DIABETES PREDICTION

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

by

V. SAIRAM **(2003A52020)**

S. SIDDHRARATHA **(2003A52016)**

S. GOPI KRISHNA **(2003A52015)**

Under the guidance of

Mr. S. Naresh Kumar

Assistant Professor, Department of CSE.

Department of Computer Science and Artificial Intelligence

Department of Computer Science and Artificial Intelligence

CERTIFICATE

This is to certify that project entitled **"DIABETES PREDICTION**" is the bonafied work carried out by **V. SAIRAM,S. SIDDHARATHA,S. GOPI KRISHNA** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** during the academic year 2022-2023 under our guidance and Supervision.

Mr. S Naresh Kumar
Asst. Professor,
S R University,
Ananthasagar, Warangal.

Dr. M.Sheshikala
Assoc. Prof. & HOD (CSE),
S R University,
Ananthasagar, Warangal.

ACKNOWLEDGEMEN

We express our thanks to Course co-coordinator **Mr. S.Naresh Kumar, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude to Head of the department CS&AI, **Dr. M.Sheshikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

ABSTRACT

In this Project ‘Diabetes Prediction’, we can measure level of a diabetic patient. Diabetes mellitus is the most common disease worldwide and keeps increasing every day due to changing lifestyle, unhealthy food habits and weight problems. There were studies handled in predicting diabetes mellitus through physical and chemical tests are available for diagnosing diabetes. In this proposed through machine learning and deep learning. Under the machine learning, we used the Various Classification Algorithms like Decision Tree, Support vector machine, neural networks, etc. Furthermore, this project aims, by detailing the full prediction & analysis, to give relaxation to user who want to check their diabetes levels.

Table of Contents

1.INTRODUCTION

1.1 OVERVIEW

Diabetes usually known as Diabetes Mellitus. It is a type of metabolic disorder in which diabetics experience blood glucose difficulties as a result of irregular insulin production and release. It's also a long-term condition marked by hyperglycemia. It is one of the world's most serious diseases, and it can lead to a variety of consequences. According to recent rises in morbidity, the global number of diabetic patients will reach 642 million by 2040, suggesting that one out of every 10 people will have the disease. This alarming figure, without a doubt, requires quick attention.

Type 1 and Type 2 diabetes are the two most common forms. Type 1 diabetes can strike anyone at any age, although it strikes teenagers and children the most commonly. The body of a person with type 1 diabetes generates very little or no insulin, requiring daily insulin injections to keep blood glucose levels under control.

Type 2 diabetes can strike anyone at any age, but it is more frequent in adults, accounting for over 90% of all diabetes occurrences. The body of a person with type 2 diabetes does not make good use of the insulin it produces. The cornerstone of type 2 diabetes care is a healthy lifestyle that includes increased physical activity and a balanced diet. The majority of people with type 2 diabetes, on the other hand, will eventually need medicines or insulin to keep their blood glucose levels in check.

Another type of diabetes is gestational diabetes, which is characterized by high blood glucose levels during pregnancy and is linked to complications for both the mother and the child. Their children are more likely to develop type 2 diabetes later in life, even though it will go away after the pregnancy.

1.2 PROBLEM STATEMENT

Diabetes mellitus, sometimes known as diabetes, is a condition characterized by a rise in blood glucose levels. If diabetes is left untreated and undiagnosed by a doctor, it can lead to plenty of problems. The traditional way of diagnosis is for patients to go to a diagnostic facility, consult their doctor, and relax for a day or more to get their results. Furthermore, they must waste their money in futile anytime they need to obtain their diagnosis report. By using this method, user can identify the level of diabetic such as normal,

type 1 or type 2 and its remedies such as food control, yoga treatment, and they get more valuable ideas from most relevant YouTube channels. User can also get the test report which is useful when they need to consult the doctor for further treatment.

1.3 EXISTING SYSTEM

Lot of work has been carried out to predict diabetic diseases using dataset.Different levels of accuracy have been attained using various machine learning techniques. The accuracy of the existing system is around 90%. The main idea behind the proposed system after reviewing the existing papers was to create a diabetic prediction system based on the required inputs. The current system employs the random forest technique, which takes a long time to train because it mixes a large number of decision trees to decide the class. It also requires a lot of computer power and resources because it creates several trees and combines their outcomes. When the number of classes is increased, many analyzers don't perform well. The existing system predict only the user has diabetic or not it does not provide the level of diabetic as well as remedies such as food control, treatment relevant videos, yoga etc. Then it does not provide the history of patient diabetic report.

Limitations of Existing System - Increased system and time complexity. - The prediction accuracy of this algorithm is lower than the other probability algorithms - It does not predict the level of diabetic

1.4 PROPOSED SYSTEM

The proposed method uses the KNN algorithm, which divides diabetes into three categories: normal, type 1 diabetic, and type 2 diabetic. There is no training time for the KNN algorithm. It saves the training dataset and only uses it to make real-time predictions in order to learn from it. As a result, the KNN approach is significantly faster than other training-based algorithms such as SVM, Linear Regression etc. The accuracy of the proposed system is achieved 94% using KNN algorithm. This system predicts the user has diabetic or not and also classify the level of diabetes then it gives remedies for the diabetic patients. After the diabetic prediction user can download the report and also see the previous diabetic reports of that particular person. Admin module is added in this project who can manage the doctors details

such as specialization, contact no, hospitals name etc. It would be much useful for the diabetic patients for their further treatment.

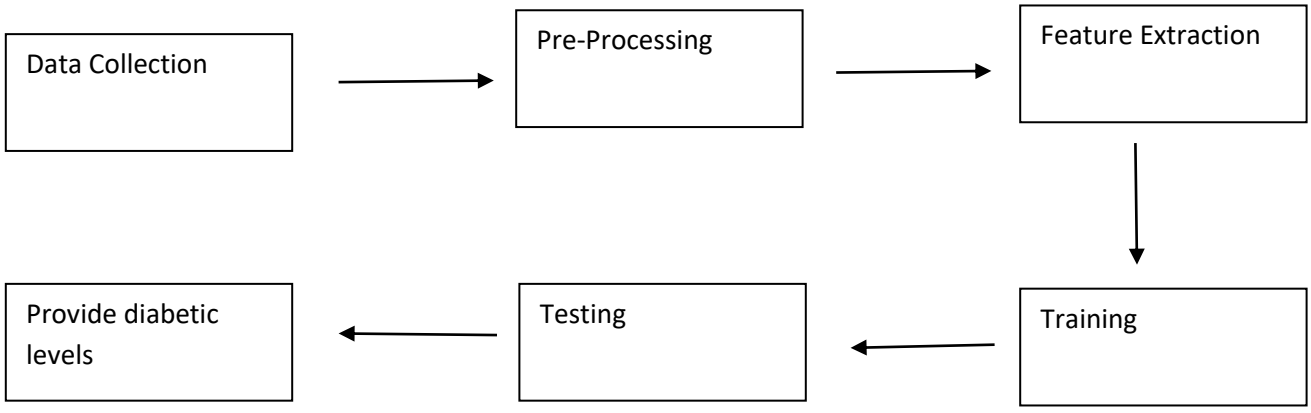
Advantages of the Proposed System

- This diabetic analysis will be helpful for remote hospitals and remote doctors to diagnose the diabetic and take correct treatment plan.
- Helps to detect the level of diabetic and also provide remedies.
- For larger datasets, it can converge faster as it causes updates to the parameters more frequently.
- Less expensive.
- It is more accurate than existing system.

1.5 OBJECTIVES DESCRIPTION

The goal of this project is to create a system that can forecast diabetes in a patient with greater accuracy using the KNN algorithm in a machine learning technique, as well as classify diabetic levels based on symptoms.

1.6 OVERALL ARCHITECTURE



2. LITERATURE SURVEY

More than 60% of Indians suffer from chronic diseases. The prediction of human diseases like diabetes, cancer, asthma, High blood pressure is done using various machine learning techniques [6]. The data presented is the complete source of information for the prediction of diseases. The tree structure of the DT algorithm explains the flow of the classification. Sisodia and Sisodia conducted a study in order to detect diabetes at early stage by using classification algorithms including Decision Tree, SVM and Naive Bayes. Results showed that Naive Bayes performed the highest accuracy of 76.30% [8]. Naive Bayes is a probabilistic classification algorithm which depends on the Bayes theory. SVM draws a line between the two datasets and classifies them accordingly.

RF is a model that may be used for both classification and regression. To find random samples, it employs the bagging method. Based on the degree of similarity or distance measure, K Nearest Neighbour predicts the category. The most appropriate feature is chosen using a modified technique, and the performance is assessed using metrics [9][10]. In order to predict diabetes mellitus, many classifying algorithms are employed and implemented. In [11], classification is performed using three classifiers: Nave Bayes, SVM, and Decision Tree (DT), and the results are assessed using several metrics. Naive Bayes was found efficient among the three with an accuracy of 76.3 %. Various studies have used different classifiers to detect diabetes and other illnesses. Random Forest Algorithm for the prediction of diabetes [1] predict the accuracy approximately 90%. Ahmad [7] described multilayer perceptron model and compared accuracy with decision tree model like ID3 and J48 computations. The J48 tree- based model performed with (89.3%) higher accuracy.

2.1.1 Document the survey done by you

INDIA has an estimated 77 million people (1 in 11 Indians) formally diagnosed with diabetes which makes it the second most affected in the world, after China.[2] Furthermore, 700,000 Indians died of diabetes, hyperglycemia, Kidney disease or other complications of diabetes in 2020. One in six people (17%) in the world with diabetes is from India.[2] (India’s population as calculated in October 2018 was about 17.5%

of the global total.[3]) The number is projected to grow by 2045 to become 134 million per the International Federation.

