



PRESENTED BY : JAYAVARDHAN AND MOUNICA

Introduction

- What is Yelp?
- Story of Yelp
- Users and Businesses on Yelp
- Growth of Yelp
- Yelp Fusion API
- Yelp OpenSource
- Yelp DataSet
- Research on Yelp Data

What is Yelp?

- Yelp is a social platform which publishes crowd-sourced reviews about small businesses.
- It also provides Online reservation service- Yelp Reservations
- Hosted on website (Yelp.com) and Yelp mobile App
- Headquartered in San Francisco, California.
- Listed on the NYSE as Yelp Inc (Price: \$45.98)
- Valuation: US\$ 5 Billion
- Annual Revenue: US\$ 713 Million

Story of Yelp

- Founded in 2004 by Jeremy Stoppelman(CEO) and Russel Simmons(Former CTO)
- Idea: Stoppelman was inspired to launch Yelp after being sick.

Stoppelman- “I got sick and needed to see a doctor. Back then there was very little information on the Internet; it was frustrating. We realized the best way to find a doctor, or other services, was by word of mouth.”

- Initial idea for Yelp was an email-based referral network
- Observation which changed Yelp forever

“People were writing unsolicited reviews of their favorite businesses just for fun.”

Users

- Everyday Individuals
 - Individual users rate and review products and service from small businesses they have visited
 - People using Yelp trust others reviews, as they are written by real people they can relate to
 - Reviews allow a person to find the best product or service
- Small Business Owners
 - Small Business owners can claim a Yelp Business page which allows them to view statistics including page views and customers leads generated by Yelp.
 - Small Businesses can also join the Yelp Ads Program which helps drive more leads and revenue for your business. Targeted ads can be placed anywhere on the Yelp website where it makes sense to do so.
- What we know about Yelp Users?
 - 53.7% are female, 46.3% are male
 - The 25-35 year old age bracket are most common users at 24.9%
 - 89 percent of customers make a purchase within one week of visiting Yelp.

Businesses on Yelp

Businesses that can be reviewed on Yelp:

- Restaurants, Nightlife/Bars
- Shopping/retail
- Beauty and Spa
- Automotive Services
- Local Services
- Home Services
- Financial Services
- Health and Medical Services Hotels and Travel Education
- Real Estate And More...

Trivia: Largest category of reviewed business on Yelp? Restaurants? No!

It's: shopping/retail

Yelp Growth

- October 12, 2004 : first review submitted to Yelp
- Summer,2006: One million User Visits per month
- May 5, 2007 : One millionth review
- May 2008: 10 million unique visitors every month.
- 2009 : Services launched in Canada, the U.K. and Ireland.
- December 2009: Offers from Google and Yahoo for 1 billion US\$
- October 2011: Siri uses Yelp data for restaurant search
- November 2011: Initial Public Offering(IPO)
- By 2011 : 65 million unique visitors per month
- As of 2016, Yelp.com had
 - 135 million unique visitors per month
 - 95 million reviews per month
 - Till date 2.8 million Businesses, 135 million reviews
- Yelp is in 27 countries worldwide.

Yelp Fusion API

- Yelp provides a HTTP based API.
- Yelp-python: A Python wrapper for the Yelp API(only works with older version of Yelp API v2)

Fusion API can be used to

- Make a geographically-oriented search
- Sort results by the best match for the query, highest ratings, or distance
- Limit results to those businesses offering a Yelp Deal, and display information about the deal like the title, savings, and purchase URL
- Identify and display whether a business has been claimed on Yelp.com

Yelp API- Authentication

- Yelp uses OAuth 1.0a for authenticating API requests

OAuth Parameter	Value
oauth_consumer_key	Your OAuth consumer key (from Manage API Access).
oauth_token	The access token obtained (from Manage API Access).
oauth_signature_method	hmac-sha1
oauth_signature	The generated request signature, signed with the oauth_token_secret obtained (from Manage API Access).
oauth_timestamp	Timestamp for the request in seconds since the Unix epoch.
oauth_nonce	A unique string randomly generated per request.

- These parameters must be passed in the HTTP (Authorization) header as URL query keys or in the POST data.

Different Functionalities in Fusion API

Search API:

- Search for local businesses using
 - Geographical Location (similar to twitter)
 - Category
 - Radius
- Limit: 40 results per search

Business API:

- Lookup business information by id.
- Returns 2 reviews for the business

Phone Search API:

- Search for businesses by phone number

- Calls to API should be made through Http GET Method.
- API Rate Limiting: By default, a client is limited to 25,000 API calls per 24 hours
- Yelp also provides GraphQL (query language) for API.

