MULTI-MODAL TOPIC SENTIMENT ANALYTICS

FOR TWITTER

A THESIS IN
Computer Science

Presented to the Faculty of the University
of Missouri-Kansas City in partial fulfillment
of the requirements for the degree

MASTER OF SCIENCE

By
SIDRAH JUNAID

B.CIT, NED University of Engineering & Technology, Karachi, Sindh 75270, Pakistan

Kansas City, Missouri
2018

MULTI-MODAL TOPIC SENTIMENT ANALYTICS

FOR TWITTER

Sidrah Junaid, Candidate for the Master of Science Degree

University of Missouri-Kansas City, 2018

ABSTRACT

Sentiment analysis has proven to be very successful in text applications. Social media is also considered a quite rich source to get data regarding user's behaviors and preference. Identifying social context would make the sentiment analysis more meaningful to the applications. Due to the limited contextual information in social media, it would be quite challenging to conduct context-aware sentiment analysis with social media. Promising frameworks such as CoreNLP, Text Blob, and Vader have been introduced to identify sentiments in the text. However, it seems to not be adequate to contextual sentiment analysis in social media like Twitter.

In this thesis, we present a contextual sentiment framework that is designed to leverage the power of the multiple models in the social context. The framework aims to classify contextual sentiment from the Twitter data as well as to discover hidden trends and topics (context) using topic modeling techniques like Latent Dirichlet Allocation (LDA). We have focused on the mismatch cases among multiple models in which different experts (models) have different opinions on social media sentiments. We have identified the five mismatch types in the social sentiment through the analysis of diverse experiments (human-machine model, and machine-machine model). We have implemented the mismatch

detection among the three models (i.e., Vader, Text Blob, and CoreNLP) and automatically corrected them by applying semantic rules to sentiment models. We compared our approach against a traditional single model approach concerning a performance metric (accuracy) and Kappa (evaluating consensus among multi-models) on three benchmarks datasets and our dataset we collected from a health dieting domain. The proposed framework showed notable performance improvement in comparison with the traditional one concerning both evaluation metrics.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Computing and Engineering, have examined a thesis titled “Multi-Modal Topic Sentiment Analytics for Twitter” presented by Sidrah Junaid, candidate for the Master of Science degree, and hereby certify that in their opinion, it is worthy of acceptance.

Supervisory Committee

Yugyung Lee, Ph.D. (Committee Chair)
Department of Computer Science Electrical Engineering

Baek-Young Choi Ph.D.
Department of Computer Science Electrical Engineering

Ye Wang, Ph.D.
Department of Communication Studies

Contents

LIST OF ILLUSTRATIONS

LIST OF TABLES

Table Page

ACKNOWLEDGMENTS

I feel very fortunate for working under the supervision of Dr. Yugyung Lee and would like to thank her for the valuable guidance and immense support throughout the research work as my advisor. Her vast experience, unparalleled knowledge, agile and prompt feedback coupled with smart ideas have helped me immensely in putting up the whole work. She is very patient in listening to all the new ideas, pragmatic in giving suggestions and always helps me in doing the reality check. Her amazing energy and enthusiasm always motivate me to go the extra mile. It has been an honor to work with her on many projects besides the thesis.

I would also like to thank the University of Missouri-Kansas City, for providing me the perfect environment to research. It provided me with many opportunities to support myself and world-class facilities to conduct research with the finest machines available without which the thesis work could not have been accomplished.

Finally, I would like to thank my family and friends who always encourage me, gave me valuable suggestions throughout the research and made sure that I pursue my dream without any problems.

CHAPTER 1. INTRODUCTION

Natural language processing (NLP) is a branch of artificial intelligence that supports computer to understand, infer and engineer human language. NLP covers many domains, including computer science and computational linguistics to fill the space between human communication and computer understanding. By extending the capabilities of Natural Language Processing, we can identify the contextual sentiment in the sentences.

Sentiment analysis is considered to be one of the important issues today. It is difficult to find the contextual sentiment of a text. The primary job of sentiment analysis is too fast-paced the process of opinion extraction from the given subject. The subject can be an excerpt from the written text, debate or day to day conversation. In sentiment analysis, we also evaluate the positive and negative intensities of symbols and words. Sentiment analysis helps to improve customer services, planning marketing strategies and manufacturing quality products.

In social media millions of active users express their opinions and interact with each other daily. Such users content in the form of posts or tweets provides a huge amount of useful information if analyzed carefully. Therefore, the data streamed from social media such as Twitter, Facebook or Instagram is so rich for researchers to perceive the users' social behavior by applying sentiment analysis on it. Twitter is one of the most significant ones among all the micro-blogging services. A huge amount of user-generated online content is freely available to the real-time monitoring of public sentiment.

The fusion of topic and sentiment has been used for sentiment analysis of a text. Topic modeling is an unsupervised statistical machine learning technique. The purpose of the topic modeling is to discover the abstract topics from the collection of documents. It is different from the rule-based approach where we use the dictionary or lexicon to search keyword. The topic modeling is used to extract and find the group of words called “topics” in huge text clusters. There are several approaches available for finding out topics from text corpus namely Term Frequency-Inverse Document Frequency and Non-Negative Matrix Factorization technique. Latent Dirichlet Allocation (LDA) is the most popular example of the topic model and is used to classify text in a document which is assumed to be a mixture of topics, to a particular topic.

Topic modeling is used to extract the topics from the unstructured twitter dataset to find the eating trends of social media user. Our approach to topic modeling is based on the framework of Latent Dirichlet Allocation (LDA) [1]. In LDA, each document is the combination of various topics where each document has a set of topics assigned by LDA. LDA is a notion of the pLSA model, which is corresponding to LDA under an unvarying Dirichlet prior distribution [2].

Online platform like social media and blogs widely used by patients and doctors to express their opinions and suggestions on healthcare issues. Moreover, sentiment analysis is also performed on food reviews and to find the food habits of the consumers. The sentiment of food consumer towards different food category is identified by investigating different sentiment cases. Particularly, in social media people are widely expressing their like dislike

towards food. Moreover, the analysis of the data is useful to decipher the eating trends and like/dislike of people regarding food.

1.1 Problem Statement

To find out the sentiment of the text at first glance is just look like a text classification problem, but when we deep dive into it, we will find out that many challenging factors involved to find an accurate sentiment. A lot of work has been done to find the sentiment in the text. However, most of the work is not efficient enough to find the sentiment in the short text. Similarly, the sentiment assignment to the Twitter text is not correct all the times, in some places, the sentiment areas like sarcasm, ambiguity, and presence of emoji, acronym and slang detection are missing.

In the sarcastic text, the negative sentiment is conveyed using positive words. It is very difficult to understand sarcasm in the text without a clear understanding of the topic, situation, and environment. Due to the continuous variation in the sentence, it is very difficult to understand sarcasm even for humans.

Positive, Negative and Neutral sentiments are the common sentiment present in the text. Some of the existing tools have strengths and weaknesses in identifying correct sentiment in the text. Some tools are proficient in identifying positive sentiment in the text and some are good in finding out the negative sentiment. Such kind of cases occurs when the context of the sentence misled by the actual meaning of the individual word in the sentence.

Ambiguity in a sentence is another blocker in sentiment analysis. Ambiguity in a sentence makes it almost impossible to assign a correct polarity to the sentence. The polarity

of the sentence strongly depends upon the context of the sentence. Due to ambiguity, some tools like CoreNLP assign negative sentiment to the neutral text.

In social media, people are extensively using emojis to show their emotion to a specific topic. They are considered as a handy and reliable indicator of sentiment. The current challenge of extracting knowledge from the unstructured text now includes the novel approach of emoji analytics. Emojis analytics can uncover the sentiment of text in a way that can not be done in text analytics. Emojis are cartoon pictures including facial expression, objects, people, animals, weather and so on. Text face or emoticons are the combinations of characters from the keyboard to create a pictorial icon that portrays emotion and sentiments. Some of the existing tools are unable to identify a pictorial representation of sentiment. Some of the sentiment tools can only recognize text face emojis.

1.2 Proposed Solution

The proposed solution focuses on the essential challenges in the evaluation phase of the text sentiment by improves the sentiment assignment of existing tools such as Text Blob, CoreNLP, and Vader by implementing contextual sentiment framework. Here, we identified five mismatched cases in Social media text specifically Twitter and leverage the capabilities of existing tools by adding semantic rules in it. These mismatch cases identified by analyzing the sentiment assignment of different tools. Identified mismatched cases are Negative, Neutral, Positive, Sarcasm, Ambiguity and Emoji presence in a sentence.

These mismatch cases are identified by a various experiment done on benchmark dataset and the Twitter food dataset among Text Blob, Vader, CoreNLP. These experiments are based on Machine-Machine, Human-Machine, Human-Human models.

The comparison of our proposed approach with the traditional approach is evaluated by kappa measurement to validate the improved performance. The measurement is used to differentiate the performance of multi-modal sentiment approach against a traditional single model approach concerning on three benchmarks datasets and our dataset we collected from a health dieting domain.

In addition to that, we used the concept of Topic discovery to find the hidden pattern and context in Twitter data. To find the optimal number of topics we evaluated different metrics like perplexity and topic coherence (U-Mass). Here, we identified the topics with relevant terms by calculating coherence score of each topic. We identified the different topics and based on their coherence score we identified the relevant one. The visualization of various topics combinations generated by using pyLDAvis to identify the frequency of terms in a specific topic.

This thesis also includes a food mood classification model as a case study to find the healthy/unhealthy food sentiment of food items addressed in Twitter data. The lexicon-based model contains various food items and on the based on the frequency of the healthy/unhealthy food names appear in a text ,it calculates the compound score. The compound score is helpful is categorizing tweet as Healthy, Unhealthy and Mixed(frequency of healthy and unhealthy items is almost the same).The model is used for the categorization of topics as Healthy food Positive sentiment, Healthy food Negative sentiment, Healthy food

Neutral sentiment, Unhealthy food Positive sentiment, Unhealthy food Negative sentiment, Unhealthy food Neutral sentiment, Mixed food Positive sentiment, Mixed food Negative sentiment and Mixed food Neutral sentiment.

CHAPTER 2. BACKGROUND AND RELATED WORK

This chapter gives background information of various components used in the thesis and gives an overview of related work that will help in understanding this work better.

2.1 Related Work

2.1.1 Sentiment Analysis

In the paper (Hu, Han, et al.)[3], they proposed and implanted an approach to label the unstructured tweets. For that they used around 6K sampled tweets from 3 million drug abused tweets. Out of 6K tweets around 5K tweets are used as training data and around 1K tweets are used as testing data. In another paper (Jangid, Hitkul, et al.)[4], they proposed and implemented aspect model and sentimental model independently and combined both models to get the most out of the tweets. For that, they first divided the tweets into one of the aspects out of 4 Level aspects, in which their system supports multiple word embeddings like Stanford GloVe [5], Google Word2Vec [6]. After that, they implemented a sentiment model in which each tweet is classified as positive, negative and neutral and they also assigned the intensity scores ranges from -1 to 1. For their experiment, they used FiQA 2018 dataset. The dataset contains 435 annotated financial headlines and 675 annotated financial tweets as the training dataset.

There is some work-related in word embeddings construction such as the word co-occurrence matrix using dimensionality reduction [7], context learning with word proximity [8] and supervised learning[9]. For building the underlying representation for word and phrase embeddings, the performance has been boosted by NLP tasks such as syntactic parsing [10] and sentiment analysis [11]. Feeding behavior on either increased or reduced

food intake is caused by external, psychological stress [12] and may lead to either increased consumption of foods leading to obesity.

Nguyen et al. [13] analyzed 80 million tweets using Machine Learning algorithms and build a national neighborhood database for well-being and health behaviors. Machine labeled as well as manually labeled tweets had a high level of accuracy: 78% for happiness, 83%for food with the F scores 0.54 and 0.86, respectively. The higher the frequency of fast food tweets was posted from big cities. The frequency of tweets about fast food restaurants was higher than the frequency of fast food mentions. Greater state-level happiness and positivity toward healthy foods, assessed via tweets, were associated with lower all-cause mortality and prevalence of chronic conditions such as obesity and diabetes controlling for state median income, median age, and percentage white non-Hispanic.

Eichstaedt et al. [14] analyzed Twitter messages using a regression model to find markers of cardiovascular mortality at the community level through the analysis of psychological correlates of mortality and demographic, socioeconomic, and health risk factors (e.g., smoking, diabetes, hypertension, and obesity). Their results showed that the Twitter-based model for predicting mortality outperformed classical risk factor-based prediction models.

According to the report of the Centers for Disease and Control Prevention (CDC) [15], young adults were half as likely to have obesity as middle-aged adults. Adults aged 45-54 years had the highest frequency (35.1%) compared to adults aged 18-24 had the lowest self-reported obesity (17.3%). Thus, social media analysis may be useful for obesity awareness and promoting healthy eating.

Paul et al. [16] presented the Ailment Topic Aspect model to analyze Twitter messages and to measure behavioral risk factors by geographic region for some medical conditions like allergies, obesity, and insomnia. They concluded that Twitter can broadly applicable to public health research. Madan et al. [17] studied the relationship between social interaction and health-related behaviors such as diet choices or long-term weight changes using sensing and self-reporting tools. Scanfeld et al. [18] analyzed Twitter data about antibiotics and determined the categories of antibiotics such as cold and antibiotics, flu and antibiotics, leftover antibiotics. There are several works on sentiment analysis with food tweets. Sentiment analysis aims to determine whether a feature of a tweet is positive, negative, or neutral.

