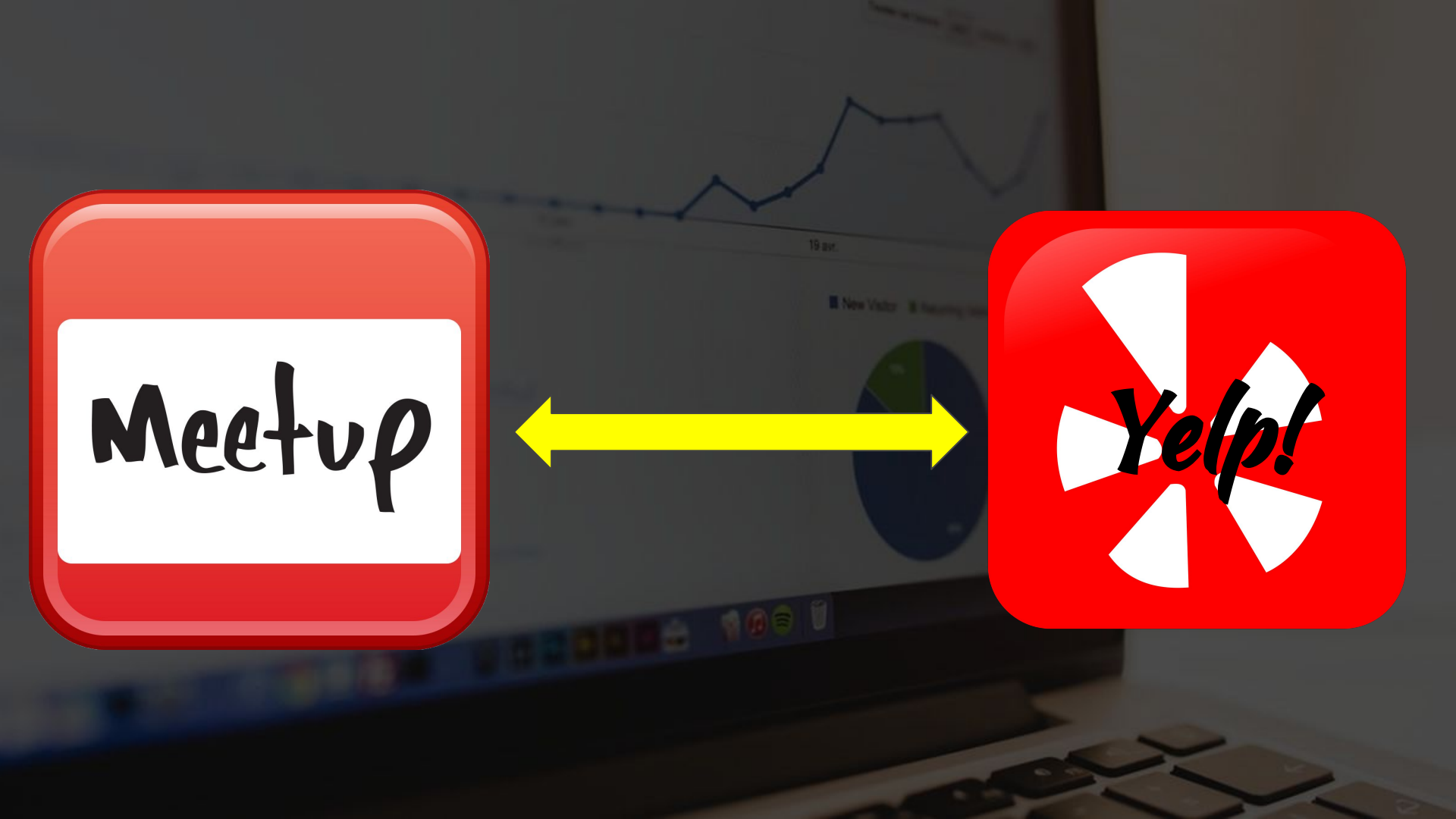
Matching Venues Across Social Networks Using Word Embedding

End Semester Project Presentation

Group No 14


Meetup
Yelp!

