UCI Heart Disease Project

Evaluation and Analysis Report

---

1. Was There a Difference in Success Among the Algorithms Used? Why?

When comparing the **test set** and **5-fold cross-validation average performances** of different algorithms, it is observed that **ensemble-based models** (Random Forest, XGBoost, LightGBM) and **Artificial Neural Network (ANN)** models generally achieve **higher accuracy and F1 scores** than classical algorithms (Logistic Regression, KNN, Naive Bayes, Decision Tree).

Reason:
Ensemble and deep learning models can better model **complex relationships and interactions** between variables, while classical methods may struggle to capture these patterns.

These differences are clearly visible in the accompanying **performance tables and confusion matrix visualizations**.

---

2. Did the Number of Features Affect Success?

Yes, the **number and quality of features** significantly influenced model performance.

- **Feature selection** and the use of **important variables** directly improved results.
- When **highly correlated variables** were used, increases in accuracy and F1 scores were observed.
- Removing **redundant or low-information features** improved the model's **generalization** capability.

A **feature importance analysis** using Random Forest confirmed the **positive impact** of selecting meaningful features.

---

3. How Did Your Own ANN Model Compare to Classical Algorithms?

The basic ANN model created with **Keras** generally achieved **higher or comparable accuracy and F1 scores** than classical algorithms.

- After thorough **data preprocessing** (e.g., scaling, balancing), the ANN model **learned complex patterns** more effectively.
- However, ANN performance was **highly sensitive** to:
    - Dataset size
    - Quality of features
    - Hyperparameter tuning

With proper preprocessing and optimization, the ANN model showed **superior performance**.

---

4. Were There Differences Between Cross-Validation and Single Test Set Results?

When comparing **5-fold cross-validation average scores** and **single test set results**, the scores were **generally close**.

- If there were large discrepancies, this could indicate:
    - **Overfitting**
    - **Random variation** due to data splitting

Conclusion:
Using both metrics together provides a **more reliable and realistic evaluation** of model performance.

---

5. Conclusion and General Evaluation

Aspect	Observation
Performance Difference	Ensemble and ANN-based models outperformed classical methods.
Number of Features	Meaningful feature selection improved accuracy and generalization.
ANN vs Classical	ANN performed better with preprocessing and tuning.
Cross-Validation	Provided consistent and trustworthy performance estimates.

Recommendation

To increase model performance:

- Apply **data preprocessing** (normalization, balancing)
- Use **meaningful feature selection**
- Perform **hyperparameter tuning**
- Always conduct **cross-validation** for reliability

For complex datasets, **ensemble** and **deep learning** models should be prioritized over classical approaches.