Yelp Review Rating Prediction

Deepika, Jinlong, Connor, Lin

Business Applications

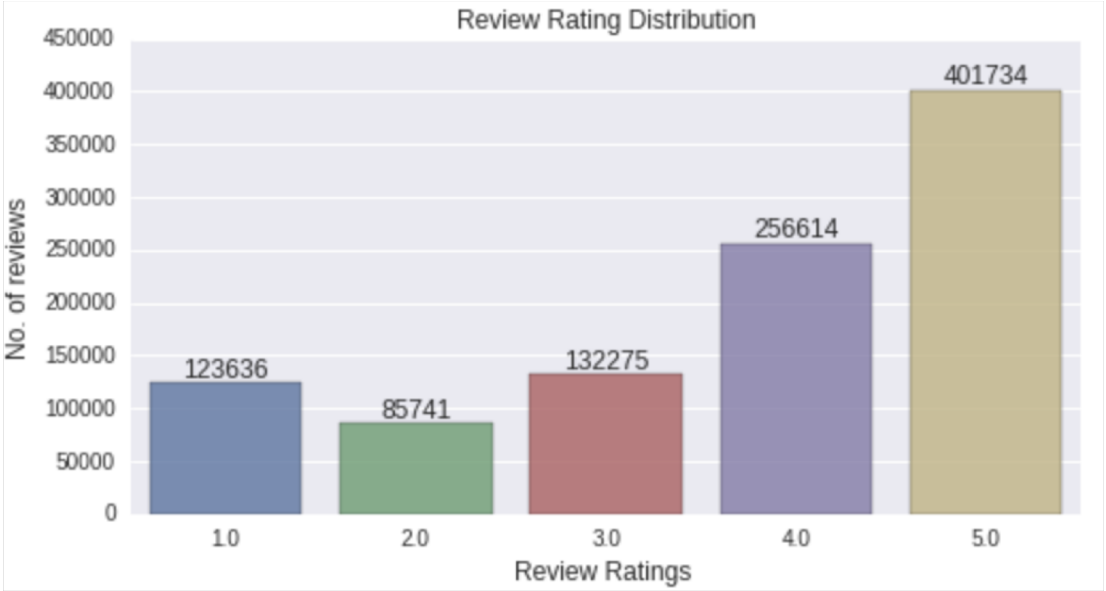
- According to Inc. Magazine, 84 percent of people trust online reviews as much as friends' recommendations when it comes to making purchasing decisions.
- Every consumer-facing business has an online presence that can be impacted by customer reviews.
- Businesses must actively manage their online reputations in order to succeed in the market.
- By building models that predict the review rating and sentiment score, businesses will be better able to manage customer complaints by intelligently allocating customer service resources.
- Our methods will reduce the burden of negative reviews by flagging customers who are likely to complain online.

Data Overview

- Dataset taken from https://www.kaggle.com/yelp-dataset/yelp-dataset.
- Tables like business, review, user, check in, tip existed.
- Combined business, review and user tables containing records for users, their reviews and the information about the businesses users have reviewed.
- 1 million rows having columns like review id, user id, business, review, city, etc. were considered.