Problem Statement

Map past events organised by groups on Meetup and use description and attributes of the venues of these events fetched from Yelp to match and recommend similar venues for future events.

Requirement

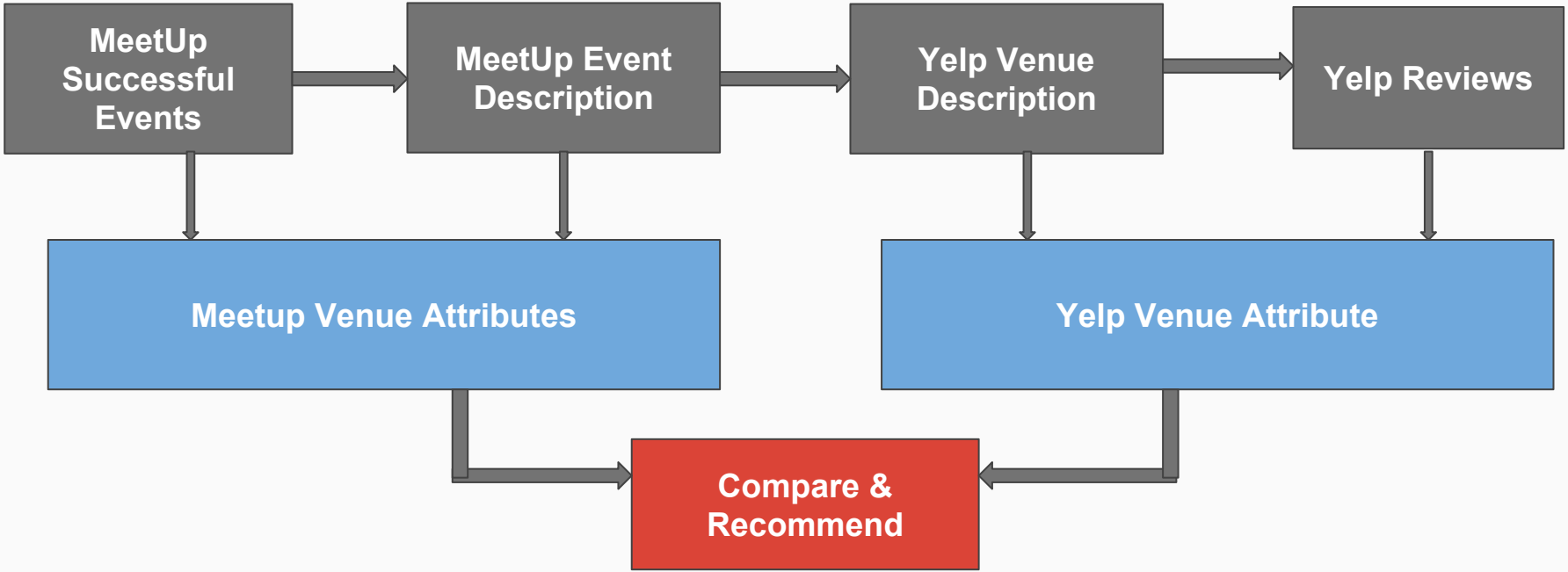
Meetup does not provide details about venue attributes where event is organised.

These details fetched from Yelp using name and location similarity match.

Attributes and Review returned by Yelp used to match venue similarity using word/paragraph vectors

