

## Capstone Project - 1

### Predicting Property Prices in a Specific Location Using Machine Learning



## Dataset Description

### File descriptions

- **train.csv** - the training set
- **test.csv** - the test set
- **data\_description.txt** - full description of each column, originally prepared by Dean De Cock but lightly edited to match the column names used here
- **sample\_submission.csv** - a benchmark submission from a linear regression on year and month of sale, lot square footage, and number of Bedroom

### Data fields

Here's a brief version of what you'll find in the data description file.

- **PropPrice** - the property's sale price in dollars. This is the target variable that you're trying to predict.
- **PropertyClass**: The building class
- **PropertyZone**: The general zoning classification
- **PropertyFrontage**: Linear feet of street connected to property

- **PropertySize:** Lot size in square feet
- **Street:** Type of road access
- **Alley:** Type of alley access
- **PropertyShape:** General shape of property
- **Elevation:** Flatness of the property
- **Amenities:** Type of Amenities available
- **LotOrientation:** Lot configuration
- **Grade:** Slope of property
- **Neighborhood:** Physical locations within Ames city limits
- **Condition1:** Proximity to main road or railroad
- **Condition2:** Proximity to main road or railroad (if a second is present)
- **BldgType:** Type of dwelling
- **PropertyStyle:** Style of dwelling
- **OverallQual:** Overall material and finish quality
- **OverallCond:** Overall condition rating
- **YearBuilt:** Original construction date
- **YearRemodAdd:** Remodel date
- **RoofStyle:** Type of roof
- **RoofMatl:** Roof material
- **Roof1Material:** Exterior covering on property
- **Roof2Material:** Exterior covering on property (if more than one material)
- **ExteriorCladdingType:** Masonry veneer type
- **ExteriorCladdingArea:** Masonry veneer area in square feet
- **ExterQual:** Exterior material quality
- **ExterCond:** Present condition of the material on the exterior
- **PropertyFooting:** Type of PropertyFooting
- **BsmntFinish:** Height of the basement
- **BsmntMaintenance:** General condition of the basement
- **BsmntVisibility:** Walkout or garden level basement walls
- **BsmntFinRat1:** Quality of basement finished area
- **BsmntFinSty1:** Type 1 finished square feet
- **BsmntFinQual1:** Quality of second finished area (if present)
- **BsmtFinSF2:** Type 2 finished square feet
- **BsmtUnfSF:** Unfinished square feet of basement area

- **BsmntSqFtage**: Total square feet of basement area
- **Heating**: Type of heating
- **HeatingEfficiency**: Heating quality and condition
- **CentralAir**: Central air conditioning
- **Electrical**: Electrical system
- **1stFlrSF**: First Floor square feet
- **2ndFlrSF**: Second floor square feet
- **LowQualFinSF**: Low quality finished square feet (all floors)
- **GrLivArea**: Above grade (ground) living area square feet
- **BsmtBath1**: Basement full bathrooms
- **BsmtBath2**: Basement half bathrooms
- **Bath1**: Full bathrooms above grade
- **Bath2**: Half baths above grade
- **BedroomUpLev**: Number of BedroomUpLevs above basement level
- **Kitchen**: Number of kitchens
- **KitchenQual**: Kitchen quality
- **CntRmsUpLev**: Total rooms above grade (does not include bathrooms)
- **Functional**: Home functionality rating
- **CntFireplaces**: Number of Fireplaces
- **QualFireplace**: Fireplace quality
- **BasementType**: Garage location
- **BasementYrBlt**: Year garage was built
- **BasementFinish**: Interior finish of the garage
- **BasementCars**: Size of garage in car capacity
- **SquareFootage**: Size of garage in square feet
- **BasementQual**: Garage quality
- **BasementSqFootage**: Garage condition
- **PavedDrive**: Paved driveway
- **WoodDeckSF**: Wood deck area in square feet
- **OpenPorchSF**: Open porch area in square feet
- **EnclosedPorch**: Enclosed porch area in square feet
- **3SsnPorch**: Three season porch area in square feet
- **ScreenPorch**: Screen porch area in square feet
- **PoolArea**: Pool area in square feet

