Yelp Review Rating Prediction

Lin Zhang
12/18/2018

Introduction

CDC 24/7
Saving Lives.
Protecting People.™

mailchimp

Panasonic
AUTOMOTIVE

Objective

Building a predictive model to forecast the review rating, purely based on reviews from yelp.

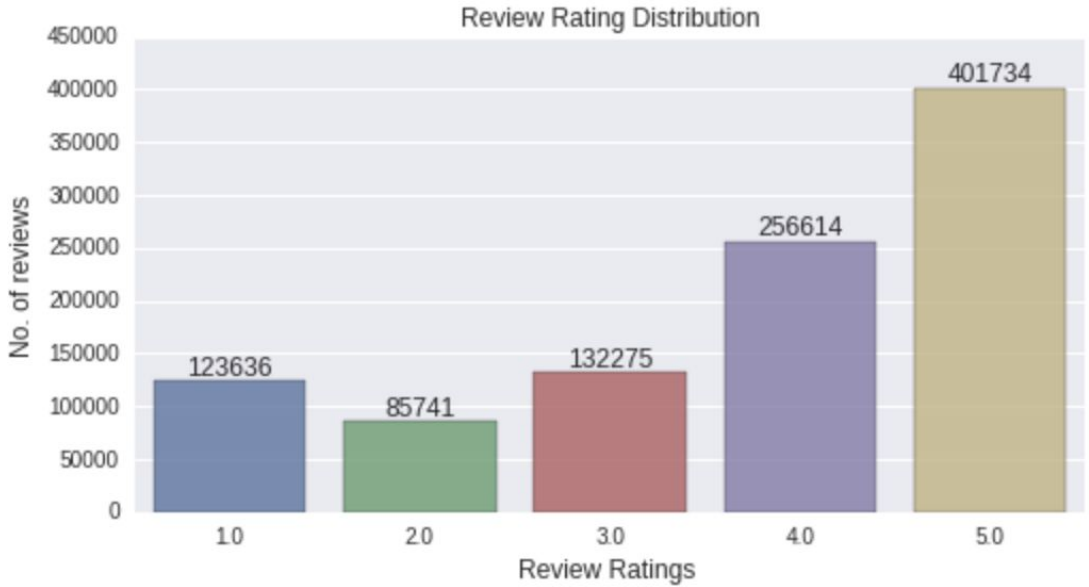
- Help businesses better understand their services from customers.
- Help businesses deal with customer complaints by intelligently allocating customer service resources and better manage their online reputation.
- Help business reduce the burden of negative reviews by flagging customers who are likely to complain online.

Data Overview

- Dataset:
  Yelp review open data source, 1 million rows of reviews and ratings
- Technique:
  Data wrangling, machine learning, natural language processing, deep learning
- Tools:
  Python library: pandas, numpy, sklearn, NLPK, tensorflow, keras

