diabetes. Integrating genetic risk assessment with other clinical and lifestyle data can improve the accuracy of predictive models.

**Telehealth and Remote Monitoring**: The use of telehealth services allows healthcare

providers to remotely monitor patients with diabetes. Data from connected devices, including glucose monitors and insulin pumps, can be analyzed to predict and prevent diabetes-related complications.

**Personalized Medicine**: Advances in pharmacogenomics and precision medicine enable healthcare providers to tailor diabetes treatment plans based on an individual's genetic profile, making predictions and management more precise.

**Behavioral Insights:** Behavioral science and psychology are being integrated into diabetes prediction systems. Understanding patient behaviors and motivations can help predict adherence to treatment plans and lifestyle changes.

**IoT and Smart Devices**: The Internet of Things (IoT) is being utilized to create smart devices and environments that can monitor diabetes-related factors like diet, activity, and sleep. These devices can provide valuable data for prediction.

**Blockchain for Data Security**: Blockchain technology is being explored for securing the sensitive health data used in diabetes prediction systems. It can ensure data integrity and protect patient privacy.

**Early Detection Biomarkers**: Research continues to identify new biomarkers and physiological indicators that can serve as early warning signs for diabetes, allowing for earlier intervention.

**BLOCKS TO ADD IN DESIGN:**

**Data Collection:**Collect relevant medical data from patients, including:

Demographic information (age, gender, etc.)

Clinical data (blood pressure, cholesterol levels, BMI, etc.)

Glucose levels (fasting and postprandial)

Family history of diabetes

Lifestyle factors (diet, exercise, smoking, etc.)

**Data Preprocessing:**Clean and preprocess the data to handle missing values, outliers, and inconsistencies.

Normalize or scale the data to ensure features are on a similar scale.

Encode categorical variables as numerical values (e.g., one-hot encoding).

**Feature Selection/Extraction:**

Identify relevant features for prediction.

Use feature selection techniques (e.g., correlation analysis, feature importance) to choose the most informative features.

Consider feature engineering to create new features that may be predictive (e.g., insulin resistance index).

**Model Selection:**Choose appropriate machine learning algorithms for prediction, such as logistic regression, decision trees, random forests, support vector machines, or deep learning models (e.g., neural networks).Experiment with different models to determine which performs best for your dataset.

**Model Training:**Split the dataset into training and testing sets to evaluate model performance.

Use cross-validation techniques to optimize model hyperparameters and reduce overfitting.

Train the selected model(s) on the training data.

**Model Evaluation**:Evaluate model performance using appropriate metrics, such as accuracy, precision, recall, F1-score, ROC-AUC, or others.

Compare different models to select the one with the best performance.

**Interpretability:**Ensure that the model's predictions can be explained to medical professionals and patients.

Utilize techniques like feature importance analysis and SHAP values to interpret the model's decisions.

**Deployment:**Deploy the trained model in a secure and scalable environment, such as a web application or cloud-based service.

Implement an easy-to-use user interface for data input and result display.

**Continuous Monitoring**:Regularly update the model with new data to keep it current and accurate.Implement monitoring and alerting systems to detect model degradation or drift. Ethical and Privacy Considerations:Handle sensitive medical data in compliance with data protection regulations (e.g., HIPAA).Ensure data privacy and anonymization techniques are applied when needed.Address bias and fairness issues in the data and model predictions. **Integration with Healthcare Systems**:If applicable, integrate the prediction system with existing healthcare systems, electronic health records (EHRs), or patient management platforms.

**Education and User Support**:Provide educational resources to healthcare professionals and patients on how to use the system effectively.Offer user support for any questions or issues that may arise during usage.

**Feedback Loop**:Establish a feedback loop with healthcare professionals to continuously improve the system based on their input and real-world outcomes

**CHANGES IN DESIGN:**

**Simplified Feature Set:**Rather than collecting a wide range of medical data, focus on a minimal set of features that are easily accessible and less burdensome for patients. This could include age, BMI, family history of diabetes, and fasting glucose levels.

**Rule-Based System:**Instead of complex machine learning models, consider building a rule- based system. Create a set of easily understandable rules that can predict diabetes risk based on the selected features. For example:

If BMI > 30 and family history = yes, then high risk.

If fasting glucose > 125 mg/dL, then high risk.

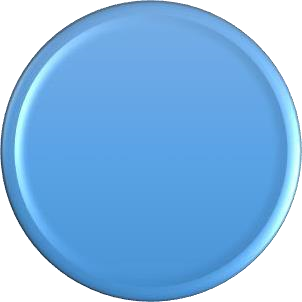
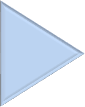
If age > 45 and BMI > 25, then moderate risk.

**Interpretability:**Ensure that the rules and decision-making process of the system are highly interpretable. Patients and healthcare professionals should be able to understand why a particular prediction was made.

