

## 6. Modeling

This section details the machine learning models used for predicting cardiovascular disease based on patient data.

### 6.1 Algorithms Used

The following machine learning models were trained and saved:

- **Support Vector Machine (SVM)**
- **KMeans Clustering** (*used for exploratory analysis or grouping, not prediction*)
- **K-Nearest Neighbors (KNN)**
- **Random Forest Classifier**
- **Logistic Regression**

These models were selected to evaluate performance across a mix of linear, tree-based, instance-based, and unsupervised methods.

### 6.2 Training Process

- Data was split into training and testing sets.
- Preprocessing steps (such as scaling and encoding) were applied before model training.
- Each model was trained using consistent features to ensure a fair comparison.
- Models were saved as `.pkl` files for later evaluation and deployment.

### 6.3 Hyperparameter Tuning

- Hyperparameter tuning (where applicable) was performed using `GridSearchCV` or `RandomizedSearchCV`.
- Models like Random Forest and SVM had tuning for parameters such as:
  - `n_estimators`, `max_depth` (Random Forest)
  - `C`, `kernel`, `gamma` (SVM)
- The best performing parameters were selected based on cross-validation results and saved model metrics.

Model performance results are organized and stored in individual JSON files in the `results/` directory.