

PROJECT DOCUMENT

INTRODUCTION

Overview:

Predictive analytics has gained a lot of reputation in the emerging technology Big data. Predictive analytics is an advanced form of analytics. Predictive analytics use statistical or machine learning method to make a prediction about future or unknown outcomes. There are many applications of predictive analytics, out of which one is health care. A most common disease now a day's is diabetes. People are suffering with it and the patient number increases day by day. The World Health Organization predicts that by 2030 there will be approximately 350 million people worldwide affected by diabetes. Mostly whatever food we eat is converted into glucose or sugar. Now, this glucose or sugar is used for energy. Glucose is transported to body cells through insulin. If the body does not produce sufficient insulin or does not make proper use of insulin then it leads to diabetes. There are several factors for developing diabetes like genetic susceptibility, body weight, food habit and sedentary lifestyle.

PURPOSE:

Diabetes is a disease in which blindness, nerve damage, blood vessel damage, kidney disease and heart disease can be developed. Undiagnosed diabetes may result in very high blood sugar level referred as hyperglycemia which can lead to complication like diabetic retinopathy, nephropathy, neuropathy, cardiac stroke and foot ulcer. So, early detection of diabetes is very important to improve quality of life of patients and enhancement of their life expectancy. By the use of predictive analytics in the field of diabetes, diabetes diagnosis, diabetes prediction, diabetes self-management and diabetes prevention can be achieved. The predictive model diagnostically predicts whether or not a patient has diabetes, based on certain diagnostic measurements related to diabetes and this early prediction is helpful in curing diabetes at early stage.

LITERATURE SURVEY:

EXISTING PROBLEM:

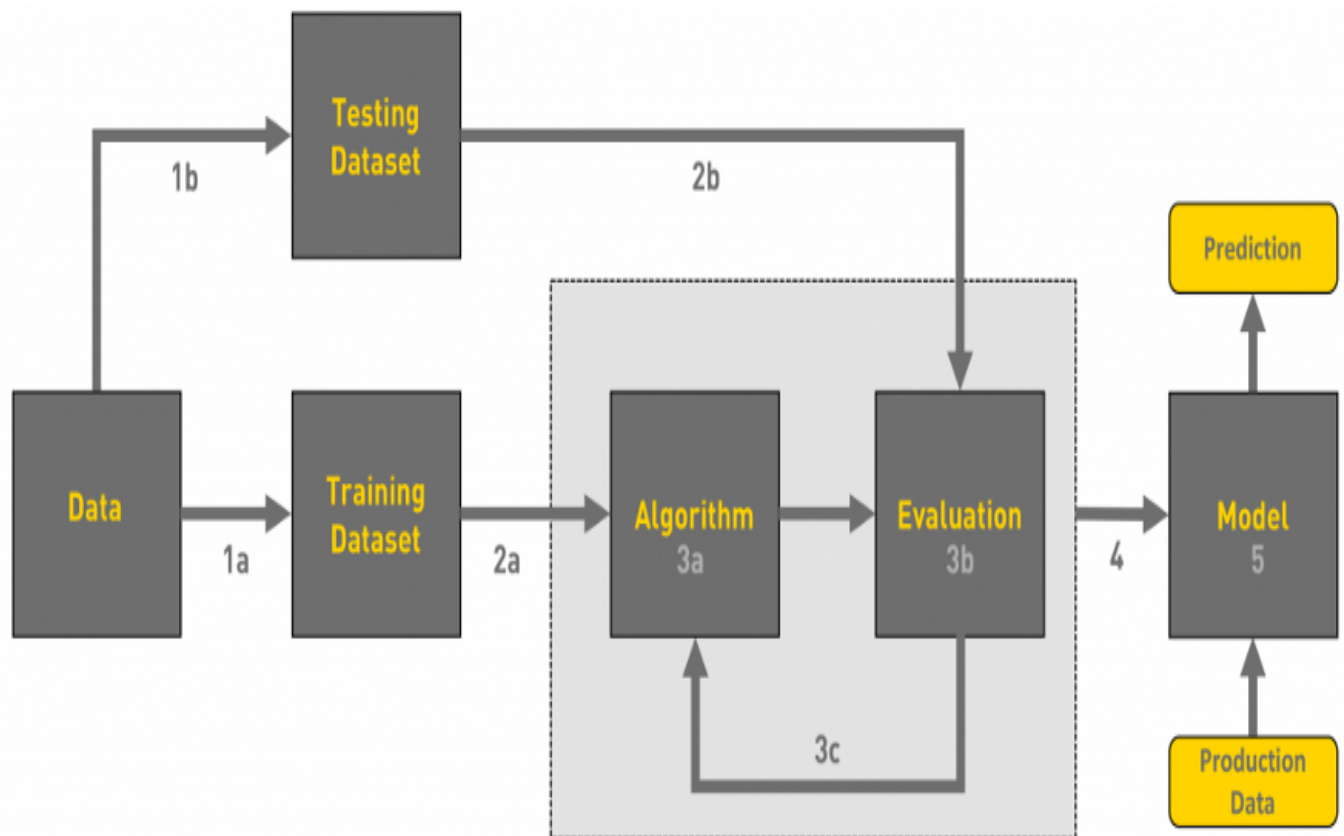
Diabetes is a major metabolic disorder which can affect entire body system adversely. Undiagnosed diabetes can increase the risk of cardiac stroke, diabetic nephropathy and other disorders. All over the world millions of people are affected by this disease. Early detection of diabetes is very important to maintain a healthy life. This disease is a reason of global concern as the cases of diabetes are rising rapidly. There are four types of diabetes which are TYPE 1, TYPE 2, GESTATIONAL, PRE DIABETES. TYPE 1 diabetes is also known as insulin dependent diabetes where the pancreas does not produce the hormone insulin. TYPE 2 diabetes is also known as non-insulin dependent diabetes where adequate insulin is produced but the body cannot make use of insulin. Gestation diabetes is a type of diabetes which occurs during pregnancy . Pre diabetes refers to a situation where blood glucose levels are higher than normal but not so high to diagnosis as diabetes. According to the National Institute of Diabetes and Digestive and Kidney Diseases, diabetes doubles your risk of heart disease and stroke. For a diabetic patient insulin can't be used effectively. This allows blood glucose levels to rise while the rest of your cells are deprived of much-needed energy. This can lead to a wide variety of problems affecting nearly every major body system. Diabetes can also damage your kidneys and affect their ability to filter waste products from your blood. Diabetes can be controlled. In some cases it can even go into remission if proper lifestyle changes are made.

2.2 PROPOSED SOLUTION:

It is important to predict diabetes at early stage this can be accomplished using a supervised learning predictive which predicts diabetes based on the diagnostic data of the person. It is a computational method for automatic learning from experience and make accurate predictions. Model is built based on diagnostic measurements included in the dataset of different patients. And all the patients here are females at least 21 years old of Pima Indian heritage. An end-to-end web application that predicts the probability of females having diabetes can be developed .