Poria et al. [19] presented an innovative method to extract features from textual and visual datasets using Deep Convolutional Neural Networks. With the use of those features and a multiple kernel learning classifier, they achieved the state of the art of multi-modal emotion recognition. Go et al. [20] trained on one million tweets in the food domain for sentiment analysis for Twitter and achieved an accuracy of 83%.

Food Mood [21] analyzed tweets for a food sentiment and social, and cultural aspects using Bayesian Sentiment classifier. Interestingly, they indicated constantly evolving food trends (e.g., meat or fast food sentiment). However, there is room for improvement in utilizing diverse data such as tweet messages and social images, to find relationships among food, sentiments, location, and obesity. In addition, real-time Analytics and interventions are not yet available for real-world applications.

Table 1:Comparative Evaluation of Sentiment Analysis Work

Paper	Objective	Dataset	Method Used	Results(Accuracy)
Leveraging Geotagged Twitter Data to Examine Neighborhood Happiness, Diet and Physical Activity(2016)	Twitter Analysis for wellbeing and health behavior	Machine and manually labeled tweets	Sentiment Learning Algorithm	78% happy,0.54 F-Score; 83% happy,0.86 F-Score
Psychological Language on Twitter Predicts County-level Heart Disease Mortality(2015)	Twitter Analysis of cardiovascular mortality with physiological factors	1,347 U.S. counties for AHD mortality rates; 50,000 tweeted words	Used regression model	Twitter-based model outperformed the prediction model
Deep Self-Taught Learning for	Label unstructured	6K sample tweets from	Used CNN & LSTM	85% accuracy

Detecting Drug Abuse Risk Behavior in Tweets(2018)	tweets related to drugs	3M drug abuse tweets		
Aspect-Based Financial Sentiment Analysis using Deep Learning(2018)	Sentiment analysis of text with respect to an aspect	FiQA 2018 (Financial Headlines and Tweets)	Used aspect and sentiment model	F-1 Score 69% , MSE 0.112

2.1.2 Topic Discovery

In this paper [22], a four generative model is proposed for the topic recommendation on micro-reviews in location-based social sites. As the micro-reviews are very short, the document pooling strategy is used to combine all micro-reviews based on venue-level or user-level. Here on a venue-level document and user-level document LDA is applied to derive VLDA and ULDA correspondingly. As both venue-level and a user-level view are essential for topic prediction, ALDA model is introduced which combines influences from both user and venue-level preferences. Moving one step further, ASLDA which consider sentiment orientation of user to identify a topic on micro-review. The model is evaluated on two real location-based social sites Yelp and FourSquare. Here 80% of a data set is used as training. ASLDA has significantly lower perplexity compared to remaining models. However, VLDA has better

performance than ULDA. Topic k=10 is used to generate the results. If PMI score of ASLDA is higher so the performance of the model is better.

In this paper [23], a Word-pair Topic-Sentiment WSTM model was introduced for short text. This model generates word-pair set for the entire corpus. The motive of the research is to introduce a weakly supervised sentiment-topic model to reduce the side effect of text sparse problem in short text. The model is applied to Chinese product reviews datasets. The dataset is split into 50% training data and 50% test data. WSTM use a How Net sentiment lexicon. Here, WSTM performs better than JST and ASUM and its accuracy is 65% on k=15 topics.

This paper[24] proposed a novel probabilistic approach based on LDA called joint sentiment/topic model (JST). The model detects topic and sentiment simultaneously from the text. The model is evaluated on Movie Review dataset. The sentiment classification accuracy is evaluated on a set of 1,50 and 100 topics. For topic=1 JST transform itself to simple LDA model with S topic and each of which represents to sentiment label. JST perform worse on topic k=1 compared to 50 and 100 topics. The limitation of the model is it represents each document as a bag of a word and condone the ordering of words.

Table 2:Comparative Evaluation of Topic Discovery Work

Paper	Objective	Dataset	Method Used	Results(Accuracy)
Sentiment-Based Topic Suggestion for	four generative models are proposed for the	Yelp and FourSquare	Venue-based and user-based topic prediction	K=10;PMI of ASLDA is higher

Micro-Reviews(2016)	topic recommendation on micro-reviews in location-based social sites			
A Short Text Sentiment-Topic Model for Product Reviews(2018)	Word-pair Topic-Sentiment WSTM model is introduced for short text	Chinese product reviews datasets	Used NetHow Sentiment Lexicon	k-15;accuracy=65%
Joint Sentiment/Topic Model for Sentiment Analysis(2009)	detects topic and sentiment simultaneously from the text	Movie Review dataset	a probabilistic approach based on LDA called joint sentiment/topic model (JST)	Good performance on k=50 and k=100;worse performance k=1

CHAPTER 3. PROPOSED FRAMEWORK

The multi-model topic sentiment analytics framework is based on a combination of sentiment topic model using LDA and sentiment analysis using improved sentiment assignment from the existing tools Text Blob[27], Vader[25], CoreNLP[26] by adding additional semantic rules.

3.1 Framework Architecture

The architecture diagram shown in Figure 1 portrays how the entire multi-modal framework works. The framework is divided into two areas. First, it is identifying hidden contextual topic-based sentiment from the Twitter dataset based on relevant terms. For topic modeling, the text will preprocess using NLP techniques so that we push it into the LDA model. Second, the sentiment analysis is performed by identifying the mismatched cases using the combination of CoreNLP, Text Blob and Vader and then improved them by adding semantic rules in it to extract the final sentiment in the short text. The mismatch cases identified by Human-Machine and Machine-Machine analysis. The weakness of the existing model improved by adding more rules that cover the mismatch cases efficiently. The performance of the Multi-Modal Topic Sentiment Analytics Framework can be computed by using interrater reliability metric Cohen Kappa.

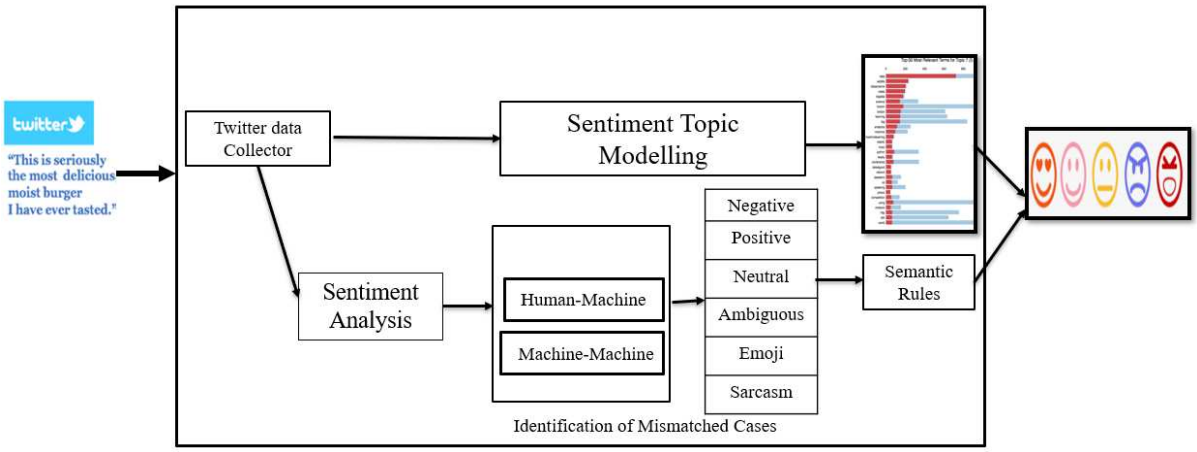

Figure 1:Multi-Modal Topic Sentiment Analytics

3.2 Topic Discovery

In this thesis, we conducted the topic modeling to extract the hidden topics in the short text. Here we used LDA which is a generative model for topic discovery. Latent Dirichlet Allocation (LDA) is an unsupervised, statistical method to document modeling that learns latent semantic topics in huge collections of text documents. LDA points out that words carry strong semantic information, and documents discussing similar topics will use a similar group of words. Latent topics are thus revealed by finding groups of words in the corpus that commonly occur together within documents.

1). Natural Language Processing:

NLP involves the following steps:

i) Tokenization: Split the sentences into words.

ii) Stop words removal: Remove all stop words and punctuation.

iii) Lemmatization: Words are normalized, words in the third person changed to first person and future and past tenses are changed into the present tense.

2). Bag of Words:

In natural language processing, a document is usually represented by a Bag of Words that are a word-document matrix. For each document, we generate a dictionary describing how many words and how many times those words appear.

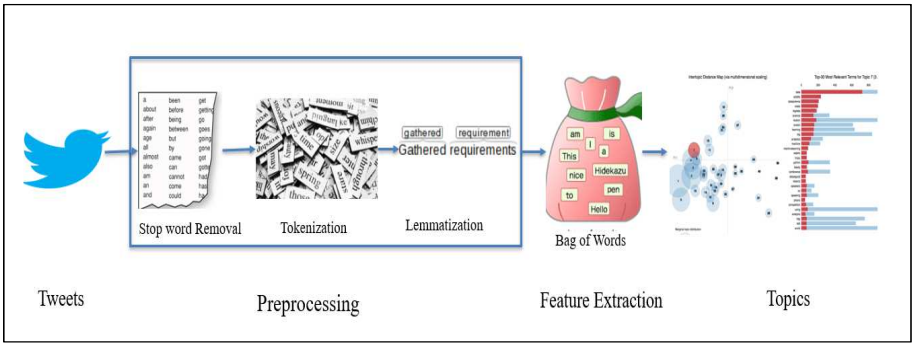

Figure 2: Topic Modelling using LDA

3.3 Sentiment Analysis using Multi-Modal Framework

Sentiment analysis is the identification of polarity of a phrase, sentence, document or speech. Here we identify that the opinion of the text is positive, negative or neutral. Here we performed the i). Human-Machine ii). Machine-Machine based analysis to identify the mismatched cases.

i) Vader (Valence Aware Dictionary for Sentiment Reasoning):

Vader [25] handles polarity and intensity and it uses human-centric based approach i.e. it uses human raters to rate the lexicon and used the wisdom of crowd so instead of taking

a single expert opinion it relies on the opinion of a group of experts. Vader[25] dictionary maps lexical features to emotional intensities. The sentiment score of Vader[25] lies on a scale -4 to +4 and 0 is neutral. The model works well with short text; especially covers acronyms, slangs, and punctuations.

ii) Stanford CoreNLP:

Stanford CoreNLP [26] compute sentiment based on how individual words change the meaning of longer phrases. It is a new type of recursive neural network that builds on grammatical structures. The sentiment treebank of CoreNLP[26] includes fine-grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences. CoreNLP[26] assign sentiments Negative, Positive, Neutral to a sentence.

ii) Text Blob:

Text Blob [27] is a python library that returns sentiment property a named tuple of the form Sentiment (polarity, subjectivity).Polarity Score of Text Blob[27] lies within range[-1.0,1.0].Subjectivity lies within range [0.0,1.0] .

3.3.1 Identification of Mis-matched cases

Here the mismatched cases identified between human, Vader, Text Blob and CoreNLP.

- Positive – some of the positive sentiments misclassified.
- Negative -some of the negative sentiments misclassified.
- Neutral - some of the neutral sentiments misclassified.
- Ambiguous –sentiment in ambiguous sentence unidentified.
- Presence of Emojis –some of the emojis unidentified in a sentence.

- Sarcasm - sarcasm is unidentified in a sentence.

i) Human -Machine:

We compared in which areas of sentiment assignment to text, the opinion of Machine and Human is different.

a) Human (Native Speaker) and Machine (VADER)

Vader is unable to identify sarcasm. Vader is identifying the presence of interjection in a sentence like Ah! *e.g. Woo! We have done up A-A-Ron! I messed up Onion bro's*

b) Human (Native Speaker) and Machine (CoreNLP)

CoreNLP is unable to identify sarcasm. CoreNLP unable to identify acronyms, emojis. *e.g. not the biggest fan of ginger beer* 😕

c) Human (Native Speaker) and Machine (Text Blob)

Text Blob is unable to identify sarcasm. *e.g. Ironically, they're very sick at tonight* 😂.Text Blob unable to identify emojis. *e.g. not the biggest fan of ginger beer* 😕

ii). Machine-Machine:

We compared in which areas of sentiment assignment to text, the opinion of Vader, CoreNLP and Text Blob is different.

a). Vader and CoreNLP:

CoreNLP usually considers sentences having negation (doesn't, don't) as negative. e.*g. Rome doesn't have Chick-fil-A .....* .Vader is assigning negative to the multiple sentence tweet as negative if the second sentence is the contradiction of the

first one, CoreNLP is assigning neutral. *e.g.: They already pay a proper wage. British don't want to do it* .CoreNLP is unable to identify slangs, emojis, and acronyms. *e.g.: not the biggest fan of ginger beer* 😔

b). Vader and Text Blob:

Vader is assigning negative to the multiple sentence tweet having negative in it. e.*g. They already pay a proper wage. British don't want to do it*. Text Blob unable to identify emojis. *e.g. not the biggest fan of ginger beer* 😔 .Vader and Text Blob is identifying slangs, acronyms, and emojis (Text face for Text Blob). e.*g. not the biggest fan of ginger beer* 😔 *sloppy scrambled eggs, tinned mushrooms, value beans - what is they're not to love LOL*

c).CoreNLP and Text Blob:

CoreNLP usually considers sentences having negation (doesn't, don't) as negative. *e.g. Rome doesn't have Chick-fil-A .....* CoreNLP is quite weak in identifying neutral sentences and usually assign them negatively. e.*g. Now playing: Red Hot Chili Peppers – Look Around* .Text Blob identifies text face emojis. e.g. *Iet's continue! Still, live! :D*

3.3.2 Semantic Rules for Improving Sentiment Assignment

To improve the sentiment assignment of existing tools new rules are added to the existing tools.

