Back To Module Home

Machine Learning System Design

0% completed

Search Module

Machine Learning Primer

Video Recommendation

Feed Ranking

Ad Click Prediction

Rental Search Ranking

Problem Statement and Metrics

Booking Model

Rental Search Ranking System Design

Mark Module as Completed

Booking Model

Learn about the booking model for the Airbnb Rental Search.

We'll cover the following

3. Model#

Feature engineering#

- Geolocation of listing (latitude/longitude): Taking raw latitude and raw longitude features is very tough to model as feature distribution is not smooth. One way around this is to take a log of distance from the center of the map for latitude and longitude separately.

- Favourite place: store user's favorite neighborhood place in 2 dimensions grid. For example, users add Pier 39 as their favorite place, we encode this place into a specific cell, then use embedding before training/serving.

Features	Feature engineering	Description
Listing ID	Listing ID embedding	See Embedding in Machine Learning Primer: Feature Selection and Feature engineering.
Listing feature	Number of bedrooms, list of amenities, listing city	
Location	Measure lat/long from the center of the user map, then normalize	
Historical search query	Text embedding	
User associated features: age, gender	Normalization or Standardization	
Number of previous bookings	Normalization or Standardization	
Previous length of stays	Normalization or Standardization	
Time related features	Month, weekofyear, holiday, dayofweek,	

	hourofday	

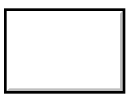
Training data#

- User search history, view history, and bookings. We can start by selecting a period of data: last month, last 6 months, etc., to find the balance between training time and model accuracy.

> In practice, we decide the length of training data by running multiple experiments. Each experiment will pick a certain time period to train data. We then compare model accuracy and training time across different experimentations.

Model architecture#

- Input: User data, search query, and Listing data.
- Output: This is a binary classification model, i.e., user books a rental or not.
- We can start with the deep learning with fully connected layers as a baseline. Model outputs a number within [0, 1] and presents the likelihood of booking.
- To further improve the model, we can also use other more modern network architecture, i.e., Variational AutoEncoder or Denoising AutoEncoder. Read more about Variational Autoencoder.



Problem Statement and Metrics

Next

Rental Search Ranking System Design

Mark as Completed

Report an Issue