**PHASE II**

**House Price Prediction Project Proposal**

**Artificial Intelligence**

**Naan Mudhalvan**

**TEAM MEMBERS**

**1.Sree Vaasini S (2021503562)**

**2.Sree Varshini S (2021503563)**

**3.Puja Sree A (2021503540)**

**4.Kiruthika Y B (2021503516)**

**5.Vidhya R (2021503060)**

**Introduction**

In this phase of the project, we aim to further improve the accuracy and robustness of the house price prediction system by exploring advanced techniques such as ensemble methods and deep learning architectures. Specifically, we will consider utilizing Gradient Boosting and XGBoost, powerful ensemble learning algorithms known for their predictive performance.

We will leverage the existing dataset "USA Housing" for experimentation and evaluation, making use of the features like location, square footage, bedrooms, bathrooms, and more.

**Proposed Approach**

1. **Data Preprocessing:**
   * Revisit and potentially refine the data preprocessing steps to ensure the dataset is suitable for the advanced techniques we plan to use.
2. **Feature Engineering:**
   * Review the feature selection process and consider additional features or modifications that might enhance the model's predictive capabilities.
3. **Ensemble Methods:**
   * Implement ensemble learning techniques like Gradient Boosting and XGBoost for regression to improve prediction accuracy.
   * Train the models and evaluate their performance using appropriate evaluation metrics (e.g., MAE, RMSE, R-squared).
4. **Deep Learning Architectures:**
   * Explore the possibility of using deep learning architectures like neural networks for regression.
   * Design, train, and evaluate various architectures to determine if they provide superior predictive performance compared to traditional machine learning models.
5. **Model Comparison and Selection:**
   * Compare the performance of ensemble methods (Gradient Boosting, XGBoost) and deep learning architectures against previously used models (e.g., Linear Regression, Random Forest Regressor).
   * Select the model that demonstrates the highest accuracy and robustness for predicting house prices.

**Implementation Steps**

1. **Refined Data Preprocessing:**
   * Reassess data cleaning and preprocessing steps to ensure compatibility with ensemble methods and deep learning architectures.
2. **Feature Engineering Enhancement:**
   * Consider new features or transformations that may improve model performance.
   * Engineer features that capture nuanced aspects of housing data.
3. **Ensemble Methods Implementation:**
   * Implement Gradient Boosting and XGBoost algorithms for regression using appropriate libraries (e.g., scikit-learn, XGBoost library).
   * Fine-tune hyperparameters to optimize model performance.
4. **Deep Learning Implementation:**
   * Set up the deep learning environment using frameworks like TensorFlow or PyTorch.
   * Design and train neural network architectures suitable for regression tasks.
   * Experiment with various architectures, activation functions, and optimizers.
5. **Model Evaluation and Comparison:**
   * Evaluate the ensemble models and deep learning models using evaluation metrics (e.g., MAE, RMSE, R-squared).
   * Compare the performance of all models and select the most accurate and robust one.
6. **Documentation and Reporting:**
   * Document the entire process, including data preprocessing, feature engineering, model implementation, evaluation, and model selection.
   * Present the findings, insights, and comparative analysis in a comprehensive report.

**Conclusion**

Through this enhancement phase, we aim to significantly improve the predictive accuracy and robustness of our house price prediction system. By exploring ensemble methods and deep learning architectures, we expect to achieve superior results compared to traditional machine learning models. The ultimate goal is to create a highly accurate and reliable model that can assist in predicting house prices effectively.