Predicting Nashville Housing Land Use case study.

Scenario

A real estate owner requested a predictive model which can predict how prospective residents want to use the land for sale based on their enquiries. Almost all residents who made enquiries were specific on the number of bedrooms and whether the land itself was sold as vacant. These residents say that land sold as vacant is trouble-free, that is why they are precise about these two features. Therefore, knowing the purpose of the land use based on these enquiries and requests can help him raise the sales prices or reduce the sales prices for the land. Although the land has values according to the acreage, building value and many others, only these two factors are crucial to his request. The owner has consulted you since you were recommended by a well-known client as one of the best Data scientists. He handed over data from the past year to help you create a machine learning model for him.

About the company

Nashville Housing is a real estate company in Nashville committed to mortgages, leasing and sale of land, buildings and parks in Nashville. Over the years it has been recognized as a reliable and dependent partner in real estate. It is well known for its legal contractual innovations and transparent transactions with clients over 30 years. The company is fully aware of technological advancements and has started to make investments into AI and machine learning. This year, to bring aboard Data Scientists and Machine learning professionals, they hosted a world-wide challenge in Data science hoping to absorb talents from all over the world. You participated in the challenge and were tasked to build a predictive model for Nashville Housing.

The challenge is meant to make every participant feel at ease, so the accuracy score allowed was lowered to be above average. The team expects that the model created should at least be 60% efficient as the next steps will determine the real accuracy score. Anyone who falls below this expected score will be disqualified for the competition.

MODEL CREATED BY: Angelica Adjei-Kwarteng.

Date: 12th November, 2024

This is my capstone project on machine learning as part of my formal training at Blossom Academy.

Data source: *https://www.kaggle.com/datasets/tmthyjames/nashville-housing-data*

Business Objective: To build a predictive model for Nashville Housing to predict the type of land use based on vacancy and number of bedrooms.

Notebook analysis and model explained.

Imported all relevant libraries.

Loaded the dataset into the notebook.

Conducted some exploratory analysis:

1. Inspected data.

2. Preprocessed: a. changed some datatypes

    b. imputed missing values with median and mode depending on the feature.

    c. dropped columns

    d. replaced No values with 'N', Yes values with 'Y' under SoldAsVacant column

    e. converted SoldAsVacant column to binary columns

    f. joined binary columns to main df and assigned a new variable data

    g. After encoding SoldAsVacant column, Yes was 0 instead of 1 so manually replaced values as shown in step40.

    h. previewed head and dropped column Y.

    i. Selected features for our model and saved as variable X and saved target variable as y.

3. Split datasets into training and testing datasets.

4. Machine learning module Decision Tree used.

5. Trained model and made predictions.

6. Evaluated the model.

7. Created a sample visualization of how our model looks like(optional)

Questions?

Do give your feedback. It is very important to me.

Do have a great time!

Thank you!