Recommend Top N similar venues for future events

Solution Model



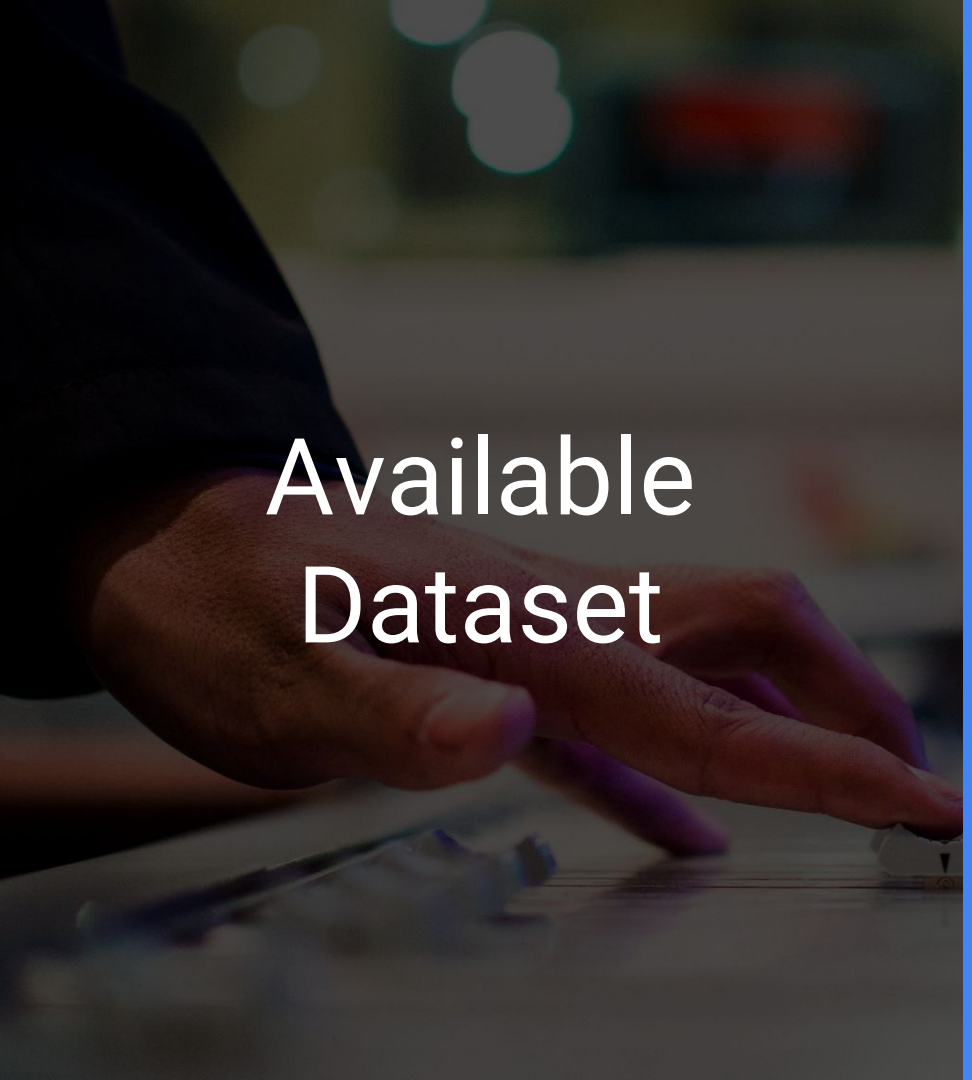
Requirement

Meetup does not provide details about venue attributes where event is organised.

These details fetched from Yelp using name and location similarity match.

Attributes and Review returned by Yelp used to match venue similarity using word/paragraph vectors

Recommend Top N similar venues for future events



Available Dataset

Meetup Event Dataset

Meetup Venue Dataset

Yelp Datasets (Venue Description & Reviews) - Not mapped to Meetup Data, Crawler Designed and Implemented

Approach

Step 1

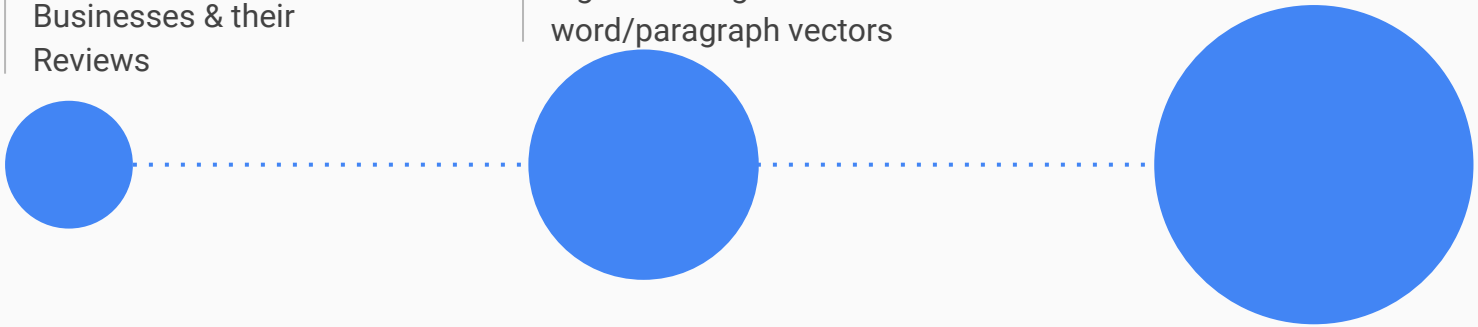
Use Meetup Venue Data and Crawl for similar Yelp Businesses & their Reviews

Step 2

Use the Business attributes & reviews to train GLOVE algorithm to generate word/paragraph vectors

Step 3

Match the returned vectors and attributes using Machine Learning against others to recommend similar venues



Generating Yelp Dataset

- Yelp Dataset Not Available for venues of Meetup
- Data crawled over 2 month time using Yelp API(3,53,299 Venue Details and 3,09,963 Reviews Crawled)
- Data Parsing and Removal of Inconsistent Data - Challenge
- Dataset does not contain enough attributes - Requirement of scraping Web Pages for additional details