In India, type 1 dtabetes is rarer than in western countries, and about 90 to 95% of Indians who were diagnosed had type 2 diabetes. Only about one-third of Type 2 diabetes in India have a Body Mass Index above 25.[4] A 2004 study suggests that the prevalence of type 2 diabetes in Indians may be due to environmental and lifestyle changes resulting from industrialization and migration to urban environment from rural.[5] This lifestyle change has led to the increased consumption of energy intake from animal foods in Asian populations.[6] This change has been seen in India where urban residents consumed 32% of energy from animal fats compared to 17% of rural residents.[7] These changes also occur earlier in life, which means chronic long-term complications are more common.

In 2020, according to the International diabetes Federation (IDF), 463 million people have diabetes in the world and 88 million people in the SouthEast Asia region. Of this 88 million people, 77 million belong to India.[8] The prevalence of diabetes in the population is 8.9%, according to the IDF. According to the IDF estimates, India has the second highest number of children with type 1 diabetes after the United States. It also contributes to the largest proportion of incident cases of type 1 diabetes in children in the SEA region.[9] Per the World Health Organization, 2% of all deaths in India are due to diabetes.[10]

The number of people with diabetes in India has increased from 26 million in 1990 to 65 million in 2016.[1] According to the 2019 National Diabetes and Diabetic Retinopathy Survey report released by the Ministry of Health and Family Welfare, the prevalence was found to be 11.8% in people over the age of 50.[11] The prevalence of diabetes is 6.5% and prediabetes 5.7% among the adults below the age of 50 years, according to the DHS survey.[12] The prevalence was similar in both male (12%) and female (11.7%) populations. It was higher in urban areas.[13] When surveyed for diabetic retinopathy, which threatens eyesight, 16.9% of the diabetic population aged up to 50 years were found to be affected. Per the report, diabetic retinopathy in the 60-69-years age group was 18.6%, in the 70-79-years age group it was 18.3%, and in those over 80 years of age it was 18.4%. A lower prevalence of 14.3% was observed in the 50-59-years age group.[11] High prevalence of diabetes is reported in economically and

epidemiologically advanced states such as Tamil Nadu and Kerala, where many research institutes which conduct prevalence studies are also present.[14]

Over the past three decades, the burden of diabetes in terms of deaths and Disability-adjusted life year (DALYs) has more than doubled in India. As per the Global Burden of Disease (GBD) Data Visualizations, the recorded death rate and DALY rate of diabetes in 2019 were 19.64 per 100,000 and 919.02 per 100,000 population, respectively, including males and females.[17] The GBD explore risk assessment framework estimated that diabetes-related DALYs attributable to high risk for Stroke, Coronary artery disease, Chronic obstructive pulmonary disease, Chronic kidney disease, Diarrhoeal diseases, Lower respiratory tract infection, Dietary iron deficiency, and Neonatal disorders.[18]

According to a survey conducted in 2016, the High Body Mass Index (BMI) appeared to be the major risk factor contributing to 36% of diabetes DALYs. Besides, other risk factors, such as dietary risk, tobacco consumption, occupational exposure to passive smoke, low physical activity, and alcohol consumption, played a significant role as contributing risk factors.[1]

The majority of diabetes cases are of type 2 diabetes.[19] In order to control diabetes in India, the Government of India initiated the National Program for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases, and Stroke (NPCDCS) in 2010.[14] It aims to set up outreach camps for opportunistic screening at all levels in the health care delivery system for early detection of diabetes, among other illnesses.[20]

Preventing diabetes in the developing nations is valued highly because of the high cost of treating it. In India, it is estimated that a diabetic person spends a median of ₹10,000 (US$130) for medical treatment. Pragmatic, cost-effective strategies for primary prevention of diabetes is necessary. Studies using information technology were tested. In one such program, SMS was used to motivate people with impaired glucose tolerance (IGT) to follow lifestyle modification (LSM). This was found to be effective with a relative risk reduction of 36% compared to the participants who had only standard care.[21]

The Indian Diabetes Prevention Program is a three-year randomized control trial that employed LSM and metformin (Met) to prevent type 2 diabetes in subjects with IGT. It concluded that LSM and Met were cost-effective interventions for preventing diabetes among high-risk individuals in India and other developing countries.[22]

The National Diabetes Control Program was initiated in 1987 in some districts of Tamil Nadu, Jammu and Kashmir, and Karnataka. Its objectives included:[23]

- Identifying high-risk individuals
- Introducing health education for the purposes of early intervention
- Aiming for early diagnosis and treatment of affected individuals
- Reducing morbidity and mortality in high-risk groups
- Preventing acute and chronic metabolic, cardiovascular, renal, and ocular complications due to the disease
- Rehabilitating people who have been handicapped due to the disease

3.DATA PRE-PROCESSING

3.1 Dataset Description:

Data Collection:

In the first step we collect data from the reliable source. Sex, age, polyuria, polydipsia, unexpected weight loss, weakness, polyphagia, genital thrush, visual blurring, itching, irritability, delayed healing, partial paresis, muscle stiffness, alopecia, obesity are some of the criteria in the diabetes record set sample. Predicted outcome class, where '1' denotes a positive diabetes patient class and '0' denotes a negative diabetes patient class.

3.2 DATA CLEANING

Pre-Processing**:**

Pre-processing is the next step. It's an important phase in the data discovery process. The majority of health-care data contains gaps in value and inconsistencies. We apply Synthetic Minority Oversampling Technique (SMOTE) in this project, which is a well-known preprocessing approach for dealing with unbalanced datasets.

Feature Extraction:

Feature extraction is the process of converting raw data into numerical features that may be processed while maintaining the information in the original data set. It yields better results than simply applying machine learning to raw data. This is an important categorizing feature.

Training**:**

During the training process we trained the machine from data source. The test data is transformed and predicts the accurate result. During the training process the machine learning automatically selects the correct learning algorithm, based on the type of target that is specified in the training data source.

Testing**:**

Training data set which will be validated using the test dataset model. The test data is transformed and predicts the accurate result is 94% is achieved.

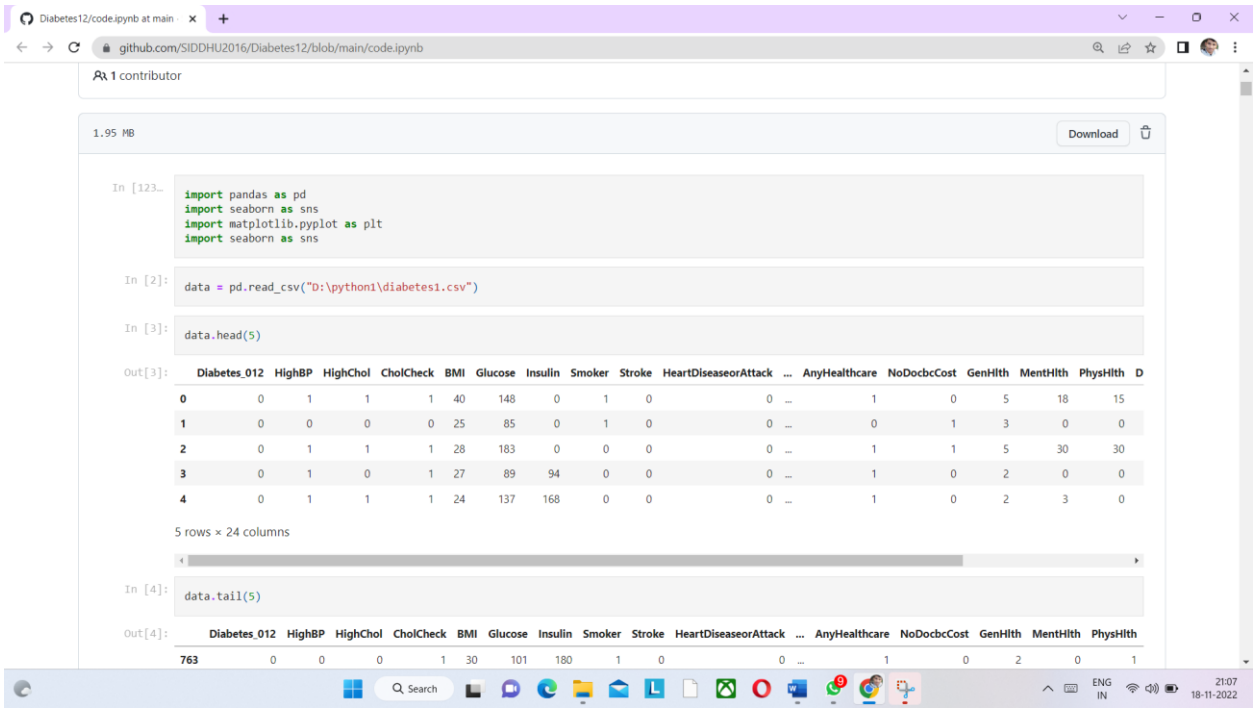


Fig 3.2.1

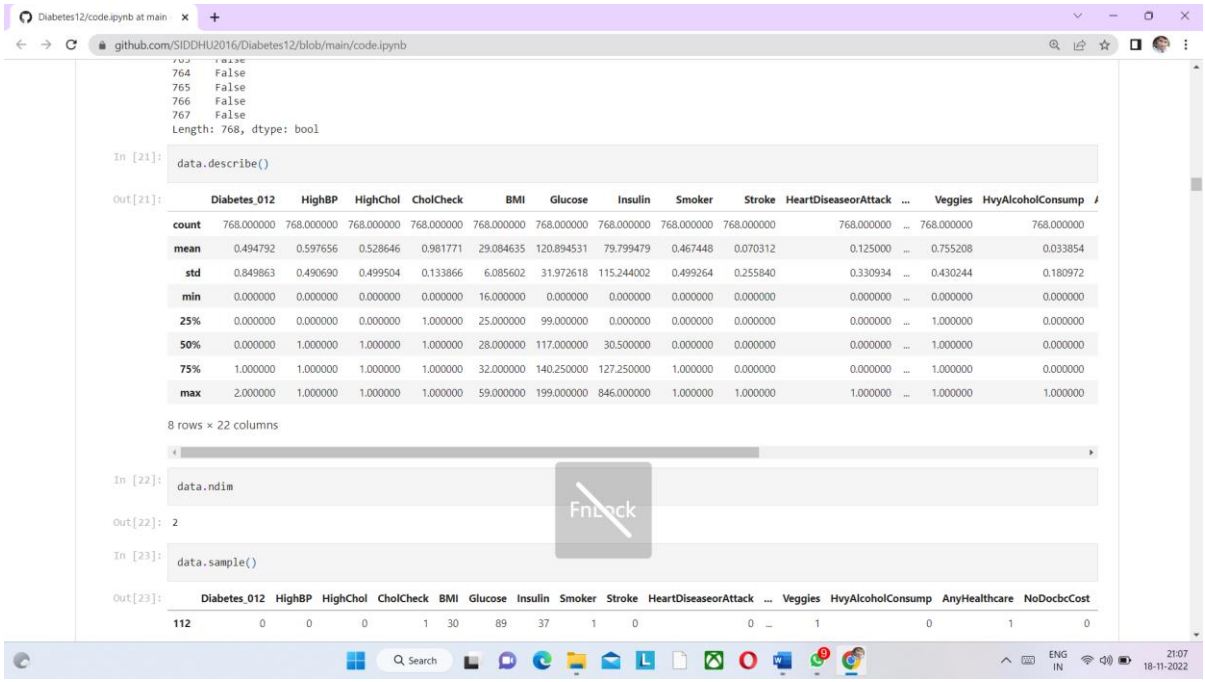


Fig 3.2.2

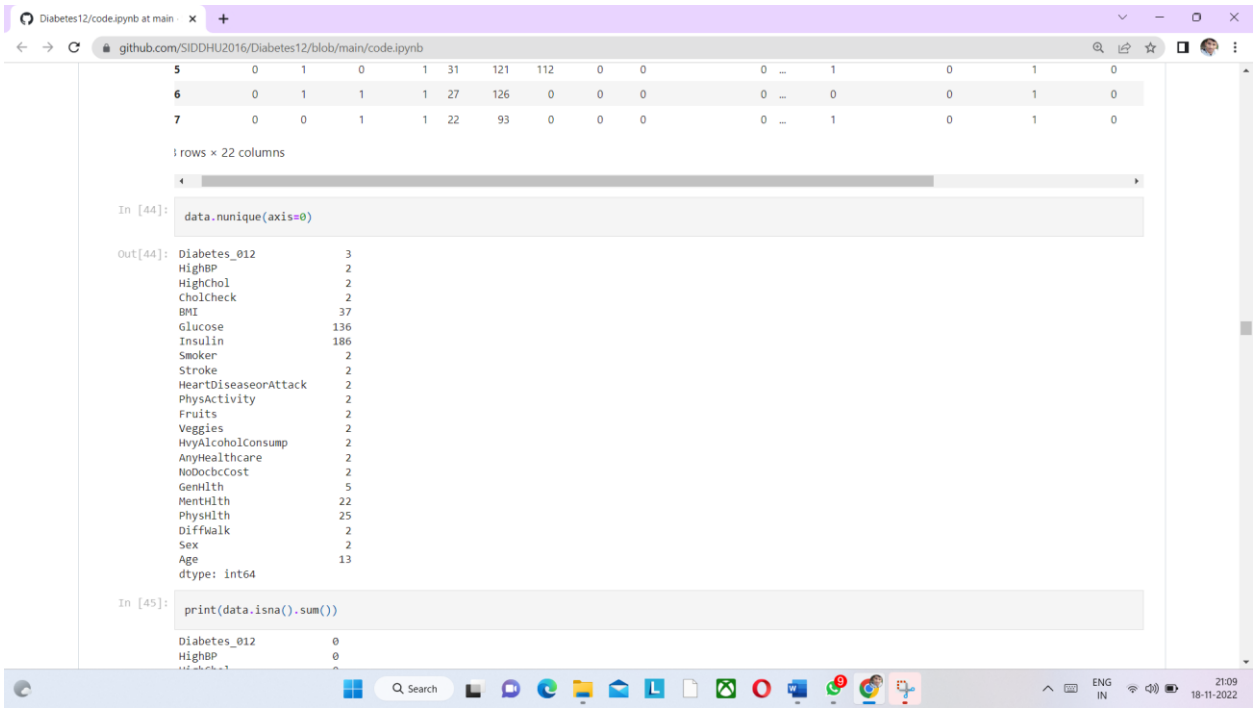


Fig 3.2.3

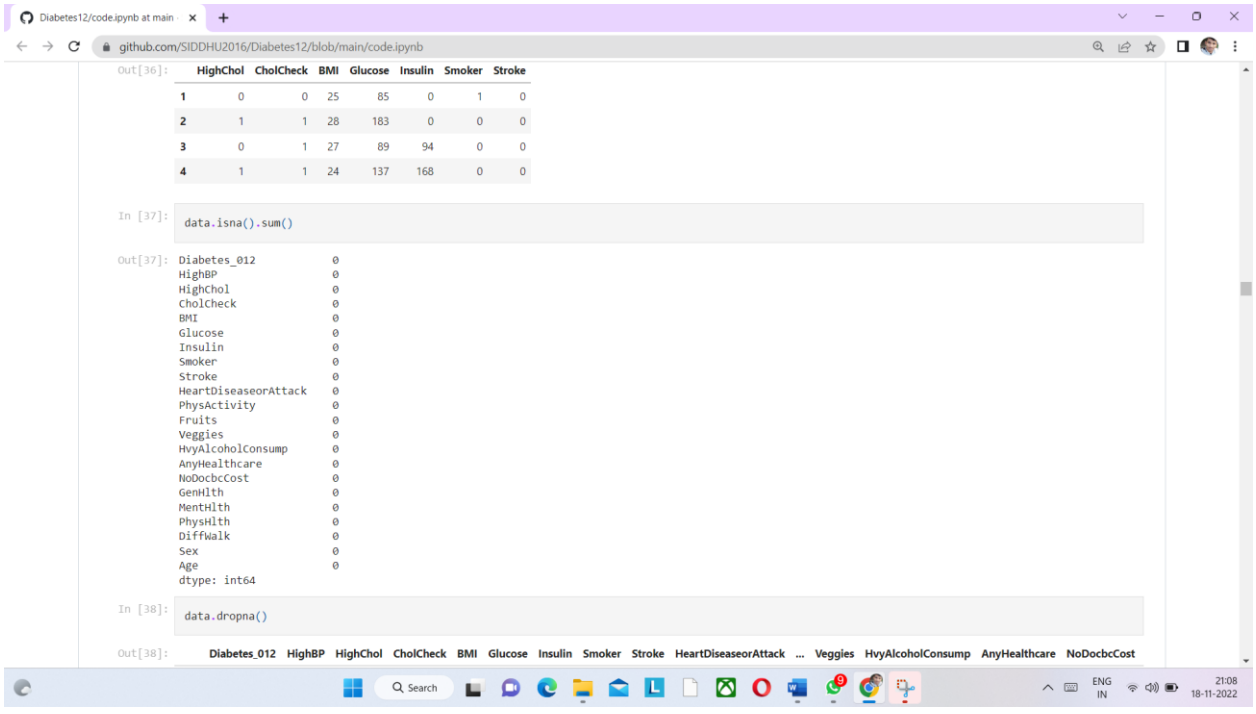


Fig 3.2.1

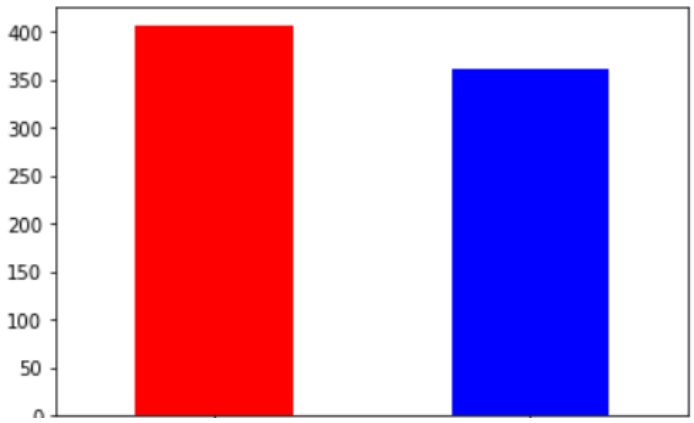
3.3 DATA AUGMENTATION

Data augmentation is a set of techniques to artificially increase the amount of data by generating new data points from existing data. This includes making small changes to data or using deep learning models to generate new data points.