Rule1:

for word in a sentence **do**

get polarity_score of word

```
        create list(polarity_score)
if list contains high positive intensities then
        W*(Vader Score)
        end
else
   print 0
```

e.g.

Tweet	Vader	Text Blob	CoreNLP
GOT MY WAVE SANDBOX INVITE! Extra excited! Too bad I have class now... but I'll play with it soon enough! ☺ #io2009 #wave	1	0	1

excited=>2.1 , ☺ =>1.6

2(1) =2 i.e. positive

```
Rule2:
for word in sentence do
        get polarity_score of word
        create list(polarity_score)
if list contains emphasis OR emoji OR punctuation then
        W*(Vader Score)
else if a sentence has high negative intensity word then
        W*(CoreNLP Score)
else
        print 0
```

e.g.

Tweet	Vader	Text Blob	CoreNLP
could time-warner cable suck more? NO.	0	1	0

Suck, ?, NO is highly intensive negatives

(-2)(0)+(-1)(0)=0

Rule 3:

if a pattern of the sentence is positive to negative ***do,***

w=1

$$\frac{W^{*}(\Sigma \text{ No. Of positive intensity words})-W(\text{No. Of negative intensity words})}{W^{*}(\Sigma \text{ No. Of positive intensity words})+W(\text{No. Of negative intensity words})}$$

e.g.

Tweet	Label	Vader	Text Blob	CoreNLP
‘‘I love being ignored all the time’’	0	4	1	0

Here love =>3.2, ignored =>1.3 $\frac{(1)(1)-(1)(1)}{(1)(1)+(1)(0)}=0$

3.4 Case Study (Food-Mood Classification)

Obesity is one of the biggest issues nowadays, it increases the chances of many chronic diseases like heart diseases and diabetes. In USA 2/3 adults are overweight and 1/3 of those overweight are obese [28]. In addition to that, social media leaves a huge impact on eating habits. Currently, social media especially Twitter is considered as the best medium to understand user’s perspective and behavior on health. Food patterns also affect the daily

sentiments of the people. According to Ypulse,63% of 13-32 years old have posted a photo on social media of food or drink they were having with their sentiments.

To identify whether food related tweet is healthy, unhealthy or mixed. The following score is used:

$$score = \begin{cases} 1 & \text{if healthy food} \\ -1 & \text{if unhealthy food} \\ \frac{\sum fs(healthy)+fs(unhealthy)}{\sum fs(healthy)-fs(unhealthy)} & \text{if mixed foods} \end{cases}$$

For instance,

"There better be wine and coffee in hell"

Food scores (fs) =['healthy':0, 'unhealthy': -2, 'compound': -1]

Compound score=$\frac{0-2}{0+2}$ =-1

CHAPTER 4. RESULTS AND EVALUATIONS

4.1 Introduction

In this section, we discuss the results and the evaluation of the proposed framework. First, we describe the results of latent features extracted using Word2vec [6] from social media (Twitter). Second, we show the accuracy results of CoreNLP[26], Text Blob [27], Vader[25] and Multi-model topic sentiment analytics model on benchmark dataset as well as Twitter food dataset.

4.2 Data Preparation

4.2.1 Twitter Data Collection

The tweets are collected from January 15th, 2018 to January 19th, 2018 (5 days duration) using Twitter Streaming API using keywords on food or obesity. We have followed standard food keywords from the following sites, defined by the USDA MyPlate (2015-20 Dietary Guidelines for Americans for children) [29] and USDA Standardized Recipe [30], Choose MyPlate [29], the most unhealthy meals in America [31], Worst Options for Restaurant Menu [32]. For Twitter analysis, we used the Healthy/Unhealthy food and disease keywords (76 healthy foods/28 unhealthy foods) as defined in Table 3. Table 4 shows the healthy dieting categories in terms of the number of healthy, mixed, unhealthy foods as well as the food mood in terms of the number of positive and negative tweets.

For the pre-processing of the twitter dataset NLP [20] techniques are applied to derive the structure from unstructured text. NLP [20] is a technique used to examine text so that machines understand how human speaks. Table 5 depicts the NLP statistics obtained from the Twitter food dataset. Originally, the corpus size was 671383 words. After applying NLP steps, the size of the corpus reduced to 528740 words. Lemmatization also used to extract the root words which is 50637 words from a corpus.

Table 3:Keywords for Data Collection

Type	Keywords	Numbers
Healthy Food Keywords	acorn squash, apple, apricot, artichoke, arugula, asparagus, avocado, baked fish, baked lentil, baked sweet potatoes, banana, basil, bean burrito bowl, beans, beef stew, bellpepper, berries, black beans and rice, black beans patty, black beans salad, blackberry, blackeyedpeas, blueberry, breadfruit, broccoli, broiled tomatoes and cheese, brussels sprouts, cabbage, cantaloupe, capsicum, carrot, cauliflower, celery, cherry, chicken casserole, chicken salad, chicken tetrazzini, chickpeas, chives, coconut, collard greens, corn pudding, cucumber, dates, eggplant, fruit salsa, garden pasta salad, garlic, ginger, gourd, grape, grapefruit, green beans potatoes, green onion omelet, greenbean, greens, guacamole, guava, honeydew, hot chilli peppers, iceberg lettuce, jackfruit, kale, kiwi, lemon, lentils, lettuce, lima bean, lime, lingon berry, maize, mandarin orange, mango, marion berry, meatloaf, melon, mint, mulberry, mushroom, mustard greens, okra, olive, onion, orange, papaya, passion fruit, patty pan squash, pea, peach, peanut, pear, peas, peppers, persimmon, pickle, pico del gallo, pineapple, plantain, plum, pluot, pomegranate, pomelo, prune, pumpkin, pumpkin soup, quince, radish, raisin muffin, raspberry, roasted potatoes, roasted salmon, roasted tilapia, rocket, romaine, rutabaga, salad, salmon, salsa, sautéed spinach and tomatoes, scallion, scalloped potatoes, seaweed, shallot, smoked turkey, sorrel, soybean, spinach, spring onion, sprouts, spuds, squash, star fruit, strawberry, string bean, succotash, sweet and sour pork, sweet potato, swiss chard, tangelo, tangerine, taro, teriyaki sauce, tofu, tomatillo, tomato, tomato salsa, tuber, tuna and noodles, turnip, ugli fruit, vegetable rice, vegetable soup, vegetable wrap, veggie burger, veggie pizza, wasabi, waterchestnut, watermelon, yam, yucca, zucchini	160
Unhealthy Food Keywords	arbys, bacon, bacon ranch beef quesadilla, baskin robins, bbq, beer, blue cheese, bread, burger, burger king, cake, captain Ds, carls jr., checkers, cheese, cheese burger omelet with pancakes, cheese curd baconburger with fries, cheese sauce, chicken pot pie, chicken tender, chickfila, chipotle, chips, chocolate cake, churchs chicken, churros, cicis pizza, cookie, cream cheese, cream gravy, creamy chicken, coffee, crispy chicken, culvers, dairy queen, del taco, dessert, dominos pizza, donut, dunkin donuts, el pollo loco, energy drink, fettuccine weesie, fish and chip, five guys burger & fries, flying gorilla drink, frenchfries, fried, fried chicken, fried fish, fried rice, fried steak, fried sweet apples, fries, fruit drink, goat cheese, hash browns, hotdog, jack in the box, jalapeno thickburger, jason sdeli, jimmy johns, ketchup, kfc, krispy kreme, krystal, little caesars, long john silvers, mac n cheese, macaroni grill chicken parmesan, margarine, mcdonalds, microwave popcorns, milkshake, nuggets, oil, onionrings, pad thai shrimp, pancake, panda express, pastry, pepperoni, pizza, potato, sausage, soda, spring roll, steak, tenders, thai curry boneless wings, ultimate smokehouse combo, vegetable oil, waffles, wine	93