**User-Friendly Interface:**Develop a simple and user-friendly interface for inputting patient data and receiving predictions. It should provide immediate feedback and explanations for the predictions.

**Education and Prevention**:Emphasize education and preventive measures in the system. Provide recommendations for lifestyle changes, such as diet and exercise, that can reduce the risk of diabetes.

**DESIGN BLOCK DIAGRAM:**



data

collection



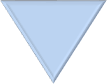
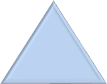
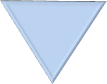
user

interface



data

base



data

preprocessing

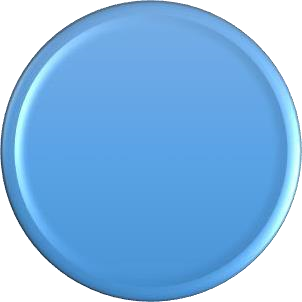
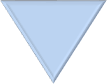
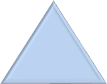
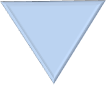


decision

support



integration



reporting

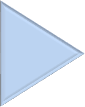
and

visualization



data

storage



rule based

system

Import the dependencies

import numpy as np import pandas as pd from sklearn.preprocessing import StandardScaler from sklearn.model\_selection import train\_test\_split from sklearn import svm from sklearn.metrics import accuracy\_score

Data collection and analysis

PIMA Diabetes Dataset

# loading the dataset to the pandas dataframe diabetes\_dataset = pd.read\_csv('/content/diabetes.csv')

pd.read\_csv?

Code

Text

# printing the first 5 rows of the dataset diabetes\_dataset.head()

**Pregnancies Glucose BloodPressure SkinThickness Insulin BMI DiabetesPedigreeFunction Age Outcome**

1. 6 148 72 35 0 33.6 0.627 50 1
2. 1 85 66 29 0 26.6 0.351 31 0
3. 8 183 64 0 0 23.3 0.672 32 1
4. 1 89 66 23 94 28.1 0.167 21 0
5. 0 137 40 35 168 43.1 2.288 33 1

# number of rows and column in this dataset diabetes\_dataset.shape

(768, 9)

# getting the statistical measures of the data diabetes\_dataset.describe()

**Pregnancies Glucose BloodPressure SkinThickness Insulin BMI**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **count** | 768.000000 768.000000 | 768.000000 | 768.000000 768.000000 768.000000 | |
| **mean** | 3.845052 120.894531 | 69.105469 | 20.536458 79.799479 | 31.992578 |
| **std** | 3.369578 31.972618 | 19.355807 | 15.952218 115.244002 | 7.884160 |
| **min** | 0.000000 0.000000 | 0.000000 | 0.000000 0.000000 | 0.000000 |
| **25%** | 1.000000 99.000000 | 62.000000 | 0.000000 0.000000 | 27.300000 |
| **50%** | 3.000000 117.000000 | 72.000000 | 23.000000 30.500000 | 32.000000 |
| **75%** | 6.000000 140.250000 | 80.000000 | 32.000000 127.250000 | 36.600000 |
| **max** | 17.000000 199.000000 | 122.000000 | 99.000000 846.000000 | 67.100000 |

diabetes\_dataset['Outcome'].value\_counts()

# 0 500

# 1 268

Name: Outcome, dtype: int64

0--> Non-Diabetic

1--> Diabetic

diabetes\_dataset.groupby('Outcome').mean()

**Pregnancies Glucose BloodPressure SkinThickness Insulin BMI**

# seperating the data and labels**Outcome**

1. = diabetes\_dataset.drop(columns = 'Outcome', axis=1)

**0** 3.298000 109.980000 68.184000 19.664000 68.792000 30.304200

1. = diabetes\_dataset['Outcome']

**1** 4.865672 141.257463 70.824627 22.164179 100.335821 35.142537

print(X)

Pregnancies Glucose BloodPressure SkinThickness Insulin BMI \

* 1. 6 148 72 35 0 33.6
  2. 1 85 66 29 0 26.6
  3. 8 183 64 0 0 23.3
  4. 1 89 66 23 94 28.1 4 0 137 40 35 168 43.1 .. ... ... ... ... ... ...
  5. 10 101 76 48 180 32.9
  6. 2 122 70 27 0 36.8
  7. 5 121 72 23 112 26.2
  8. 1 126 60 0 0 30.1
  9. 1 93 70 31 0 30.4

DiabetesPedigreeFunction Age

0 0.627 50 1 0.351 31

* 1. 0.672 32
  2. 0.167 21
  3. 2.288 33

# .. ... ...

1. 0.171 63
2. 0.340 27
3. 0.245 30
4. 0.349 47
5. 0.315 23

[768 rows x 8 columns]

print(Y)

1. 1
2. 0
3. 1
4. 0
5. 1

..