3.4 DATA VISUALIZATION

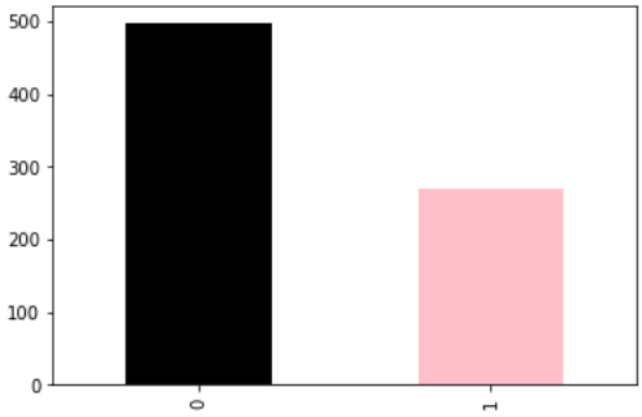
Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

```
In [47]: data["HighChol"].value_counts().plot(kind="bar", color=["red", "blue"]);
```



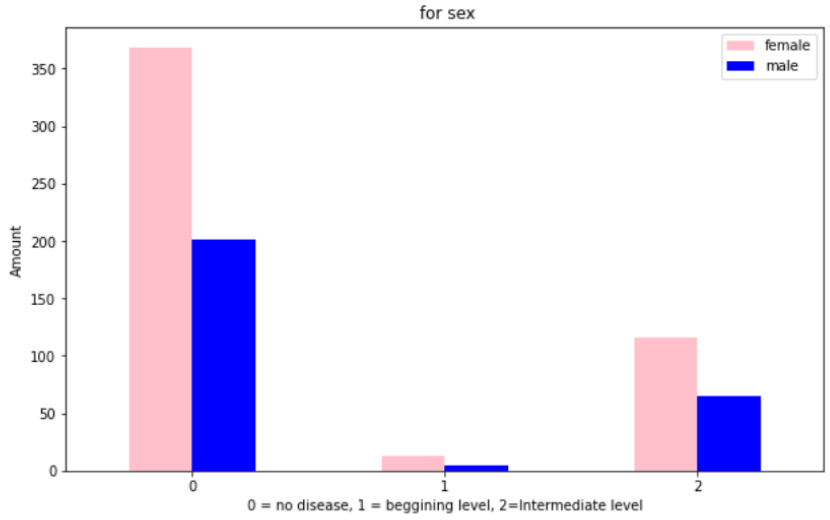
The above figure shows about the high cholesterol how many people are having it is shown in red bar and how many people are not having it is shown in blue bar from our collected data.

In [49]:
```
data["Sex"].value_counts().plot(kind="bar", color=["Black", "Pink"]);
```



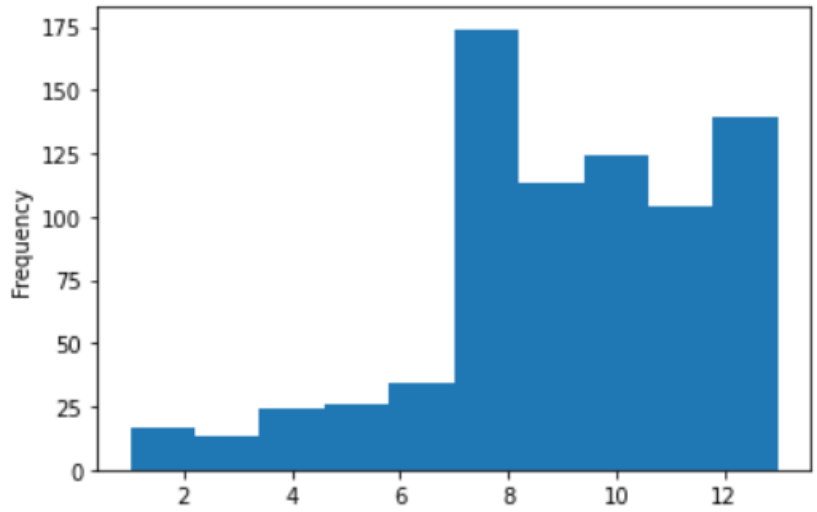
The above figure shows about the sex how many people are male it is shown in black bar (0) and how many people are female it is shown in pink bar(1) from our collected data.

In [50]:
```
pd.crosstab(data.Diabetes_012, data.Sex).plot(kind="bar",
                                    figsize=(10,6),
                                    color=["pink","blue"])
plt.title(" for sex")
plt.xlabel("0 = no disease, 1 = beggining level, 2=Intermediate level")
plt.ylabel("Amount")
plt.legend(["female", "male"])
plt.xticks(rotation=0);
```



The above figure shows There are levels in diabetes 0=no disease, 1=beginning level, 2=intermediate level it is compared with sex how many people having 0 level or 1 level or 2 level of diabetes.

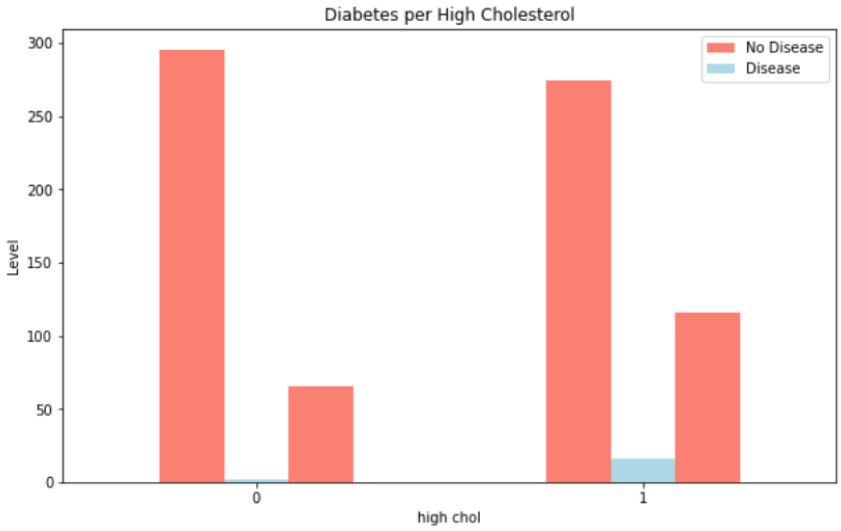
```
data.Age.plot.hist();
```



The above figure shows about the distribution of age column with a histogram.

In [52]:

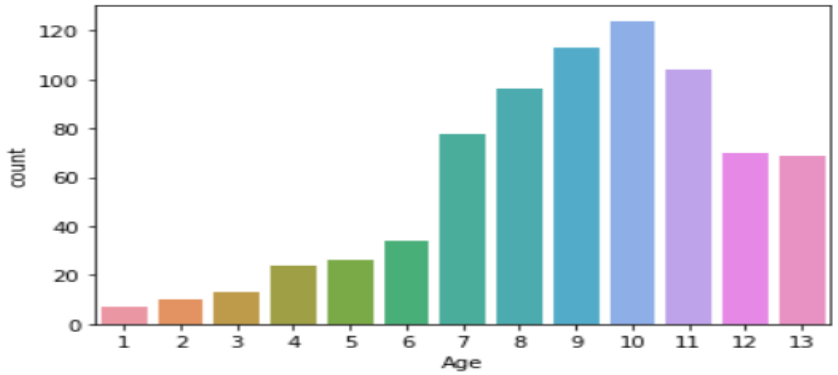
```
pd.crosstab(data.HighChol, data.Diabetes_012).plot(kind="bar",
                                    figsize=(10,6),
                                    color=["salmon","lightblue"])

# Add some communication
plt.title("Diabetes per High Cholesterol")
plt.xlabel("high chol")
plt.ylabel("Level")
plt.legend(["No Disease", "Disease"])
plt.xticks(rotation=0);
```



The above figure shows about the compared between the no diabetes people or diabetes people and high cholesterol having people or not having people.

```
sns.countplot(data['Age'])
plt.show()
```

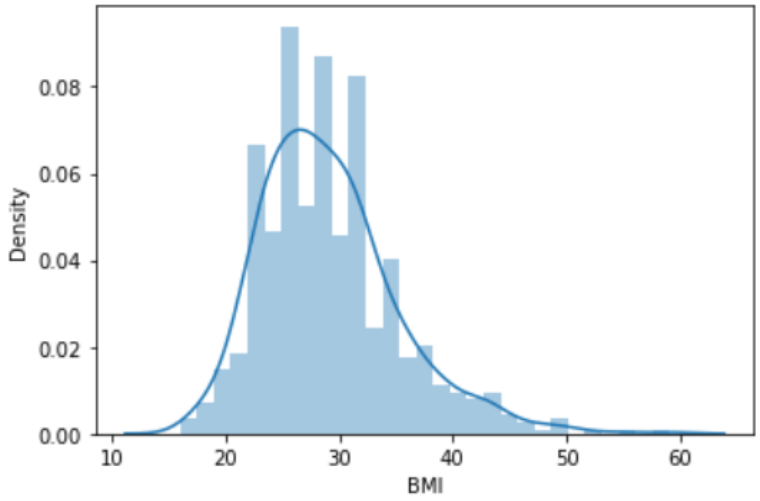
```
C:\Users\SAI SIDDHARTHA\anaconda3\lib\site-packages\seab
rsion 0.12, the only valid positional argument will be `
nterpretation.
  warnings.warn(
```



The above figure shows about the how many peoples are having at an age.