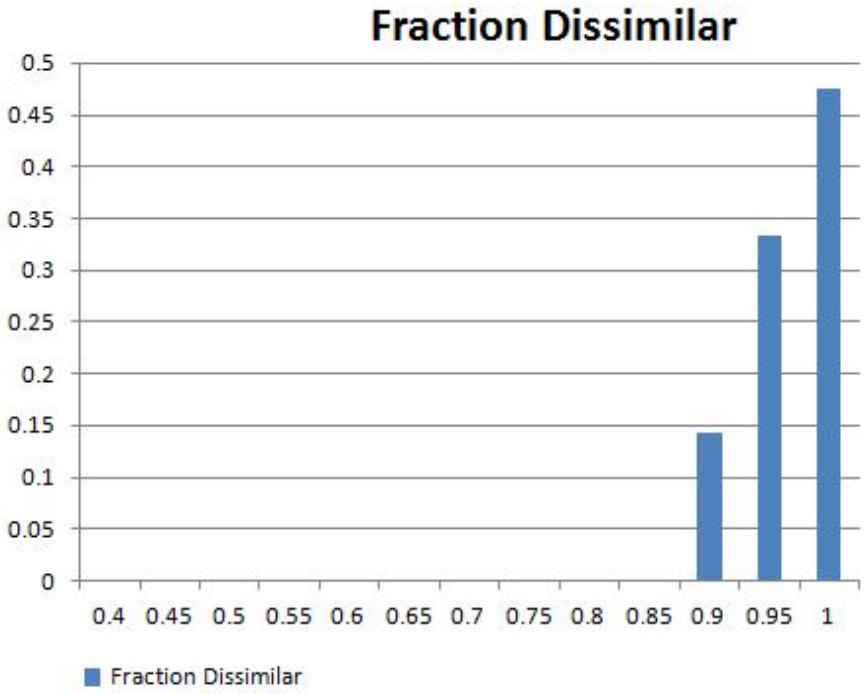
Generating Vectors

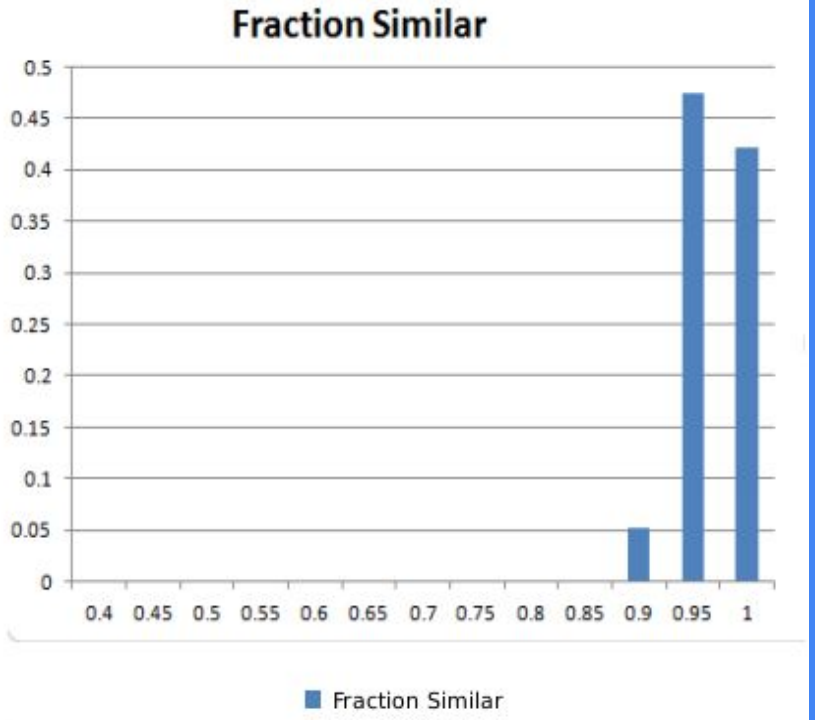
- Averaging Technique
  Used Pre-trained GloVe vectors
- GloVe Algorithm
  Training Data
    : Wiki-Dump (150 MB,500 MB)
    : Venue Reviews
    : Venue Reviews + Business Attributes

GloVe: Global Vectors for Word Representation

GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.

