**Name: Ishaan Barde**

**Roll No: 116**

**M.SC.IT FINAL PROJECT**

**RESEARCH METHODOLOGY:**

The research methodology for a house price prediction project involves a systematic approach to collecting, pre-processing, analysing, and modelling data to develop accurate prediction models. A general outline of the research methodology is used

* **Problem Definition and Data Collection:**

Clearly define the scope of your project, including the geographical area, type of properties, and target variables (e.g., sale price). Collect relevant data from reputable sources, including property listings, real estate databases, government sources, and economic indicators. This data should encompass property attributes, location information, economic factors, and any other relevant features.

* **Data Pre-Processing:**

Clean the dataset by handling missing values, duplicates, and outliers. Impute missing values using appropriate methods or consider removing records with substantial missing information. Convert categorical variables into numerical representations through techniques like one-hot encoding or label encoding.

Normalize or scale numerical features to ensure they have similar ranges and distributions.

* **Feature Engineering:**

Create new features that could potentially enhance predictive power, such as feature interactions, ratios, or aggregations. Incorporate geographic features like distance to key locations (schools, hospitals, transportation) using geospatial calculations.

Extract relevant information from textual data using techniques like natural language processing (NLP) for property descriptions.

* **Exploratory Data Analysis (EDA):**

Perform data visualization to understand the relationships between variables, identify trends, and spot outliers. Analyse correlations between features and the target variable to identify potential predictors.

* **Model Selection:**

Choose a variety of predictive models suitable for regression tasks. Common choices include linear regression, decision trees, random forests, gradient boosting, support vector machines, and neural networks.Consider ensemble techniques to combine the strengths of multiple models.

* **Model Training and Evaluation:**

Split the dataset into training, validation, and test sets. Training data is used to train the model, validation data helps tune hyperparameters, and the test data evaluates final model performance. Train the selected models on the training data and fine-tune hyperparameters using cross-validation or grid search.

Evaluate model performance using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.

* **Model Interpretation:**

Analyse feature importance to understand which variables contribute most to the predictions.

Use techniques for values to provide insights into how individual features affect predictions.

* **Model Comparison:**

Compare the different machine learning algorithms used and the deep learning algorithms. Check which is best suited for the problem and has higher accuracy in terms of prediction.

* **Continuous Improvement:**

Monitor model performance over time and retrain the model periodically to account for changing market dynamics and new data.