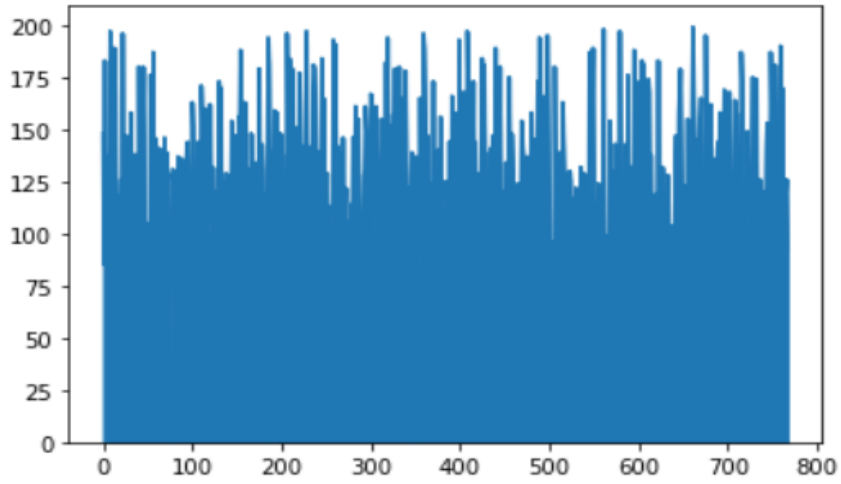
```
sns.distplot(data['BMI'])
```

```
C:\Users\SAI SIDDHARTHA\anaconda3\lib\site-packages\seabor
ved in a future version. Please adapt your code to use eit
l function for histograms).
  warnings.warn(msg, FutureWarning)
```



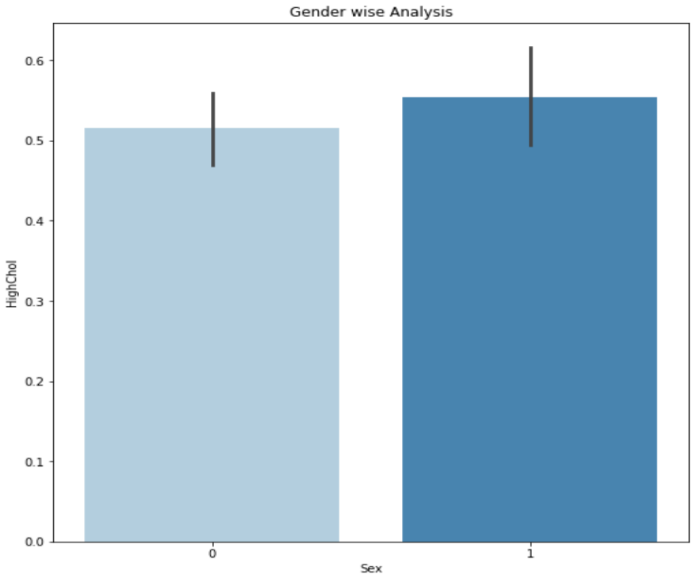
The above figure shows about the density of BMI column.

```
data.Glucose.plot.area()
```



The above figure shows about the area of glucose column.

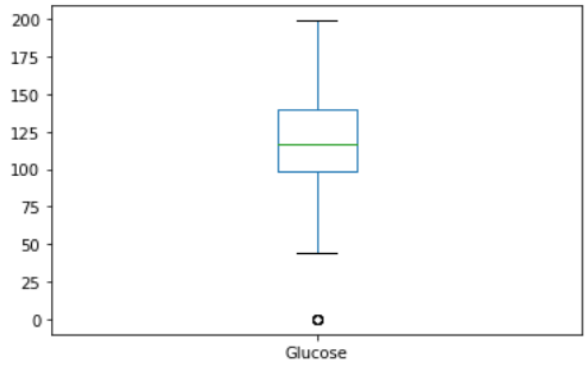
In [72]:
```
plt.figure(figsize=(9,9))
sns.barplot(x='Sex',y='HighChol',data=data,palette="Blues")
plt.title('Gender wise Analysis')
plt.xlabel("Sex")
plt.ylabel("HighChol")
plt.show()
```



The above figure shows about the high cholesterol how many people are having it is shown in red bar (1) and how many people are not having it is shown in blue bar(0) from our collected data.

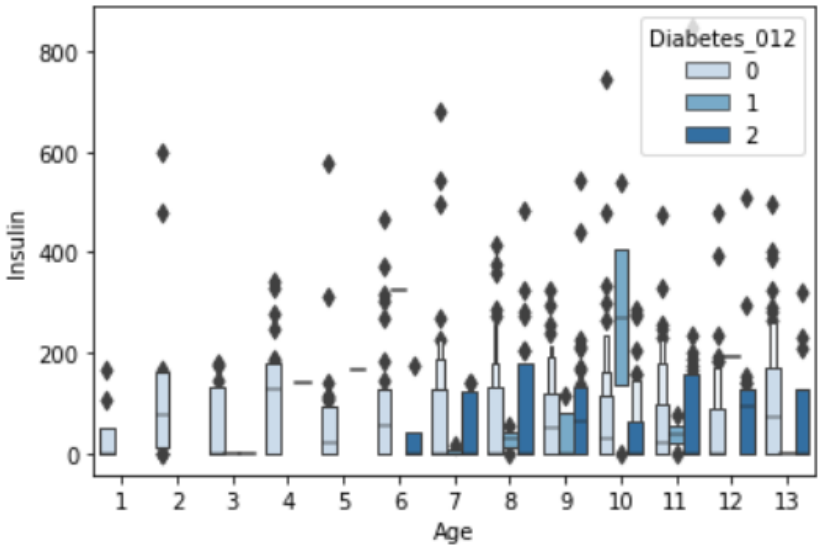
```
In [67]: data.Glucose.plot.box()
```

Out[67]:



The above figure shows about the glucose column in the box plot but use it in a consideration when using this chart is that the box and the whiskers can overlap, which is very common when plotting small sets of data.

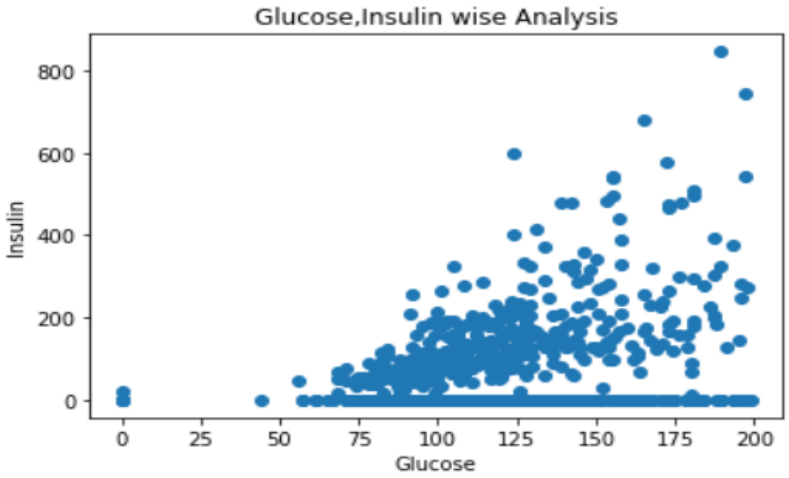
```
sns.boxenplot(x='Age',y='Insulin',hue='Diabetes_012',data=data,palette="Blues")
plt.xlabel("Age")
plt.ylabel("Insulin")
plt.show()
```



Compared between the age and insulin and diabetes levels.

In [69]:

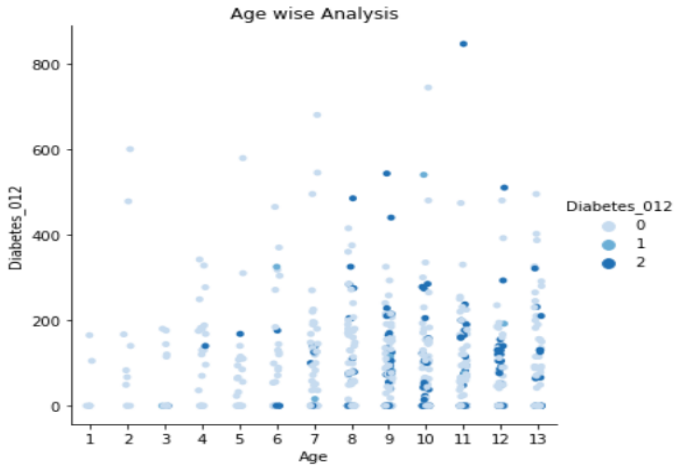
```
plt.title('Glucose,Insulin wise Analysis')
plt.xlabel("Glucose")
plt.ylabel("Insulin")
plt.scatter(data.Glucose,data.Insulin)
plt.show()
```



Compared to analysis the insulin column and glucose column.

