**Basic Understanding Questions:**

**1. What is house price prediction, and why is it useful?**  
House price prediction helps estimate the selling price of a house based on different factors like size, location, and market trends. It is useful for:

* **Homebuyers** – Helps them decide if a house is within budget.
* **Sellers** – Helps set a fair selling price.
* **Real Estate Agents** – Helps analyze the market and suggest prices.
* **Financial Institutions** – Helps in mortgage loan approvals.

Example: If a house in **New York** is listed at $500,000 and our model predicts it should be around **$450,000**, a buyer may negotiate a better price.

**2. How does machine learning help in house price prediction?**  
Machine learning analyzes past house sales and detects patterns. Instead of manually estimating prices, ML models can:  
✔ Process large datasets efficiently.  
✔ Identify complex relationships between features (size, location, etc.).  
✔ Give accurate price predictions in real-time.

Example: If a model is trained on thousands of past sales, it can predict a **3-bedroom house in Los Angeles** will cost around **$600,000** based on historical trends.

**Dataset & Features Questions:**

**3. What dataset is used for training the model?**  
The dataset consists of:

* **Features (Inputs)**:
  + Location
  + Number of bedrooms
  + Number of bathrooms
  + Area (square feet)
  + Year built
* **Target Variable (Output)**:
  + House Price ($)

Sample dataset:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **House ID** | **Location** | **Bedrooms** | **Bathrooms** | **Area (sq ft)** | **Price ($)** |
| 1 | New York | 3 | 2 | 1,500 | 500,000 |
| 2 | Texas | 4 | 3 | 2,200 | 300,000 |
| 3 | Florida | 2 | 1 | 1,000 | 200,000 |

**Machine Learning Model Questions:**

**4. Why is Random Forest the best model for this project?**  
Random Forest is chosen because:  
✔ It handles large datasets well.  
✔ It reduces overfitting by using multiple decision trees.  
✔ It performs better on real-world data compared to Linear Regression.

Model comparison:

|  |  |  |
| --- | --- | --- |
| **Model** | **Accuracy (%)** | **Mean Absolute Error (MAE)** |
| Linear Regression | 78% | $30,000 |
| Decision Tree | 85% | $20,000 |
| **Random Forest** | **92%** | **$12,000** |

Random Forest performs **better** because it combines multiple decision trees to get a **more stable and accurate** prediction.

**5. What preprocessing steps are needed before training?**  
To prepare the data, we:  
✔ Remove missing values.  
✔ Normalize numerical features (square feet, price, etc.).  
✔ Encode categorical features (convert location into numbers).

Example:

* **New York → 1**
* **Texas → 2**
* **Florida → 3**

**Implementation Questions:**

**6. How is the project implemented using React and Flask?**  
✔ **Frontend (React.js):** Users enter house details (size, location, etc.).  
✔ **Backend (Flask):** Processes data and calls the ML model.  
✔ **Machine Learning (Python):** Predicts house prices.

Example Flow:  
1️⃣ User enters house details in the React form.  
2️⃣ React sends the data to Flask via an API request.  
3️⃣ Flask processes the data and calls the Random Forest model.  
4️⃣ The predicted price is sent back to the React frontend and displayed.

**7. Why did we use Flask instead of Django?**  
Flask is used because:  
✔ It is lightweight and easy to integrate with ML models.  
✔ It requires fewer resources than Django.  
✔ It is ideal for simple APIs like this project.

**Evaluation & Future Enhancements Questions:**

**8. How do we evaluate the model’s performance?**  
We use these evaluation metrics:

|  |  |
| --- | --- |
| **Metric** | **Value** |
| Accuracy | 92% |
| MAE (Error) | $12,000 |

The lower the MAE (Mean Absolute Error), the **better** the model’s predictions.

**9. How can this project be improved?**  
Future improvements:  
✔ Adding **real-time market trends** for better accuracy.  
✔ Using **deep learning models** for more complex patterns.  
✔ Integrating **Google Maps API** to consider nearby amenities (schools, parks, etc.).

**Tools & Technologies Questions:**

**10. Why did we use React.js for the frontend?**  
✔ React.js provides a **fast and interactive user interface**.  
✔ It is component-based, making the frontend **scalable and reusable**.  
✔ It works well with Flask for API communication.

**11. Why did we choose Python for machine learning?**  
✔ Python has powerful libraries like **Scikit-Learn** for ML.  
✔ It supports **data preprocessing and visualization** easily.  
✔ It integrates well with Flask for API development.