

# Visualizing and Classifying Diabetes Data Using SVM and Visualization Techniques

## Introduction

The second project focused on visualizing and classifying the same diabetes dataset using Support Vector Machines (SVM) and data visualization techniques. This project emphasized the importance of data visualization in understanding and preprocessing data for machine learning models.

## Data Exploration and Visualization

- **Loading Data:** The dataset was read into a Pandas DataFrame and basic information about the dataset, such as data types and memory usage, was extracted.
- **Visual Analysis:** Visualization tools like Matplotlib and Seaborn were employed to explore relationships between variables.
  - A subset of the data was reshaped into an 8x8 matrix for visualization, illustrating the importance of proper data format for image-based tasks.
  - Color maps and heatmaps provided insights into feature distributions and correlations, guiding feature selection and engineering.

## Model Building and Training

### 1. Support Vector Machine (SVM)

- **Data Preparation:** The dataset was split into features ( $x$ ) and labels ( $y$ ), followed by a train-test split.
- **Training:** An SVM classifier was trained on the training data. This model is known for its effectiveness in high-dimensional spaces.
- **Evaluation:** The performance was measured using accuracy scores, confusion matrices, and classification reports. These metrics helped assess the classifier's precision, recall, and F1-score across different classes.

### 2. Visualizing Model Output

- **Error Analysis:** Misclassifications were visualized using confusion matrices, providing insights into where the model struggled.
- **Model Interpretation:** Visual tools were used to understand the SVM decision boundaries, helping identify feature interactions contributing to the model's predictions.

## Results and Insights

The SVM model's performance was compared with simpler models like logistic regression and decision trees. Despite its complexity, the SVM provided comparable accuracy with more robustness to feature scaling and outliers.

## **Conclusion**

This project highlighted the role of visualization in data preprocessing and model evaluation. It showed that SVMs, when paired with thorough data exploration, can serve as powerful tools for classification tasks. Future work could explore kernel methods and hyperparameter tuning to enhance model performance.