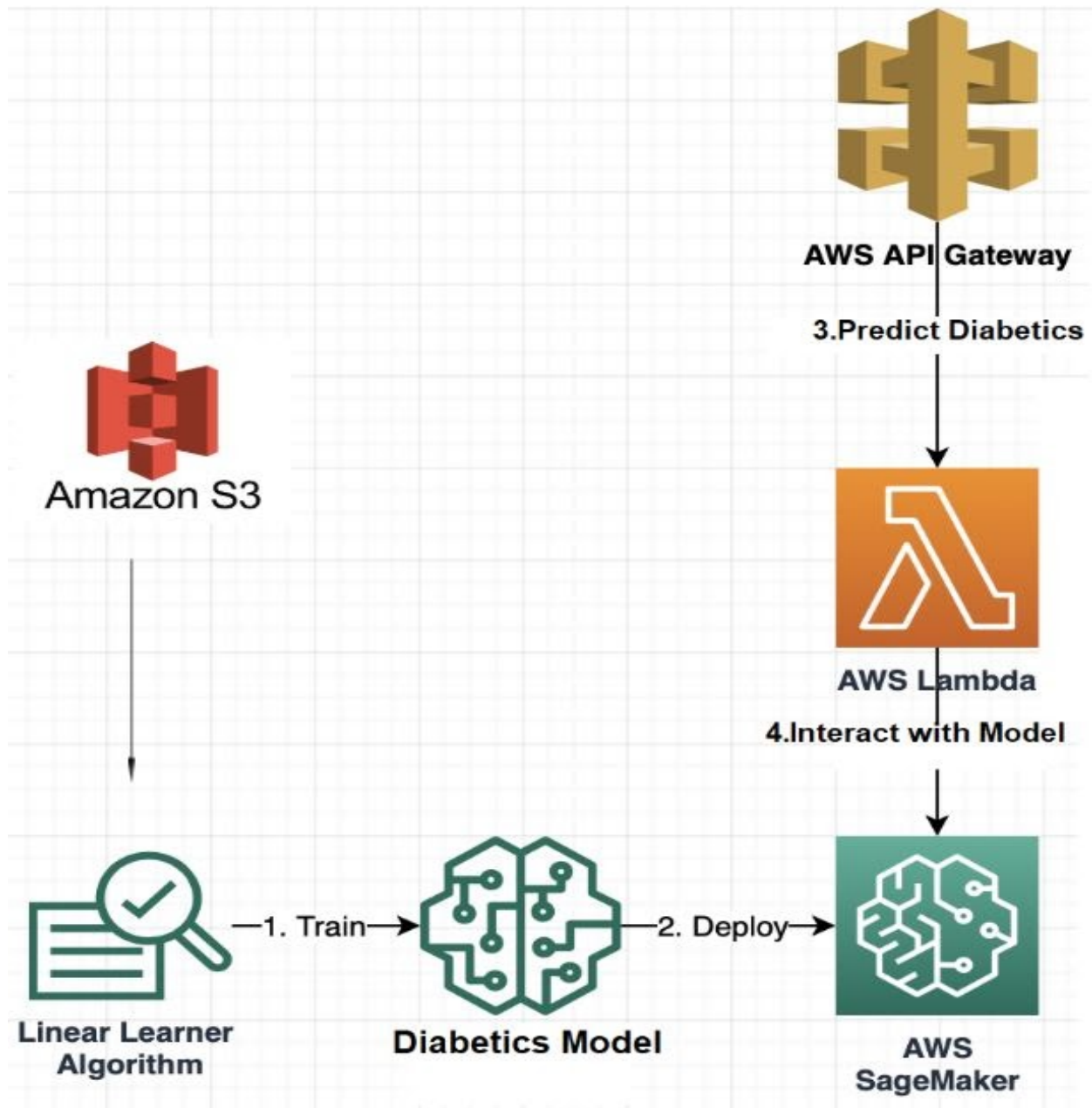
THEORITICAL ANALYSIS:

BLOCK DIAGRAM:



SOFTWARE/HARDWARE DESIGNING:

