CREDIBILITY OF REVIEWS

By Metin Senturk

GOAL AND IMPORTANCE

In a customer's perspective, reviews are a key factor while **choosing** a service from a business.

In a business' perspective, customer opinion may enlighten more **insights** about how well a business operating.

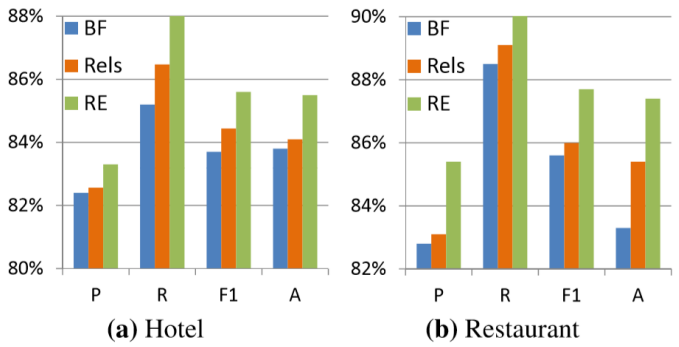
For both, the **quality** of the content matters in order to seek the need from reviews.

"Objective is to find characteristics of fake/ spam reviews."

LITERATURE REVIEW

Previous studies divided into two main studies.

1. Feature generation through linguistic and behavioral characteristics
2. Domain-centric approach. Hotels, restaurants, etc.



(a) Hotel (b) Restaurant

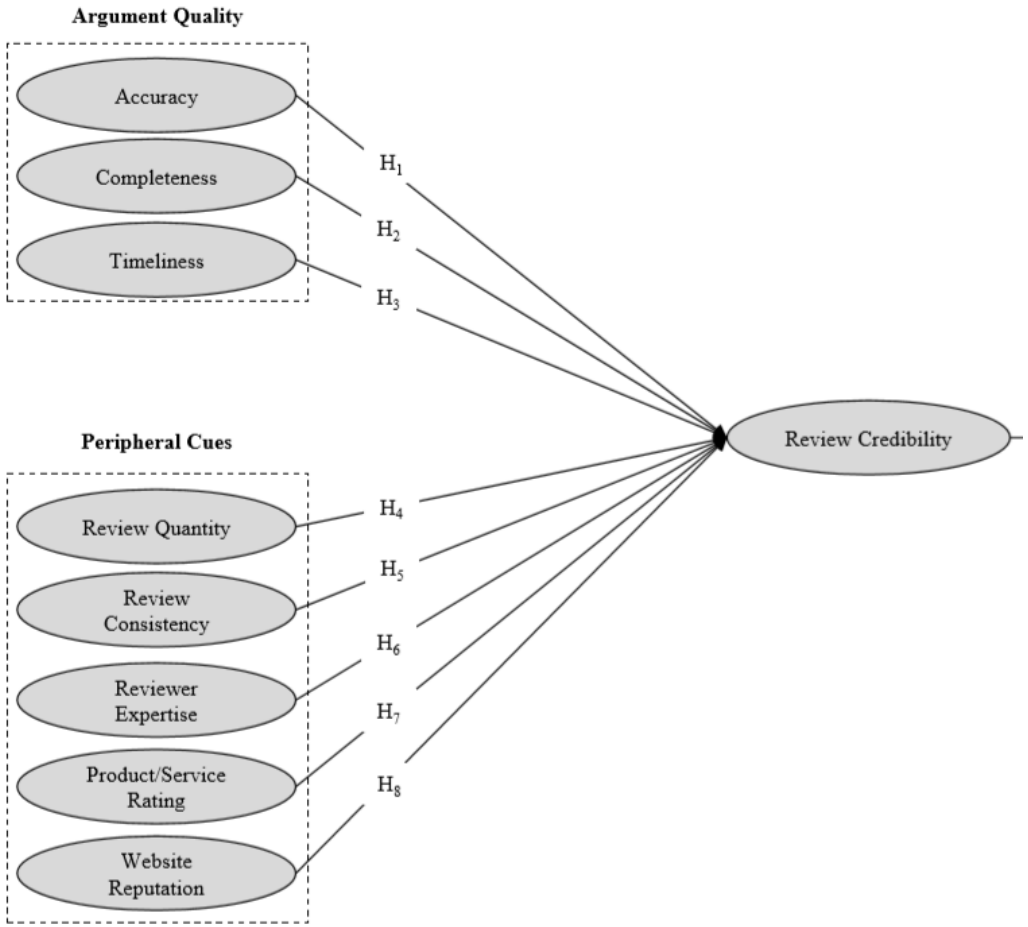


Figure 1: Conceptual Research Model

ANALYZING REVIEW CREDIBILITY PROCESS

EDA

DATA

DATA CLEANSING & PROCESSING

RESULTS

MODELS

DATA AND STACK




6,685,900 reviews

192,609 businesses

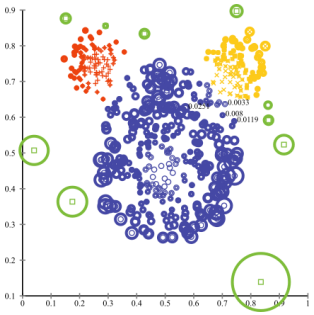
200,000 pictures

10 metropolitan areas

1,223,094 tips by 1,637,138 users
Over 1.2 million business attributes like hours, parking, availability, and ambience
Aggregated check-ins over time for each of the 192,609 businesses

