

CS411: AI

Group Project 2: Real Estate Pricing

Team – *CHAT+*: Nick, Adam, Jace, Thomas

Business Context:

The real estate industry is highly competitive, and accurate house price predictions are critical to making profitable decisions. Real estate agents, investors, and homebuyers rely on market trends and various factors like house location, size, and features to estimate property values. However, market fluctuations, buyer behavior, and local variables can create significant uncertainty in price estimates.

This project asks you and your team to create a machine learning model to predict house prices based on extensive datasets that describe various attributes of homes sold in the past. The predictive model will help real estate companies and homebuyers make better-informed decisions by accurately pricing homes according to market conditions and individual house features. This can optimize the sale process, attract more buyers, and prevent undervaluation or overvaluation of properties.

You will engage with the broader economic context of real estate, learning how to combine statistical analysis and machine learning to provide insights that support financial decisions.

Dataset Link: <https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>

Project Objectives:

1. **Data Preprocessing:** Handle missing values, encode categorical variables, and perform exploratory data analysis.
2. **Feature Engineering:** Create new features (e.g., price per square foot) that can improve prediction accuracy.
3. **Modeling:** Train **multiple** models and tune them for optimal performance.
4. **Model Evaluation:** Use appropriate regression metrics to evaluate models.
5. **Business Insight:** Provide recommendations to reduce customer churn based on model findings.

Project Flow:

1. **Data Preprocessing:** A clean dataset ready for modeling.
2. **Feature Engineering:** A dataset with added features to enhance prediction power.
3. **Model Selection and Training:** A notebook with trained models.
4. **Model Evaluation:** A report comparing the models and suggesting the best one for price prediction.
5. **Final Deliverables:** A final notebook, presentation, and report explaining the models and their impact on business.

Hint:

You can improve model performance by transforming the target variable (*SalePrice*). Applying a **log transformation** can help reduce skewness in the target and improve the linear relationship between features and the target. Don't forget to handle **missing data** and use **one-hot encoding** for categorical variables like neighborhood and house style.

Instructor: Swapnil Chhabra (s.chhabra@snhu.edu)