## Artificial Intelligence

## and

## Machine Learning

Project Abstract

Semester-IV (Batch-2022)

House price predictor

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajiv Bhardwaj Name: Suman Ranjan

Roll Number:2210990872

Group:G-31

Saniya Singla (2210990789),

Sanchit Prashar (2210990783),

Sonakshi(2210990862)

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**Title:** Predicting Home Prices using Artificial Intelligence and Machine Learning

**Abstract:**

The real estate market is complex, influenced by numerous factors ranging from location and economic conditions to architectural trends and demographic shifts. Predicting home prices accurately is crucial for buyers, sellers, and investors to make informed decisions. This project proposes the development of an Artificial Intelligence (AI) and Machine Learning (ML) model for predicting home prices, leveraging advanced algorithms and data analytics techniques.

The objective of this project is to create a robust predictive model that can analyze various features of a property and its surrounding environment to estimate its market value accurately. The model will utilize a diverse dataset encompassing information such as property characteristics (size, age, amenities), neighborhood demographics, economic indicators, and historical sales data.The dataset is using supervised learning, Supervised learning is a type of machine learning approach where the algorithm is trained on a labeled dataset, meaning that each input data point is paired with an output label. The goal of supervised learning is to learn a mapping from inputs to outputs, such that the algorithm can generalize well to unseen data.

Key components of the project include data collection and preprocessing, feature engineering, model selection, training, and evaluation. Various ML algorithms, including linear regression, random forests, and gradient boosting, will be explored to identify the most suitable approach for the predictive task. Feature selection techniques will be employed to determine the most relevant attributes impacting home prices.

Moreover, the project will focus on the development of a user-friendly interface to enable stakeholders to interact with the predictive model easily. This interface will allow users to input property details and receive instant price estimates based on the trained AI model's predictions.

The potential impact of this project is significant, as it can empower buyers, sellers, real estate agents, and investors with valuable insights into property valuation. By leveraging AI and ML techniques, this project aims to revolutionize the way home prices are predicted, providing more transparency and accuracy in real estate transactions.

**Supervised Learning**: Supervised learning is a type of machine learning where the algorithm learns to map input data to a target output based on example input-output pairs. In supervised learning, the training data is labeled, meaning each example consists of an input and the corresponding correct output.

**Unsupervised Learning**: Unsupervised learning is another type of machine learning where the algorithm learns to find patterns or structure in input data without explicit supervision or labeled responses.

Supervised Learning is used in House Price Prediction

I am using supervised learning. Supervised learning is a type of machine learning where a model is trained on labeled data, meaning the input data is paired with the corresponding correct output. The model learns from this labeled data to make predictions or classifications on new, unseen data

The respective columns in my project, such as beds, baths, size, size\_units, lot\_size, lot\_size\_units, zip\_code, and price, can be used as features and the target variable for training the supervised learning model

**LIBRARIES**

I have imported the following libraries for the project:

**pandas (pd):** Pandas is a powerful library for data manipulation and analysis. It provides data structures and functions to efficiently work with structured data, such as tables or spreadsheets. It is commonly used for tasks like data cleaning, transformation, and exploration.

**numpy (np):** Numpy is a fundamental library for numerical computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays efficiently. It is widely used for scientific computing and data analysis tasks.

**matplotlib.pyplot (plt)**: Matplotlib is a popular plotting library in Python. The pyplot module provides a simple and convenient interface for creating various types of plots, such as line plots, scatter plots, histograms, and more. It is often used in conjunction with Numpy and pandas to visualize data.

**seaborn (sns):** Seaborn is a data visualization library built on top of matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics. Seaborn simplifies the process of creating complex visualizations and offers additional functionality compared to matplotlib.

**Scikit-learn:** sklearn is a popular open-source machine learning library in Python. It provides a wide range of tools and algorithms for various machine learning tasks, including classification, regression, clustering, dimensionality reduction, and more.

**To install the following libraries:**

pip install pandas

pip install numpy

pip install matplotlib

pip install seaborn

pip install scikit-learn

**Regression Algorithms:**

Linear Regression: A simple and commonly used regression algorithm that models the relationship between the independent variables and the dependent variable by fitting a linear equation to the observed data.

Lasso regression: short for Least Absolute Shrinkage and Selection Operator regression, is a type of linear regression technique used for variable selection and regularization.

Ridge regression: also known as Tikhonov regularization, is a type of linear regression technique used for modeling and prediction. It is similar to ordinary least squares (OLS) regression but includes a regularization term to prevent over fitting.

**GRAPHS:**









