



# CS145 Project Introduction

# Help Yelp!

Instructor: Yizhou Sun

TAs: Junheng, Shengming, Yunsheng

10/08/2018



- Background & Motivation
- Project Task
- Dataset
- Evaluation
- Project Deadlines and Grading



Based on the user and business information, can you predict the rating of a user that visits a new restaurant?

Recommender System

Multi-information

Various types of data

A good fit for our class

15539 reviews

Write a Review Add Photo Share Save

\$\$\$ · Italian, Bakeries, Bi

7th St / Metro Center  
W 8th St

700 S Grand Ave  
Los Angeles, CA 900  
Downtown

Get Directions  
(213) 802-1470  
bottegalouie.com  
Send to your Phone

**Sam V.**  
Brea, CA  
170 friends  
37 reviews  
1 photo

Share review  
Embed review  
Compliment  
Send message  
Follow Sam V.

5 stars 10/5/2018  
Sometimes the most amazing service just makes your day!  
Had the pleasure of being served by an amazing young man named Tomas.  
Food was perfect and his attitude and attention to our table was just outstanding.  
Will be coming back loveddd this place

Was this review ...?  
Useful Funny Cool

**Kennedy B.**  
Los Angeles, CA  
496 friends  
80 reviews  
69 photos  
Elite '18

3 stars 10/7/2018  
I love their bakery, but have no interest in ever dining there again.  
My boyfriend and I went to Bottega for date night. I ordered one of their pasta dishes and my boyfriend ordered a pizza. The food was good, but it took forever to come out.

"Overall, i would in 1417 reviews"

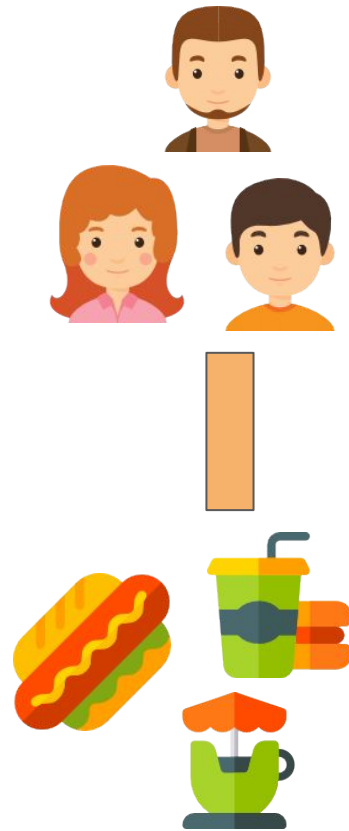
"To end out nigh sauce." in 214 re



Based on data of

- Business: id, name, neighborhood, address, stars, review\_count etc.
- User: id, name, review\_count, yelping\_since, friends, average\_stars etc.
- Review: id, user, business, date, text(comment)

We aim to predict the rating of a user coming to a new business [1,2,3,4,5]





Training part:

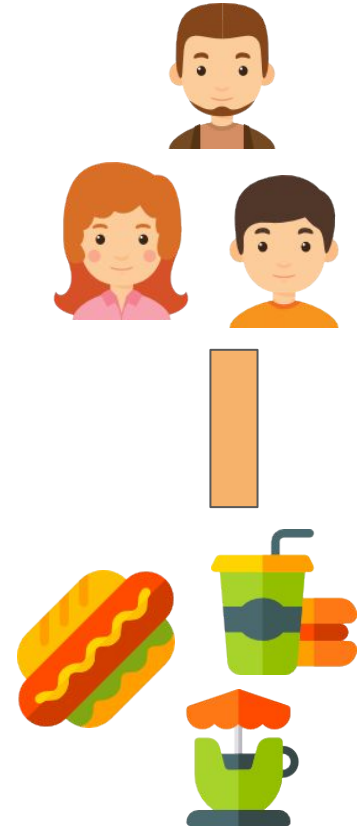
- Review information that users have ever left
  - Include ratings
- User and business profile information

Based on the training data, build a multi-class classification model that can:

Given a {user\_id, business\_id} pair, predict the rating

Pairs to be predicted are unseen before

User and business were observed during training





Original Yelp Challenge Dataset

Contains

~6M reviews

~180K businesses

~1.5M users

For Our Course Project:

Sample a subset of the original Yelp Challenge Dataset

Every user has only one comment for one business



Try your model on the Kaggle competition

See your score on the leaderboard

Evaluation:

RMSE: Root Mean Squared Error

See details in the description of kaggle competition

<https://www.kaggle.com/c/yelpratingprediction>



- Midterm Report
- Final Report
- Performance on Kaggle
  - A small amount of score (~10%-15%) will be influenced by the performance of your project on the leaderboard.





- **No longer than 3 pages**
- Current progress about project, including
  - Data processing and transformation
  - Designed & tested models / methods
- Discussion and future project plan
  - Some conclusions and findings
  - Analysis of current models and techniques
  - Timeline of future project plan (around 4 weeks)

***Details about midterm report guidelines will be released later!***



- **No longer than 5 pages!**

- Must include:

- Group member information
- Data selection and pre-processing
- Model and techniques
- Evaluation and conclusion
- \*Current leaderboard rank and score
- References and credit (papers, other's codes, maximum 1 page)
- Related work (maximum ½ page)
- **Task distribution form**

- Must NOT include:

- Background or too much description on given original datasets
- Any source code

***Details about final report  
guidelines will be released later!***



Task	People
Data processing	Student A
Implementation: Algorithm 1	Student B, C
Implementation: Algorithm 2	Student B, D
Implementation: Algorithm 3	Student A, D
Writing final report	Student C



- **Oct. 10:** Group formation due
- **Nov. 12:** Midterm project report due
- **Dec. 9:** Final project report due (together with all codes)

**UCLA**



**Q & A**

**UCLA**



**Thank you!**

*Enjoy "mining" and good luck!*