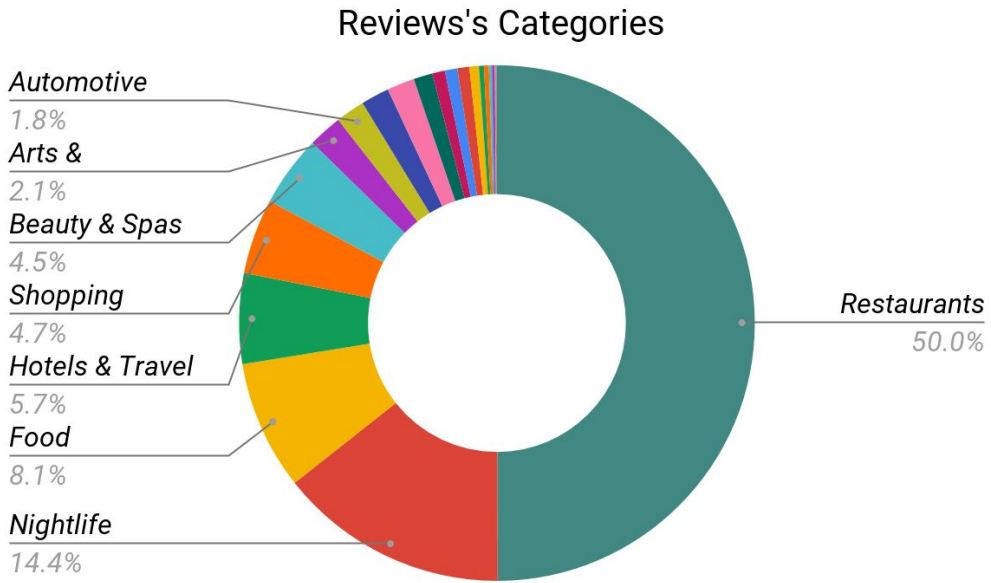
Exploratory Data Analysis

Review rating distribution



Exploratory Data Analysis

Popular business categories



Exploratory Data Analysis

Most common words in reviews

Methodology

Feature Engineering

POS
- Review Length(raw)
- Review Length(clean)
- # Punctuations
- # Noun
- # Adjective
- # Verb

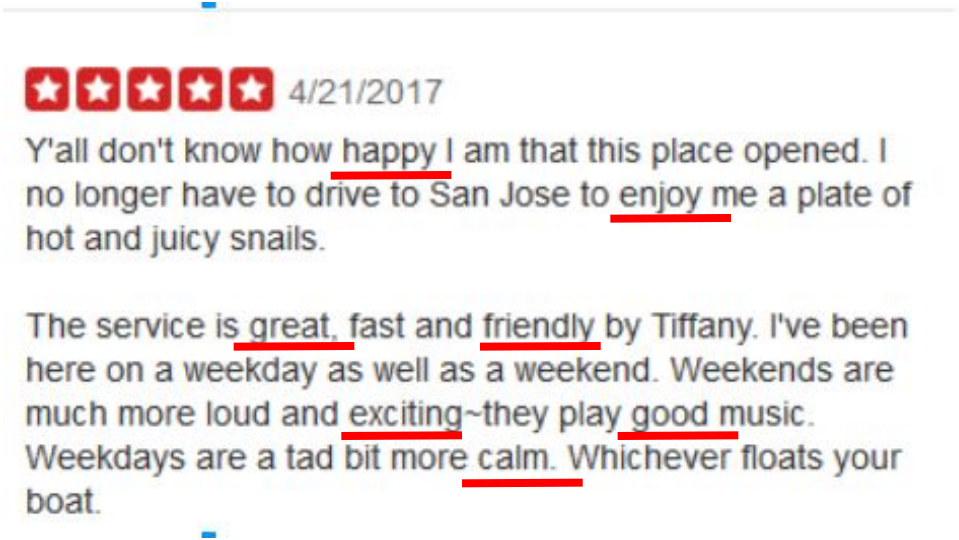
Sentiment analysis
- Positive score
- Neutral score
- Negative score
- Compound score

LDA Model
- Topic 1
- Topic 2
- Topic 3
- Topic 4
- Topic 5

ML algorithms
- Logistic regression
- SVM
- Naive bayes
- Decision Tree
- Random Forest
- KNN

Deep learning:
- LSTM

Methodology



Text cleaning first:

- Remove punctuations
- Remove stop words
- Lemmatization
- Tokenization

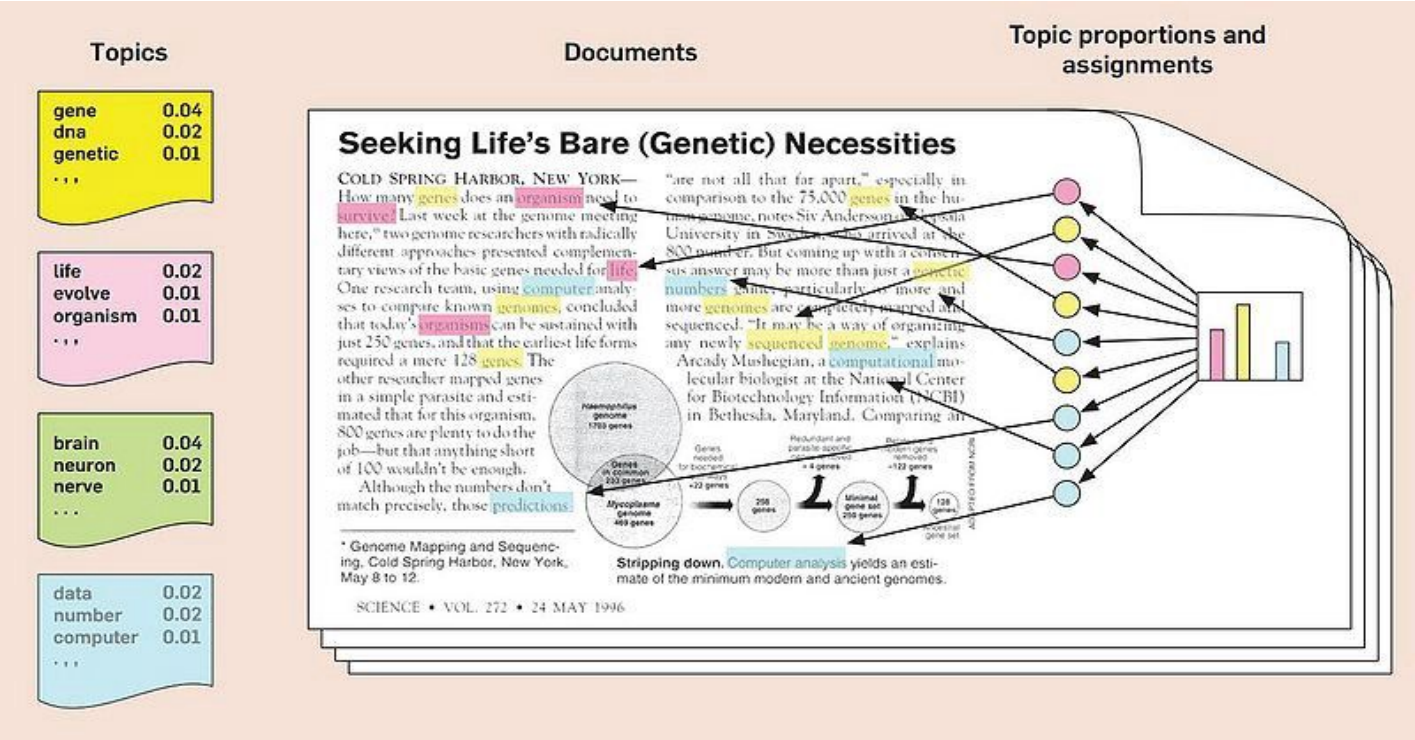
Part of Speech:

- How many adj, verb or noun?

