User-Input Text to Restaurant Recommendations

Samantha Chu
Capstone Presentation
DSIR-2221
5/14/2021

Let's Talk about Yelp

What is Yelp?

Yelp was founded in 2004 by former PayPal employees Russel Simmons and Jeremy Stoppelman.

It has since grown in usage and raised several rounds of funding in the years following. By 2010, it had $30 million in revenue, and the website had published about 4.5 million crowd-sourced reviews. From 2009 to 2012, Yelp expanded throughout Europe and Asia. Yelp became a public company via an initial public offering in March 2012 and became profitable for the first time two years later.

Yelp connects people with great local businesses. Users have contributed reviews of almost every type of local business, from restaurants, boutiques and salons to dentists, mechanics, plumbers and more.

184 million reviews worldwide

That's a lot of reviews!

44th most visited website in the US

According to Alexa.

92 million unique mobile users per month

On average.

178 million unique visitors monthly across mobile/desktop/app platforms

45% of customers likely to check Yelp reviews before visiting a business

35% of people searching on Yelp will make a visit to a business they checked within 24 hours

POINT:
TON of User Engagement

Want to utilize that engagement to bring better recommendations to the user.

> **What is it like to choose a restaurant on Yelp?**

More Features

General Features

☐ Yelp Delivery   ☐ Waiter Service
☐ Yelp Takeout   ☐ Wheelchair Accessible
☐ Reservations   ☐ Coat Check
☐ Waitlist   ☐ Has TV
☐ Open At: 3:44 PM   ☐ Dogs Allowed
☐ Open Now 6:47 PM   ☐ Liked by 20-somethings
☐ Takes Reservations   ☐ Hot and New
☐ Accepts Credit Cards   ☐ Offers Military Discount
☐ Outdoor Seating   ☐ Gender-neutral restrooms
☐ Good for Kids   ☐ Open to All
☐ Good for Groups   ☐ Accepts Apple Pay

Search

5

" What is it like to choose a restaurant on Yelp?

More Features ✕

☐ Outdoor Seating   ☐ Gender-neutral restrooms
☐ Good for Kids   ☐ Open to All
☐ Good for Groups   ☐ Accepts Apple Pay

Alcohol ⌃

☐ Full Bar   ☐ Good For Happy Hour
☐ Beer & Wine Only

Meals Served ⌄

Music ⌄

Parking ⌄

Search

"

What is it like to choose a restaurant on Yelp?

More Features   ✕

☐ Beer & Wine Only

Meals Served   ⌄

☐ Breakfast   ☐ Dinner
☐ Brunch   ☐ Dessert
☐ Lunch   ☐ Late Night

Music   ⌄

Parking   ⌄

Wi-Fi   ⌄

Smoking

Search

"

What is it like to choose a restaurant on Yelp?

More Features ✕

- ☐ Breakfast
- ☐ Brunch
- ☐ Lunch
- ☐ Dinner
- ☐ Dessert
- ☐ Late Night

Music ⌃

- ☐ DJ
- ☐ Juke Box
- ☐ Karaoke
- ☐ Live

Parking ⌄

Wi-Fi ⌄

Smoking ⌄

Search

"

What is it like to choose a restaurant on Yelp?

More Features ✕

Music ⌃

☐ DJ   ☐ Karaoke
☐ Juke Box   ☐ Live

Parking ⌃

☐ Street   ☐ Private Lot
☐ Garage   ☐ Validated
☐ Valet

Wi-Fi ⌄

Smoking ⌄

Search

"

What is it like to choose a restaurant on Yelp?

More Features ✕

☐ DJ   ☐ Karaoke
☐ Juke Box   ☐ Live

Parking ⌃

☐ Street   ☐ Private Lot
☐ Garage   ☐ Validated
☐ Valet

Wi-Fi ⌃

☐ Free   ☐ Paid

Smoking ⌄

Search

"

Problem Statement:

It can be <u>overwhelming</u> choosing from all these features

Sometimes people <u>do not know</u> what they want

--

Not really capturing the experience/ambience of the restaurant that the user wants.

While these features are helpful, it's much easier if a user can directly describe what type of restaurant experience they are looking for.

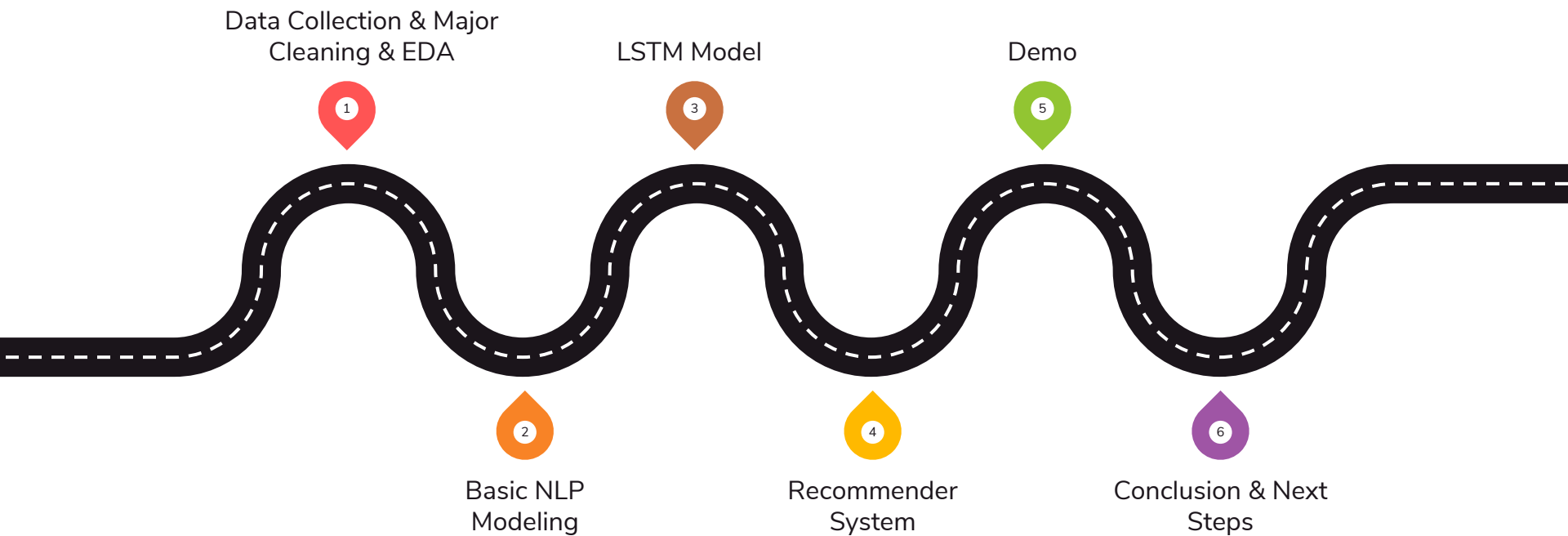*Suggest an enhancement to Yelp by adding user-input section.*

Goal:

Using the TON of user engagement we saw earlier…
- Train a model to use restaurant reviews to accurately predict whether a restaurant has a Casual and/or Classy Ambience
- That language model will then be used on user inputted text to predict what type of experience they want (Casual and/or Classy) and recommend the appropriate restaurants.
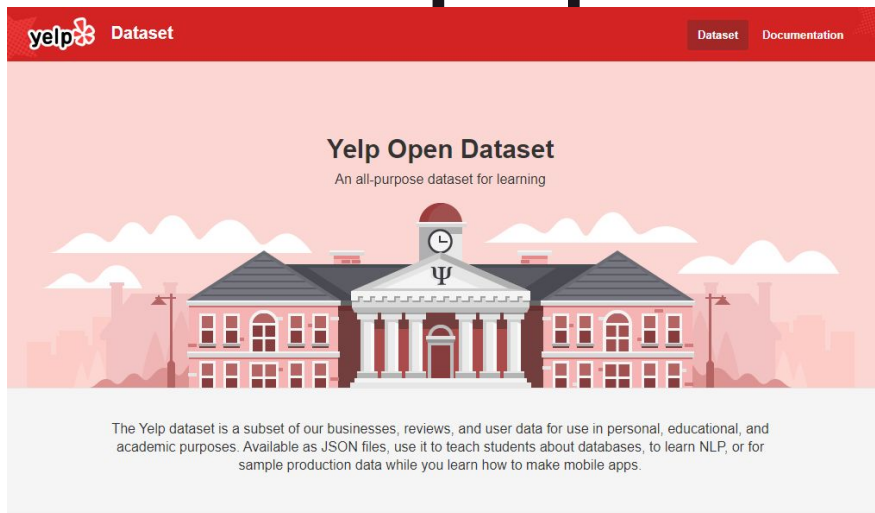
Roadmap:

Data Collection & Major
Cleaning & EDA

LSTM Model

Demo

1

3

5

2

4

6

Basic NLP
Modeling

Recommender
System

Conclusion & Next
Steps

1

Data Collection & Cleaning

Data Source: Yelp Open Dataset

Data Extracting/Cleaning:

Data Extraction:

- Convert json → csv files
- Unpack dictionaries
- Unpack dictionaries within a dictionary
- Pull out labels
- Make them binary variables

Data Cleaning: (2,023,319 reviews)

- Filtered for open businesses/restaurants
- Filtered for restaurants categorized under: Food, Restaurant, Bar
- Created Multi-Class target:
  - ***Neither casual nor classy***
  - ***Casual only***
  - ***Classy only***
  - ***Both casual and classy***
- Undersampling the majority class (Casual Ambience)
- Subset: State = Ohio (114,110)
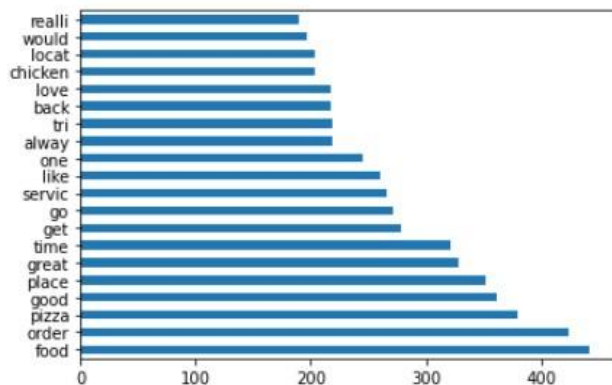 - Did NOT subset a city -- restaurants not evenly spread
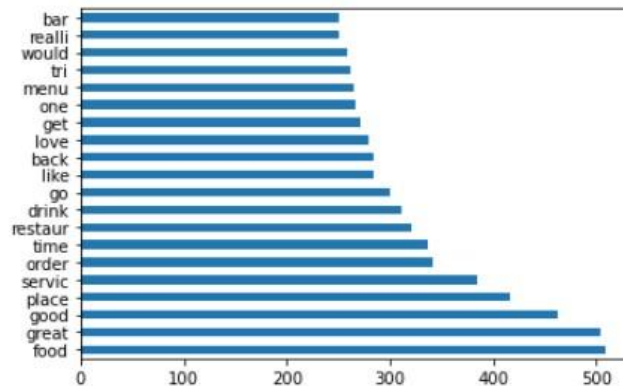- Stemmed words

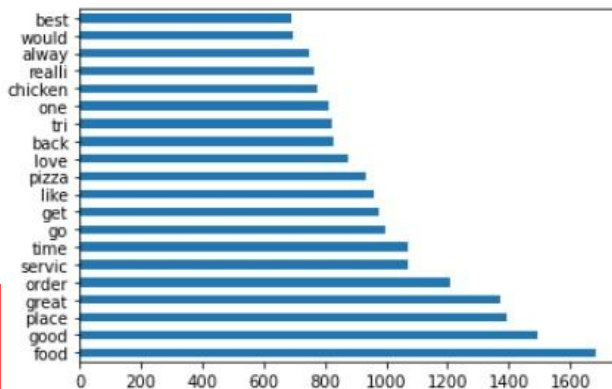# Exploratory Data Analysis
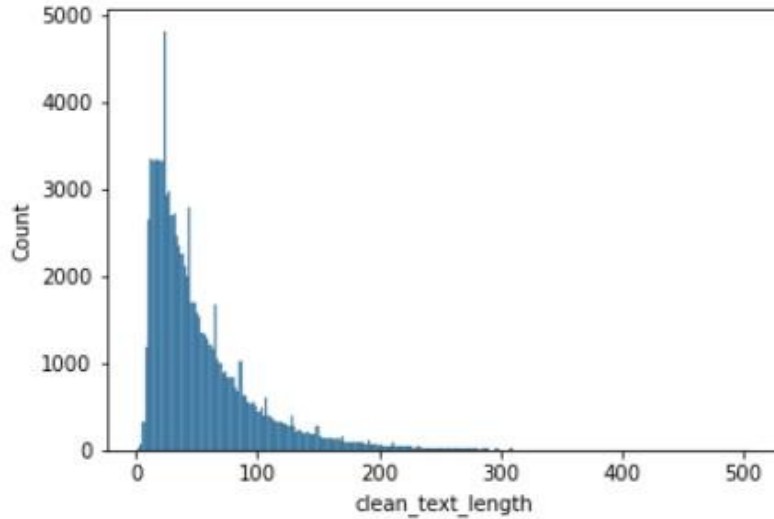
Exploratory Data Analysis: Common Words
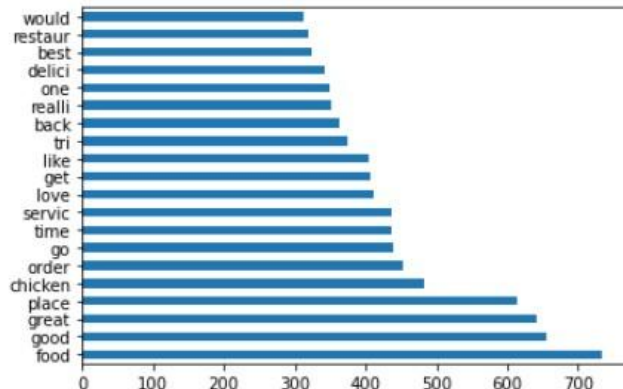
Neither casual nor classy