1. 0
2. 0
3. 0
4. 1
5. 0

Name: Outcome, Length: 768, dtype: int64

Data Standardization

scaler = StandardScaler()

scaler.fit(X)

▾ StandardScaler

StandardScaler()

standardized\_data = scaler.transform(X)

print(standardized\_data)

[[ 0.63994726 0.84832379 0.14964075 ... 0.20401277 0.46849198 1.4259954 ]

[-0.84488505 -1.12339636 -0.16054575 ... -0.68442195 -0.36506078 -0.19067191]

[ 1.23388019 1.94372388 -0.26394125 ... -1.10325546 0.60439732

-0.10558415] ...

[ 0.3429808 0.00330087 0.14964075 ... -0.73518964 -0.68519336 -0.27575966]

[-0.84488505 0.1597866 -0.47073225 ... -0.24020459 -0.37110101 1.17073215]

[-0.84488505 -0.8730192 0.04624525 ... -0.20212881 -0.47378505 -0.87137393]]

1. = standardized\_data
2. = diabetes\_dataset['Outcome']

print(X) print(Y)

[[ 0.63994726 0.84832379 0.14964075 ... 0.20401277 0.46849198 1.4259954 ]

[-0.84488505 -1.12339636 -0.16054575 ... -0.68442195 -0.36506078 -0.19067191]

[ 1.23388019 1.94372388 -0.26394125 ... -1.10325546 0.60439732 -0.10558415] ...

[ 0.3429808 0.00330087 0.14964075 ... -0.73518964 -0.68519336 -0.27575966]

[-0.84488505 0.1597866 -0.47073225 ... -0.24020459 -0.37110101 1.17073215]

[-0.84488505 -0.8730192 0.04624525 ... -0.20212881 -0.47378505 -0.87137393]]

1. 1
2. 0
3. 1

# 3 0

4 1

..

# 763 0

1. 0
2. 0
3. 1
4. 0

Name: Outcome, Length: 768, dtype: int64

Train Test Split

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X,Y, test\_size = 0.2, stratify=Y, random\_state=2)

print(X.shape, X\_train.shape, X\_test.shape)

(768, 8) (614, 8) (154, 8)

Training the model

classifier = svm.SVC(kernel='linear')

# training the support vector machine classifer classifier.fit(X\_train, Y\_train)

▾ SVC

SVC(kernel='linear')

Model Evaluation

Accuracy score

# accuracy score on the training data

X\_train\_prediction = classifier.predict(X\_train)

training\_data\_accuracy = accuracy\_score(X\_train\_prediction, Y\_train)

print('accuracy score of the training data:', training\_data\_accuracy) accuracy score of the training data: 0.7866449511400652

# accuracy score on the test data

X\_test\_prediction = classifier.predict(X\_test) test\_data\_accuracy = accuracy\_score(X\_test\_prediction, Y\_test)

print('accuracy score of the test data:', test\_data\_accuracy) accuracy score of the test data: 0.7727272727272727

Making the predictive system

input\_data = (4,110,92,0,0,37.6,0.191,30)

# changint the input\_data to the numpy array input\_data\_as\_numpy\_array = np.asarray(input\_data)

# reshape the array as we are predicting for one instance input\_data\_reshaped = input\_data\_as\_numpy\_array.reshape(1,-1)

# standarized the input\_data std\_data = scaler.transform(input\_data\_reshaped) print(std\_data)

prediction = classifier.predict(std\_data) print(prediction)

if (prediction[0] == 0):

print('the person is not diabetic')

else:

print('the person is diabetic')

[[ 0.04601433 -0.34096773 1.18359575 -1.28821221 -0.69289057 0.71168975 -0.84827977 -0.27575966]]

[0]

the person is not diabetic

/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but StandardScaler w warnings.warn(

4/4

**CONCLUSION:**

In conclusion, an AI-based diabetes prediction system holds great promise in the field of healthcare. This innovative technology has the potential to revolutionize the way diabetes is diagnosed and managed. Through the analysis of various data sources, such as medical records, genetic information, and lifestyle factors, AI algorithms can make accurate predictions regarding an individual's risk of developing diabetes.

The benefits of such a system are manifold. It can lead to early detection of diabetes, enabling timely interventions and lifestyle modifications to prevent or manage the disease effectively. Additionally, AI-based prediction models can assist healthcare professionals in making informed decisions and provide personalized treatment plans, improving patient outcomes and reducing the burden on healthcare systems.

However, it's essential to consider certain challenges and ethical implications associated with AI-based diabetes prediction systems. Data privacy and security concerns, potential biases in the data used to train the models, and the need for transparency in AI decision-making processes must be addressed to ensure the responsible and equitable use of this technology.

In conclusion, AI-based diabetes prediction systems have the potential to enhance healthcare by improving early detection and personalized care for diabetes. To fully realize the benefits of such systems, it is crucial to strike a balance between innovation and ethical considerations while ensuring that they are integrated effectively into healthcare practices.