Cipheart - Heart Disease Classification with Homomorphic Encryption

Team: Sarnika Sanjiv Kumar (22PD31) and Valincia John (22PD39)

Introduction:

Heart disease remains one of the leading causes of mortality worldwide, making early prediction vital for proactive healthcare management. This project, Cipheart, utilizes machine learning combined with homomorphic encryption to build a privacy-preserving application for heart disease classification. By integrating encrypted data processing, Cipheart ensures sensitive health information remains secure even during analysis, addressing key concerns in medical data confidentiality. This project is particularly useful for healthcare providers and research institutions that manage sensitive health data, as it allows for accurate predictions and analysis without compromising patient privacy.

ML Models Used:

Our approach involves two main classification models: **logistic regression** and **Naive Bayes**.

- **Logistic Regression**:
    - Effective for binary classification tasks.
    - Provides high interpretability, making it suitable for understanding key predictors.
- **Naive Bayes**:
    - Offers a probabilistic framework for classification.
    - Can handle smaller datasets efficiently, making it robust even with limited data.

Together, these models create a solid foundation for identifying patterns associated with heart disease risk factors.

Dataset:

The UCI Heart Disease dataset is used to develop and validate our models. This dataset includes attributes such as age, cholesterol levels, and exercise-induced angina, which are critical for predicting heart disease risk. Essential data preprocessing steps, including imputation for missing values and standardization, are applied to ensure the dataset is model-ready, handling real-world data challenges effectively.

CKKS Homomorphic Encryption:

Cipheart integrates CKKS (Cheon-Kim-Kim-Song) homomorphic encryption using the TenSEAL library to maintain data privacy, which is essential when dealing with sensitive healthcare information. CKKS is specifically designed for efficient encrypted computations, allowing for operations on encrypted floating-point numbers that are common in machine learning tasks. This encryption scheme enables Cipheart to perform model predictions directly on encrypted data without decrypting it, ensuring that sensitive health data remains confidential throughout the entire process.

The CKKS encryption method is particularly well-suited to this project because:

1. **Privacy Compliance**: It addresses the stringent privacy requirements in healthcare, allowing the processing of sensitive medical data without exposure, which is crucial for building trust with patients and meeting regulatory standards.
2. **Efficient for Machine Learning**: CKKS supports operations on real numbers, which aligns well with the mathematical requirements of logistic regression and Naive Bayes. This compatibility with machine learning models makes it an effective choice for real-time, encrypted predictions.
3. **Secure Insights without Sacrificing Accuracy**: By enabling encrypted data processing, CKKS allows Cipheart to provide accurate heart disease predictions without compromising data security, creating a balance between patient confidentiality and clinical utility.

Conclusion:

Cipheart demonstrates that homomorphic encryption can be successfully integrated with machine learning models for healthcare applications, achieving high predictive accuracy while preserving privacy. This combination of secure data processing and effective classification presents a promising solution for real-time, confidential healthcare analytics. Cipheart serves as a step forward in developing AI-driven, privacy-conscious tools for healthcare, meeting the critical need for secure, reliable predictive models in medical diagnostics.