Project Introduction: Help Yelp

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

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

https://www.yelp.com/biz/bottega-louie-los-angeles
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]
Task

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
Dataset

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
Submit your results to Kaggle!

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
Project Grading

- 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.
Project Midterm Report

• 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!
Project Final Report

● **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 Distribution Form: Example

<table>
<thead>
<tr>
<th>Task</th>
<th>People</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data processing</td>
<td>Student A</td>
</tr>
<tr>
<td>Implementation: Algorithm 1</td>
<td>Student B, C</td>
</tr>
<tr>
<td>Implementation: Algorithm 2</td>
<td>Student B, D</td>
</tr>
<tr>
<td>Implementation: Algorithm 3</td>
<td>Student A, D</td>
</tr>
<tr>
<td>Writing final report</td>
<td>Student C</td>
</tr>
</tbody>
</table>
Important Dates & Milestones

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

Enjoy “mining” and good luck!