Table 4: Healthy Dieting Twitter Categories

Healthy Food	**Unhealthy Food**	**Mixed Food**	**Total**
30,953	39,281	7,163	77,397
Emotions	Positive	Negative	Total
	62,485	14,912	77,397

Table 5: NLP Statistics on Twitter Food Dataset

Actual size of a dataset	Before NLP application	671383 words
After removing stop words	a, about, above, after, again, against, or, other, some, so, then etc.	528740 words
Tokenization	food, beer, cake, party, cheese, eat, lunch etc.	50637 words
Lemmatization	assault, ginger, sniffle, merchandise, haunt, thwart, ache, quench etc.	81791 words

4.2.2 Benchmark Dataset

a) Sentiment 140 Dataset

The datafile of Sentiment-140[33] dataset is in CSV format, it contains tweets with emoticons removed. The polarity of the tweets is:0 = negative, 2 = neutral, 4 = positive. The dataset creator claimed that the approach used for tweet annotation is automatic rather than manual. In tweet dataset, the tweets with positive emoticons like ☺ are considered as positive and tweets with negative emoticon☹ are considered as negative. Twitter API with a

keyword search is used for the collection of tweets. The size of the dataset is 77.6 MB. The training dataset is around 1865.66 MB large and test dataset is 73KB. The dataset overall comprises 1.6M tweets.

b) Amazon Reviews

This Amazon Reviews [34] dataset consists of a few million Amazon customer reviews which are around 142.8 M during a period of May 1996-July 2014. Each review has star rating: __label__1 corresponds to1-and 2-star reviews, and __label__2 corresponds to 4-and 5-star reviews. The datasets include reviews (ratings, text, votes), metadata and links.

c) IMDB Movie Reviews

The dataset is for binary sentiment classification. The dataset is labeled with respect to their overall sentiment polarity or subjective rating. The IMDB dataset[35] contain lowercase English reviews. The reviews were originally released in 2002 but the cleaned and refined version was released in 2004. Author of the dataset named it “Polarity dataset”. Dataset [35] contains movie reviews in two folders one for negative reviews and other is positive reviews. The dataset provides the set of 25,000 reviews for training and 25,000 reviews for testing.

4.3 Evaluation Metrics

4.3.1 Sentiment Evaluation Metrics

i. Precision

In binary classification, precision (also called positive predictive value) is the fraction of related instances among the retrieved instances.

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Precision[39] is used to determine when the cost of false positive is high. Precision is basically the ability of classifier not to label as positive a sample which is negative[40].

ii. Recall

In binary classification, recall (also known as sensitivity) is the ratio of correctly identified instances over the total amount of relevant instances.

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Recall helps to determine when the cost of false negative is high[39]. The recall is the ability of the classifier to find all negative samples[40].

iii. F-Score

F-score considered as one of the most popular performance metrics. It is also called balanced F-score or F-measure. It is a harmonic mean of recall and precision[40]. It is used to test accuracy. It is the consideration of both precision and recall [39].F1 score considered perfect when the value is 1 and considered as a complete failure when the value is 0 [39].Precision and Recall contributed to the F1 score equally[40].

$$\frac{2\text{xPrecisionxRecall}}{\text{Precision} + \text{Recall}}$$

iv. Interrater reliability

It is a degree of agreement between raters. It gives the score of how much consensus there is in the rating given by the judges. Inter-rater reliability can be evaluated by using

several different statistics[41]. Some of the more common statistics include percentage agreement, kappa, product-moment correlation, and intraclass correlation coefficient[42]. High inter-rater reliability values refer to a high degree of agreement between two examiners[42]. Low inter-rater reliability values refer to a low degree of agreement between two examiners[42].

"Cohen Kappa" is a form of correlation for measuring agreement on two or more investigative categories by two or more method. The raters involved in the analysis rate the items individually and independently. Kappa can be defined as the proportion of agreements after chances agreement is removed.

$$\frac{\text{Observed Agreement -Agreement by chance}}{\text{1-Agreement by chance}}$$

Here the observed agreement is how much agreement is there among raters and agreement by chance is how much agreement would be expected to be present by chance alone [40].Here we compute agreement by chance for positive and agreement by chance for negative. The denominator in the above formula standardizes the score.

If, Kappa 0=>agreement is no better than chance. Kappa 1=>agreement is perfect. Kappa negative => Less agreement what you'd expect by chance.

Table 6: Kappa Score and Interpretation

Kappa Score	**Interpretation**
0.0-0.20	No or slight agreement
0.21-0.40	Fair
0.41-0.60	Moderate

0.61-0.80	Good
>0.80	Very Good

4.3.2 Topic Evaluation Metrics

To find the optimal topics for LDA we can calculate perplexity or Coherence score.

i) Perplexity

It is an indicator of the generalized performance of a model.

$$L(D') = \frac{\sum log_2 p(w_d;\theta)}{count\ of\ tokens}$$

Here w_d represents the unseen data in the holdout set and θ ids the parameters learned by the model. The first equation computes the log-likelihood; the probability of observing some hidden data given a model encountered earlier. This checks whether the model captures the distribution of the held-out set. If it doesn't then the perplexity is very high suggesting that model is bad [38]. The disadvantage of perplexity is it is to strongly correlate to human judgment.

ii) Topic Coherence

Topic models learn topics typically represented as sets of important words automatically from unlabeled documents in an unsupervised way. But topics are not guaranteed to be well interpretable, therefore, coherence measures have been proposed to distinguish between good and bad topics[36].