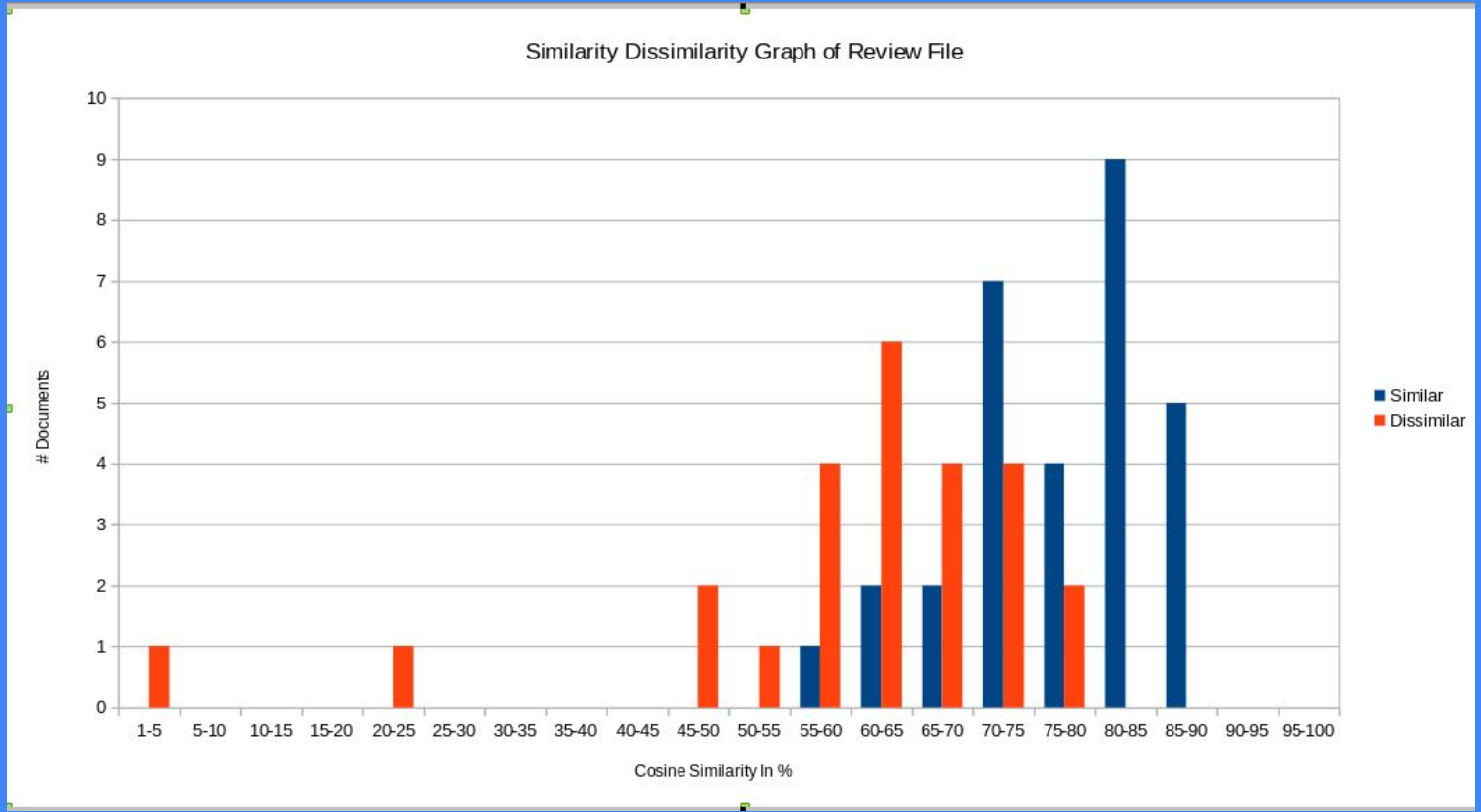
Average Technique Results



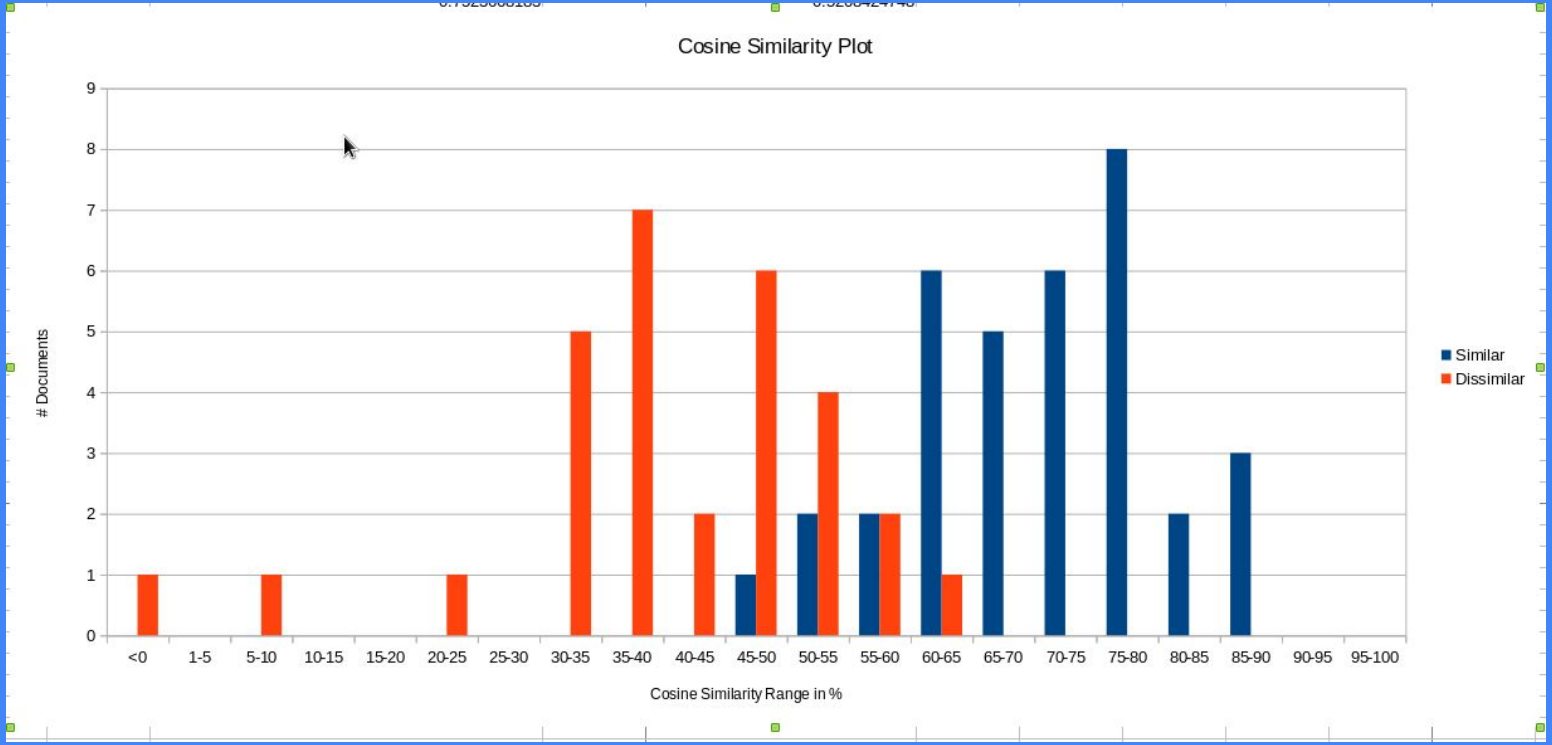


X-axis - Cosine similarity
Y-axis - No. Of similar Documents

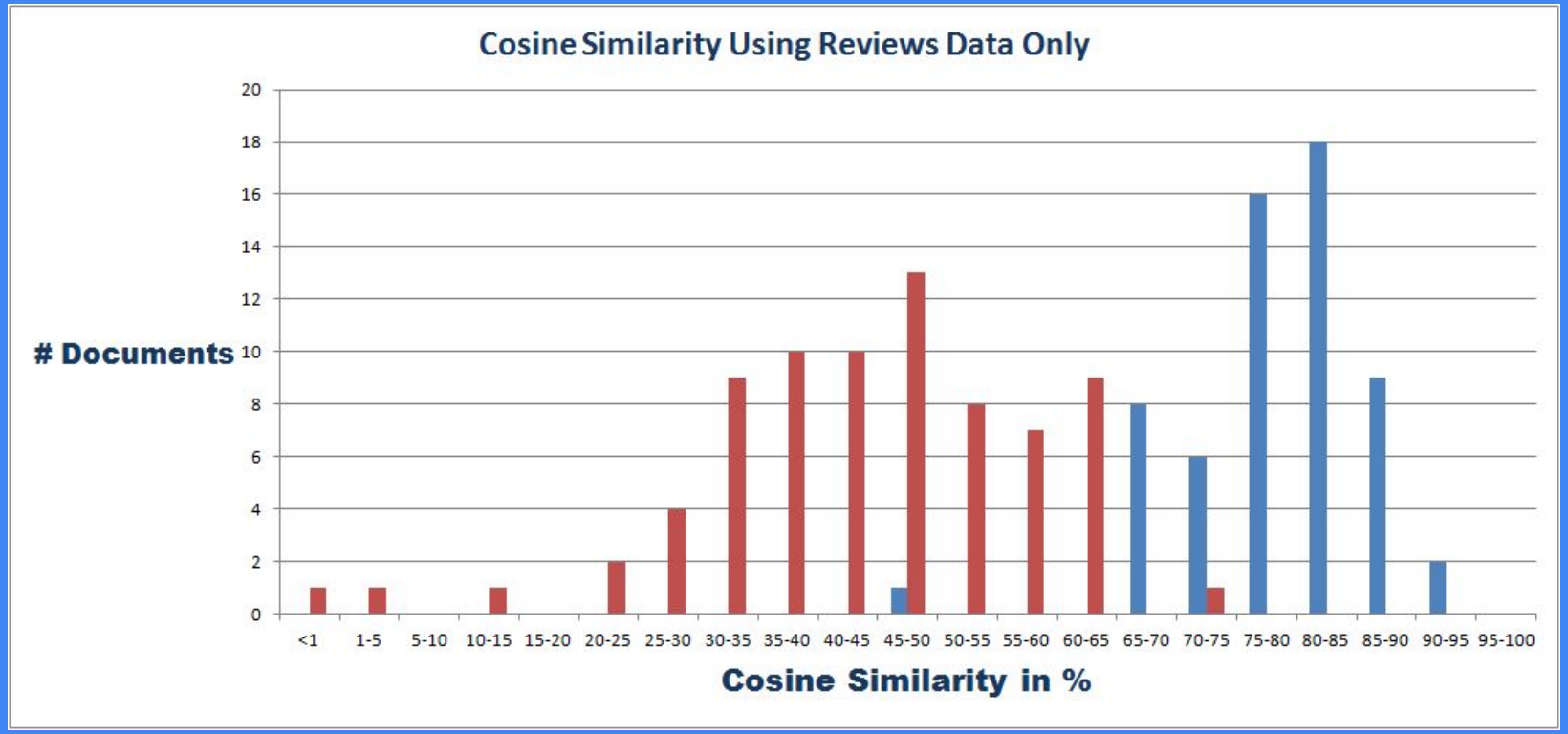
Training On Review+Business



