Machine Learning Canvas

PREDICTIONS

OBJECTIVES

DATA

End-user

Who will use the predictive system / who will be affected by it?

- Primary users: Equity Residential property management team, real estate market analysts, marketing and sales department.
- Secondary users: Potential tenants, investors.

Value proposition

What are we trying to do for the system's users? (e.g. spend less time on X, increase Y...)

- Reduce the time spent determining optimal rental rates.
- Increase the accuracy of rent price forecasting.
- Maximize rental income.
- Decrease vacancy rates.
- Enhance tenant satisfaction through competitive and transparent pricing.

Data sources

Where do/can we get data from? (internal database, 3rd party API, etc.)

Web scraping data from the Equity Residential website and other rental property websites.

Problem

Question to predict answers to (on behalf of user)

Question: What will be the optimal rent for a specific apartm

Input (i.e. question "parameter")

- Price - Beds - Baths - sq.ft - Floor - Move_in_date - buildin

Possible outputs (i.e. "answers")

Predicted rent price.

Type of problem (e.g. classification, regression, recommendation...)

Regression.

Baseline: simple, alternative way of making predictions (e.g. manual rules)

Performance evaluation

Domain-specific / bottom-line metrics for monitoring performance in production

- Vacancy rate.
- Rental income.

Prediction accuracy metrics (e.g. MSE if regression; % accuracy, #FP for classification)

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE).
- Coefficient of Determination (R^2).

Offline performance evaluation method (e.g. cross-validation or simple training/test split)

Data preparation

How do we get training data (inputs, and outputs if supervised learning)? How many data points?

- Historical data obtained via web scraping from Equity Residential.
- Amount of data: 62,810 records.

Input features (extracted from data sources). If too many, list types of features and mention key ones.

PREDICTIONS

Manual rules based on historical data and current market conditions.

OBJECTIVES

- Cross-validation.
- Simple train-test split.

DATA

- Price
- Beds
- Baths
- sq.ft
- Floor
- Move_in_date
- building_id
- unit_id
- URL
- Day_Recorded
- Amenity
- Apartment Name
- Address
- City
- Units
- Northern_Exposure
- Southern_Exposure
- Eastern_Exposure
- Western_Exposure
- Balcony
- Walk_In_Closet
- Fireplace
- City_Skyline
- Kitchen_Island
- Stainless_Appliances
- Renovated
- Office_Space
- Days_Till_Available
- Day_of_the_week_recorded
- Unique_ID
- Estiamted_Vacancy

Using predictions

GRAT When do we make predictions and how many?

- Predictions are made upon receiving a new listing and during weekly data updates.
- The number of predictions varies depending on the number of listings.

Learning models

When do we create/update models? With which data / how much?

- Models are updated annually or when there are significant market changes.
- Data from the past year is used for model updates.

PREDICTIONS

What is the time constraint for making those predictions?

Predictions should be made in real-time or within a few seconds after the request.

How do we use predictions and confidence values?

- Predicted rent prices are used for setting rates on the website.
- Confidence values can be used for automatic rate checking and adjustment, as well as for management decision-making.

OBJECTIVES

What is the time constraint for creating a model?

The model should be created within a few hours to ensure timely updates.

Criteria for deploying model (e.g. minimum performance value — absolute, relative to baseline or to previous model)

- Minimum model performance: MAE < $50.
- $R^2 > 0.85$.
- Performance must be better than or equal to the baseline model and previous versions.

DATA

Reset Form

Machine Learning Canvas v0.1