Topic coherence, which is grouped into 4 following dimensions:

1.Segmentation

2. Probability Estimation

3. Confirmation Measure

4. Aggregation

Coherence = $\sum$ score (w_a, w_b) where a<b

In each topic select the top n frequently appearing terms. Calculate the pairwise score for the selected words. To calculate the coherence score of topic combines all computed pairwise scores.

a.Intrinsic Measure

It is represented as UMass[37]. It measures to compare a word only to the prior and following words respectively, so it requires an ordered word set. It is utilized as pairwise score function which is the heuristic conditional log-probability with leveling count to deflect calculating the logarithm of zero.

b. Extrinsic Measure

It is represented as UCI[37]. In UCI measure, every single word is harmonizing with every other single word. The UCI coherence uses pointwise mutual information (PMI).

Both Intrinsic and Extrinsic measure calculate the coherence score c (sum of pairwise scores on the words w1, ..., wn used to define the topic).

To draw out relevant words with respect to a specific topic[43], here the terms which are appearing in multiple topics are considered less significant.

$$R(w , k|\lambda)=\lambda \log(\phi kw)+(1- \lambda)\log(\frac{\phi kw}{pw})$$

λlog(ϕkw) is the overall frequency of words appeared in the topic. If the word appeared in multiple topics so the word will not be considered as good to distinguish the topic well. (1- λ)log($\frac{\phi kw}{pw}$) is the rareness of term in a topic. The unique term decreases the score of terms occurring frequently, but on the other hand, it upsurges the score of unique terms that occur in a topic. Through empirical research, it is suggested that an ideal value of the weight, ƛ is usually kept around at 0.3[43].

4.4 Results

4.4.1 Sentiment Evaluation Results

Vader, CoreNLP, Text blob and multi-modal topic sentiment analytics (MMTSA) applied on benchmark datasets and the Twitter food datasets to compare their accuracies.

Table 7 : Accuracy of Vader on Datasets

Datasets	**F-Score Vader**		
	Positive	Negative	Neutral
Sentiment-140	74.52%	71.4%	67.36%
Amazon Reviews	74.67%	58.94%	-
IMDB Reviews	77.28%	68.9%	-
Twitter Food	63.63%	38.46%	40%

Table 7 shows Vader performs well on positive sentiment on all dataset. The accuracy of Vader on positive assignment is more than 70% on the benchmark and Twitter food datasets. The negative and neutral sentiment assignment of the datasets is promising. Vader is considered weak in identifying negative sentiments in long texts i.e. Amazon Products and IMDB reviews.

Table 8: Accuracy of Text Blob on Datasets

Datasets	F-Score Text Blob		
	Positive	Negative	Neutral
Sentiment-140	69.0%	60.35%	63.93%
Amazon Reviews	76.66%	59.62%	-
IMDB Reviews	74.87%	57.92%	-
Twitter Food	58.53%	33.33%	45.75%

Table 8 shows Text blob has good accuracy on Amazon and IMBD reviews datasets. The accuracy of Sentiment-140 dataset is comparatively low. Text blob is considerably weak in identifying negative in short as well as long texts. The assignment of neutral to twitter text is more than 60%.

Table 9: Accuracy of CoreNLP on Datasets

Datasets	F-Score CoreNLP		
	Positive	Negative	Neutral
Sentiment-140	38.09%	58.33%	40%
Amazon Reviews	68.96%	66.66%	-
IMDB Reviews	73.33%	64%	-
Twitter Food	42.42%	48.48%	35.29%

Table 9 shows Text blob has good accuracy on negative sentiment in Amazon and IMBD reviews datasets. The text of the review datasets is long so CoreNLP assignment of negative sentiments to those texts are considerably good. However, the negative assignment to the short text is still low.

Table 10: Multimodal Topic Semantic Analytics (MMTSA)

Datasets	F-Score Vader		
	Positive	Negative	Neutral
Sentiment-140	80%	79%	75.25%
Amazon Reviews	90.00%	88.43%	-
IMDB Reviews	94.23%	89.05%	-
Twitter Food	77.2%	79%	64.5%

MMTSA in Table 10 performed very well in positive, negative and neutral sentiments in all four datasets. In Amazon and IMDB reviews the positive assignment of the Amazon and IMDB reviews are more than 90%.In addition to that, the negative sentiment assignment is more than 80%.

Table 11:MMTSA Confusion Matrix on Sentiment-140

	Negative	**Neutral**	**Positive**
Negative	119	28	30
Neutral	7	111	21
Positive	10	17	155

Table 12:MMTSA Confusion Matrix on Amazon Product Reviews

	Negative	**Neutral**	**Positive**
Negative	195	-	44
Neutral	-	-	-
Positive	7	4	248

Table 13:MMTSA Confusion Matrix on IMDB Movie Reviews

	Negative	**Neutral**	**Positive**
Negative	197	-	8
Neutral	-	1	-
Positive	2	-	280

Table 14:MMTSA Confusion Matrix on Twitter Food

	Negative	Neutral	Positive
Negative	10	1	2
Neutral	2	10	-
Positive	2	-	280

Table 11-14 shows the confusion matrix of MMTSA model on different datasets. The number of misclassified entries are low for the datasets.

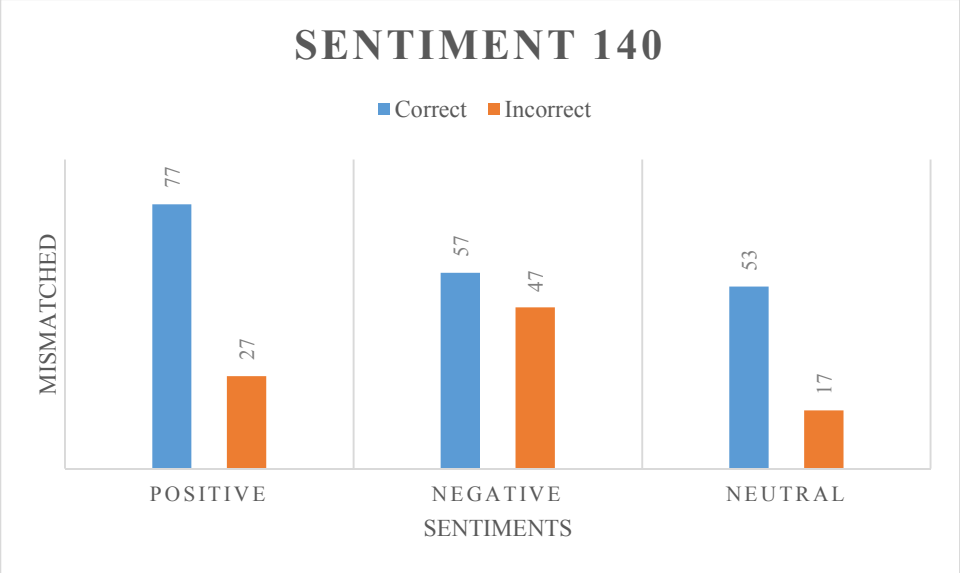

Figure 3 Mismatched of Sentiment-140 Cases Corrected by MMTSA

In figure 3 MMTSA able to identify 77% mismatched cases and 27% remains misclassified in positive sentiment assignment in Sentiment-140. Moreover, the negative sentiment assignment of the model is a bit better as the number of misclassified cases are around 40%.