Trained And Tested On Review + Business Attributes



Training on Reviews



Future Work

Improve Mapping Method to increase recall probability

Testing of Prediction Model and recommending most common venues.

Individual Contribution

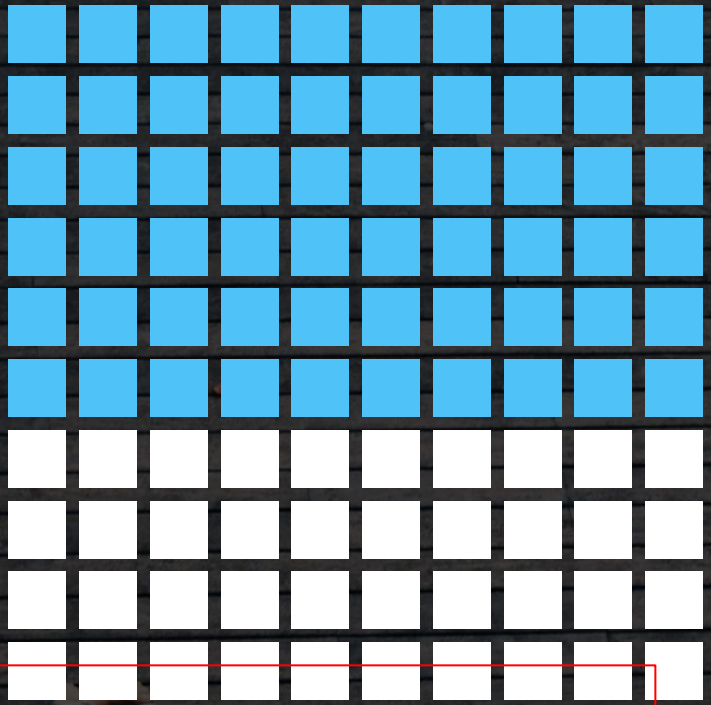
Dataset Generation & Cross Mapping

Saurabh Singh
Abhipsa Basu

Features & Vector Extraction & Mapping

Prishni Rateria
Satinder Kaur
Preeti Meena

Thank You!



Project Repository@Github
https://github.com/saurabhima/NLP_PROJECT_CODE