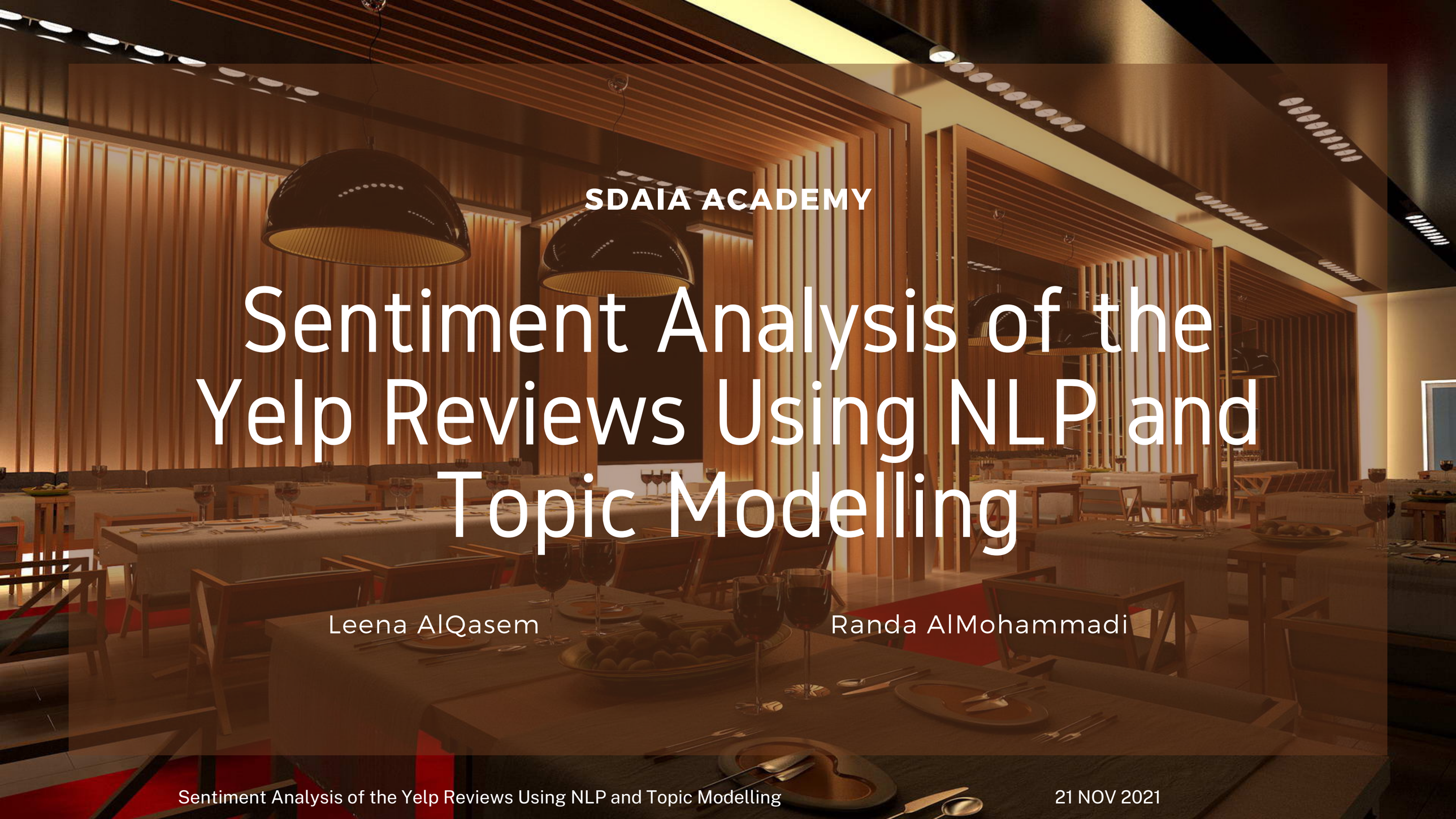


SDAIA ACADEMY

Sentiment Analysis of the Yelp Reviews Using NLP and Topic Modelling

Leena AlQasem

Randa AlMohammadi

TABLE OF CONTENT

INTRODUCTION

Brief Description:

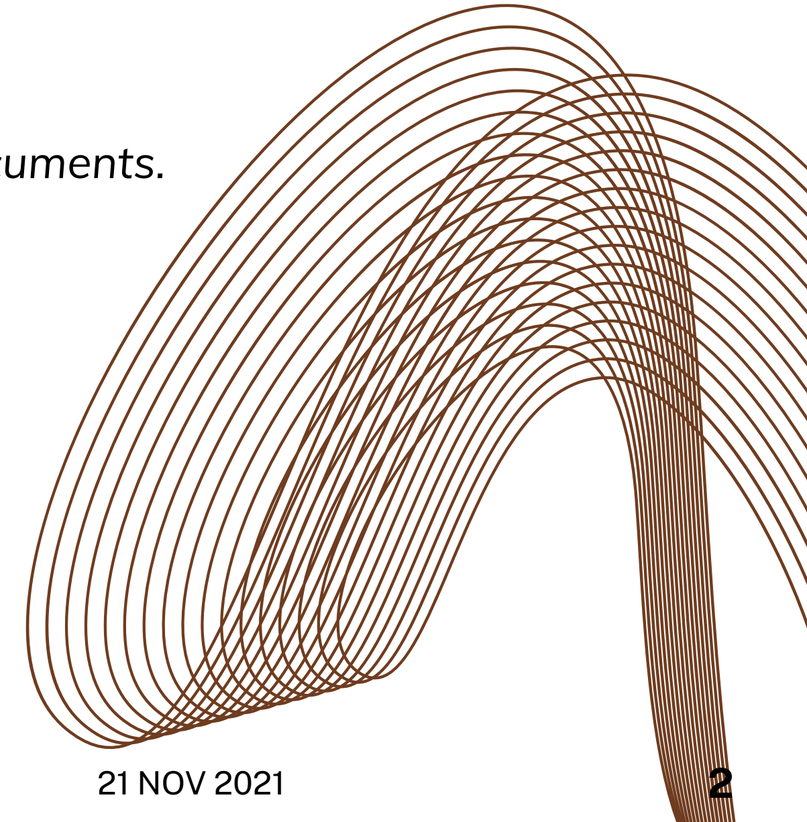
- *YELP publishes crowd-sourced reviews about local businesses.*

- *NLP computer program to understand human language.*

- *Topic modeling analyzes text data to determine cluster words for a set of documents.*

Project Objectives:

- Understanding large thinking and emotions of restaurant reviews



DATASET

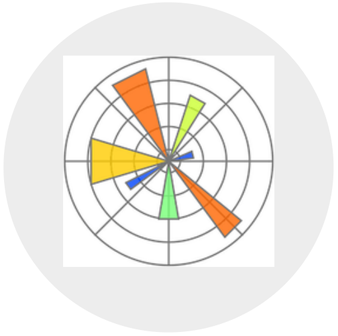
- *Yelp Dataset is on customers reviews of different restaurants were obtained from Kaggle*

6911 ROWS
10 COLUMNS

Split data into positive and negative for analysis

Columns:
Text customers' Reviews, Ratings, Business ID, Review ID, etc.

