**Assignment #2 (Git, GitHub, ML)**

**Somya (100901887)**

**Introductions:**

This Jupyter Notebook explores a breast cancer dataset, performing EDA, statistical analysis, outlier management, data normalization, duplicate line removal, and providing insights through visualizations. The dataset is then split into training and testing sets for machine learning modeling.

I have used both GitHub desktop version and PowerShell to make changes in my local machine and push it to remote server.

I did crate 3 branches and pushed identical ipynb file and data set however I have only created Readme file in one branch to show the working process of keeping changes apart from main branch.

**Step 1: Install git in your local machine (If it is already done, skip this step)**

A screenshot of a computer

Description automatically generated

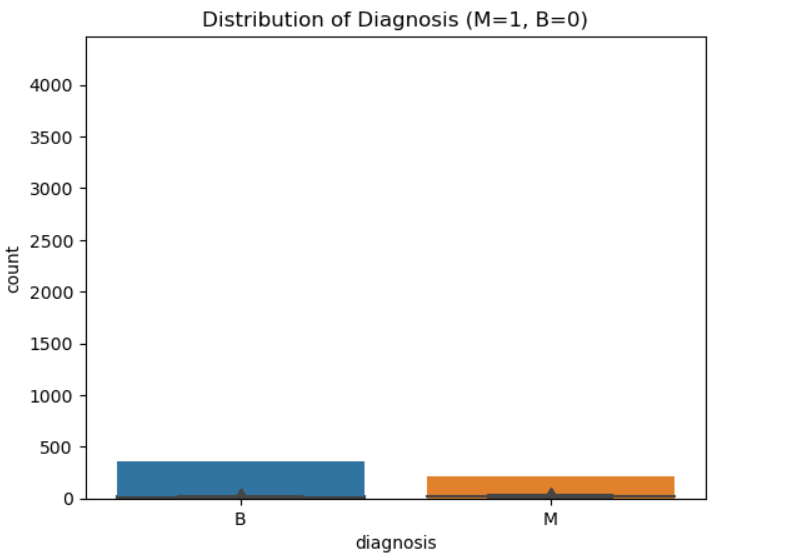
**A picture containing text, screenshot, font

Description automatically generatedStep 2: Build a ML model for Breast Cancer Wisconsin (Diagnostic) Data set in jupyter notebook**

**A screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a computer screen

Description automatically generated with low confidenceStep 3: Please provide screenshots for various stages of the design process (importing data, training, evaluation …)**



Steps break down -

1. Prepare data: We split the dataset into input features (X) and target variable (y). Then, we scale the input features using StandardScaler() from the sklearn library.
2. Perform PCA: We use PCA() from the sklearn library to perform PCA on the scaled input features. Here, we set n\_components=2 to reduce the number of features to 2 principal components.
3. Create logistic regression model: We create a logistic regression model using LogisticRegression() from the sklearn library. Here, we set random\_state=42 for reproducibility.
4. Perform cross-validation to compare algorithms: We use cross\_val\_score() from the sklearn library to perform cross-validation. Here, we set cv=10 to perform 10-fold cross-validation, scoring='accuracy' to use accuracy as the evaluation metric, and log\_reg as the model to evaluate. We print the accuracy for each fold and the average accuracy across all folds. A picture containing text, screenshot, font, line

   Description automatically generated

A picture containing text, font, screenshot, line

Description automatically generated

**Step 4: Upload your model (Python script, let’s called it <yourname>\_model\_v1) to GitHub. Provide screenshot of all your git commands and your command prompt showing success of commit of your model files in the remote host.**