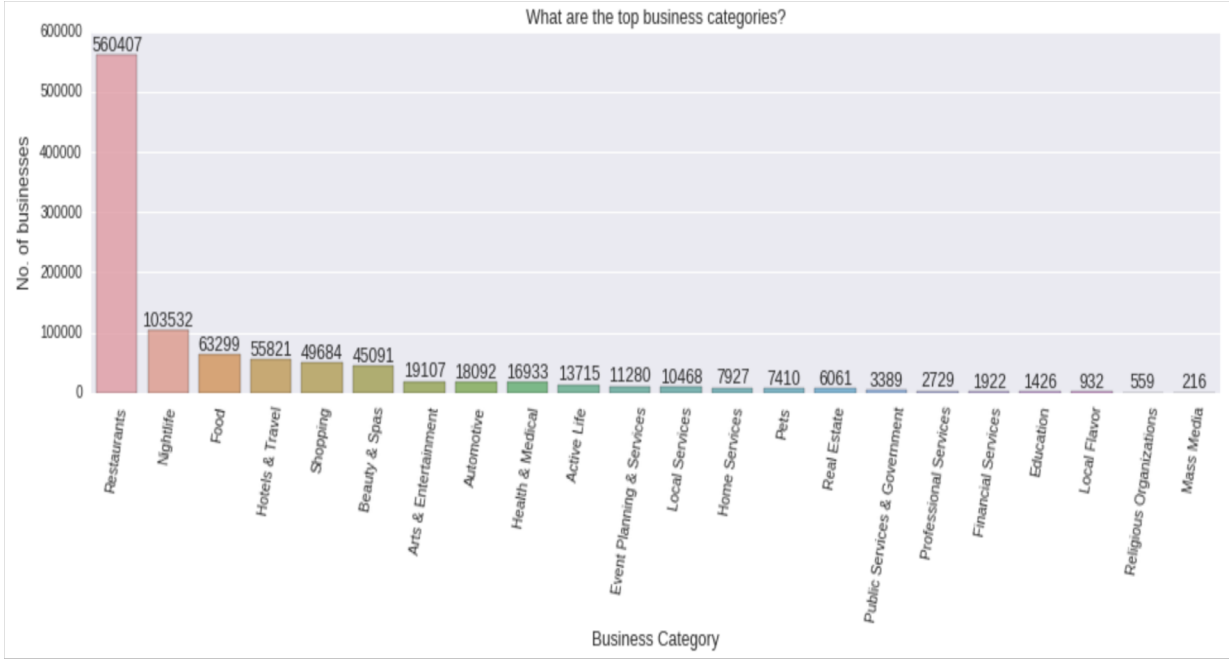
Exploratory Data Analysis

- No. of reviews by rating



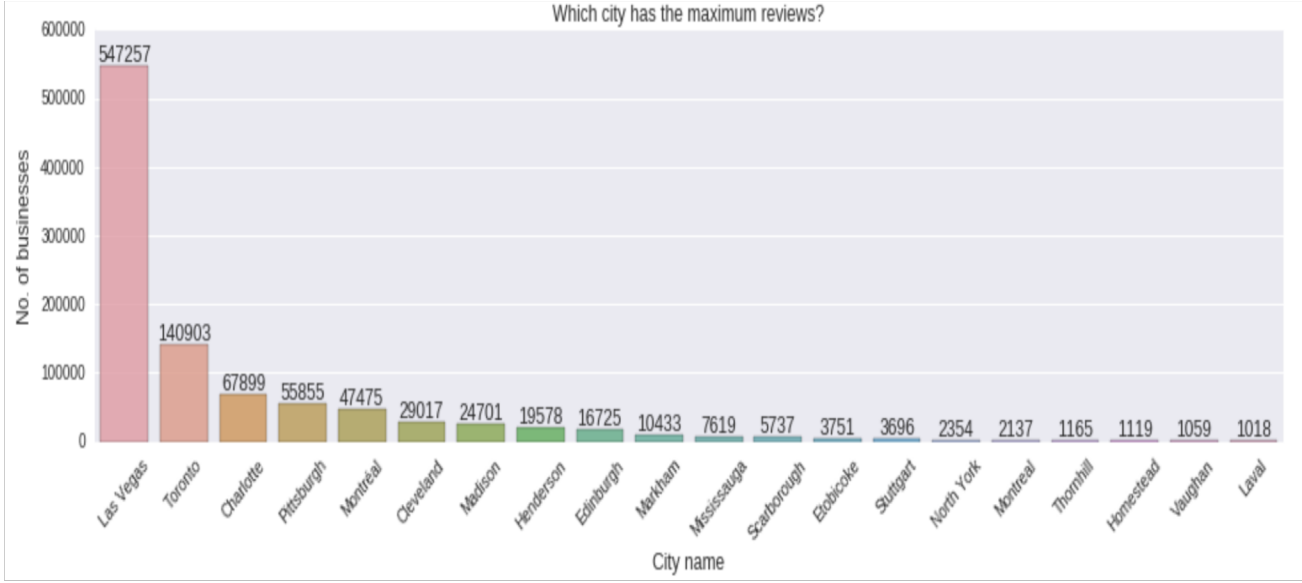
Exploratory Data Analysis

- Popular business categories



Exploratory Data Analysis

- No. of reviews by cities

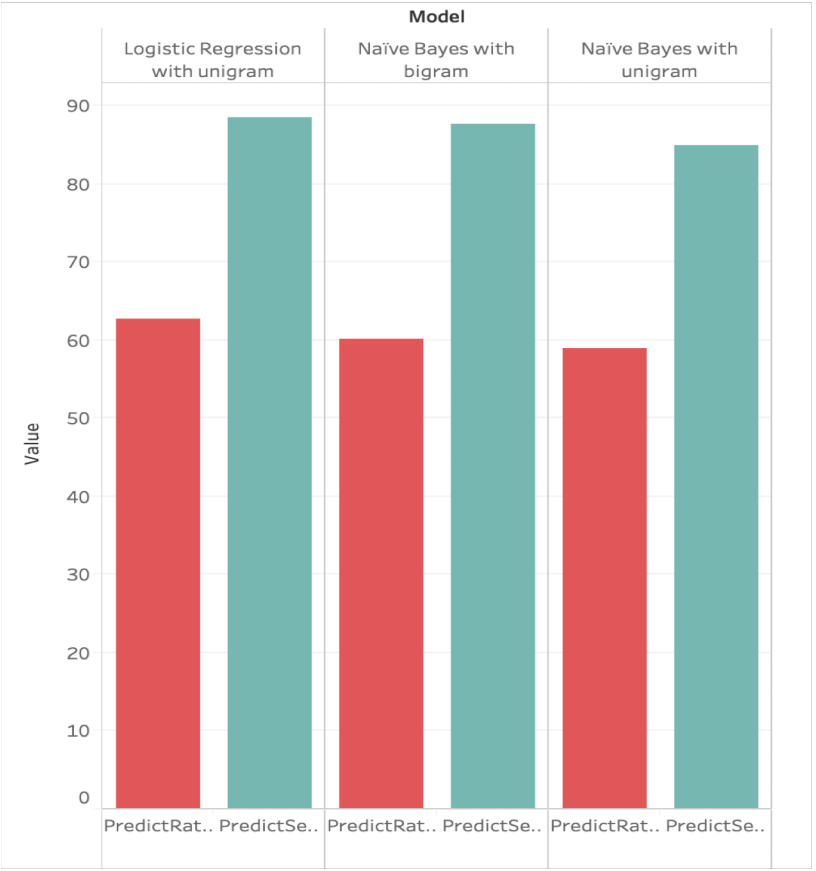


Exploratory Data Analysis

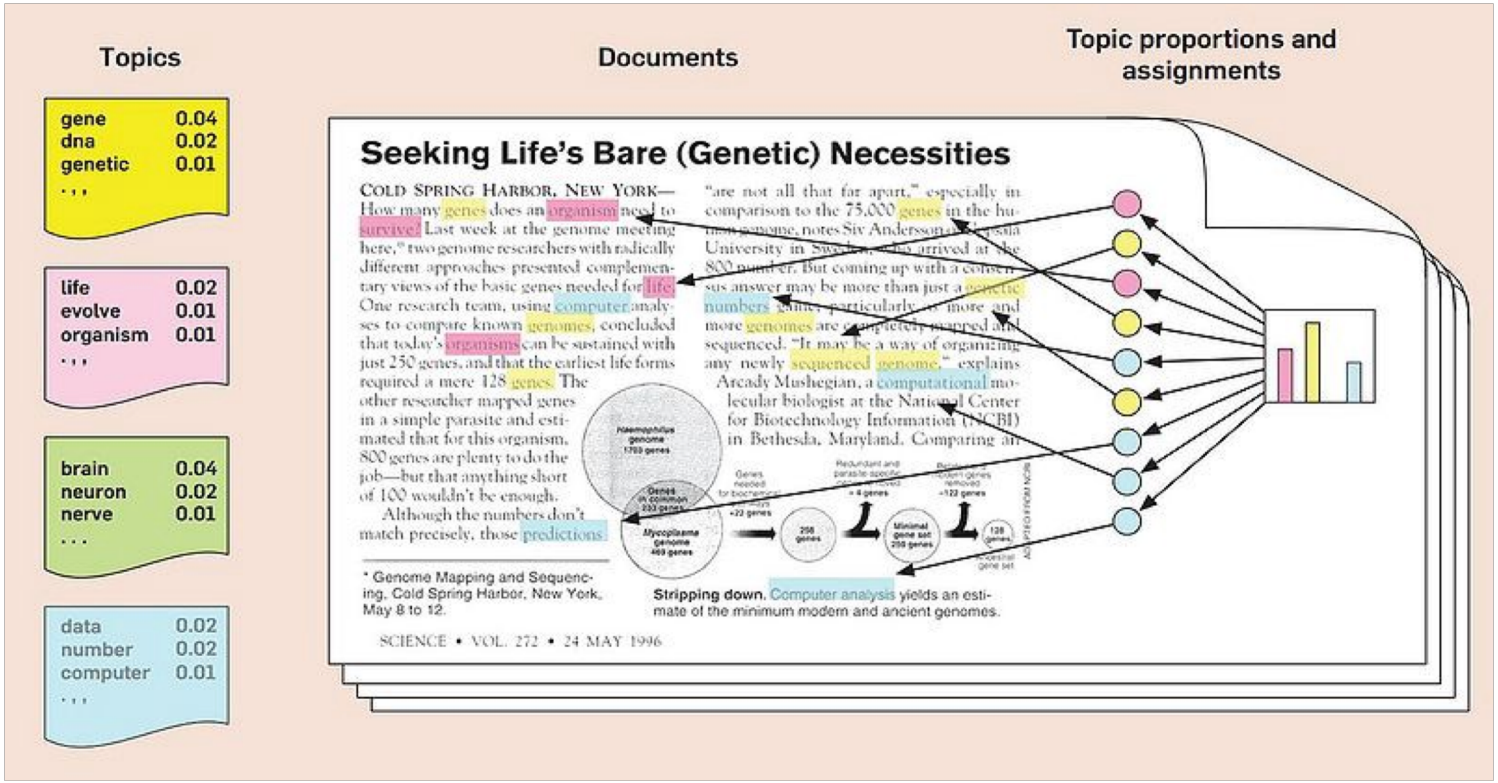
- Most common words in reviews

N-Gram

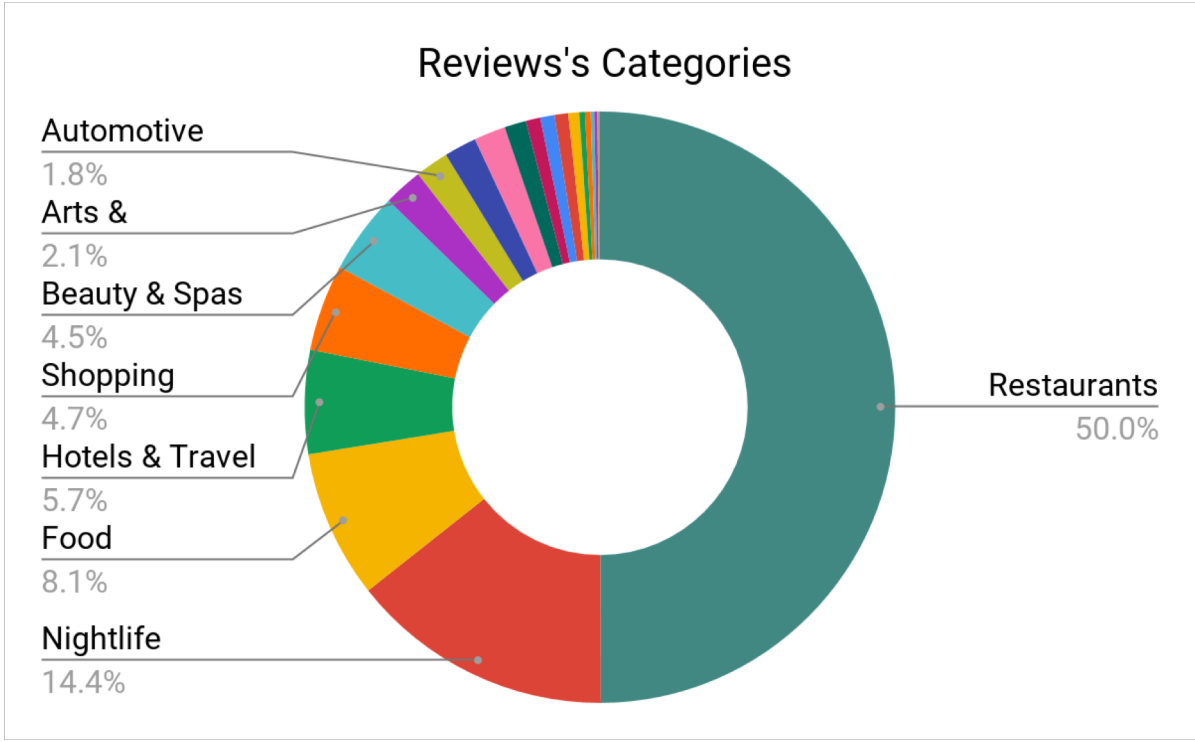
Model	PredictRating(%)	PredictSentiment(%)
Naïve Bayes with unigram	58.95	84.86
Logistic Regression with unigram	62.65	88.52
Naïve Bayes with bigram	60.16	87.58



