CSP-571 Project Plan and Detail

Group Members:

- Vaishnavi Mule (A20516627)
- Neha Gawali Gawali (A20523722)
- Atharva Nirali (A20517247)
- Vikas Reddy- Project Lead (A20539316)

Sr.no	TASK NAME	DUE DATE	ASSIGNED TO
1	**Project Initiation**		
2	Define the research goal and objectives.	10/19/23	Vaishnavi Mule
3	Identify the dataset source and relevant references.	10/19/23	Neha Gawali
4	Formulate research questions.	10/19/23	Vaishnavi Mule
5	**Literature Review**		
6	Review relevant papers and articles on housing price prediction	10/25/23	Vaishnavi Mule/ Neha Gawali
7	Analyze data analysis techniques and modeling approaches for insights	10/28/23	Vikas Reddy/ Atharva Nirali
8	**Data Collection and Preprocessing**		
9	Obtain the "California Housing Prices" dataset from Kaggle.	10/19/23	Neha Gawali
10	Perform data cleaning:		
11	Handle missing values through imputation or removal.(Handling through code)	11/08/23	Vikas Reddy
12	Identify and handle outliers using techniques like Z-score or IQR.(Handling through code)	11/10/23	Atharva Nirali
13	Feature engineering:		

15	creating ratios, combining features, or transforming variables to capture more meaningful information	11/11/23	Atharva Nirali
16	Perform binarization for categorical features.	11/11/23	Vaishnavi Mule
17	Apply scaling or normalization to numeric features.	11/11/23	Vikas Reddy
18	Categorical data encoding:	11/11/23	
19	Utilize one-hot encoding or label encoding, as appropriate.	11/11/23	Neha Gawali
20	Construct a data processing pipeline for consistency.	11/10/23	Vikas Reddy
21	**Data Splitting and Sampling**		
22	Utilize stratified sampling to divide the dataset into training and testing sets	11/12/23	Atharva Nirali
23	**Model Selection and Evaluation**		
24	Decide regression models: Linear regression, Random forest regression	11/04/23	All
25	Perform hyperparameter tuning:		
26	Utilize grid or random search for optimal hyperparameter values in the random forest model	11/12/23	Vaishnavi Mule/ Neha Gawali
27	Cross-validation:		
28	Assess models' performance using cross-validation techniques	11/12/23	Atharva Nirali
29	Define Root Mean Squared Error (RMSE) as the primary evaluation metric	11/12/23	Vikas Reddy
30	**Baseline Model**		
31	Train and evaluate a linear regression model to establish a baseline	11/14/23	Vikas Reddy
32	**Model Comparison**		
33	Compare the performance of decision tree and random forest regression models against the baseline model	11/14/23	Atharva Nirali

34	**Software Packages and Tools**		
35	Utilize R and RStudio for data analysis and modeling	11/17/23	Neha Gawali
36	Employ R packages, including:		
37	tidyverse for data manipulation and visualization	11/17/23	Vaishnavi Mule
38	caret for streamlining model creation	11/17/23	Vikas Reddy
39	randomForest for building regression models	11/17/23	Atharva Nirali
40	**Project Documentation**		
41	Maintain comprehensive documentation for all project activities	NA	All
42	Record findings, results, and insights from the data analysis and model selection	11/27/23	All
43	**Project Implementation**		
44	Code development:		
45	Implement the data preprocessing steps	11/19/23	Vikas Reddy
46	Create scripts for training and evaluating regression models	11/19/23	Vaishnavi Mule
47	Collaborate with team members for code review and integration	11/20/23	All
48	**Testing and Validation**		
49	Conduct rigorous testing of the predictive model	11/22/23	Neha Gawali
50	Validate model performance against known data and benchmarks	11/22/23	Atharva Nirali
51	**Final Report and Presentation**		
52	Summarize research findings in a comprehensive report.	11/29/23	Vaishnavi Mule
53	Prepare materials for a clear and informative presentation	11/29/23	Vikas Reddy
54	Discuss the implications and potential real-world applications of the model	11/29/23	Neha Gawali

55	**Project Review and Future Enhancements**		
56	Identify limitations and assumptions of the model	11/29/23	Vikas Reddy
57	Suggest further improvements and potential data sources for enhancing accuracy.	11/29/23	Vaishnavi Mule
58	**Project Deliverables**		
59	Deliver the predictive model for housing price estimation.	11/30/23	All
60	Present a well-documented research report with detailed findings and insights.	11/30/23	All
61	Share documentation of data preprocessing and modeling steps.	11/30/23	All
62	Maintain the project codebase for reproducibility and future enhancements	11/30/23	All