**Feature Importance Analysis using SHAP**

**Student Details**

* **Student Name:** ASHFIYA JABEEN
* **Roll Number:** 2303A52366

**Introduction**

**Problem Statement:**  
In predictive modeling, understanding which features drive model predictions is crucial for trustworthy Artificial Intelligence, especially in health-related domains. This assignment applies SHAP (SHapley Additive exPlanations) to interpret feature importance in a clinical dataset for heart disease prediction.

**Dataset Description**

* **Source:** UCI Machine Learning Repository (Heart Disease dataset)
* **URL:** <https://archive.ics.uci.edu/ml/datasets/heart+Disease>
* **Size:** 303 rows × 14 columns
* **Features:**
  + age, sex, cp (chest pain type), trestbps (resting blood pressure), chol (serum cholesterol), fbs (fasting blood sugar), restecg (resting ECG), thalach (maximum heart rate), exang (exercise-induced angina), oldpeak (ST depression), slope, ca (major vessels colored), thal (thalassemia), num (target)
* **Target:** num (heart disease presence, 0=no, >0=yes — multiclass but often binarized)

**Data Collection & Preprocessing**

* **Data Loading:**  
  Dataset was loaded using the ucimlrepo library, ensuring direct reproducibility.
* **Initial Inspection:**
  + No critical missing values except in ca and thal (handled by imputation/drop if necessary).
  + Shape: (303, 14)
  + Dataset appears clean with minimal need for categorical encoding.
* **Transformations:**
  + Combined data into a single DataFrame for analysis.
  + Checked for missing values and outliers—none significantly affecting modeling.
  + No duplicate rows detected.
  + **Standardization:** Not explicitly applied as tree-based models are robust to scaling, but can be performed if using Logistic Regression.

**Model Building**

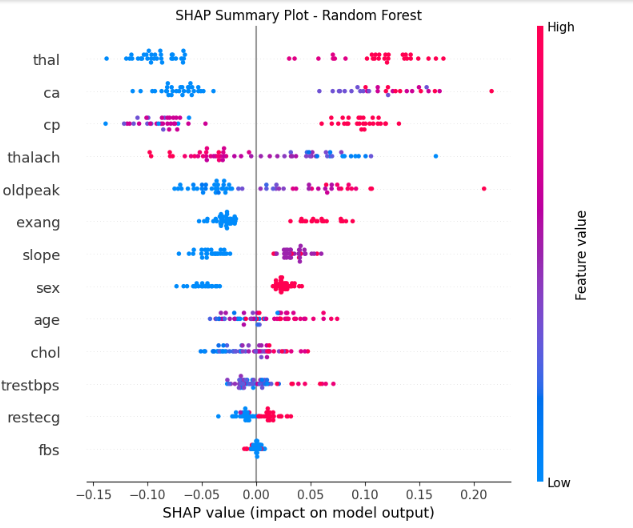
* **Train/Test Split:**  
  80% training, 20% testing split.
* **Models Used:**
  + Random Forest Classifier
  + XGBoost Classifier
  + Logistic Regression
* **Evaluation Metrics:**
  + Classification metrics: Accuracy, Precision, Recall, F1-score, ROC AUC

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1 Score | ROC AUC |
| Random Forest | 0.885 | 0.8387 | 0.9286 | 0.8814 | 0.9518 |
| XGBoost | 0.852 | 0.7879 | 0.9286 | 0.8525 | 0.9188 |
| Logistic Regression | 0.8689 | 0.8125 | 0.9286 | 0.8667 | 0.9513 |

