**Ideation Phase**

**Defining the Problem Statements**

|  |  |
| --- | --- |
| **Date** | **27-09-2023** |
| **Team ID** | **696** |
| **Project Name** | **House pricing forecasting using (ML)** |

**House Price Prediction using Machine Learning**

**Problem Definition and Design Thinking**

**Introduction**

Our objective is to create a machine learning model capable of accurately predicting house prices using pertinent features. Predicting house prices is a prevalent challenge within the real estate sector, offering diverse applications, including aiding buyers in informed decision-making and supporting real estate experts in determining competitive listing prices. In this document, we will delineate the problem statement, the steps required for its resolution, and the design thinking approach that will steer our project.

**Problem Statement**

Objective: The aim is to create a highly accurate machine learning model for predicting house prices.

Data: We possess a dataset comprising diverse house features (such as size(square feet of house) , location, bedroom and bathroom count, etc.) paired with their respective sale prices. This dataset will serve as the foundation for training and assessing our machine learning model.

**Key Challenges:**

1. **Data Quality**: Obtaining a comprehensive and clean dataset with sufficient data points can be difficult. Inaccurate or incomplete data can lead to poor model performance.

2. **Feature Selection:** Choosing the right set of features that have a significant impact on house prices is crucial.

3. **Outliers**: Housing market data often contains outliers, which can skew model predictions. Handling outliers appropriately is essential to ensure accurate predictions.

4. **Data Imbalance**: In some cases, there may be an imbalance in the distribution of house prices, with a significant number of houses at a certain price range. This can affect the model's ability to predict prices accurately.

5. **Deployment:** Crafting a user-friendly interface or API that facilitates predictions for end-users.

**Design Thinking Approach**

**Empathize:**

Before embarking on the journey to solve this problem, it is of utmost importance to cultivate empathy and gain a profound understanding of the needs and perspectives of our primary users: potential homebuyers and real estate professionals. To do this effectively, we must engage in a process of gathering insights that illuminate what factors hold the greatest significance for them when contemplating house prices. Additionally, we need to delve into how precise and reliable price predictions can directly benefit them in their respective roles and decisions.

**Actions:**

1. **User Engagement:** Engage potential users through surveys and interviews to capture their unique perspectives and requirements. This direct interaction enables us to gather valuable insights into what factors matter most to them when considering house prices.
2. **Market Trend Analysis:** Examine historical real estate market data to pinpoint pivotal pricing factors. by analyzing past trends and patterns, we can identify critical variables that have consistently influenced house prices.
3. **Expert Consultation:** Collaborate with domain experts within the real estate industry. These experts possess specialized knowledge and can provide invaluable feedback on the intricacies of the market, ensuring that our predictive model aligns with real-world dynamics.

**Define:**

Our primary objective is to develop a machine learning model that can predict house prices with a high level of accuracy, catering to the requirements of both potential homebuyers and real estate professionals. Success will be measured by achieving a model that consistently provides precise predictions, ultimately aiding users in making informed decisions and enhancing the effectiveness of real estate professionals in their pricing strategies.

**Objectives:**

- The primary goal is to develop a machine learning model that can predict house prices with a high degree of accuracy. To measure this, we aim to achieve a Mean Absolute Error (MAE) on the test data that is less than a specified threshold value, denoted as $X

- In addition to the model's performance, it's equally essential to create a user-friendly web application. This application will serve as an interface for users to input details about a house, such as its size, location, amenities, and other relevant features. Upon submission, the application will use the trained machine learning model to provide price predictions promptly.

**Ideate:**

Brainstorm potential solutions and approaches to address the problem. This phase involves thinking creatively and considering various algorithms and techniques for house price prediction.

**Actions:**

- To build an effective house price prediction model, it's crucial to explore various machine learning algorithms. This involves implementing and experimenting with algorithms such as linear regression, decision trees, random forests, and neural networks

- Feature engineering plays a vital role in model development. We need to experiment with various techniques to enhance the model's performance. This can involve creating new features from existing data, transforming variables to make them more suitable for modeling, and selecting the most informative features

- To further improve the accuracy of our predictions, we should consider integrating external data sources. Information such as neighbourhood crime rates, school quality, proximity to public transportation, and economic indicators can provide valuable context and additional features that may have a substantial impact on house prices.

**Prototype**

Create a prototype of the machine learning model and the user interface for price prediction.

**Actions:**

- Develop a Jupyter Notebook or Python script for data pre-processing, model training, and evaluation.

- Create a simple web interface using tools like Flask or Django to allow users to input house details.

- Test the prototype with a subset of the dataset to ensure it meets performance objectives.

**Test**

Evaluate the model's performance using appropriate metrics and gather feedback from users.

**Actions:**

- Split the dataset into training and testing sets.

- Train the model on the training set and evaluate it on the testing set.

- Use metrics such as MAE, Root Mean Square Error (RMSE), and R-squared to assess model performance.

- Collect user feedback on the web interface for usability and accuracy.

**Implement**

Once the prototype meets the defined objectives and receives positive feedback, proceed with full implementation.

**Actions:**

- Train the final machine learning model on the entire dataset.

- Deploy the model as part of a production-ready web application.

- Conduct thorough testing to ensure the application is robust and user-friendly.

**Iterate**

Continuous improvement is essential. Gather user feedback and iterate on the model and interface to enhance accuracy and usability.

**Actions:**

- Monitor the model's performance and retrain it periodically with updated data.

- Address user feedback and make necessary improvements to the web interface.

- Stay informed about advancements in machine learning and real estate pricing models for potential enhancements.

**Conclusion:**

In this document, we have meticulously outlined our strategy for addressing the challenge of house price prediction through machine learning. We commenced by precisely defining the problem, followed by a discerning identification of key challenges inherent in this undertaking. Our approach is anchored in design thinking principles, encompassing a sequence of crucial steps. We begin with empathizing with the users, which allows us to intimately understand their needs and perspectives. Subsequently, we set forth clear and measurable objectives, delineating the path to success. Ideation, an integral phase, entails brainstorming potential solutions and strategies. Prototyping enables us to materialize our concepts into tangible forms for testing and validation. Upon validating and refining our prototypes, we proceed to the implementation phase. Continuous iteration and improvement remain central to our process, ensuring the final product is not only accurate but also user-friendly. Our ultimate aspiration is to forge a solution that not only meets but exceeds expectations, delivering valuable insights that empower both homebuyers and real estate professionals within the dynamic housing market. By adhering to this meticulous and structured approach, we aim to contribute significantly to the real estate industry with a dependable tool that fosters positive impact.