Breast Cancer Diagnostic Models

# AI in Enterprise – Lab 2

# Shahabuddin Syed – 100895169

Steps followed to build the Machine learning Models

## Importing data:

In this step, the dataset of the Breast Cancer Wisconsin (Diagnostic) dataset is imported into the Jupyter Notebook. This involves reading the dataset file (e.g., CSV) using appropriate libraries such as pandas or NumPy. The data is loaded into a data structure that allows for further exploration and analysis.

A screenshot of a computer

Description automatically generated

EDA and preprocessing:   
Exploratory Data Analysis (EDA) is performed to gain insights into the dataset and understand its structure. This includes examining the distribution of features, identifying missing values or outliers, and visualizing relationships between variables. Preprocessing techniques are then applied, such as handling missing data, feature scaling, and encoding categorical variables, to prepare the data for model training. ­A picture containing screenshot, rectangle, square, parallel

Description automatically generated

­A screenshot of a video game

Description automatically generated with medium confidence

Model building:   
In this step, the ML model(s) are constructed using algorithms such as logistic regression, random forest, support vector machines, or neural networks. The features from the preprocessed dataset are used as inputs, and the target variable (diagnosis) is defined as the output variable for supervised learning. The model is trained using an appropriate training algorithm and hyperparameters are tuned to optimize performance.

A screen shot of a computer program

Description automatically generated with low confidence

Model evaluation and best model selection:   
Once the model(s) are trained, they are evaluated using evaluation metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve. This step assesses how well the model performs on unseen data and helps determine the model's effectiveness. Multiple models may be evaluated, and the best model is selected based on the chosen metric(s) and performance criteria.

A screenshot of a computer

Description automatically generated with medium confidence