

## Task

Predict destination of new orders given its corresponding users, start\_time, start\_loc, and some history data.

## Data

Six raw features, 3,214,096 training records, 2,002,996 test records.

orderid	userid	bikeid	biketype	starttime	geohashed_start_loc	geohashed_end_loc
1893973	451147	210617	2	2017/5/14 22:16	wx4snhx	wx4snhj
4657992	1061133	465394	1	2017/5/14 22:16	wx4dr59	wx4dqz

## Challenge

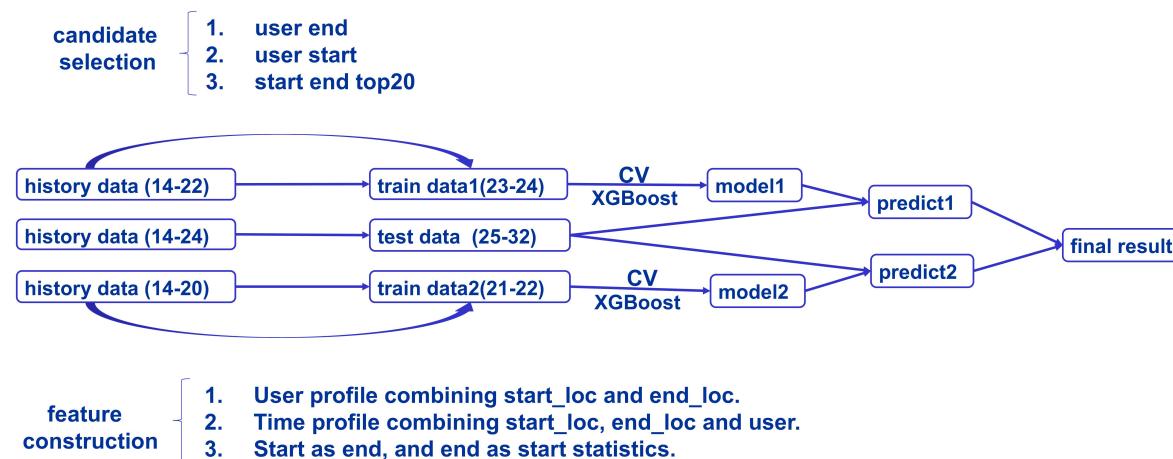
Few raw categorical features; 52% new users in test set

## Procedure

- Step 1: split data into train and test.
  - Step 2: allocate candidate destinations for each order.
  - Step 3: construct new features using history data statistics.
  - Step 4: make feature combinations.
  - Step 5: train the binary model using training dataset.
  - Step 6: predict the test dataset to say yes/no; get top3 for each order.
  - Step 7: model ensemble using probabilistic weight.

1

## Overview of our solution



2

## Feature construction and importance

We split the train/test data into two part, and one part is for providing statistical quantity for another. In this way, every order is represented by history information with more sense. More importantly, due to the limited expressive power of single feature, we exploit feature combination by means of CV for filtering useful combinations, whose importance is shown below.

