PROJECT REPORT

DIABETES PREDICTION MODEL FOR INDIAN WOMEN

SREE VAISHNAVI.G

PROJECT: DIABETES PREDICTION FOR INDIAN WOMEN

PROJECT DOMAIN: MACHINE LEARNING

CONTENTS:

1.INTRODUCTION

1.1 OVERVIEW

Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. According to International Diabetes Federation 382 million people are living with diabetes across the whole world. By 2035, this will be doubled as 592 million. Diabetes mellitus or simply diabetes is a disease caused due to the increase level of blood glucose. Various traditional methods, based on physical and chemical tests, are available for diagnosing diabetes. However, early prediction of diabetes is quite challenging task for medical practitioners due to complex interdependence on various factors as diabetes affects human organs such as kidney, eye, heart, nerves, foot etc. Data science methods have the potential to benefit other scientific fields by shedding new light on common questions. One such task is to help make predictions on medical data. Machine learning is an emerging scientific field in data science dealing with the ways in which machines learn from experience. The aim of this project is to develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning techniques. This project aims to predict diabetes via three different supervised machine learning methods including: SVM, Logistic regression, ANN. This project also aims to propose an effective technique for earlier detection of the diabetes disease.

2.LITERATURE SURVEY:

2.1EXISTING PROBLEM:

In this, we need to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

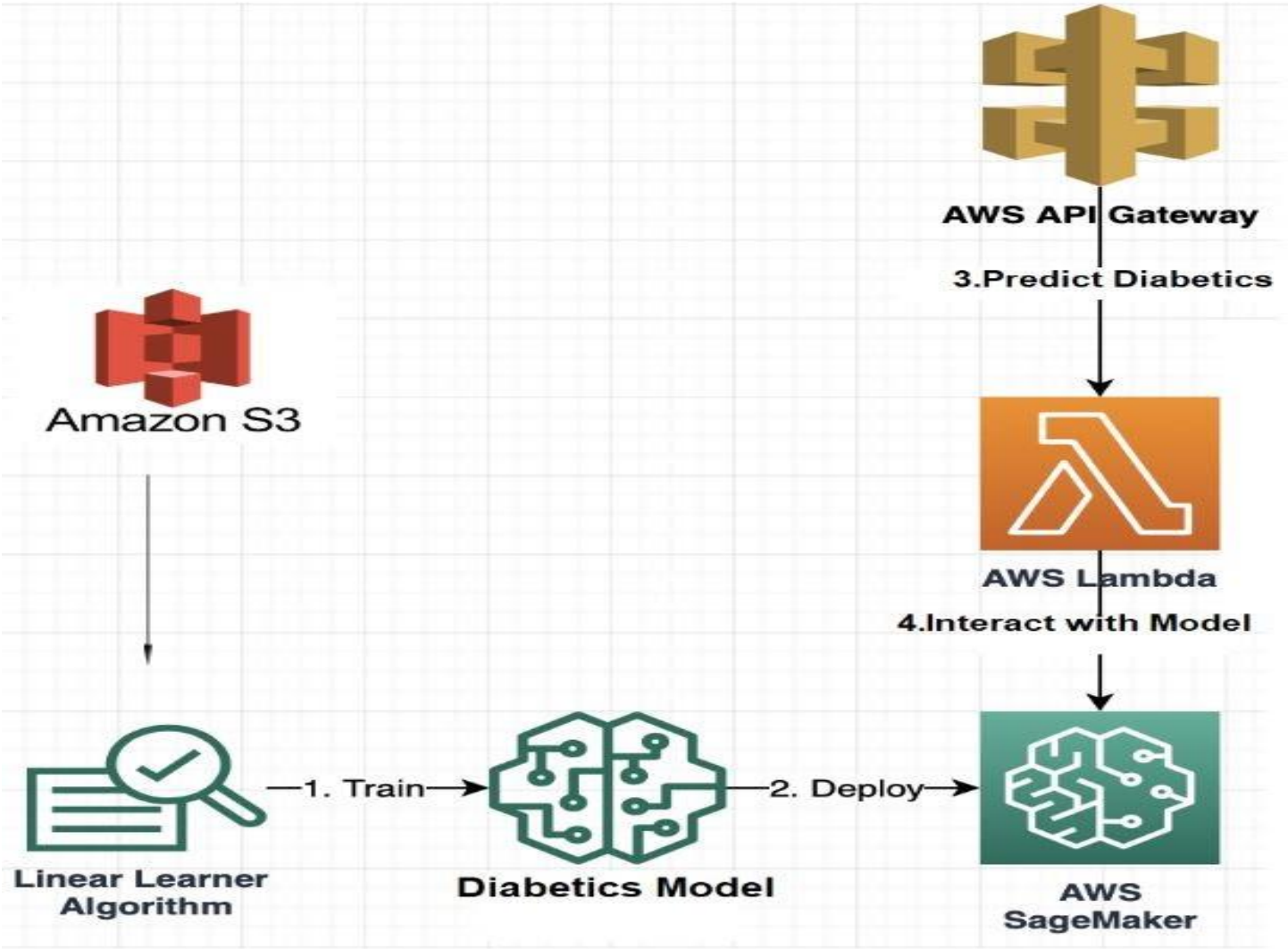
The datasets consist of several medical predictor variables and one target variable, Diabetes. Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

