ML'S PROJECT

Section:- F

NAME :-

- **MOHIT KUMAR**
- **MOHAMMED SHEHZAAD KHAN**

5TH SEM

SRN :-

- **PES2UG23CS350**
- **PES2UG23CS349**

Project Title:-

Predicting Airbnb Rental Prices Using Machine Learning.

Problem Statement:-

Setting the right price for an Airbnb rental is challenging for hosts, as it directly impacts demand and revenue, while guests often struggle to assess whether a listing is fairly priced. Without clear pricing benchmarks, both parties face uncertainty. The challenge is to develop a machine learning model that predicts optimal Airbnb listing prices using property features, host characteristics, and customer reviews. By experimenting with models such as linear regression, tree based methods, support vector regression, K-means clustering, and neural networks, the system aims to provide data-driven price predictions to help hosts maximize occupancy and earnings while giving guests transparency into fair pricing.

High-Level Architecture:-

- Data Acquisition & Cleaning:
  Airbnb listing data is processed for analysis, with missing values handled and relevant features engineered.
- Feature Engineering:
  Numerical/categorical attributes are encoded and selected for modeling.
- Model Pipeline:
  1. Linear Regression – baseline regression model.
  2. Random Forest – ensemble for non-linear patterns.
  3. Support Vector Regression (SVR) – supports complex boundaries.
  4. Neural Network – deep learning-based price estimator.
  5. K-Means Clustering – segments listings by price to identify market groups.
- Evaluation:
  Regression models compared via R^2 and RMSE. Clustering is assessed by average prices per segment.
- Data Collection → Data Preprocessing → Modeling (LR, RF, SVR, NN, K-Means) → Model Evaluation & Segmentation → Price Recommendations

Results:-

Model	**R^2**	**RMSE ($)**

Linear Regression	0.299	250.09
Random Forest	0.476	216.38
SVR	0.216	264.58
Neural Network	0.343	242.23

Cluster	**Mean Price ($)**
0	458.59
1	190.50
2	305.63
3	161.91
4	180.20

K-Means Clustering: Price Segmentation

- Best Model: Random Forest (R^2=0.476, RMSE=$216.38) has the highest predictive accuracy.
- K-Means Insight: Discovered distinct pricing tiers in the market; enables strategic host positioning.