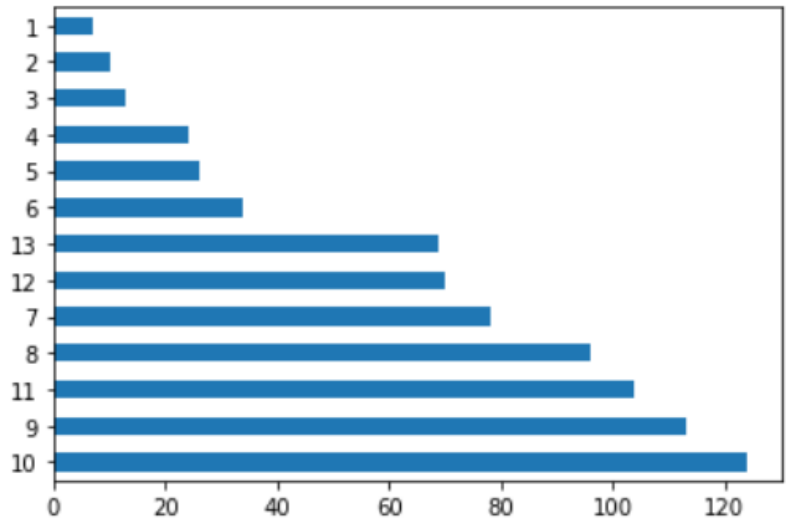
Glucose

In [66]:
```
sns.catplot(x='Age',y='Insulin',hue='Diabetes_012',data=data,palette="Blues")
plt.title('Age wise Analysis')
plt.xlabel("Age")
plt.ylabel("Diabetes_012")
plt.show()
```



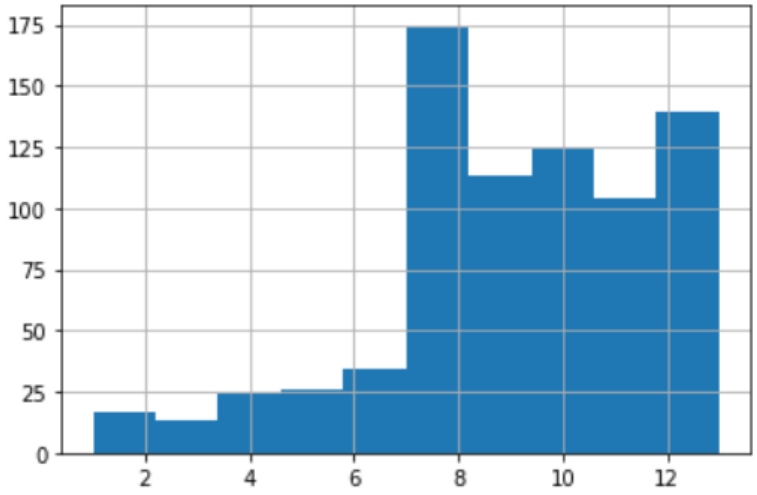
Compared between the age and insulin and diabetes levels.

```
data.Age.value_counts().plot.barh()
```

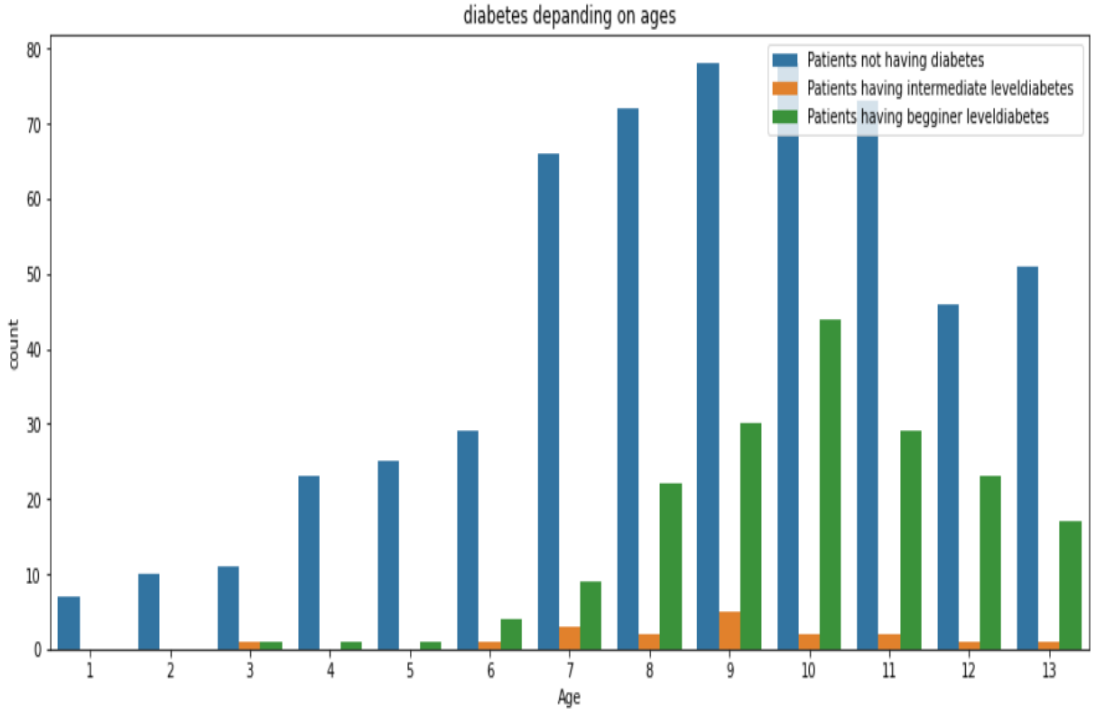


The above figure is which age of people are highest visitors and lowest.

```
data.Age.hist()
```



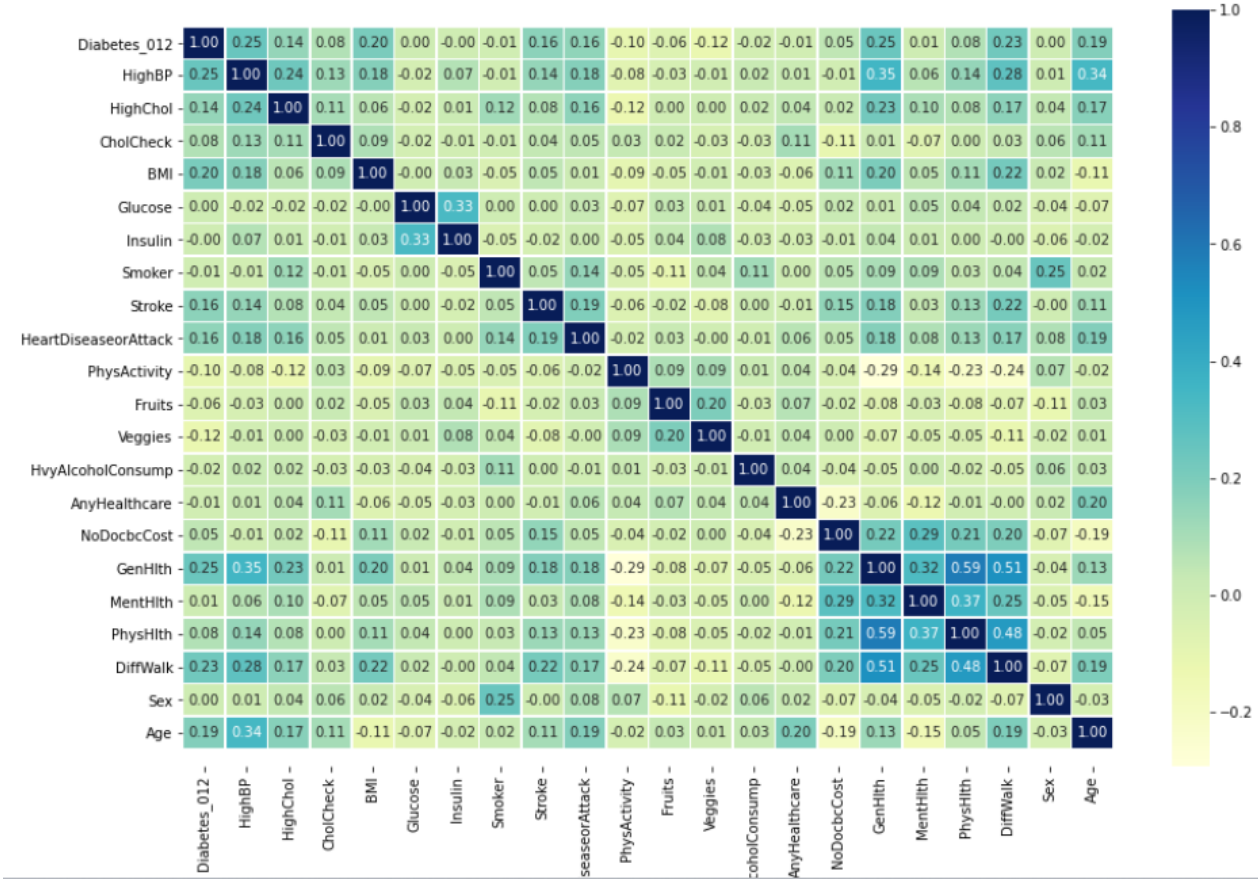
```
plt.figure(figsize=(15,6))
sns.countplot(x = 'Age', hue = 'Diabetes_012', data = data)
plt.title("diabetes depanding on ages")
plt.legend(["Patients not having diabetes ","Patients having intermediate leveldiabetes ","Patients having begginer leveldiabetes "], loc= "upper righ
```



The above figure shows There are levels in diabetes 0=no disease, 1=pre diabetes, 2=diabetes. it is compared with sex how many people having 0 level or 1 level or 2 level of diabetes.

```
corr_matrix = data.corr()
fig, ax =plt.subplots(figsize=(15,10))
ax = sns.heatmap(corr_matrix,
                 annot =True,
                 linewidths=0.5,
                 fmt=".2f",
                 cmap="YlGnBu");
bottom, top = ax.get_ylim()
ax.set_ylim(bottom+0.5,top-0.5)
```