Python

Relational Databases



Machine Learning
- Outliers
- Clustering

DATA CLEANING AND PREPROCESSING

Cleaning:

- Users less than 10 reviews
- Users with more than 10 friends
- Users with reviews ratings greater than/ less than
- Reviews longer than 600 words
- Reviews with more than 10 funny/cool/useful tags

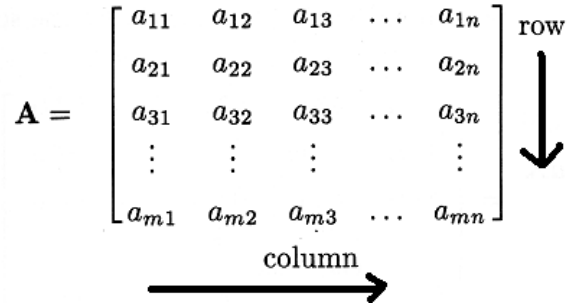
Preprocessing:

Most features are mostly continuous, or categorical.

- Transformation with one-hot-encoding & min-max standardization

Reduced the dataset down to

- Users —> 166,781 Users
- Reviews —> 1,245,408 Reviews



EXPLORATION – DAILY REVIEW COUNTS

More than 2 reviews

Less than 2 reviews

EXPLORATION – REVIEW POSITIVITY

Users with more than 80% positive reviews

Users with less than 80% positive reviews

EXPLORATION – GIVEN AVERAGE STAR RATING

Rating being more than 4

Rating being less than 4

RESULTS

Using behavioral features of users, linguistic characteristics of reviews can shed light on credibility.

Using clustering algorithms, fake reviews can be detected.

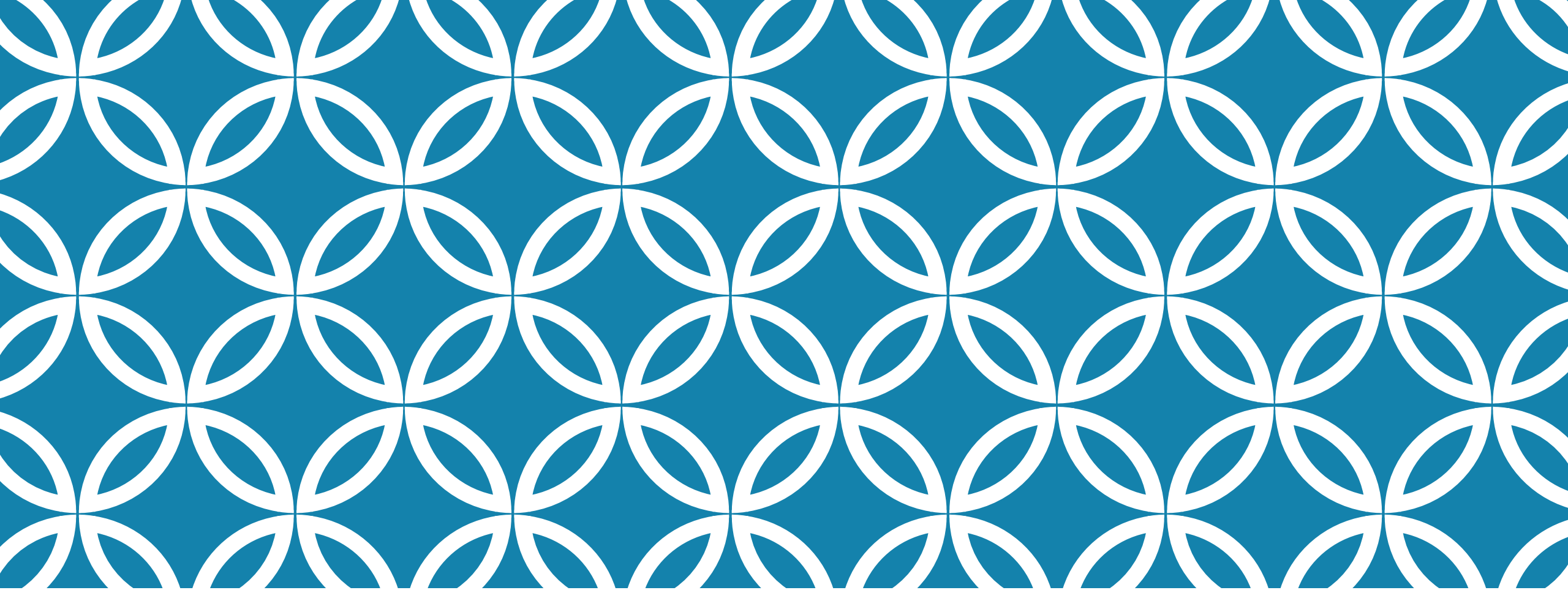
DISCUSSION

Observations and setbacks

- Having no ground-truth and struggle with visualization
- Complexity of relationships
- Computational struggle

Future Work

- ✓ Many example of good reviews, therefore, training a good review detector and reversing the set.
- ✓ Cleaning documents from most common words (ones with less words or none could be fake).
- ✓ Word embeddings integration



THANK YOU!

Feel free to contact with my regarding any questions about the presentation, project, and studies.