A screen shot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

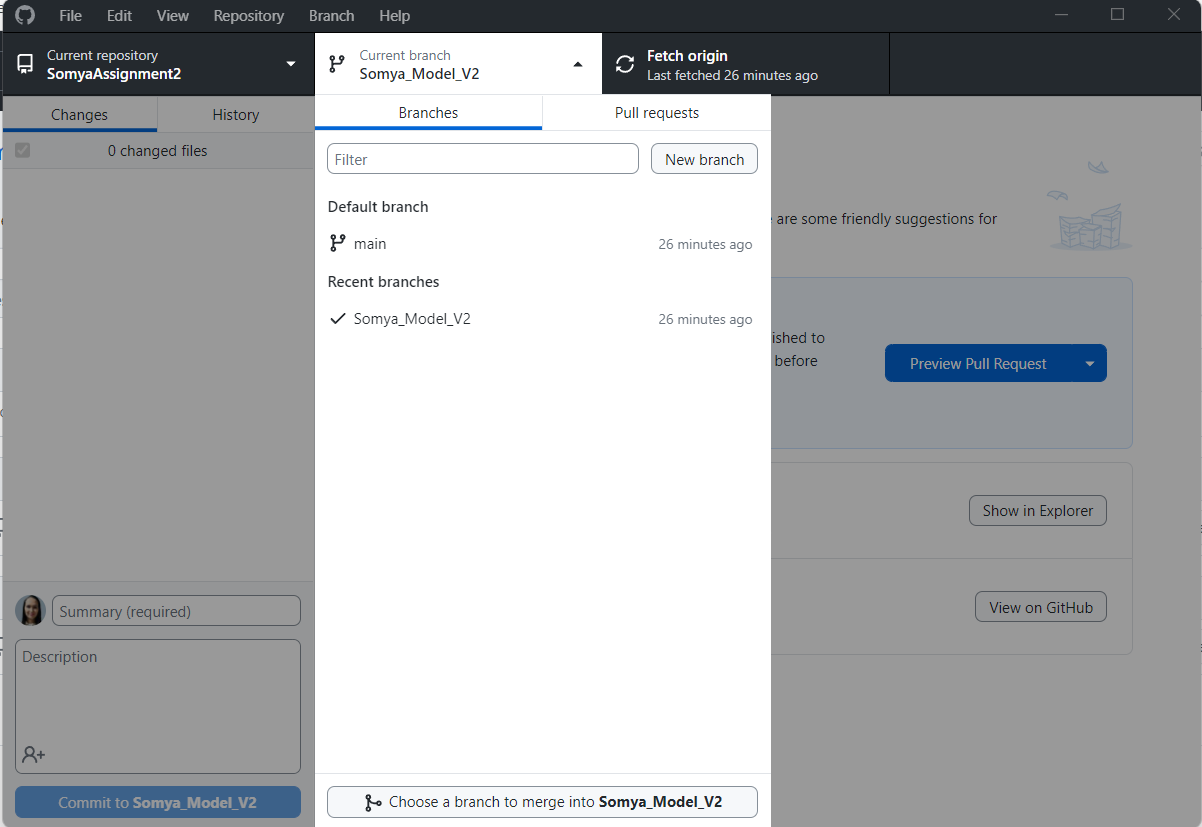
On Local Machine

A screenshot of a computer

Description automatically generated

Cloned on Github

**Step 5: Create a branch in your repo and upload another ML model (may be using a different algorithm and named the file: <yourname>\_model\_v2) of your choice for the same dataset into that branch.**



A screenshot of a computer

Description automatically generated

**Step 6: Navigate to your newly created branch and provide a screenshot showing status of your repo.**

A screenshot of a computer program

Description automatically generated with medium confidence

**Step 7: Provide a screenshot showing your log of activities and perform your final commit.**

A screenshot of a computer program

Description automatically generated with medium confidence

**Step 8: Provide a description of your program in the README.md file.**

1. Created a Readme.md file.
2. Put it in the local folder.
3. Pushed it in clone branch (Somya\_Model\_V2)
4. Commit the changes in main branch.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

**Step 9: Make your repo public and share the link of your repo for check.**

<https://github.com/Sachan25/SomyaAssignment2>