

# AIML

Sneha Pareek

June 2024

I have successfully developed a Flask web application that enables users to predict six different diseases: heart disease, cancer, kidney disease, hepatitis, lung cancer, and diabetes. For this project, I gathered datasets for each disease and constructed classification models. After evaluating the performance of each model, I selected the ones with the highest accuracy.

## 1 Data sources and Model training

- Relevant datasets for each disease were gathered from Kaggle.
- Each dataset contained different features, resulting in varied accuracies for the trained models.
- For each disease, six classification models were trained: Logistic Regression, k-NN, kernel SVM, Decision Tree, SVM, and Random Forest.
- After the training process, a feature selection technique was applied to remove features that did not contribute to improving accuracy.
- The Chi-square test was used for feature selection as the datasets included both categorical and numerical features.
- Model performance was evaluated using the cross-validation technique, specifically k-fold cross-validation.
- After selecting the best training models for each disease, the hyperparameters were fine-tuned to enhance accuracy.

## 2 User Interface

- The web app was built using HTML, CSS, and Bootstrap for the front end, and Flask for the back end.
- The trained models were deployed and are used to provide predictions to users based on their input.

### 3 Instructions for running the application

- The folder **Disease\_web\_app** contains all the necessary files related to the project.
- The folder **model\_training\_process** includes the training process involved in selecting the best models for each disease.
- To run the app, one can simply execute the command **python3 app.py** in the terminal.
- To run the models separately, one can navigate to the **trained\_models** folder and open the Jupyter notebook to evaluate the models.