Sentiment Analysis

0 0.05 0.95

Methodology---LDA Model

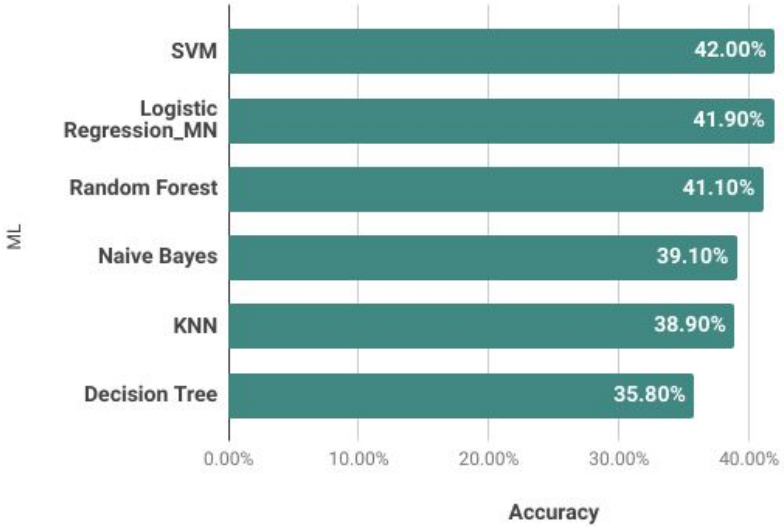


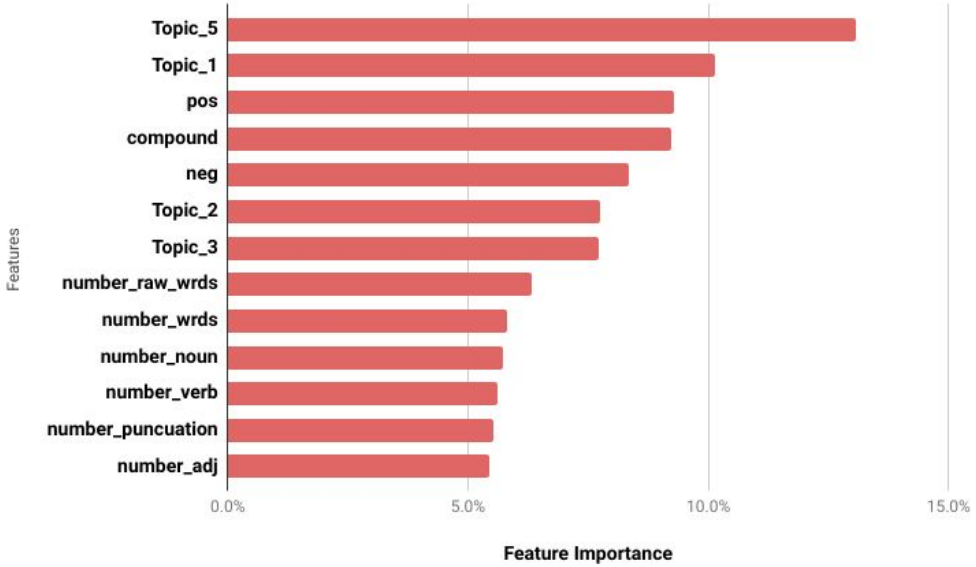
LDA Topics in reviews of restaurant

“ 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.”

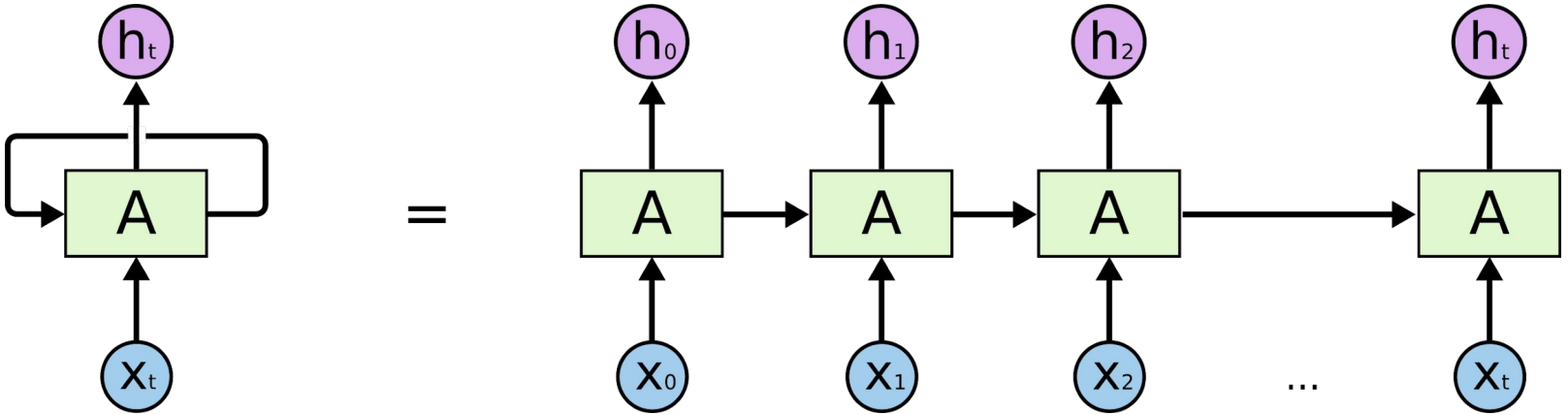
0.63	0.17	0.14	0.05	0.01
Place, Food, Good, Great, Service, Love, Really, Price, Best, Delicious	Chicken, Fry, Burger, Sauce, Sushi, Rice, Roll, Soup, Order, Dish	Pizza, Cheese, Dessert, Steak, Bread, Salad, Sauce, Dish, Wine, Potato	Buffet, Pour, Cest, Mai, Nicht, Station, Plu, Trè, Dan, Sehr	Order, Time, Food, Like, Wait, Would, Table, Even, Came, Didn’t