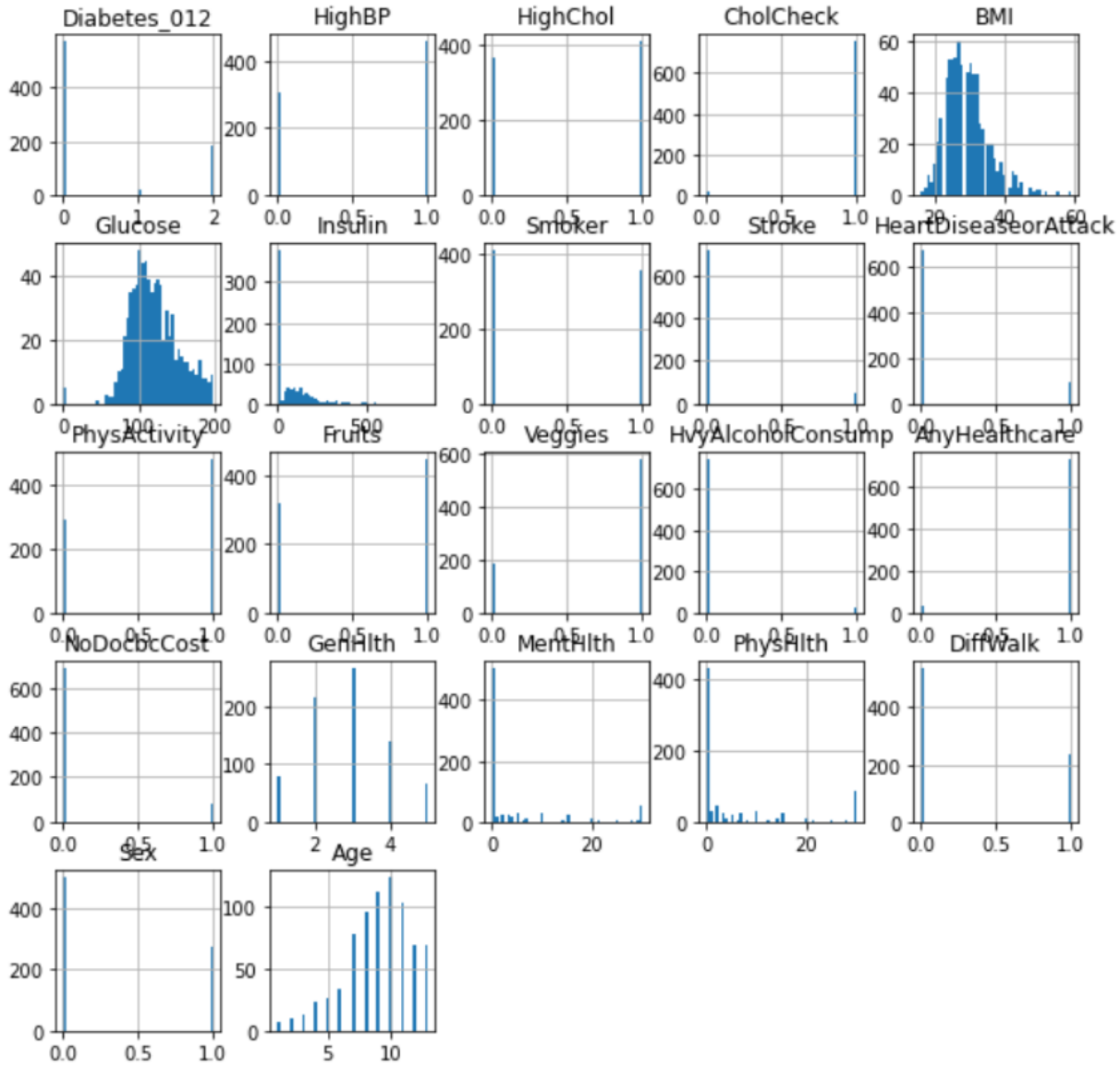
```
(22.5, -0.5)
```



A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. A correlation matrix is used to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses. From above figure we plotted the correlation matrix and we took the necessary measures to plot it.

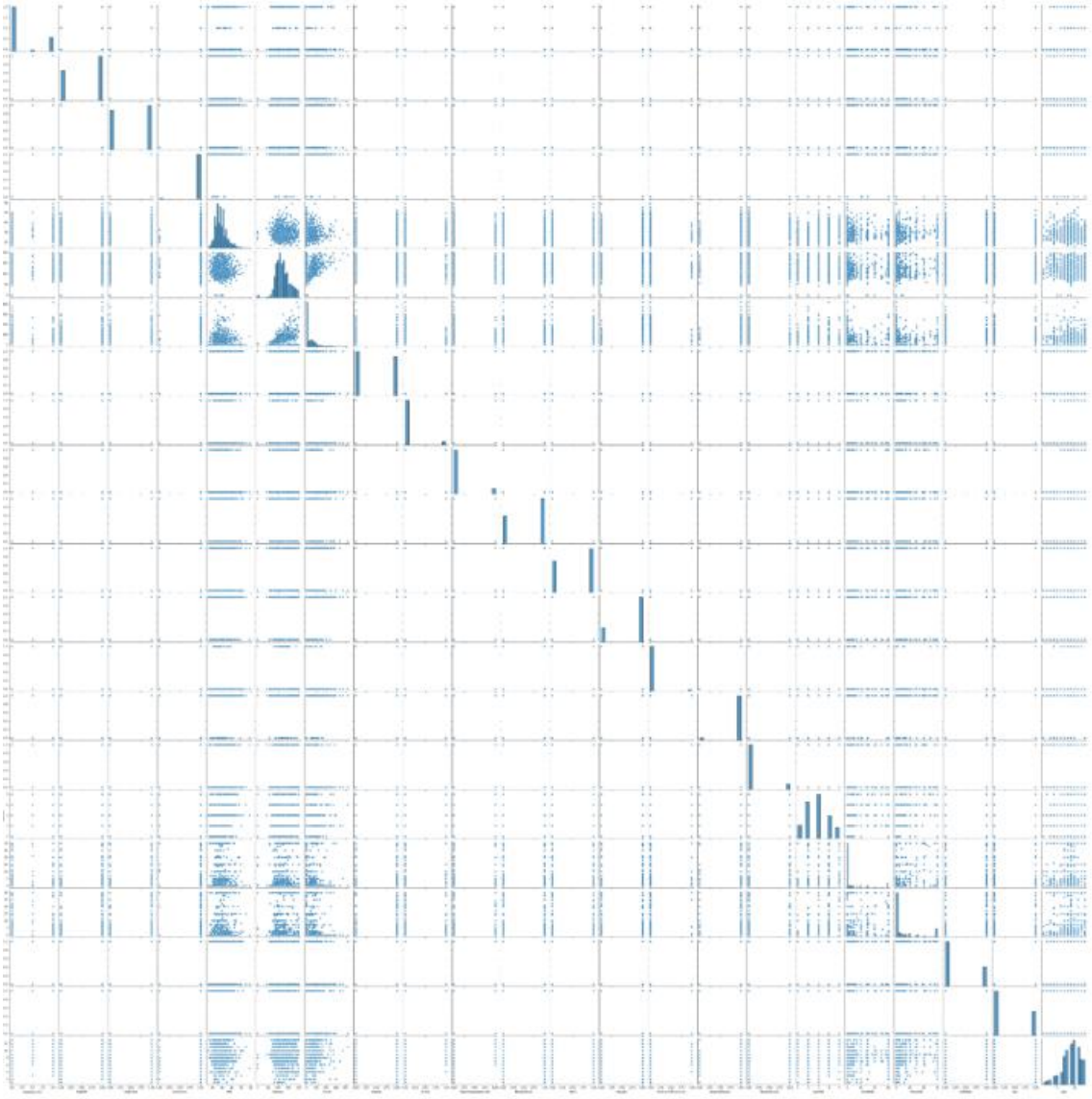
We see that there are so many similar values and with this we can evaluate the model easily and visualize predominantly.

```
data.hist(figsize=(10,10),bins=50)
plt.show()
```



The above Histogram describes about variation of a bar chart in which data values are grouped together and put into different classes and there are 22 values and 22 figures showing different visualizations

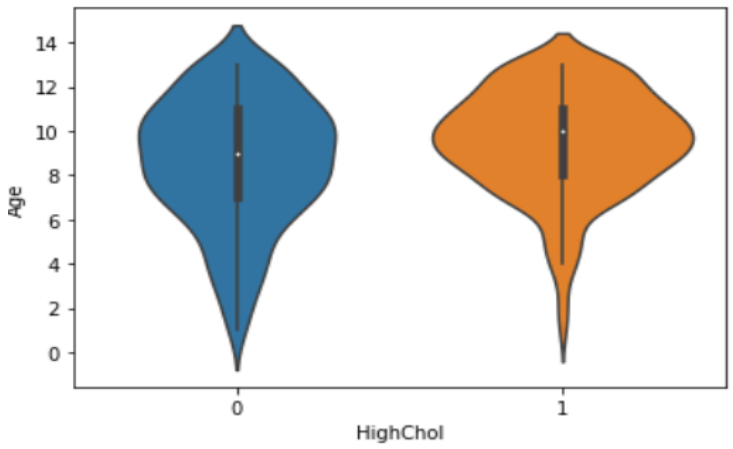
```
sns.pairplot(data)
```



Pairplot visualizes given data to find the relationship between them where the variables can be continuous or categorical. Plot pairwise relationships in a data-set. Pairplot is a module of seaborn library which provides a high-level interface for drawing attractive and informative

statistical graphics. By this pair plot hotel or resort management can visualize there issues and sort it out in a proper manner

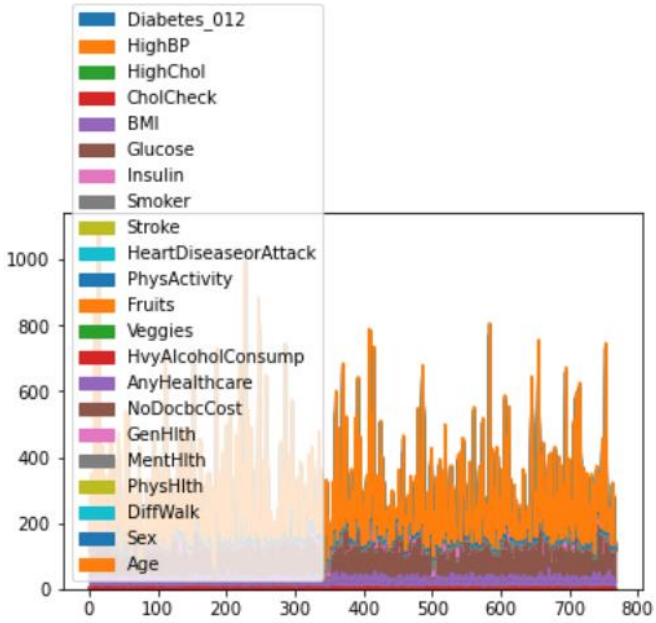
```
sns.violinplot(x='HighChol', y='Age', data=data)
```



The above figure shows about the high cholesterol and the age Violin plots and they are used to visualize data distributions, displaying the range, median, and distribution of the data.

