Previous work:

Yelp Dataset

## Team 47: Yelp Rating Analysis

## Function:

Vertical search engines, specific topics

Track user experience lead to greater customer traffic and sales for the business

Business classify their reviews and filter to identify weakness, provide business owner with categorized feedback

Areas of opportunity and growth, better understand their client(?)

How reviews help business?

every star in a review leads to a 5-9% additional revenue

## Implementation:

Two classifications: goods and service=> manually labelled => cause some error?

Tokenization: bag of words, NLTK

Classification: naïve-Bayes, LinearSVC

Precision, recall, and F1

Sentiment analysis: NLTK, scikit-learn libraries

Visualization: data distribution

## Future work:

More categories for the classifier

Specialize categories by business sector

Target specific wants to certain group of customers

Web browser tool

## Hooman Understanding the Sentiment of Restaurant Reviews Based on Text Content

## Function:

Sentiment analysis

## Implementation:

TF-IDF

Classification: LinearSVC, random forest, logistic regression, multinomial naïve Bayes

Visualization: word cloud, pie chart

## Team Rob Text based analysis to determine the usefulness of an online review

## Function:

Predictive model which indicate the usefulness of a review

For user to pick out the key comments that are useful regardless of the rating

## Implementation:

Time range: 2012 – 2017

Good identifier: number of words, number of reviews completed by the reviewer, high sentiment score

Useful score(not exactly)

Sentiment analysis: <https://cran.r-project.org/web/packages/sentimentr/README.html>

## Future work:

More reviews, additional analysis of the review text

Other Hotel Review Dataset

## Airbnb Review Analysis

Dateset:

Chicago, Composite ratings by Accuracy, cleanliness, location, check-in and value over 0-10

## Function

Language identification: NLTK => English

Normalized the positive sentiment scores

Bigrams

## Airbnb Search System With Text Mining

<https://hcxyeah.github.io/MCS-BetterSearchHotel/>

## Function:

Twitter app

Keyword extraction using TF-IDF

## Implementations:

Break into frequent phrases

Sort the list by sentiment score

Topic modeling

Limitation: small set of data

Other resources:

## Understanding Customer Frustrations in the Airline Industry with Aspect-based Sentiment Analysis

* Document-Level Sentiment Analysis
* Aspect-based Sentiment Analysis

http://blog.aylien.com/understanding-customer-frustrations-in-the-airline-industry-with-aspect-based-sentiment-analysis/

Papers

<https://pdfs.semanticscholar.org/dc8e/272b9f56b935b926603bdf42eb033c6e94a3.pdf>

Our dataset:

<https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe/home>

booking.com

Concerns:

analyzing the sentiment of text allows us to get an idea of whether a piece of text is positive, negative or neutral.

What if we’ve already known the comment is positive or negative

 a document-level sentiment analysis model would just look at the entire document and add up whether the overall sentiment was mostly positive or negative.

* Fit a regression model on reviews and score to see which words are more indicative to a higher/lower score
* Perform a sentiment analysis on the reviews
* Find correlation between reviewer's nationality and scores.
* Beautiful and informative visualization on the dataset.
* Clustering hotels based on reviews
* Simple recommendation engine to the guest who is fond of a special characteristic of hotel.

Yelp Sentiment Analysis

identify, extract, quantify, and study affective states and subjective information

Previous work

Wiki:

The following is a list of a few open source sentiment analysis tools.

1. [GATE](https://en.wikipedia.org/wiki/General_Architecture_for_Text_Engineering) plugins
   1. SEAS(gsi-upm/SEAS)
   2. SAGA(gsi-upm/SAGA)
2. Stanford Sentiment Analysis Module (Deeply Moving: Deep Learning for Sentiment Analysis)
3. LingPipe (Sentiment Analysis Tutorial)
4. TextBlob (Tutorial: Quickstart)[[3]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-3)
5. Opinion Finder (OpinionFinder | MPQA)
6. Clips pattern.en (pattern.en | CLiPS)

Open Source Dictionary or resources:

1. SentiWordNet
2. Bing Liu Datasets (Opinion Mining, Sentiment Analysis, Opinion Extraction)
3. General Inquirer Dataset (General Inquirer Categories)
4. MPQA Opinion Corpus (MPQA Resources)
5. WordNet-Affect ([WordNet](https://en.wikipedia.org/wiki/WordNet) Domains)
6. SenticNet
7. Emoji Sentiment Ranking

<https://medium.com/@datamonsters/sentiment-analysis-tools-overview-part-1-positive-and-negative-words-databases-ae35431a470c>

<https://medium.com/@datamonsters/sentiment-analysis-tools-overview-part-2-7f3a75c262a3>

Stanford CoreNLP – Natural language software

https://stanfordnlp.github.io/CoreNLP/

Other languages:

Java:

1. [GATE](https://en.wikipedia.org/wiki/General_Architecture_for_Text_Engineering) plugins
   1. SEAS(gsi-upm/SEAS)
   2. SAGA(gsi-upm/SAGA)

<http://www.b-eye-network.com/view/9516>

<https://www.talkwalker.com/blog/best-sentiment-analysis-tools>

1) Useful software toolkits for processing text data or building text data applications.

2) Emerging new applications of text retrieval or analysis.

3) New techniques for text retrieval or analysis.

Project proposal:

* What is the function of the tool?
* Who will benefit from such a tool?
* Does this kind of tools already exist? If similar tools exist, how is your tool different from them? Would people care about the difference?
* What existing resources can you use?
* What techniques/algorithms will you use to develop the tool? (It's fine if you just mention some vague idea.)
* How will you demonstrate the usefulness of your tool.
* What is exactly the function of the tool that you would like to develop?
* Is the envisioned tool really useful? Who will benefit from such a tool? Note that the usefulness of the tool is an important grading factor, so make sure that you invest your effort on developing a really useful tool. It's fine if the tool is a very specialized tool, or only useful for a few people.
* Does this kind of tools already exist? If similar tools exist, how is your tool different from them? Would people care about the difference? Why is this difference important?
* Do you have some rough idea about how the target function might be achieved? (The instructor and TAs can help you answer this question, so please feel free to propose problems that you don’t yet have a clear idea about how to solve it)
* How do you plan to show people that the tool to be developed is indeed useful? (You could make a demo of its function and/or report some quantitative evaluation figures.)
* What is the minimum goal to be achieved during this semester? (Try to drop everything non-essential and only keep the function that is truly novel.)

Tourism, experience economy and understand the emotional and experiential elements of tourist activities

How big data can be used in tourism context

Thoughts:

Paper Work: