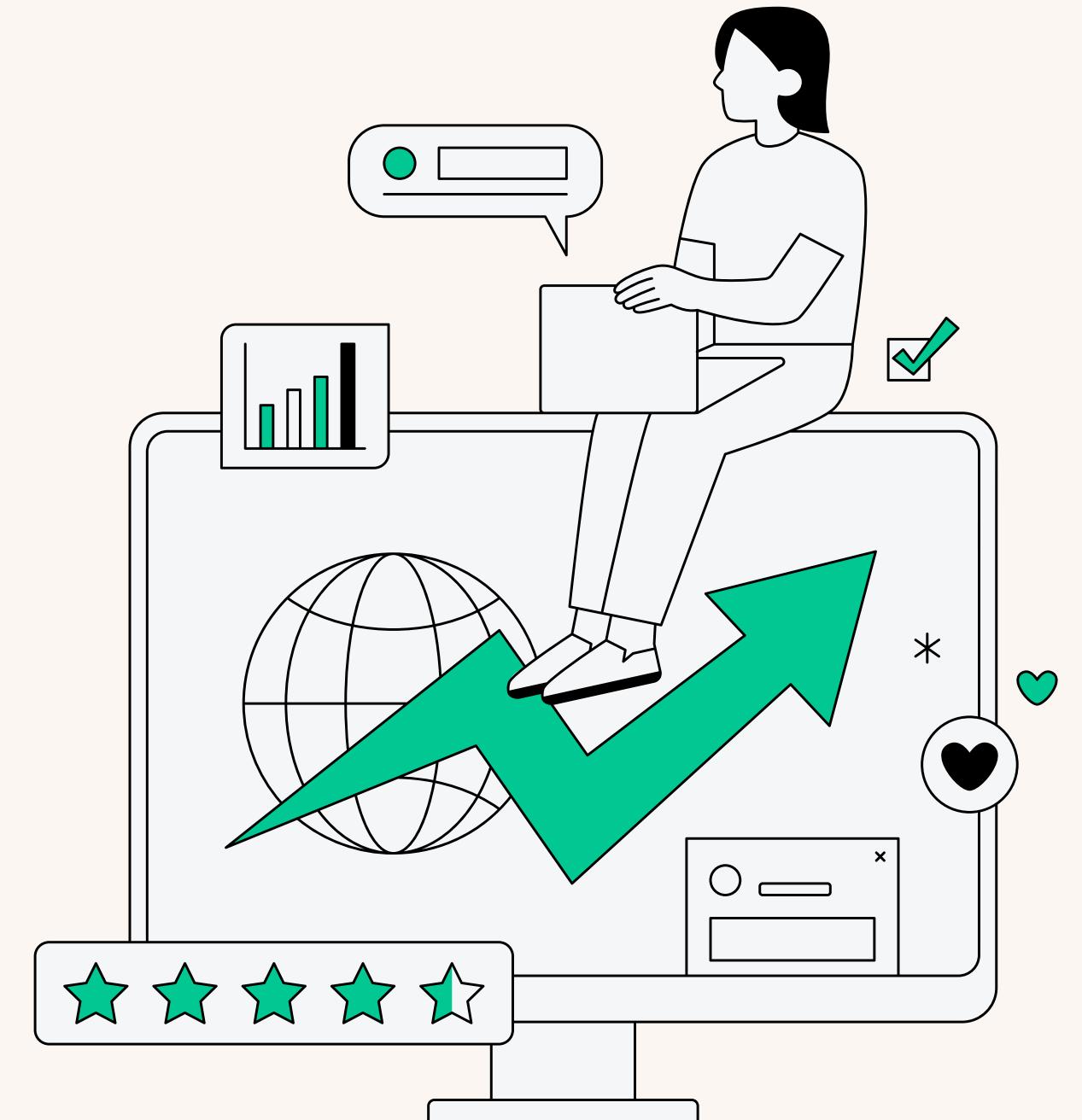


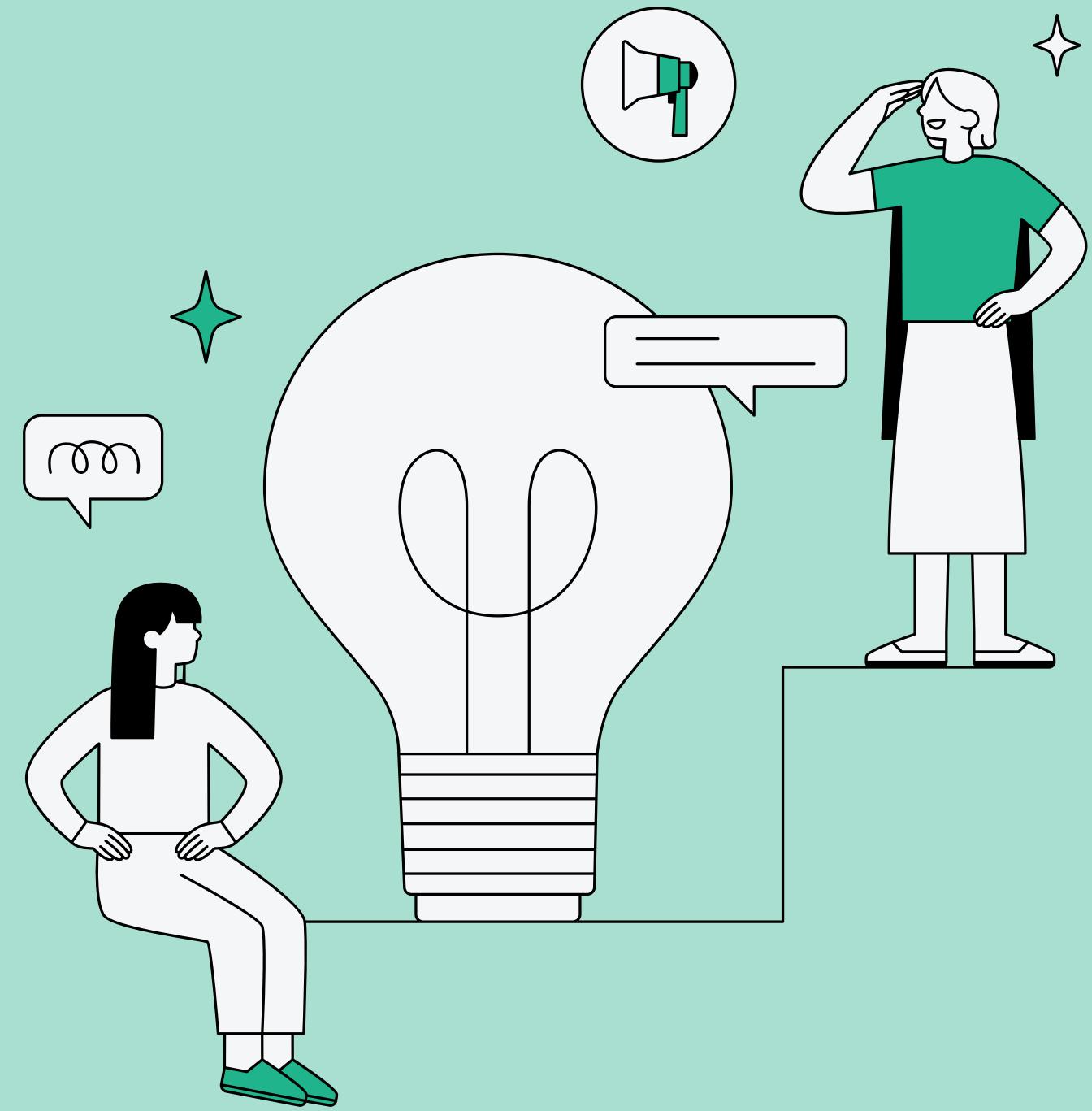
Presented by group 1

A Comprehensive Approach to Housing Price Classification and Prediction”



Problem Statement

- Objective:
 - To analyze and predict California housing prices using machine learning classification models.
- Scope:
 - Understanding key factors influencing housing prices.
 - Implementing classification techniques to categorize data effectively.
- Challenges:
 - Handling missing or inconsistent data.
 - Identifying and engineering critical features for accurate classification



Data Sources:

Kaggle - California Housing Prices (Dataset)

ML Frameworks:

- Numpy, Pandas, Seaborn, Matplotlib
- Scikit-learn (Classification Models)
- JobLib (Pipelining)

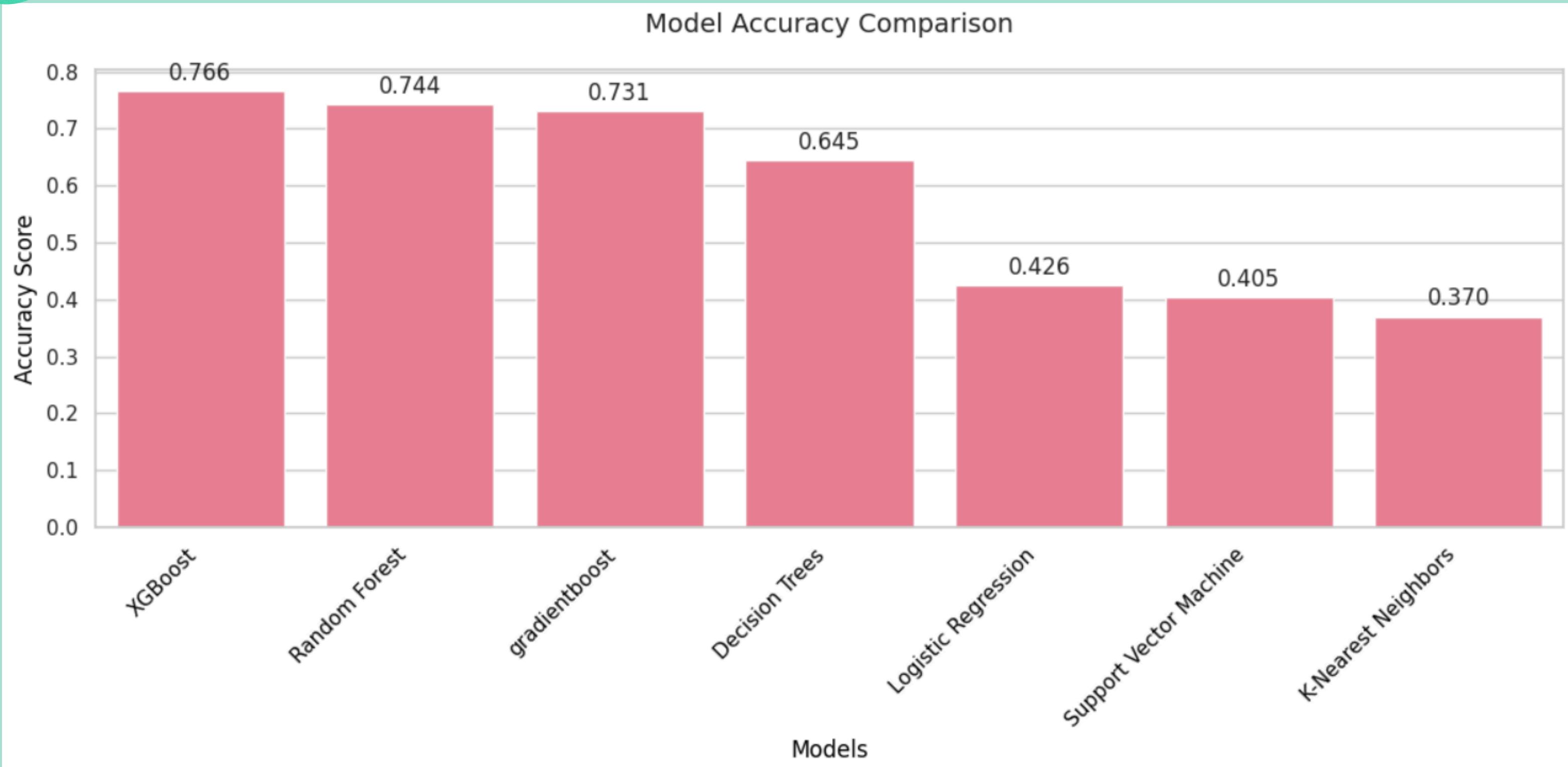
Categories :

Category	Price Range (Percentile)	Description
Least Expensive	\leq 25th Percentile (0.25)	Houses that fall within the lowest 25% of house prices.
Affordable	$25\text{th Percentile} < \text{Price} \leq 50\text{th Percentile}$ (0.25 < value \leq 0.50)	Houses that fall between the 25th and 50th percentiles.
Expensive	$50\text{th Percentile} < \text{Price} \leq 75\text{th Percentile}$ (0.50 < value \leq 0.75)	Houses that fall between the 50th and 75th percentiles.
Luxury	$> 75\text{th Percentile}$ (0.75)	Houses that fall within the top 25% of house prices.

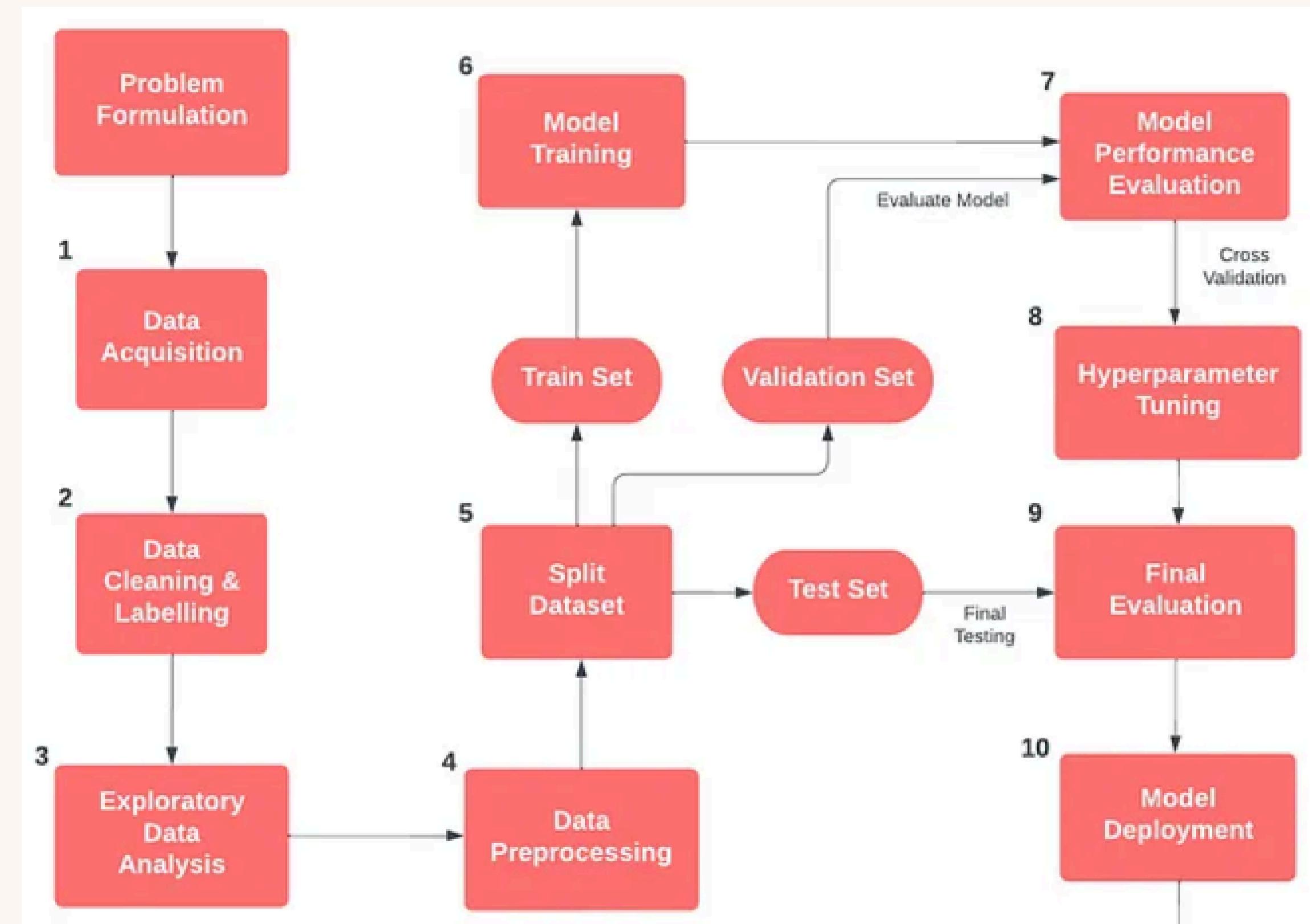
Features :

Feature Names	Data Types	Unique Values
longitude	float64	844
latitude	float64	862
housing_median_age	float64	52
total_rooms	float64	5926
total_bedrooms	float64	1923
population	float64	3888
households	float64	1815
median_income	float64	12928
median_house_value	float64	3842
ocean_proximity	object	5

Performances/Accuracy (Classification Models)



Flow Chart:



Contributions

01.

Sakshi Agrawal
: Data cleaning and preprocessing.

02.

• Ankur Das:
Feature engineering and selection.

03.

Unnati Agarwal :
Model implementation and evaluation.

04.

• Aashwin Sharma:
Visualization and data analysis.