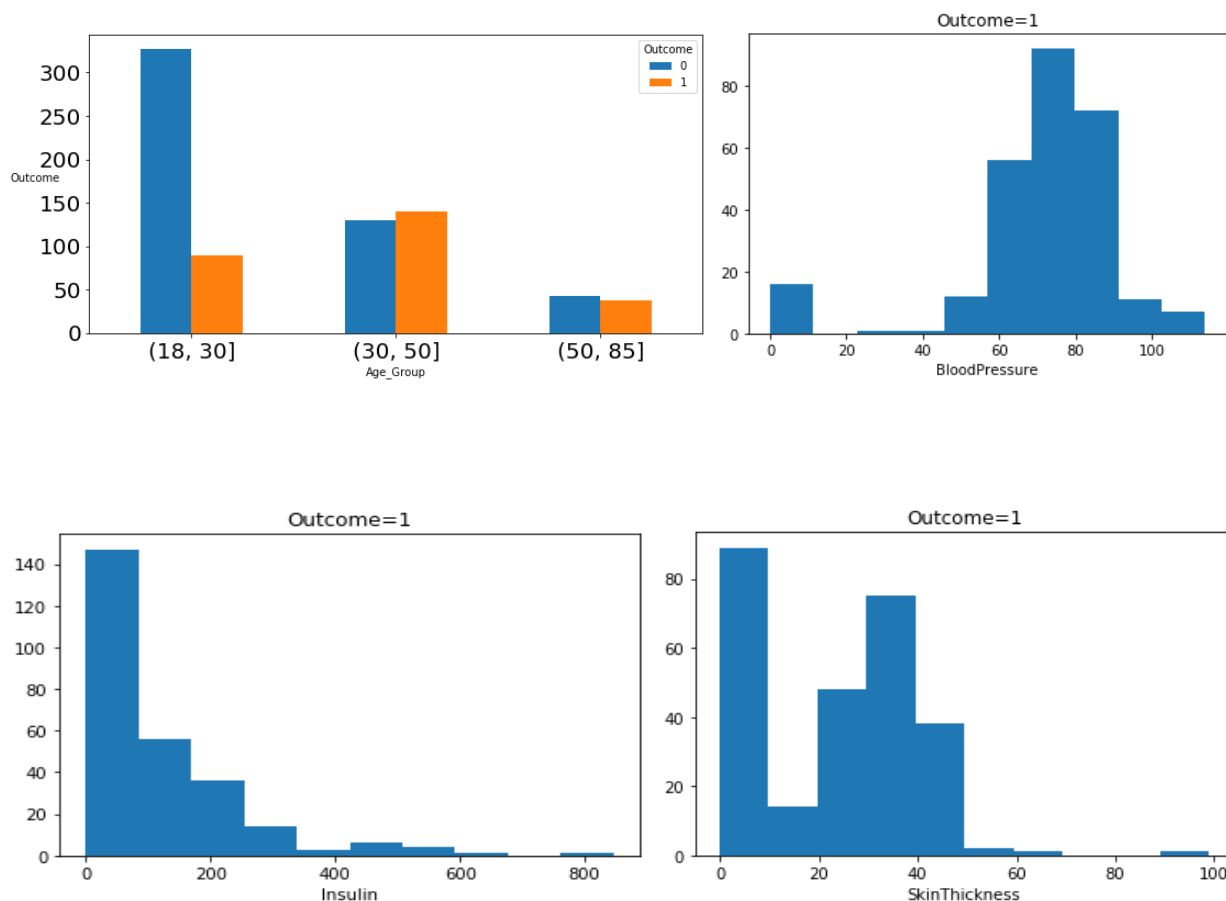
- The data set is collected and using data preprocessing techniques redundant, null and extreme values are removed and the data is prepared for training.
- Identifying the influential factors on target variable and the variation of target variable with independent variables.
- Splitting the data into train and test for training and validation.
- The algorithm used is XGBOOST Linear regressor as it is a classification problem.
- Building the model and training regressor model using train data and validating the model using test data.
- Analysing the performance using precision, recall.