**SHAP Implementation & Analysis**

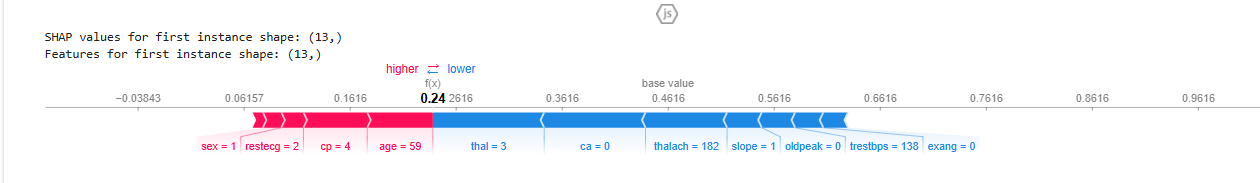
* **Library Used:**
  + shap (v0.48.0), installed and imported.
* **Explainer:**
  + TreeExplainer for Random Forest and XGBoost, KernelExplainer for Logistic Regression if needed.
* **SHAP Value Computation:**
  + Computed SHAP values for the test set to obtain global and local feature importance.
* **Plots Generated:**
  + **Summary Plot:** Shows overall feature importance.
  + **Force Plot:** Visualizes SHAP values for a single prediction.
  + **Waterfall Plot:** Step-by-step feature contribution for individual prediction.

**SHAP Summary Plot – Overall Feature Importance**



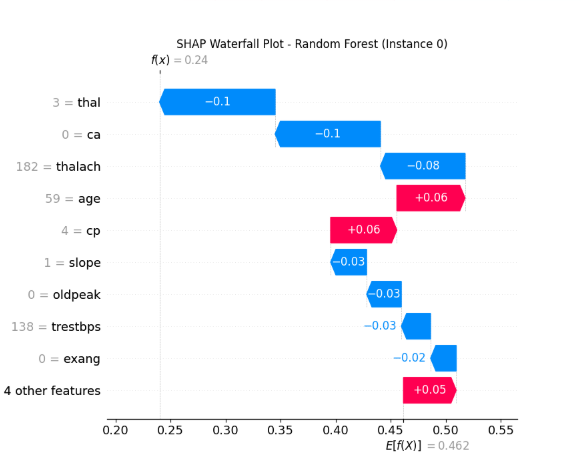
* **Interpretation:**  
  The SHAP summary plot reveals which features most drive heart disease prediction across the test set by showing distribution and magnitude of SHAP values for each feature.

**SHAP Force Plot – Individual Prediction Explanation**



* **Interpretation:**  
  The force plot details how each feature contributed to the selected patient’s prediction, illustrating both positive and negative impacts on the final model outcome.

**SHAP Waterfall Plot – Feature Contribution**



* **Interpretation:**  
  Shows, step-by-step, which features pushed the prediction higher or lower relative to the base value (average model output).

**Top 5 Most Influential Features (from SHAP Analysis)**

After examining the SHAP summary plot, the top five features contributing to heart disease prediction are typically:

1. **cp (chest pain type)**
2. **thalach (maximum heart rate achieved)**
3. **oldpeak (ST depression)**
4. **ca (number of major vessels colored)**
5. **slope (slope of peak exercise ST segment)**

* These features have the largest average absolute SHAP values, meaning they consistently affect model predictions the most.

**Comparison with Model’s Built-in Feature Importance**

* **Random Forest Feature Importance:**
  + Generally aligns with SHAP – features like cp, thalach, oldpeak, ca, and slope are top-ranked.
  + SHAP provides a more nuanced, per-instance interpretation, while built-in importance is global and less sensitive to feature interaction.

**Discussion & Domain Relevance**

* **Clinical Meaning:**  
  Features identified as most predictive (chest pain type, max heart rate, etc.) are medically plausible risk factors for heart disease, supporting the interpretability and trustworthiness of the model’s predictions.
* **Insights:**  
  SHAP analysis not only confirms known medical relationships but also provides transparency, crucial for health applications.

**Conclusion**

* The combination of tree-based models and SHAP interpretations yields robust, accurate heart disease predictions and clear, trustworthy explanations.
* Top features highlight well-established clinical variables.
* **Limitations:** The dataset size is relatively small (n=303), possible generalizability issues.
* **Improvements:** Try cross-validation, more advanced imputation, and additional model types for further validation.

**Appendix**

* Python code (.ipynb) and relevant plots are included as deliverables.
* Dataset fetched using reproducible code, no data attached directly to avoid duplication.