TOOLS AND LIBRARIES



METHODOLOGY

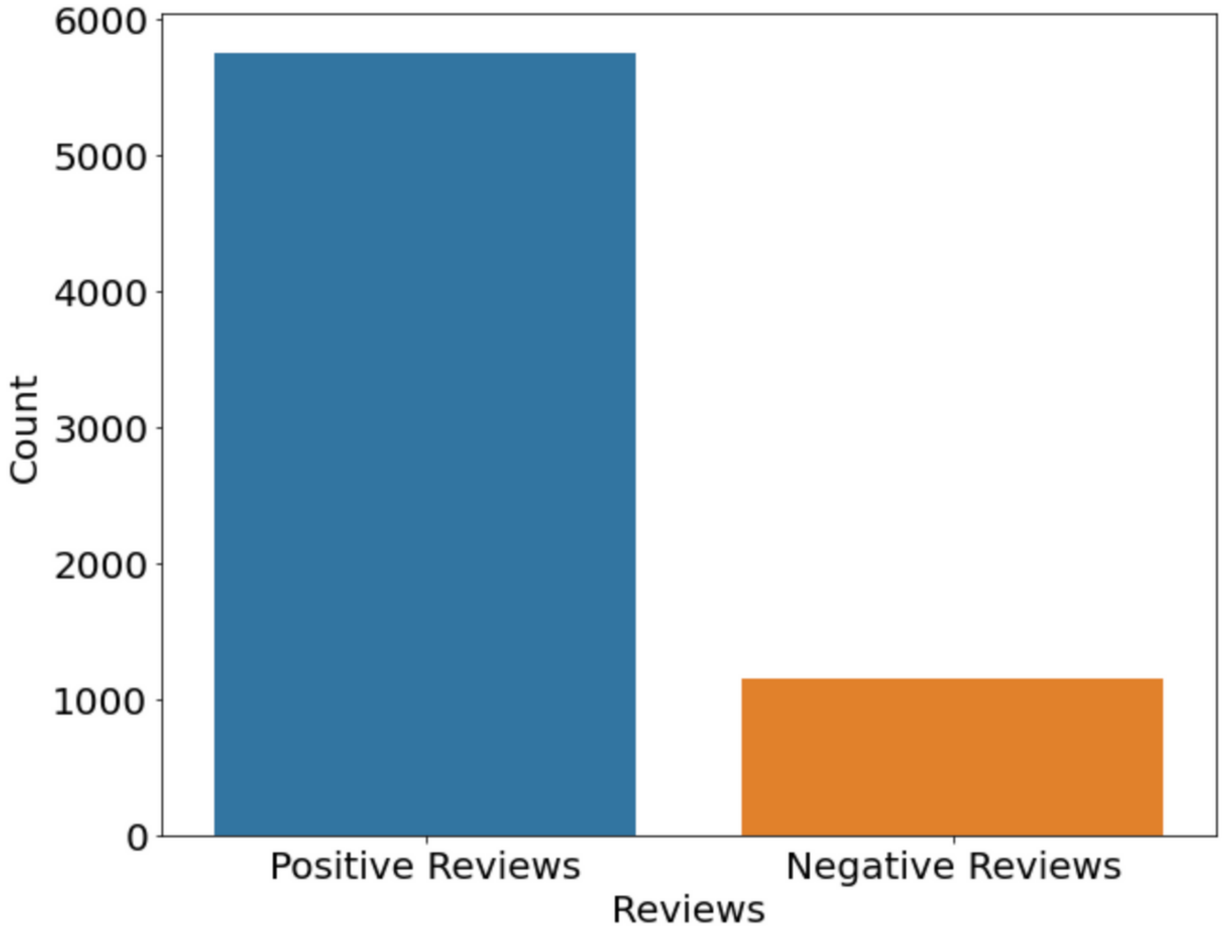
Gathering data

Text Preprocessing (NLP)

EDA

Topic Modeling

EXPLORATORY DATA ANALYSIS



Visualize the Number of both the Positive and Negative Reviews.

The stars with 3 and above are considered as positive reviews AND stars with 2 and 1 are considered as negative.

EXPLORATORY DATA ANALYSIS

Visualize positive words

Visualize negative words

TEXT PREPROCESSING

NLP

01 CLEAN TEXT

Remove: Punctuation, lowercasing, non-alphabetic, URLs, stopwords,
apply: lemmatization

02 TOKENIZE

TF-IDF, not only focuses on the frequency of words present but also provides the importance of the words.

SPLITTING DATA

03 SPLIT DATA

Split data into 90% Training and testing 10%

TOPIC MODELING

LDA	n_components=5 num_words=15
LSA	TruncatedSVD(5) no_top_words = 10
NMF	n_components=3 num_words=10 (positive) num_words=6 (negative)

The Topics for positive & negative datasets are:

Atmosphere, Food, Service

RESULTS AND ANALYSIS

	atmosphere	service	Food	dominant_topic
Doc0	0.138	0.005	0.016	0
Doc1	0.079	0.000	0.040	0
Doc2	0.088	0.006	0.055	0
Doc3	0.082	0.015	0.050	0
Doc4	0.060	0.008	0.135	2
Doc5	0.125	0.000	0.065	0
Doc6	0.046	0.000	0.021	0
Doc7	0.044	0.000	0.093	2
Doc8	0.086	0.000	0.000	0
Doc9	0.117	0.000	0.065	0

0	**385**
2	**142**
1	**49**

Experiments the NMF model on the Positive Reviews (Testing)

Comparing the topics with the documents. Found out that our model predicts 70% of topics right (7 out of 10)

RESULTS AND ANALYSIS

	Food	Service	atmosphere	dominant_topic
Doc0	0.070	0.232	0.196	1
Doc1	0.174	0.117	0.000	0
Doc2	0.106	0.357	0.037	1
Doc3	0.000	0.145	0.087	1
Doc4	0.098	0.372	0.051	1
Doc5	0.044	0.349	0.014	1
Doc6	0.054	0.000	0.285	2
Doc7	0.295	0.000	0.138	0
Doc8	0.000	0.264	0.165	1
Doc9	0.092	0.121	0.284	2