Yelp Opensource

- Yelps loves open source and has been giving back to open source community by open sourcing internally developed projects.
- Some popular Open Source projects from Yelp
 - paasta - An open, distributed platform as a service
 - elastalert - Easy & Flexible Alerting With ElasticSearch
 - osxcollector - A forensic evidence collection & analysis toolkit for OS X
 - dockersh - A shell which places users into individual docker containers
- Yelp has open sourced 79 projects till date.
- Github Page: <https://yelp.github.io/>

Yelp Dataset Challenge

- Conduct research and analysis using Yelp's datasets and share the findings with Yelp
- 2 datasets and 1 database
 - Yelp Photos dataset
 - Yelp Reviews dataset
 - Yelp Local Graph
- 10 best works given \$5000 each. Additional \$1000 if the research is published
- Currently, Round 10 of dataset challenge

Yelp Dataset Challenge - Yelp Reviews

- 5 million reviews and ratings from 1.1 million users on 150,000 businesses
- 1.2 million business attributes like hours, parking, availability, ambience
- What is Yelp looking for?
 - Infer sentiment towards the business
 - Infer business attributes (what the customer should expect): Eg: ambience (casual, hippy), good for (lunch, dinner).
 - Understand user interests
- Highly used for sentiment analysis and recommendation systems

Yelp Dataset Challenge - Yelp Photos

Yelp Photos

- 200,000 photos with their captions and labels
- What is Yelp looking for?
 - Photo classification: Eg: Identify food in the the photos.
 - Is the photo beautiful or not?
 - Infer business attributes from user-submitted photos

Yelp Local Graph

- Database of local businesses, business categories, reviews and users
- What is Yelp looking for?
 - Understand trends and how they start
 - Usage patterns

Academic Research

- Recommendation systems
- Sentiment Analysis
- Social networks
- Opinion Mining

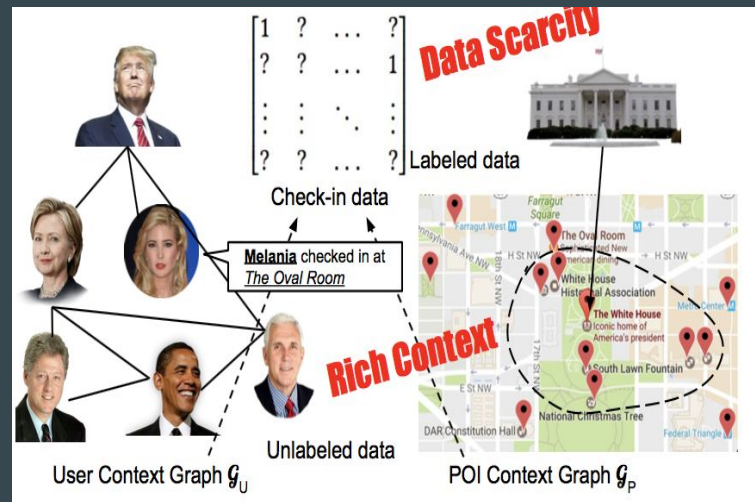
Majority of the research is done with Yelp reviews dataset

Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text

- By researchers of Stanford
- Understand the tastes of users as well as the product properties to predict user rating
- Eg: To predict if a user will like Harry Potter, we need to identify that the book belongs to fantasy genre as well as the user's interest in fantasy
- “Hidden Factors as Topics” model: Variant to LDA which learns rating parameters (product properties) and review parameters (topics discussed in the review) together.
- Understands user interests, discovers clean genres and identifies impactful reviews

Bridging Collaborative Filtering and Semi-Supervised Learning: A Neural Approach for POI Recommendation

- By researchers of UIUC
- Point of interest / location recommendation
- Challenges:
 - Data Scarcity : Eg: Check-in info is scarce
 - Various context: Users preference is based on dynamic factors. Eg: location, time
- Semi-supervised learning with heterogeneous context graphs
- Neural networks to jointly learn user context, POI context and user-POI interactions



Aspect Based Recommendations: Recommending Items with the Most Valuable Aspects Based on User Reviews

- By researchers of NYU
- Aspect based sentiment analysis: “Food is good but service is so bad”
- Identifies valuable aspects of product and predicts not just the overall sentiment but also the sentiment towards each aspect.
- Sentiment Utility Logistic Model (SULM): Logistic regression model to predict the rating of the review along with impact of each aspect towards overall rating.
- Objective function:

$$Q(\theta) = -\alpha \cdot l_r(R|\theta) - (1 - \alpha) \cdot l_s(S|\theta_s) + \frac{\lambda_r}{2} \cdot \|\theta_r\|^2 + \frac{\lambda_s}{2} \cdot \|\theta_s\|^2$$

- Training using Stochastic Gradient Descent

Survival Analysis for Modeling Critical Events that Communities May Undergo in Dynamic Social Networks

- By researchers of University of Quebec
- Survival analysis: Analyze duration of time until event happens. Eg: Remission duration of a disease
- Critical events for communities in social network (expand, merge, split, constant)
- Predict the next critical event based on how the communities in the networks changes temporally
- Used user network of Yelp

Finding local experts from Yelp

- By researchers of UIUC
 - Model prominent location of users using Gaussian mixture models
 - Identify topical authority using user-specific features (reviews, friends, categories for which reviews are posted)
-



THANK YOU :)
QUESTIONS?