LDA Model Introduction



Reviews segmentation based on categories

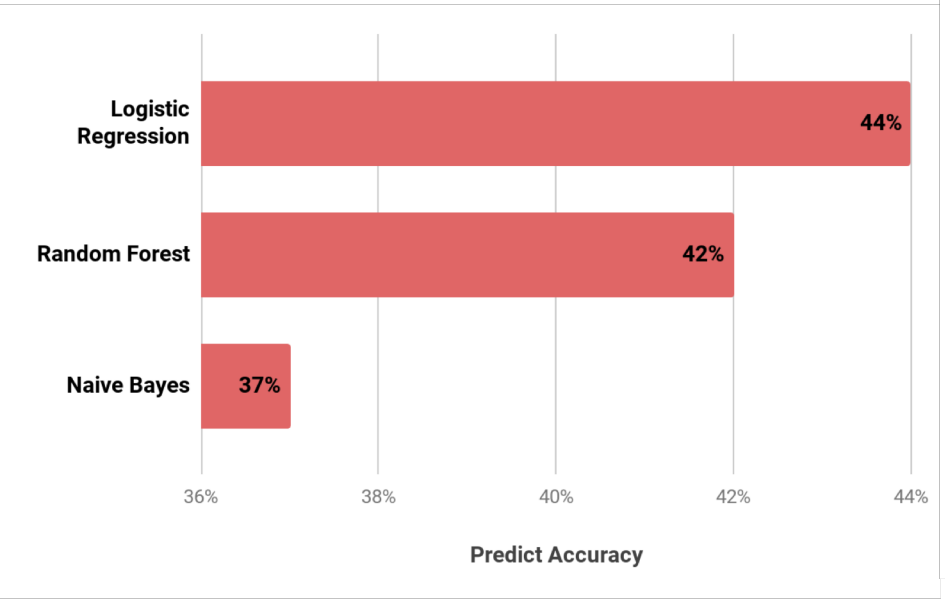


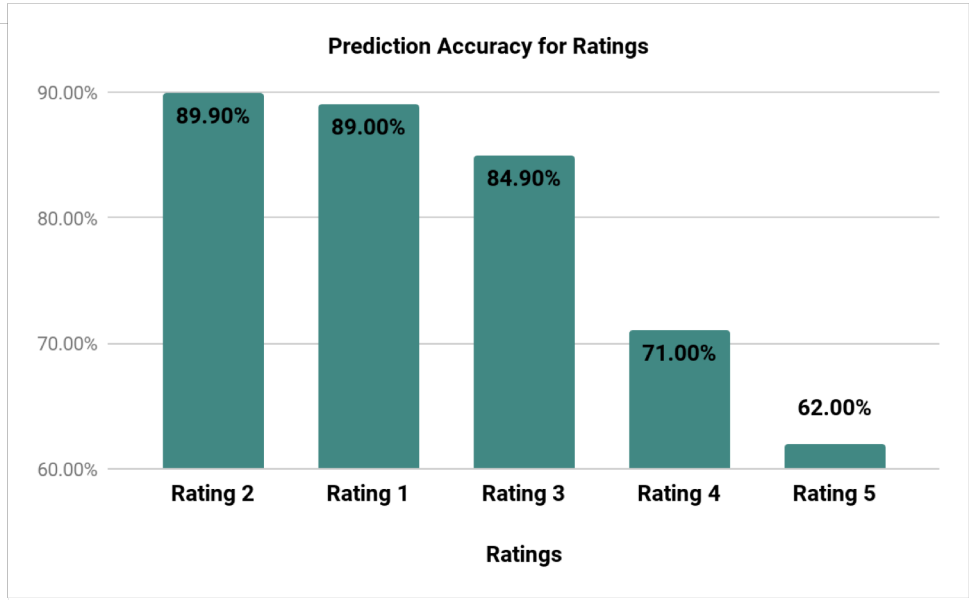
LDA Topics

“ Super simple place but amazing nonetheless. It's been around since the 30's and they still serve the same thing they started with: a bologna and salami sandwich with mustard. Staff was very helpful and friendly.”