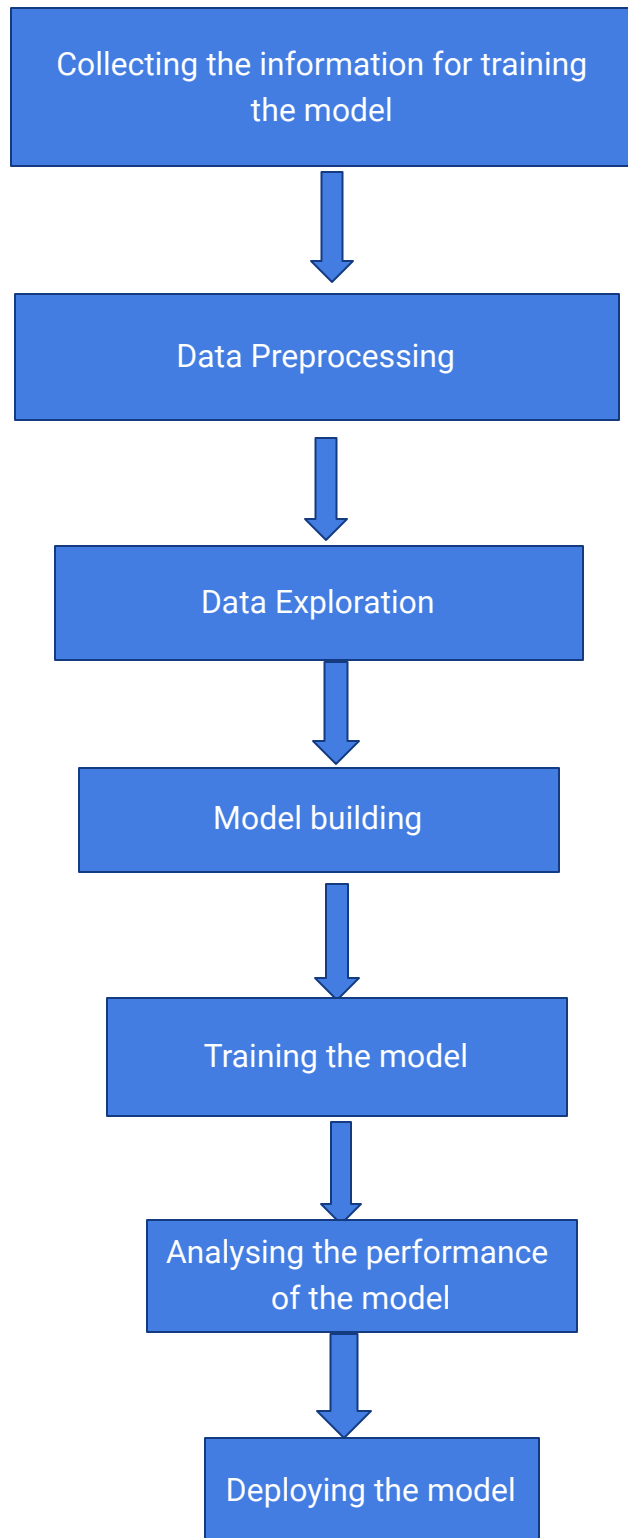
Software/Hardware: Amazon S3, AWS API Gateway, AWS Lambda, Amazon SageMaker

EXPERIMENTAL INVESTIGATIONS:

- 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.
- Influential factors on target variable are analysed using exploratory data analysis.
- It is observed women in the age group of 30-50 are more prone to diabetes.
- Similarly the variation of other dependent variables is observed. Findings indicate a strong association of diabetes with body mass index (BMI) and with glucose level,