- **PoolQC**: Pool quality
- **BoundaryFeatures**: BoundaryFeatures quality
- **AddFeatures**: Miscellaneous feature not covered in other categories
- **AddVal**: \$Value of miscellaneous feature
- **SaleMon**: Month Sold
- **YrSold**: Year Sold
- **SaleType**: Type of sale
- **SaleCond**: Condition of sale

**Background:** The real estate market is highly dynamic and can be influenced by various factors such as location, property size, amenities, neighborhood, and other related factors. Predicting the accurate price of a property is a crucial task for real estate agents, buyers, and sellers. Machine learning has proven to be a useful tool in predicting property prices. Therefore, this capstone project aims to develop a machine learning model that can accurately predict property prices in a specific location.

## Objectives:

1. To collect and clean real estate data from a specific location.
2. How can ordinal and nominal columns be handled separately in property price prediction using the metadata sheet provided to identify which columns are ordinal or nominal?
3. How can scaling, PCA, and fillna() techniques be used in property price prediction to handle missing data and improve the accuracy of the model?
4. To perform exploratory data analysis (EDA) on the collected data to identify key variables that influence property prices.
5. What is the proper encoding technique to be used for ordinal and nominal variables in property price prediction, based on the requirements of the model?

- 6.To develop a machine learning model that can predict property prices based on the selected variables.
- 7.To evaluate the performance of the model and compare it with other machine learning algorithms.
- 8.To present the findings and insights from the project in a clear and concise manner.

**Note:1.Try not to drop the columns rather than Property ID.**

**2.Expected R2 score is between 75%-85% whereas more than 85% is much better.**

## **Methodology:**

- 1.Data Collection: The first step is to collect data on various features of properties in a specific location. This can be achieved by scraping data from real estate websites or collecting data from local real estate agents.
- 2.Data Cleaning: The collected data will be preprocessed and cleaned to remove missing values, outliers, and other errors.
- 3.Exploratory Data Analysis (EDA): The cleaned data will be analyzed using EDA techniques to identify important features that influence property prices.
- 4.Feature Engineering: After identifying the significant features, new features will be created based on domain knowledge or statistical techniques to enhance the predictive power of the model.
- 5.Model Selection: Various machine learning algorithms, including linear regression, decision trees, and random forests, will be evaluated to determine the best model for predicting property prices.
- 6.Model Training and Testing: The selected machine learning algorithm will be trained on a subset of the data and tested on the remaining data to evaluate its performance.
- 7.Model Evaluation: The performance of the model will be evaluated using various metrics such as mean absolute error (MAE) and root mean squared error (RMSE).
- 8.Model Deployment: The final model will be deployed to predict property prices for new data.

## **Expected Outcomes:**

- 1.A machine learning model that accurately predicts property prices in a specific location based on selected variables.
- 2.Insights into the significant features that influence property prices in the specific location.
- 3.A comprehensive report detailing the methodology, results, and insights from the project.

**Conclusion:** This capstone project aims to develop a machine learning model that can predict property prices accurately in a specific location. The project will involve collecting and cleaning data, performing EDA, feature engineering, model selection, training, testing, and evaluation. The final outcome will be a comprehensive report detailing the methodology, results, and insights from the project.

## **Make a Project Summary Report**

The Project synopsis should be the in the following format:

1. Introduction to the Project
2. Objectives of the Project
3. Flow Chart of operations
4. Python Codes
5. Screenshot of the Outputs
6. Report on EDA (include pictures of the graphs)
7. Learning Outcomes
8. Conclusion
9. Citations – Books and Websites used for research.

## **The following files must be submitted:**

- 1.Jupyter Notebook
- 2.Project Summary report (Word File)