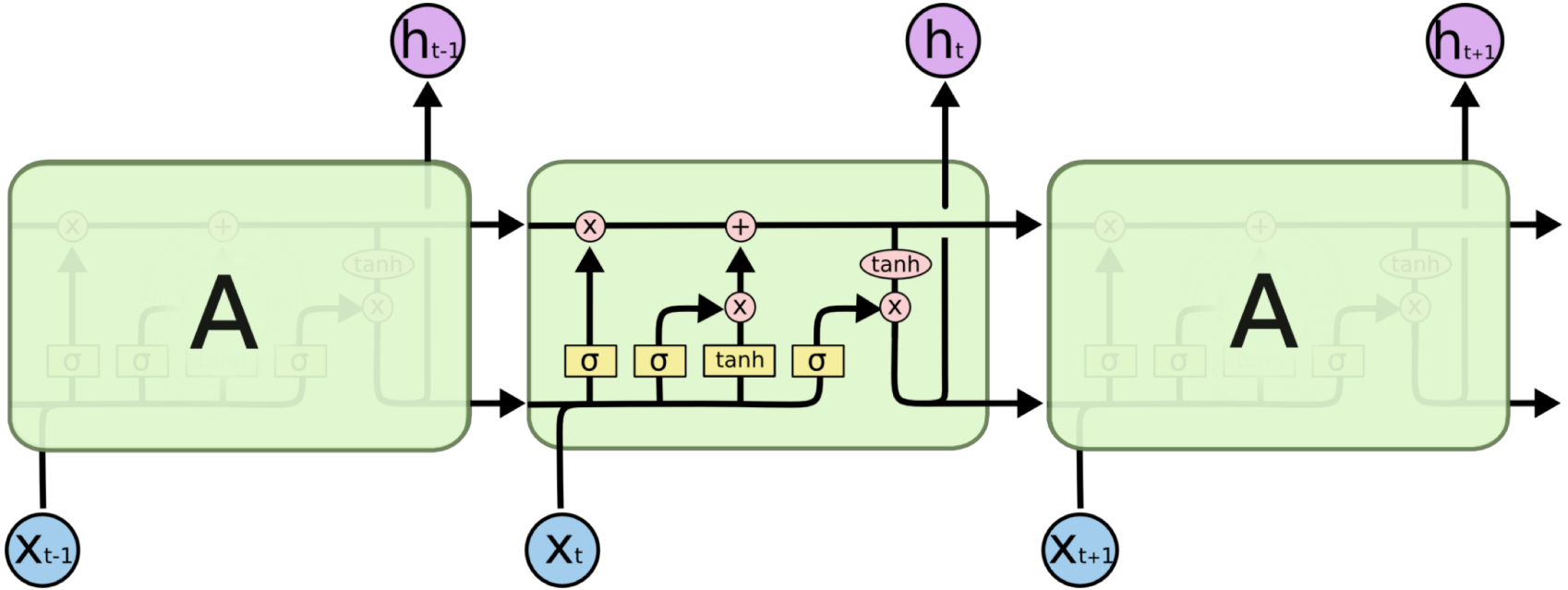
63%	17%	14%	5%	1%
Place, Food, Good, Great, Service, Love, Really, Price, Best, Delicious	Chicken, Fry, Burger, Sauce, Sushi, Rice, Roll, Soup, Order, Dish	Pizza, Cheese, Dessert, Steak, Bread, Salad, Sauce, Dish, Wine, Potato	Buffet, Pour, Cest, Mai, Nicht, Station, Plu, Trè, Dan, Sehr	Order, Time, Food, Like, Wait, Would, Table, Even, Came, Didn’t

