

CSP-571 Project Plan and Detail Fall 2022

Members

Utkarsha Malegaonkar (A20493621)

Saurabh Jadhav (A20491831)

Rutika Dhoka (A20501969)

TASK NAME	DUE DATE	ASSIGNED TO
Data Gathering		
Kaggle Dataset : Breast Cancer Prediction	9/23/22	Utkarsha
Data Preparation		
Description of data	10/10/22	Rutika, Saurabh
Importing the data	10/12/22	Rutika, Saurabh
Checked missing values and duplicate entries	10/14/22	Rutika, Saurabh
Data Exploration and Visualisation		
Identify the predictors which we are going to use from dataset	10/15/22	Utkarsha, Saurabh
Divided data into Training and Test data	10/20/22	Utkarsha, Saurabh
create the model matrix with all the predictors	10/22/22	Utkarsha, Saurabh
Data Cleaning		
Selecting most relevant data	10/28/22	Rutika, Utkarsha
Exploring features	10/29/22	Rutika, Utkarsha
Normalizing Data	10/31/22	Rutika, Utkarsha
Building Model		
Building RandomForest (RF), Gradient Boosting Machine (GBM)	11/1/22	Rutika, Saurabh
Building Light Gradient Boosting Machine (lightGBM) and XGBoost	11/7/22	Saurabh, Utkarsha
Evaluating The Model		
Identify the best model of the above	11/10/22	Rutika, Utkarsha
Test our models and find the RMSE for the training vs test data	11/10/22	Rutika, Utkarsha
Deploy the model		
Perform PCA on the data	11/11/22	Saurabh
Finds patterns in the data: t-SNE transform	11/12/22	Utkarsha
Reviewed model test results and select the final model	11/14/22	Rutika
Run the selected model on the entire training dataset and apply it to the test data	11/15/22	Saurabh, Utkarsha, Rutika
Documentation		
Collect plots and performance matrix	11/23/22	Rutika
Write conclusion and insights	11/24/22	Utkarsha
Final report generation	12/1/22	Saurabh
Final project presentation	12/1/22	Saurabh, Utkarsha, Rutika

DUE DATE

