**Assignment 1: Building an AI Chatbot using prompts**

**Tools:** We'll utilize ChatGPT's code interpreter, Python, TensorFlow, and Matplotlib for visualization.

We'll follow the CRISP-DM framework (Cross Industry Standard Process for Data Mining) to guide our analysis, which includes six phases:

**1. Business Understanding**

The primary aim is to create a machine learning model using a AI for automatically predicting house prices.

**2. Data Understanding**

In this phase, we'll explore the dataset used for model training. The focus is on familiarizing ourselves with the data, assessing quality, and identifying any issues that need addressing.

**Key Objectives:**

* **Data Collection:** Import and examine the dataset to understand its structure and contents.
* **Descriptive Statistics:** Analyze the dataset's size, distribution, and check for missing values.
* **Visualization:** Create samples of the data for a better grasp of patterns and challenges in recognition.
* **Quality Assessment:** Detect any missing, corrupt, or irrelevant data that may affect model performance.

**3. Data Preparation**

Here, we'll prepare the dataset for training, which includes necessary pre-processing steps to predict the predict correctly using AI model.

**Key Steps:**

* **Normalization:** Scale pixel values to a range between 0 and 1 to enhance neural network performance.
* **Reshaping:** Ensure prices are scaled correctly for the model, typically with standard or normal scaler
* **Train-Test Split:** Although we may treat this dataset as a test set, we could also split data into training and validation sets as needed.
* **Data Augmentation (Optional):** Employ techniques like rotation, flipping, or zooming to help the model generalize better by expanding the dataset artificially.**5. Model Evaluation**

After training the CNN model, the next critical step is to evaluate its performance, assessing how well it generalizes to unseen data and if it meets project objectives.

**Key Metrics for Evaluation:**

* **Accuracy:** Percentage of correctly classified digits among all samples.
* **Confusion Matrix:** Detailed breakdown of classification results, indicating where the model confuses certain digits.
* **Precision, Recall, and F1-Score:** Metrics that evaluate performance for each digit class, providing insights into false positives and negatives.
* **Loss and Accuracy Curves:** Visuals illustrating model performance over training epochs.

**6. Model Deployment**

After training and evaluation, deployment is the final phase, integrating the model into a production environment. This step allows the model to make predictions in real-time or through batch processing.

**Common Deployment Options:**

* **Web Service:** Serve the model via a REST API to enable applications to send image data and receive predictions.
* **Mobile/Embedded Application:** Export the model to a format compatible with mobile or edge devices, such as TensorFlow Lite.
* **Cloud Deployment:** Utilize cloud services like AWS, Google Cloud, or Azure for scalable model deployment.