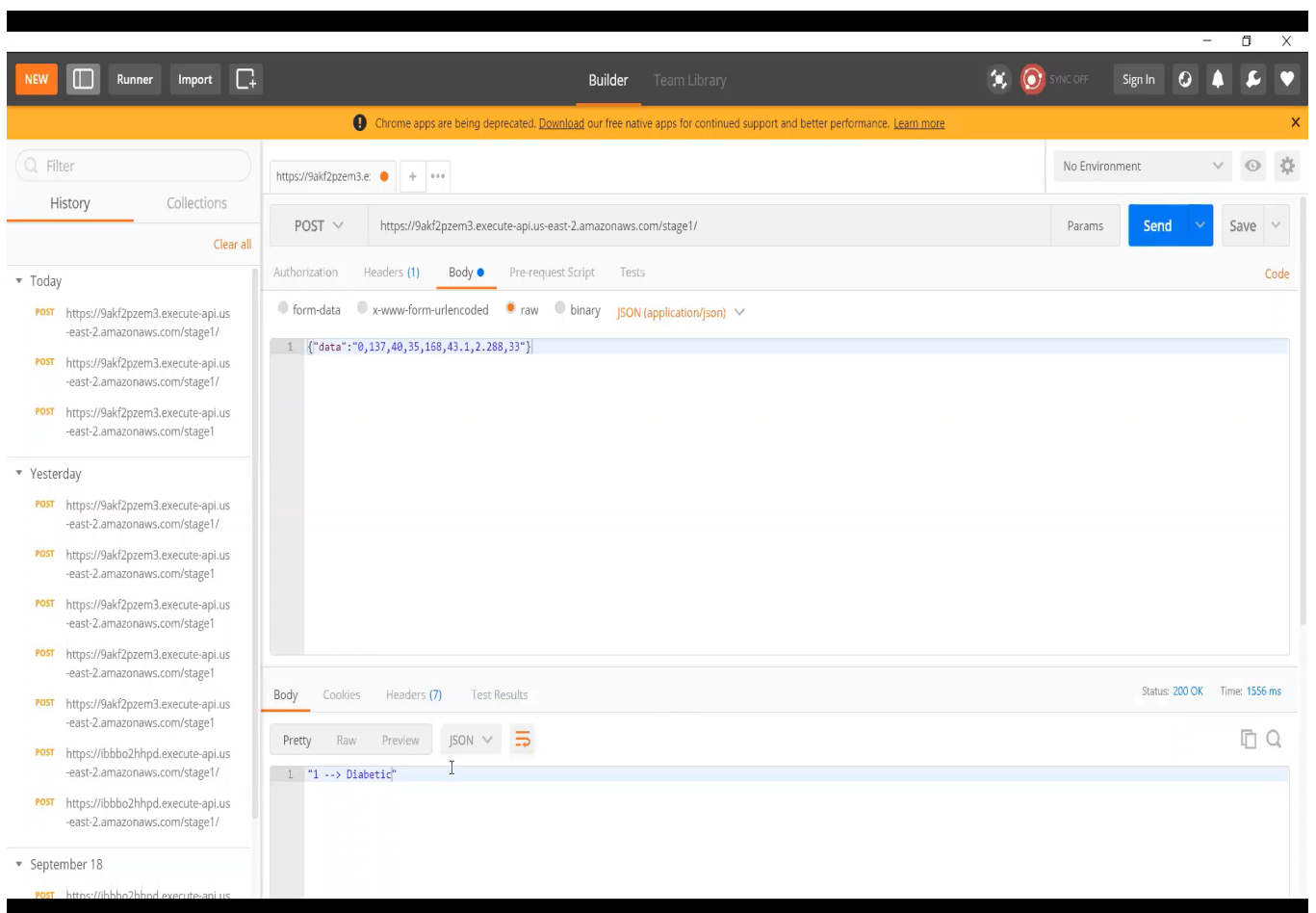
FLOWCHART



RESULT:

Overall Classification Rate: 74.0%

Predicted	NO	YES
Observed		
No Purchase	76% (130)	32% (19)
Purchase	24% (41)	68% (41)



ADVANTAGES:

Diabetes is a common, chronic disease. Prediction of diabetes at an early stage can lead to improved treatment. Data mining techniques are widely used for prediction of disease at an early stage.

Early prediction avoids acute decompensation, prevent or delay the appearance of late disease complications, decrease mortality, and maintain a good quality of life.

APPLICATIONS:

- The ability of the model to predict patients with diabetes using some commonly used lab results is high with satisfactory sensitivity.
- These models can be built into an online computer program to help physicians in predicting patients with future occurrence of diabetes and providing necessary preventive interventions.

CONCLUSION:

To conclude this predictive model predicts the probability of females having diabetes and improves the life expectancy of people.

FUTURE SCOPE:

Using the predictive model it would be easy to predict the diabetic patient easily and to control the adverse effects caused by the disease.

BIBLIOGRAPHY

- 1.<https://www.emerald.com/insight/content/doi/10.1016/j.aci.2018.12.004/full/html>
- 2.<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0082-7>