

# DATA SCIENCE PROJECT PRESENTATION

By Chaitanya Putcha

# TABLE OF CONTENTS

- ABSTRACT
- METHODOLOGY
- WEB-APP
- IDE'S & LANGUAGES
- CONCLUSIONS

# INTRODUCTION

This project is a web-based diabetes prediction system that uses machine learning to assess an individual's risk of diabetes. By entering health parameters such as glucose, BMI, and age, users receive an instant prediction powered by a trained model. The application demonstrates how data science and web technologies can be combined to support early health risk detection.

# METHODOLOGY

## DATASET

The project uses the Pima Indians Diabetes dataset, which contains medical and personal information from female patients of Pima Indian heritage. The dataset includes features such as number of pregnancies, glucose level, blood pressure, skin thickness, insulin level, BMI, diabetes pedigree function, and age. The target variable indicates whether the patient has diabetes (1) or not (0).

### **Methods applied for dataset preparation**

#### **1. Handling Duplicates:**

Duplicate rows are detected and removed to ensure data quality.

#### **2. Outlier Capping:**

Outliers in numeric columns are capped using the Interquartile Range (IQR) method to reduce their impact.

### 3. Feature Scaling:

All feature values are standardized using StandardScaler so they have mean 0 and variance 1.

## ML MODEL

This project uses a Logistic Regression model trained on the Pima Indians Diabetes dataset to predict diabetes risk. Logistic Regression is a popular machine learning algorithm for binary classification tasks. In this project, the model achieved an accuracy of approximately 87% on the test data, indicating reliable performance in predicting whether a person has diabetes based on their medical information..

# WEB-APP

This web application provides an easy-to-use interface for predicting diabetes risk using machine learning. Users can enter their health information—such as glucose level, blood pressure, BMI, and age—into a simple form. The app processes this data and instantly predicts whether the user is likely to have diabetes, helping to support early detection and awareness. The backend is powered by a trained machine learning model and built using the Flask framework.

## Diabetes Prediction

Pregnancies:

Glucose:(mg/dl)

Blood Pressure:(mm Hg)

Skin Thickness:(mm)

Insulin:(microU/mL)

BMI:(kg/m<sup>2</sup>)

Diabetes Pedigree Function:

# IDE'S & LANGUAGES

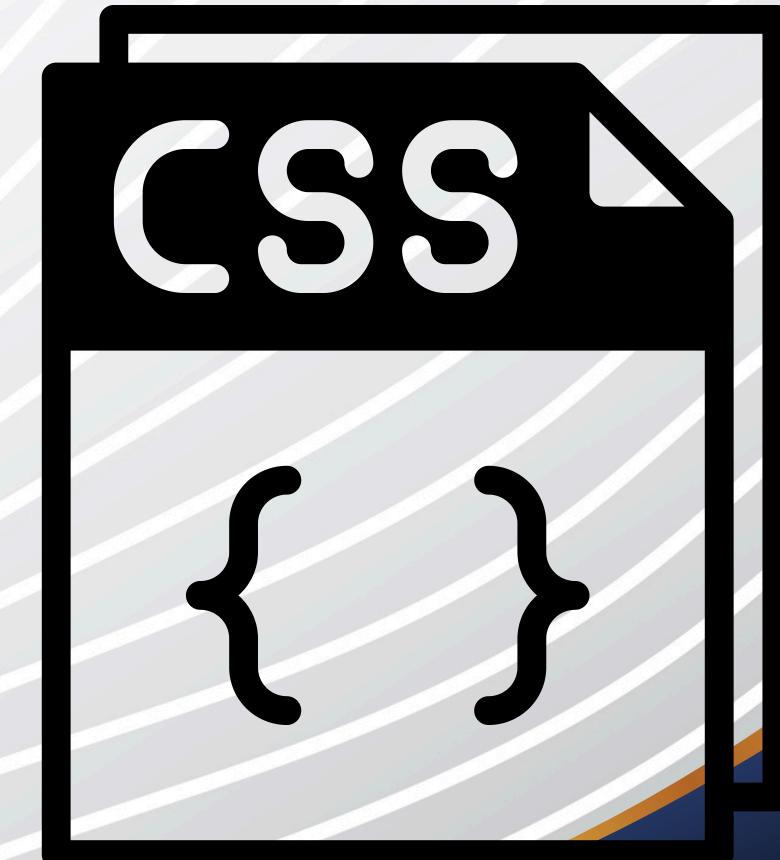
## HTML



Visual Studio Code



PYTHON



# CONCLUSION

This project successfully demonstrates how machine learning can be integrated with a web application to provide quick and accessible diabetes risk predictions. By combining effective data preprocessing, a trained classification model, and a user-friendly interface, the system offers valuable support for early diabetes detection and awareness.

# THANK YOU