Machine Learning Algorithms

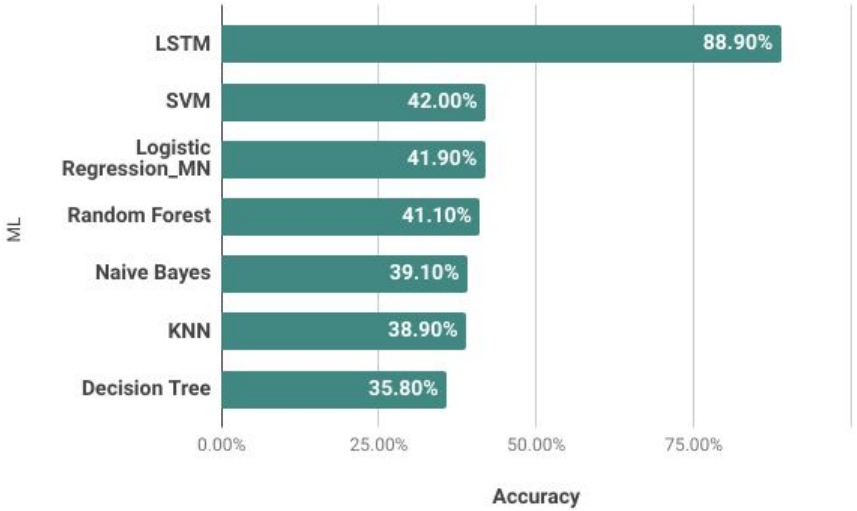


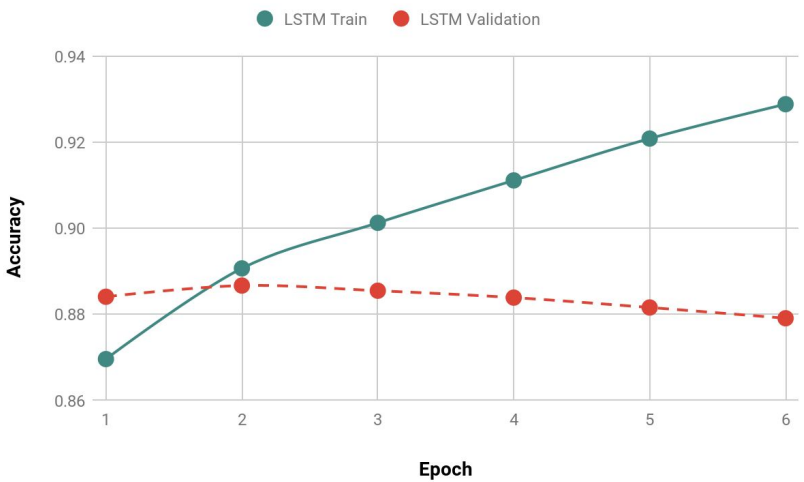


Deep Learning: LSTM



LSTM & Conclusion





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