

# Yelp Rating Prediction Visualization

Jiawen Liu  
Keke Wu  
Wei Miao  
Xu Han  
Yawen Zhang

## ABSTRACT

### Your motivating problem, what you did, and what you found

Yelp is a crowd-sourced local business review and social networking community, which has hundreds of thousands of users contribute their data every day. Based on users' reviews and ratings, good local businesses stand out among their categories on top of the list, acting as a word-of-mouth reference. However, tons of user data doesn't make any sense unless we make good use of it. In this case, we decided to look into the problem that how we might get as much as useful information from our data with a particular interest in how the users' rating behavior are influenced by different factors, and what kind of prediction we can make out of the rating pattern we found. With various features extracted from Yelp data, we conducted feature selection, best split and finally built a Decision Tree model which predicts users' rating based on those features. Our visualization includes the complete process of this typical machine learning method, which provides insights about the inner mechanism of how the rating prediction is conducted.

## KEYWORDS

Yelp; Rating Prediction; Decision Tree

### ACM Reference Format:

Jiawen Liu, Keke Wu, Wei Miao, Xu Han, and Yawen Zhang. 2018. Yelp Rating Prediction Visualization. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUCTION

### The motivating Problem, why it's interesting or important

Rating prediction plays an important role in the recommendation system as to capture users' preference for specific product. In Yelp, user can post their rating and review for a restaurant they visited, and these data as well as user/restaurant related attributes, e.g., user average rating, restaurant location or category, can be well gathered for conducting the rating prediction. If a user's rating for a restaurant can be accurately predicted, Yelp would be able to recommend high rating product to the user which meets their preferences.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference'17, July 2017, Washington, DC, USA

© 2018 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

The challenge in this problem lies in the feature selection and modeling processing. With multiple features related to a user's rating, picking up the effective features is quite important, which would significantly influence the prediction performance. Additionally, the modeling process which selects the best machine learning model for rating prediction is also critical, and it's necessary to gain insights about how the model functions, i.e., the inner mechanism of model. Previously, most machine learning processes have been conducted in a "black box", by just showing input and output. In our project, we aim at visualizing the whole process, including feature selection, best split, and model tree. The Decision Tree model is selected because its straightforward idea in classification, which would be beneficial for visualization.

## 2 RELATED WORK

Summarize research related to your projects, minimum eight citations.

## 3 DESIGN PROCESS

Justifications for any design elements

## 4 MODELING PROCESS

Decision Tree

## 5 RESULTS

Our storyline

## 6 DISCUSSION

## REFERENCES