Machine Learning Canvas

PREDICTIONS	OBJECTIVES	DATA

IDEA

Context

Who will use the predictive system / who will be affected by it? Provide some background.

Airbnb hosts and platform admins.

Hosts need to set optimal rental prices to maximize occupancy and revenue. Platform admins can benefit from insights into pricing trends and patterns to improve service offerings.

Value Proposition

What are we trying to do? E.g. spend less time on X, increase Y...

Optimize rental pricing to increase occupancy rates and revenue for Airbnb hosts.

Data Sources

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

Kaggle dataset containing Airbnb listings data.

SPECS

Problem

Question to predict answers to (in plain English)

How much should an Airbnb host charge for a rental property to maximize revenue and occupancy?

Input (i.e. question "parameter")

Property features; Time-related data; Market conditions.

Possible outputs (i.e. "answers")

Optimal rental price for a given property at a specific time.

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

Regression problem (predicting a continuous variable: price).

Baseline

What is an alternative way of making predictions (e.g. manual rules based on feature values)?

Hosts use personal experience and historical data to set prices.

Performance evaluation

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

Increase in average daily revenue, occupancy rate improvement.

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

Mean Squared Error (MSE): Measures average squared difference between predicted and actual prices.
Mean Absolute Error (MAE): Measures average absolute difference between predicted and actual prices.

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

Cross-Validation: Ensures model robustness and generalizability.
Train/Test Split: Provides initial performance benchmarks.
Dataset

Dataset

How do we collect data (inputs and outputs)? How many data points?

Extract data from Kaggle dataset, preprocess and clean.

Initial dataset contains 74,111 rows with 29 columns of various features.

Features

Used to represent inputs and extracted from data sources above. Group by types and mention key features if too many to list all.

Numerical: Price, number of rooms, number of beds, etc.
Categorical: Property type, room type, neighborhood.
Text: Descriptions and reviews
Datetime: Booking date, listing date, etc.
Key Features: Location, number of rooms, property type, amenities, season, reviews.

DEPLOYMENT

Using predictions

When do we make predictions and how many?

Predictions are made whenever a host lists or updates a property.

What is the time constraint for making those predictions?

Predictions should be quick to ensure a seamless user experience.

How do we use predictions and confidence values?

Provide suggested prices with confidence intervals to hosts for final decision-making.

Learning predictive models

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

Models should be updated periodically (e.g., monthly) or when significant new data is available. The most recent data is used.

What is the time constraint for creating a model?

Few hours ideally.

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

Model must meet or exceed our success criteria metrics. Model must perform consistently across different time periods and property types.

Reset Form

Machine Learning Canvas v0.1