LDA Topics used for Machine Learning





LSTM

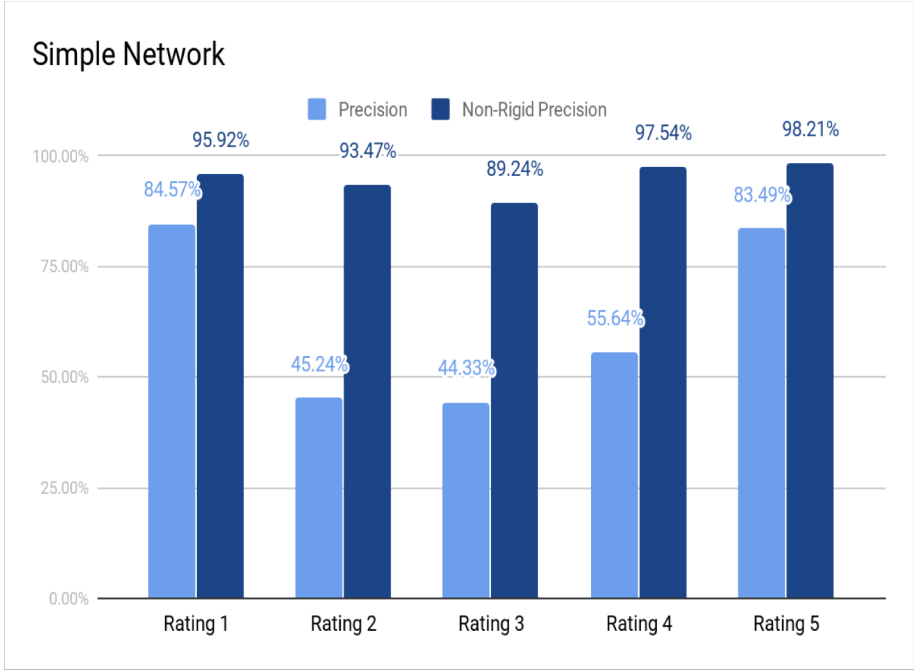


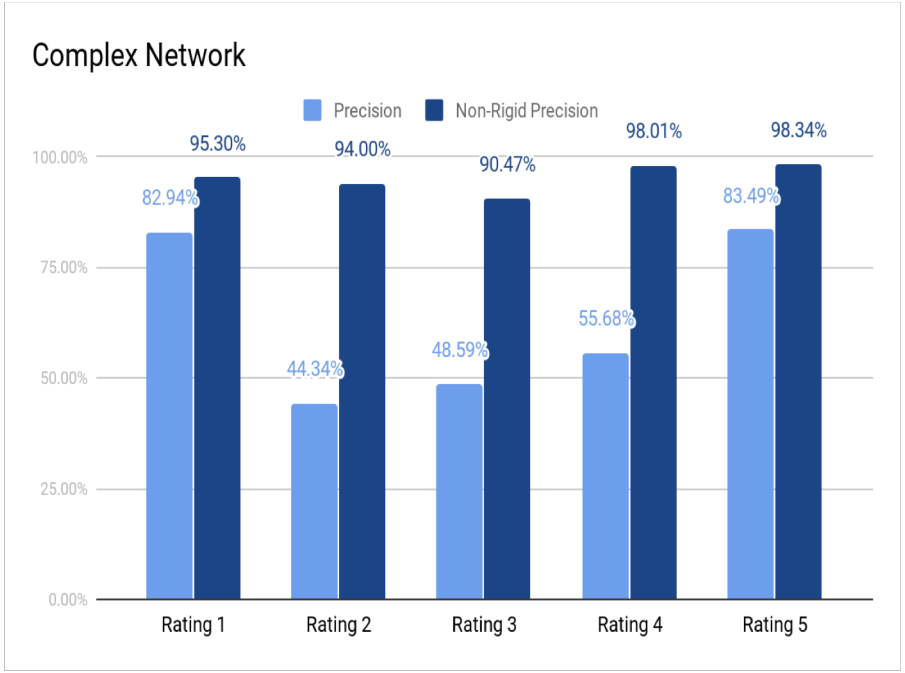
LSTM

01	Vocabulary Size	• **50K most frequently used words**
02	Training Dataset	• **25K reviews** for training
03	Review Trim	• Maximum of **200 words** for each review

01	Vocabulary Size	• **80K most frequently used words**
02	Training Dataset	• **1 million reviews** for training
03	Review Trim	• Maximum of **300 words** for each review

Performance Evaluation





References

[1]. Alexandr, B. (n.d.). Rating prediction with sentiment analysis.

[2]. Asghar, N. (n.d.). Yelp Dataset Challenge: Review Rating Prediction.

[3]. Ganu, G., Elhadad, N., & Marian, A. (n.d.). Beyond the Stars: Improving Rating Predictions using Review Text Content.

[4]. Jong, J. (n.d.). Predicting Rating with Sentiment Analysis.

[5]. Kavousi, M. (n.d.). Estimating the Rating of Reviewers Based on the Text.

Q & A