Classy only

Casual only

Both casual and classy

2

Modeling

- Basic NLP Modeling
- LSTM Model

Basic NLP Modeling with Accuracy

Baseline: 0.3287	Decision Tree	Support Vector Machines	Bagging	Random Forest	AdaBoost
Train Accuracy	0.5429	1.00	1.00	0.4357	0.4873
Test Accuracy	0.3286	0.3287	0.3287	0.3287	0.3287
Test Accuracy (after pruning: max depth = 5)				0.3286	

Need a more powerful model…
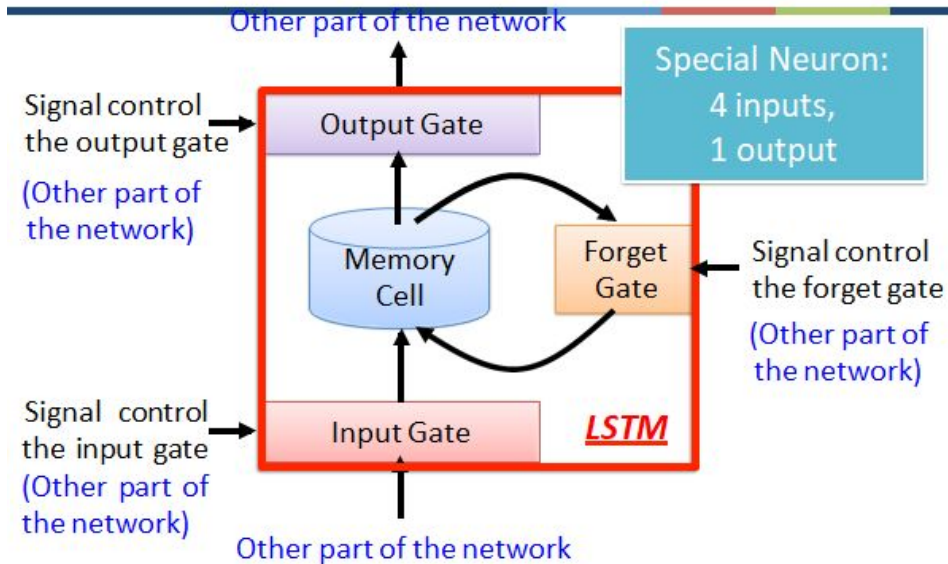- **That can account for word order/sequences. ----> LSTM**

LSTM: Long-Short Term Memory

- A variety of a recurrent neural network (RNN) architecture used deep learning

- LSTM has feedback connections

- Learns long-term dependencies

- Good for sequence prediction problems

- A property of selectively remembering patterns for long durations of time.

LSTM: Long-Short Term Memory

Long Short--term Memory (LSTM)

Other part of the network

Signal control the output gate → Output Gate

Special Neuron: 4 inputs, 1 output

(Other part of the network)

Memory Cell

Forget Gate ← Signal control the forget gate

(Other part of the network)

Signal control the input gate → Input Gate   *LSTM*

(Other part of the network)