2	**55**
1	**40**
0	**21**

Experiments the NMF model on the Negative Reviews (Testing)

Comparing the topics with the documents. Found out that our model predicts 80% of topics right (8 out of 10)

CONCLUSION

NLP

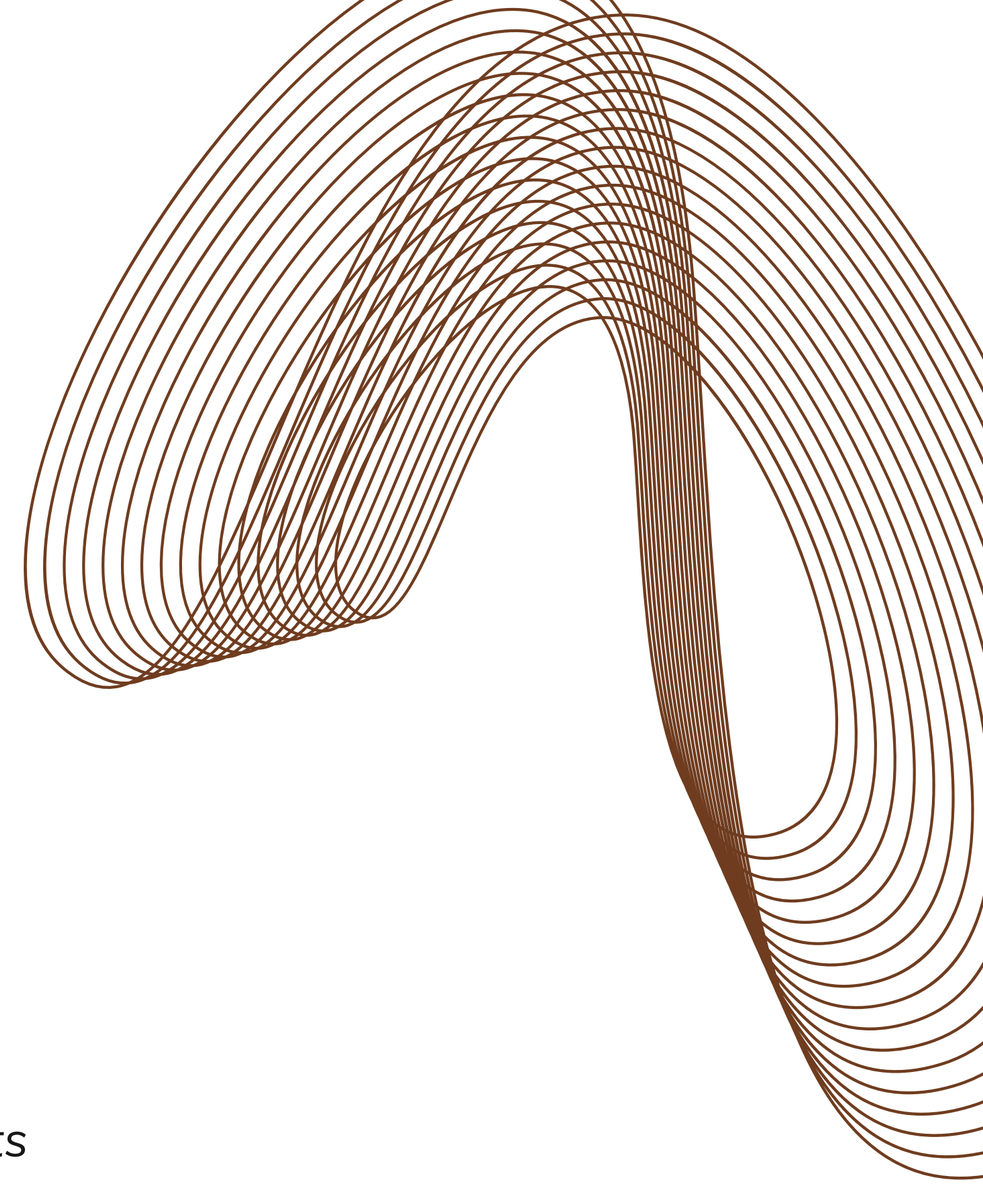
- *Prepare text data for analysis*

Topic modeling

- *NMF gives most meaningful topics.*

Recommendations:

- *Sentiment analysis for one business.*
- *Understanding customers satisfaction level.*
- *Decision-based on customer reviews for improvements highlighting strength and weaknesses from a customer point of view.*



THANK YOU!

Any Questions?