2.2PROPOSED SOLUTION:

Develop an end-to-end web application that predicts the probability of females having diabetes. The application must be built with Python-Flask or Django framework with the machine learning model trained & deployed on AWS Sagemaker. Create an API Endpoint for the model with the help of API Gateway and AWS Lambda Service.

3.THEORITICAL ANALYSIS:

3.1. BLOCK DIAGRAM:



3.2. SOFTWARE DESIGNING:

- Amazon S3
- AWS API Gateway
- AWS Lambda
- Amazon SageMaker

4.EXPERIMENTAL INVESTIGATIONS:

Aws Cloud:

Aws Cloud Provides Many Services Such as Sagemaker,lambda and Api Gateway,etc..

Sagemaker:

Amazon SageMaker is a fully managed service that provides every developer and data scientist with the ability to build, train, and deploy machine learning (ML) models quickly. SageMaker removes the heavy lifting from each step of the machine learning process to make it easier to develop high quality models.

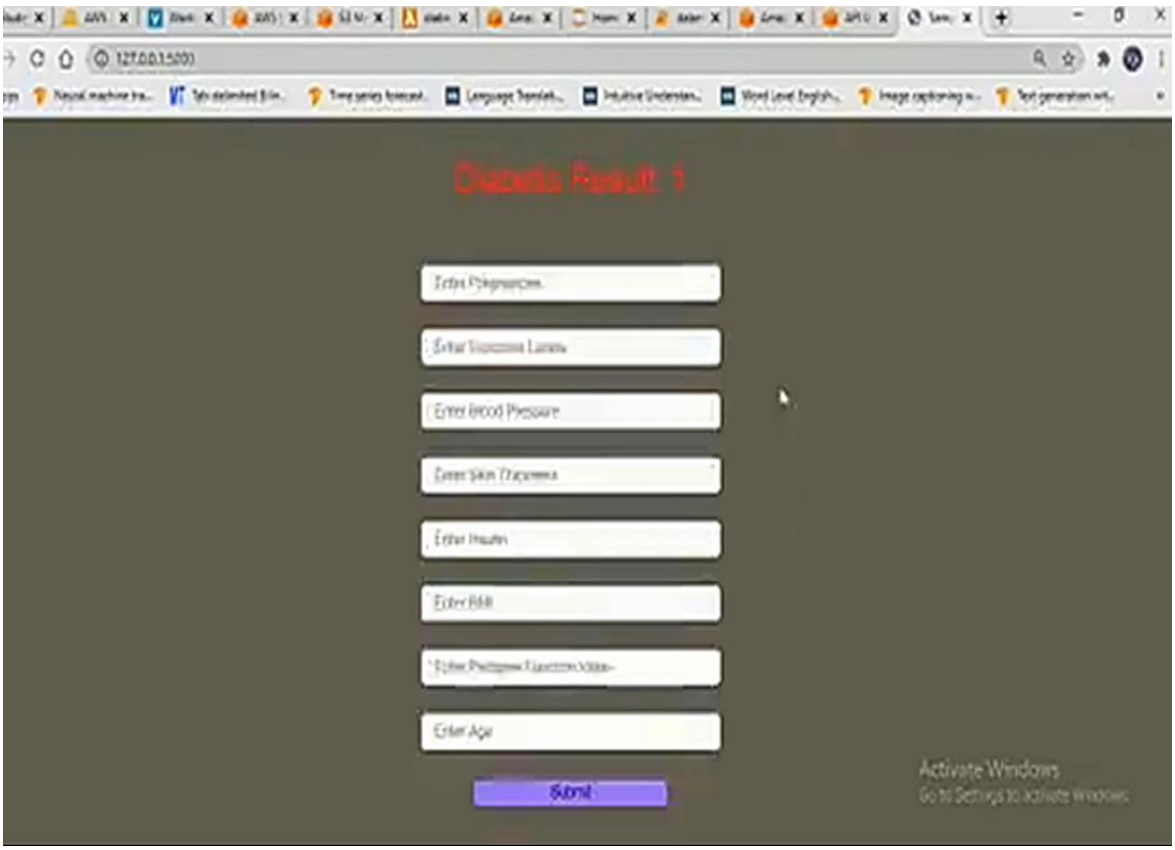
Lambda:

With Lambda, you can run code for virtually any type of application or backend service - all with zero administration. Just upload your code and Lambda takes care of everything required to run and scale your code with high availability. You can set up your code to automatically trigger from other AWS services

Api Gateway:

Amazon API Gateway is an AWS service for creating, publishing, maintaining, monitoring, and securing REST, HTTP, and WebSocket APIs at any scale. API developers can create APIs that access AWS or other web services, as well as data stored in the AWS Cloud APIGateway creates RESTful APIs that: Are HTTP-based

5.RESULT:



6.ADVANTAGES :

- Easy to understand and efficient training algorithm(xgclassifier algorithm).
- Order of training instances has no effect on training
- Pruning can deal with the problem of overfitting
- Always find a “good solution”

7.APPLICATIONS:

- Used in early prediction for diabetes for women.
- We can also use it for predicting heart disease, phenomena, kidney disease
- We can also use it to predict the medical health condition of the people.

8.FUTURE SCOPE:

This study can be conducted with a large hospital sample size in rural and urban community settings in multiple states across India. This would allow determining if the readmission factors Differ based on patient geographical location or it similar traits are observed nationwide. In addition, this would strengthen both urban and rural models while assessing the importance of age categorization.

This research study has only targeted patients with diabetes. Readmission prediction model needs to be generated for other key health conditions also such as Heart disease, Phenomena, kidney disease etc. in Indian Healthcare system. In the future studies, planned and unplanned(emergency) readmission needs to be considered.

Various other key features in the medical records, like family history (to find hereditary information), emotional status (depression), socioeconomic status, and lifestyle habits (exercise), smoking status and season of readmission need to be collected and analyzed. It will be interesting to perform a more exhaustive exploration of additional features in the dataset and study their relevance towards predicting the risk of readmission.

Living with diabetes is challenging and distressful. Diabetic patient's condition cannot be understood only from his with medical charts. There is a need to collect and analyze both subjective and objective patient information in order to fully understand the occurrence of readmission of patients with diabetes. Subjective data can be captured by interviewing patients or by conducting surveys which will enrich the depth of patient information. The conversation between doctor and patient can also be collected and analyzed which could help to extract important features corresponding to patient's willpower and attitude by test-mining techniques. This information might improve the intelligent models to identity patients at high risk or readmission.

9.CONCLUSION:

Machine learning has the great ability to revolutionize the diabetes risk prediction with the help of advanced computational methods and availability of large amount of epidemiological and genetic diabetes risk dataset. Detection of diabetes in its early stages is the key for treatment. This work has described a machine learning approach to predicting diabetes levels. The technique may also help researchers to develop an accurate and effective tool that will reach at the table of clinicians to help them make better decision about the disease status.

10.BIBILOGRAPHY:

- Komi, Zhai. 2017. Application of Data Mining Methods in Diabetes Prediction
- Alan Siper, Roge Farley and Craig Proceedings of Student/Faculty Research Day, CSIS, Pace University, May 6th, 2005
- Devi, M. Renuka, and J. Maria Shyla. "Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus." International Journal of Applied Engineering Research 11.1 (2016): 727-73