Predicting Diabetes with Logistic Regression

Diabetes is a chronic disease with a significant impact on people's health and quality of life. In this presentation, we will explore how to create a machine learning model to predict diabetes using logistic regression.

What is Diabetes?

Definition

Diabetes is a chronic condition that affects the way your body processes blood sugar. There are two types of diabetes: Type 1 and Type 2.

Symptoms

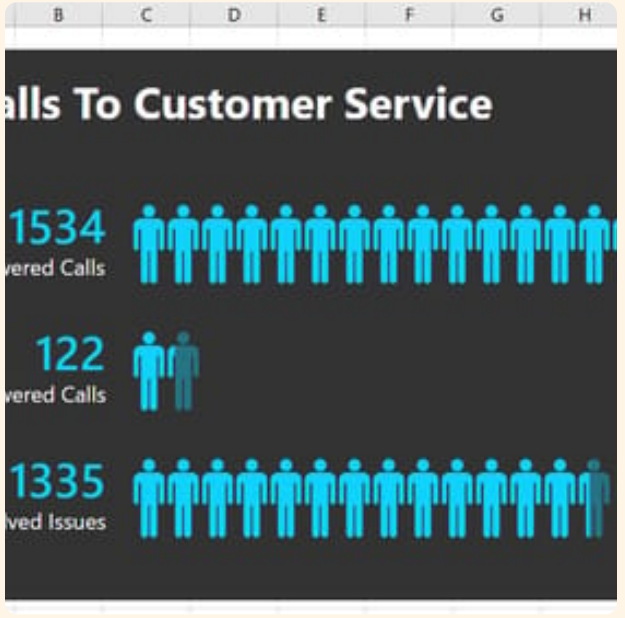
Frequent urination, excessive thirst, unexplained weight loss, blurred vision, slow wound healing and fatigue are common symptoms of diabetes.

Complications

Some of the complications of diabetes include heart disease, nerve damage, blindness, kidney failure, and amputations.

Logistic Regression Overview

$$\sqrt{x^2} = \left(x^2\right)^{\frac{1}{5}} = x$$



Definition

Logistic Regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome.

Applications

Logistic Regression is commonly used to model binary outcomes such as classifying an email as spam or not spam or predicting if a patient has diabetes or not.

Data Preprocessing and Feature Selection

1 Data Cleaning

Eliminating missing data and handling outliers can improve the accuracy of our model.

2 Feature Engineering

Creating new features from existing data can provide additional insights and improve the performance of our model.

3 Feature Selection

Removing irrelevant or redundant features can improve the interpretability and performance of our model.

Building the Logistic Regression Model

1 — Data Splitting

We split our data into training and testing sets to evaluate the performance of our model.

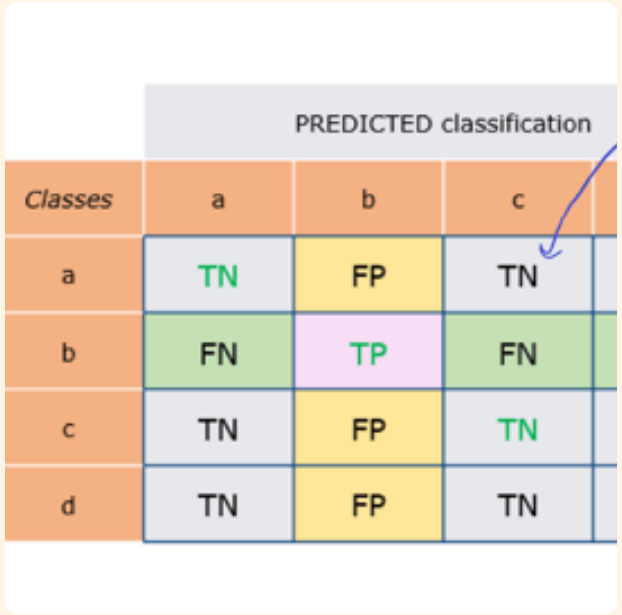
2 — Model Training

We fit our logistic regression model to the training data and adjust its parameters to optimize its performance.

3 — Model Validation

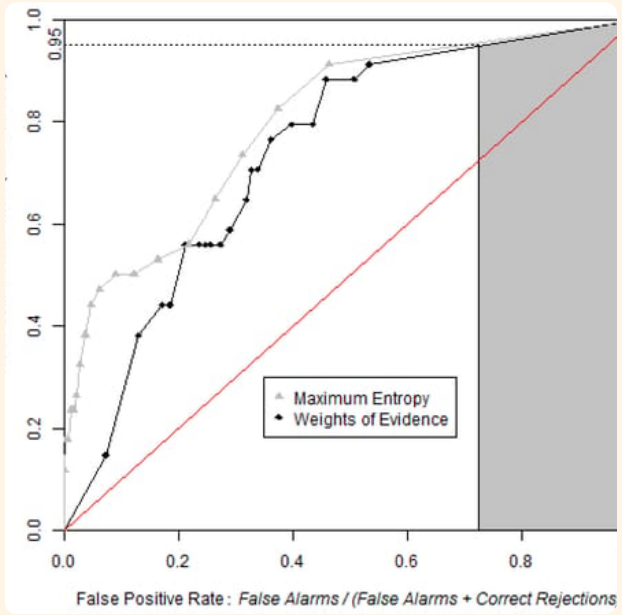
We validate our model with the testing data and metrics such as accuracy, precision, recall, and F1 score.

Evaluating and Interpreting the Model Results



Confusion Matrix

A confusion matrix presents the performance of our model in terms of true positives, false positives, false negatives, and true negatives.



ROC Curve

The ROC curve is a graphical representation of the trade-off between sensitivity and specificity of our model.

Applications of the Model for Diabetes Prediction

Healthcare

Our model can be used in clinics and hospitals to screen patients for diabetes and provide early intervention.

Insurance

Our model can be used by insurers to calculate premiums and adjust risk levels for their policyholders.

Research

Our model can be used in epidemiological studies to identify risk factors and patterns of incidence of diabetes in specific populations.

Conclusion and Next Steps

1 Conclusion

Logistic Regression is a powerful tool for predicting diabetes and other binary outcomes. With proper data preparation, feature selection, and model optimization, we can create accurate and interpretable models.

2 Next Steps

To improve our model's performance, we can explore other machine learning algorithms, gather more data, and refine our feature selection and engineering process.

Presenting To:

EXPOSYS LAB

Presented by:

SABBANI SHARATH CHANDRA

sharathsabbani05@gmail.com

College name : Jawaharlal Nehru Technological University, Hyderabad

THANK YOU