In [82]:
```
data.plot.area()
```

Out[82]:



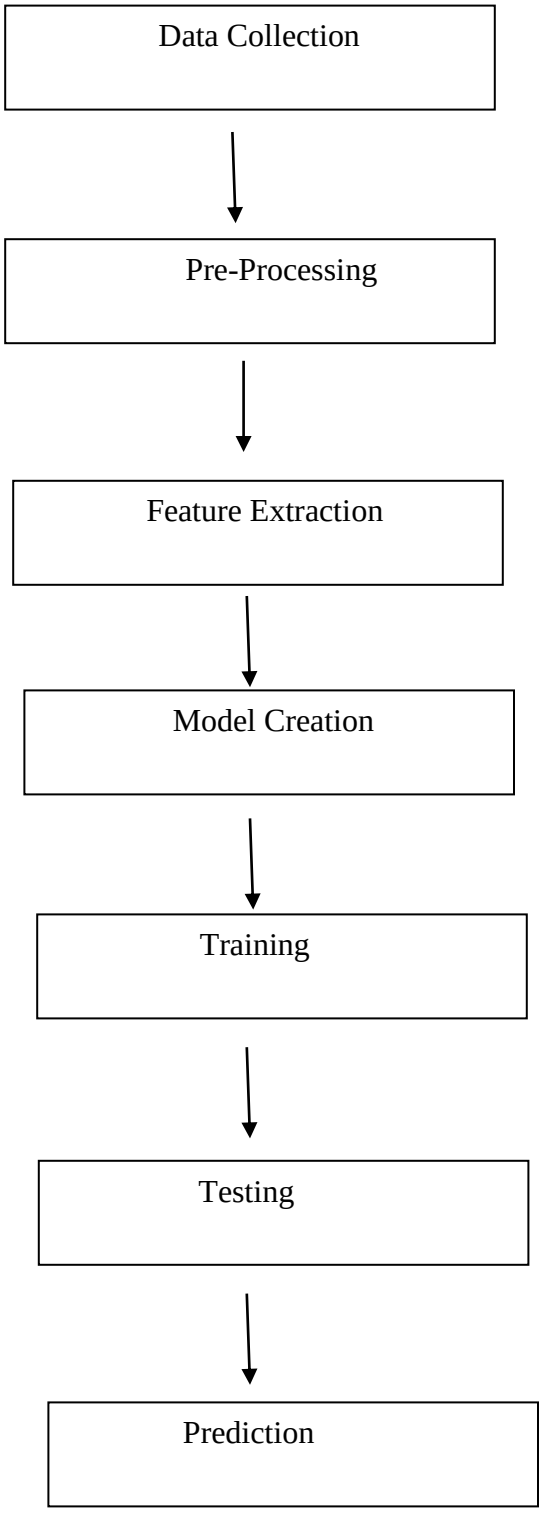
It shows area of an each column by different colour .

4.Methodology

4.1 Procedure

First we loaded the dataset into jupyter notebook. uploaded the path of our csv file in jupyter notebook. And we performed the basic methods and feature engineering and split into categorical into numerical ,found the outliers and completed the data pre processing and imputing the variables .

4.2Model Architecture.



Data Collection:

In the first step we collect data from the reliable source. Sex, age, polyuria, polydipsia, unexpected weight loss, weakness, polyphagia, genital thrush, visual blurring, itching, irritability, delayed healing, partial paresis, muscle stiffness, alopecia, obesity are some of the criteria in the diabetes record set sample. Predicted outcome class, where '1' denotes a positive diabetes patient class and '0' denotes a negative diabetes patient class.

Pre-Processing:

Pre-processing is the next step. It's an important phase in the data discovery process. The majority of health-care data contains gaps in value and inconsistencies. We apply Synthetic Minority Oversampling Technique (SMOTE) in this project, which is a well-known pre processing approach for dealing with unbalanced datasets.

Feature Extraction:

Feature extraction is the process of converting raw data into numerical features that may be processed while maintaining the information in the original data set. It yields better results than simply applying machine learning to raw data. This is an important categorizing feature.

Model Creation:

The KNN algorithm, which stands for K Nearest Neighbour, is used in this project. It's a classification and regression supervised machine learning algorithm. Every single accessible example is categorized according to a resemblance measure in the KNN algorithmic calculation (e. separation functions). A case is assigned to the class with the most votes from its neighbours, with the case being relegated to the KNN class with the most votes calculated using a separation function. Examining the set of information data determines the estimation of k (the positive whole number). Cross-validation is another method for reflectively determining a good value for k by using a free data set to cross-validate the k.

Training :

During the training process we trained the machine from data source. The test data is transformed and predicts the accurate result. During the training process the machine learning automatically selects the correct learning algorithm, based on the type of target that is specified in the training data source.

Testing:

Training data set which will be validated using the test dataset model. The test data is transformed and predicts the accurate result is 94% is achieved.

Prediction:

This module predicts the user is suffer from diabetic or not using KNN algorithm. And also predict the diabetic level normal, type 1 and type 2 using the diabetic symptoms. After training the model we had measured with different parameters within the datasets and accuracy rate of KNN with 94% (k=3) is achieved. In the future, this hierarchical framework combined with machine learning algorithms could be used to predict or analyse various disorders. Other ML computations can be used to enhance and improve the job for diabetes examination.

4.3 Software Description

Python is a high-level, general-purpose programming language. Its design philosophyemphasizes code readability with the use of significant indentation.

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library.

Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

5.RESULT AND DISCUSSION

In this project, the SVM algorithm is used to predict diabetes, with the outcome being categorized as Yes or No. If the result is 'No', the persons predicted diabetic level is 'Normal'. If the result is 'Yes' there are two possibilities of diabetics level such as type1 with early stage diabetic and type 2 with established diabetics. The patient also gets the remedies and test result in timely manner without any cost. A cost-effective disease prediction can be found by analysing these supervised learning techniques. The results reveal a high level of diabetic illness diagnosis accuracy when non-clinical factors are used. The diabetic detection and the type of diabetic level is detected using SVM algorithm and the accuracy is achieved at 96%. The result shows enforced methodology and the SVM algorithm performs comparatively well.

```
from sklearn.svm import SVC
model =SVC()
```

```
# metrics
from sklearn import metrics
from sklearn.metrics import confusion_matrix
```

```
model=model.fit(x_train,y_train)
```

```
model.score(x_test,y_test)
```

```
0.9675324675324676
```

```
model.score(x_train, y_train)
```

```
0.9788273615635179
```

```
y_pred=model.predict(x_test)
y_pred
```

```
array([0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0,
       0, 2, 0, 2, 2, 2, 0, 2, 0, 0, 2, 0, 2, 0, 0, 0, 0, 0, 2, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 2, 2, 0, 0,
       0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 2, 0, 0, 2,
       0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 2, 0, 0, 0, 2, 0, 2, 0,
       0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 2,
       2, 2, 2, 0, 2, 0, 0, 0, 2, 2, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0],
      dtype=int64)
```

```
import numpy as np
print(" %.2f" % np.mean(((y_pred - y_test)** 2)**0.5))
```

```
 0.03
```

6.CONCLUSION AND FUTURE SCOPE

One of the most clinically difficult tasks nowadays is detecting diabetes in its early stages. The goal of this project is to create a framework for getting a timely diabetes prognosis. The diabetic detection and the type of diabetic level is detected using SVM algorithm and the accuracy is achieved at 96%. Any user can use this system with their login id and password and user can create an account also. Our system can provide better interface which will provide better user system communication and thereby reducing complexity. The

remedies also given in this site which are useful for the diabetic patient and the medical filed as a whole. The remedies include the yoga treatment, food habits, and they can watch relevant diabetic videos.

An admin module included to manage the doctors details such as the name, specialization, hospital details, contact number etc. which would be helpful for the diabetic patients for their further treatments. The goal of this project is to use machine learning techniques to create a system that can accurately predict diabetes in a patient early on.

In future we can add more algorithms to find outputs and algorithms can be compared to find the efficient algorithm. We can add visitor query module, where visitors can post queries to administrator and admin can send reply to those queries. In admin module we can add more about doctors booking, treatment etc. So, the diabetic patient can book the doctors for their further treatment can view treatment module, where doctors upload treatment details for corresponding patient. This work can be extended by involving different datasets for any other kind of predictions. In future, this structured framework with the ML algorithms can be utilized to anticipate or analyze different illnesses. The work can be expanded and improved for diabetes examination by utilizing other ML calculations.

7.REFERENCES

https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset

https://github.com/Anillingampalli/j-component/blob/main/Defml%20project%20(1)%20(1).ipynb