

Data Analysis on Diabetes Dataset using Machine Learning Algorithms

Introduction:

In this project, we leverage the influential Pima Indians Diabetes Database to develop a robust diabetes prediction model. Our approach involves comprehensive data analysis, exploratory data analysis (EDA), and model building using a diverse set of nine advanced machine learning algorithms.

1. Data Analysis:

We meticulously analyze the dataset, uncovering patterns and insights that form the foundation for our predictive model.

2. Exploratory Data Analysis (EDA):

Through sophisticated visualizations and statistical techniques, we extract meaningful patterns, identify outliers, and understand feature relationships.

3. Model Building with Nine Algorithms:

We employ a diverse set of nine cutting-edge machine learning algorithms, including ensemble methods, deep learning, and advanced boosting, to benchmark and identify the most effective model for diabetes prediction.

4. Saving the Best Model:

Utilizing industry-standard serialization techniques, we save the best-performing model, ensuring seamless deployment for real-time predictions using the 'pickle' library.

Conclusion:

This project integrates data science principles and advanced analytics, showcasing the power of nine machine learning algorithms in creating a robust diabetes prediction model. Through a concise yet comprehensive journey, we provide a deployable solution for accurate diabetes diagnosis.

Dataset link: [Diabetes Dataset](#)

Technology Used:

- Linear Regression
- Logistic Regression
- Knn
- Decision Tree
- Random Forest
- XGBoost
- CatBoost
- Gradient Boost
- Artificial Neural Network

Team Members: Brinda Nathvani(20011268)

Shrey Shah (20009523)

Saumil Trivedi(20011349)