Other part of the network

LSTM: Long-Short Term Memory
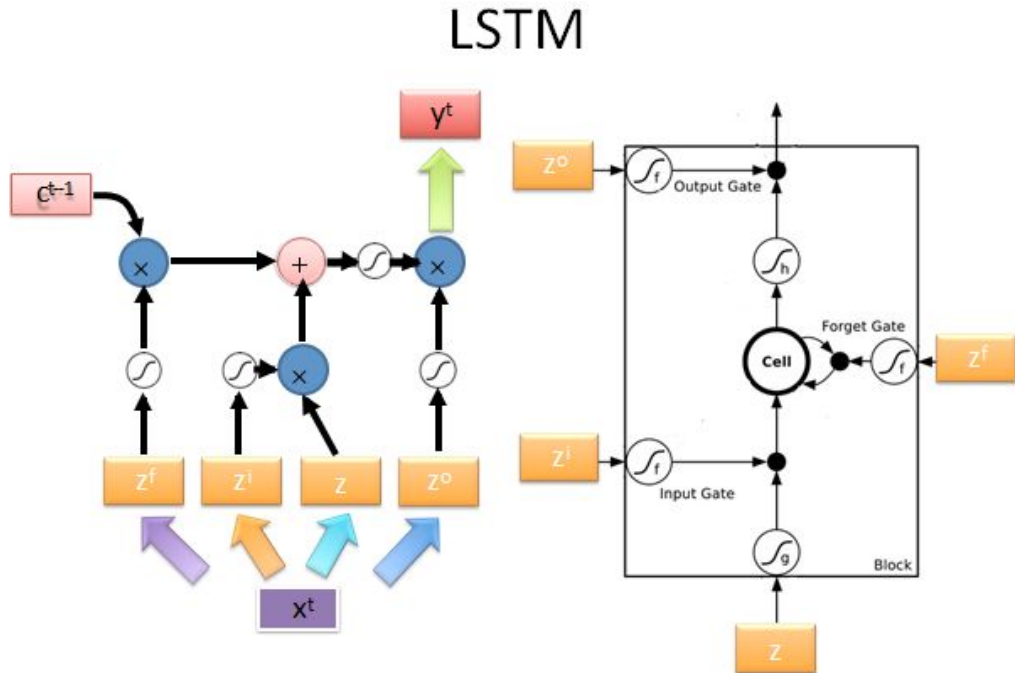
LSTM

LSTM Models

Baseline: 0.3287

	LSTM Topography #1	LSTM Topography #2	LSTM Topography #3	LSTM Topography #4	LSTM Topography #5	LSTM Topography #6	LSTM Topography #7	LSTM Topography #8
Embedding Layer	32 neurons	32 neurons	32 neurons	32 neurons	32 neurons	32 neurons	32 neurons	32 neurons
LSTM Layer	64 neurons	128 neurons	64 neurons (Bidirectional)	64 neurons	64 neurons	64 neurons	64 neurons	64 neurons
Dense Hidden Layer	-	-	64 neurons	64 neurons	64 neurons	64 neurons (Dropout: 0.5)	64 neurons	64 neurons (Dropout: 0.5, Regularized)
Dense Hidden Layer	-	-	-	-	32 neurons	8 neurons (Dropout: 0.5)	8 neurons	8 neurons (Dropout: 0.5, Regularized)
Output Layer	4 neurons	4 neurons	4 neurons	4 neurons	4 neurons	4 neurons	4 neurons	4 neurons
Train Accuracy	0.8293	0.8502	0.8836	0.8360	0.8333	0.6479	0.8533	0.6755
Test Accuracy	0.6208	0.5937	0.6043	0.5995	0.5911	0.5467	0.6058	0.5849

3

Recommender System

- Cosine Distances

Recommender System:
Cosine Distances

Allow users to choose attributes.

These chosen attributes were those I thought would help distinguish a person's experience/ambience in particular at a restaurant:

- GoodForKids
- GoodForGroups
- OutdoorSeating
- TakesReservations
- HasAlcohol
- RestaurantsTableService
- Meal Type (Lunch, Dinner, Both, Other)
- **User Inputted Text → Ambience prediction (neither, casual only, classy only, both)**

Cosine distances were found between these vectors for restaurants → Restaurant Recommendations

26

4

Demo

- To streamlit...

5

Conclusions
&
Next Steps

Conclusions and Next Steps...
because this is just a pitch or framework for the idea of including user-inputted text.

🍋 Future Model Improvements

The LSTM model performed okay, and good enough for pitching the user-input text idea. Want to improve in future iterations:

- Expand to other NLP models (BERT)

🧊 Gather More/Different Data

- Use on reviews specifically written for restaurant ambience
- Gather more city-specific restaurants

🍉 Expand To Specific Cities

- Expand all cities in Ohio
- Expand to 1 major city in Ohio
- Expand to 1 major city in a different state

🌽 Recommender Improvements

- Expand features...
- Expand model to predict more ambience classes
  - Trendy, upscale, touristy, hipster, divey, intimate, romantic

Sources

- Data Source: https://www.yelp.com/dataset
- https://www.linkedin.com/company/yelp-com/
- https://review42.com/resources/yelp-statistics//
- https://en.wikipedia.org/wiki/Yelp
- https://intellipaat.com/blog/what-is-lstm/
- https://en.wikipedia.org/wiki/Long_short-term_memory
- https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/

Thanks!

- Special thanks to Chuck, Varun, and Grant
- Thank you to all my classmates - it's been a fun ride.