Table 15: Mismatched Cases of Existing Models and MMTSA of Sentiment -140

Sentiment -140				
@siratomofbones we tried but Time Warner wasn't being nice so we recorded today. :)				
Actual Label	**Vader**	**Text Blob**	**CoreNLP**	**MMTSA**
Negative	Positive	Positive	Negative	Negative
wth..i have never seen a line this loooong at time warner before, ugh.				
Negative	Neutral	Neutral	Positive	Neutral
By the way, I'm totally inspired by this freaky Nike commercial				
Positive	Positive	Neutral	Neutral	Positive
@uscsports21 LeBron is a monsta and he is only 24. SMH The world ain't ready.				
Positive	Negative	Positive	Neutral	Negative

In Table 15 the CoreNLP is unable to identify slangs (wth, lol etc.) and assigned the tweet containing it as neutral. MMTSA can identify sarcasm in the tweet.

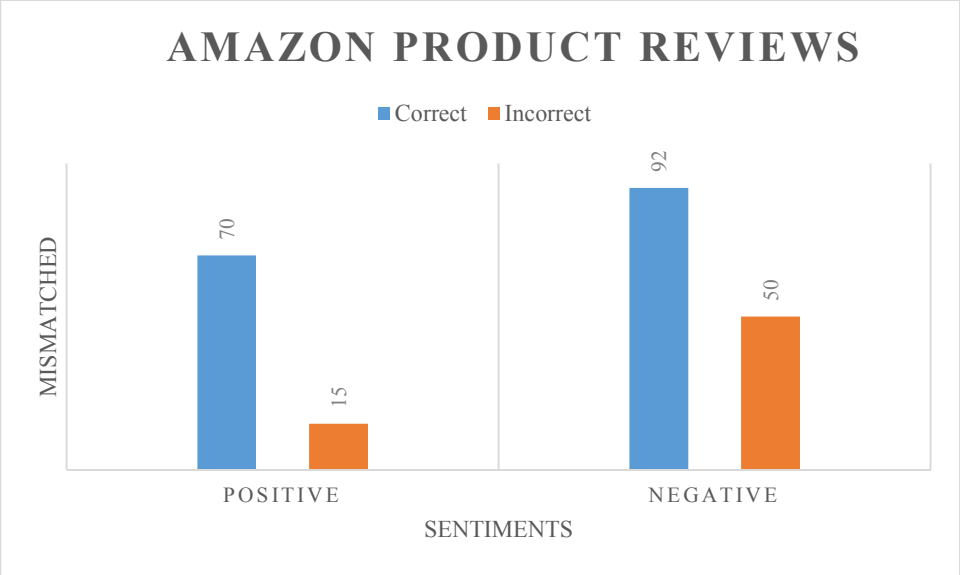

Figure 4:Mismatched of Amazon Product Reviews Cases Corrected by MMTSA

In figure 4 MMTSA able to identify 70% mismatched cases and 15% remains misclassified in positive sentiment assignment in Amazon Product Reviews. Moreover, the negative sentiment assignment of the model is extremely performed well as the number of misclassified cases are quite low.

Table 16: Mismatched Cases of Existing Models and MMTSA of Amazon Product Reviews

Amazon Product Reviews				
Kingston Technology KVR133X64C3/ 256 PC133 256MB 32MX64: Works great. Just snap it into the slot and turn on my pc. In a few seconds, I was up and running and my pc working much faster than before				
Actual Label	**Vader**	**Text Blob**	**CoreNLP**	**MMTSA**
Positive	Positive	Positive	Negative	Positive
Life changing!: This book has changed my life. It opened my eyes about how we think ourselves into misery. I'm want to buy copies to away give to people in my life.				
Positive	Negative	Neutral	Neutral	Neutral
Can I shoot myself now??: A few years back I was forced to read "Tess" for my English class. Nobody in my class could ever finish reading the book... only a few of us could ever even finish the Cliff Notes for it. Our Professor tried solving the situation by renting us the video, but all of us were fast asleep half an hour into it.Tess is a bimbo with no brain and an insult to women, ……..				
Negative	Positive	Neutral	Neutral	Neutral
The book was boring, because of its Victorian ideals.: Tess was a very boring book. From an analytical point, the book's major theme was FATE. However, you must be interested in the Victorian era, in order to enjoy the book. It is the same thing with the Scarlet Letter, to enjoy the book you must know about the time era. The problems that the characters are faced with in the book are laughable by today's standards.				
Negative	Positive	Negative	Negative	Negative

Table 16 depicts Vader is struggling in identifying the negative in a sentence especially a long text and assigned it Positive. In addition to that Text Blob is assigning it neutral.

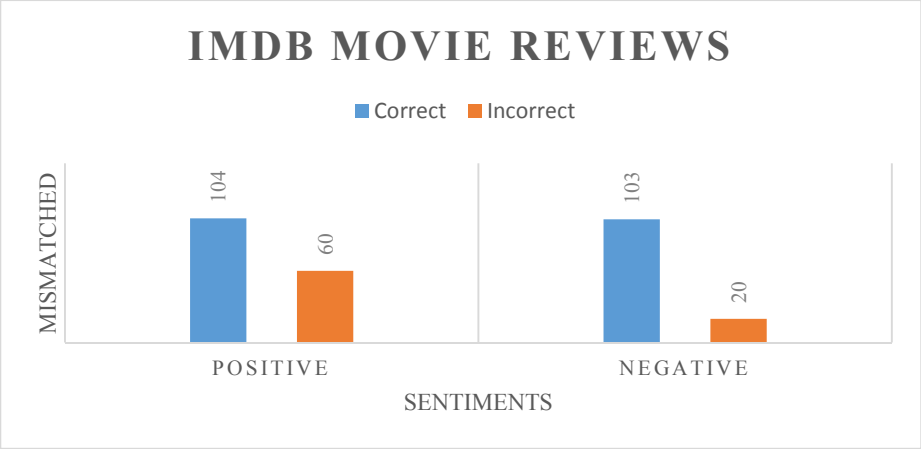

Figure 5: Mismatched of IMDB Movie Reviews Cases Corrected by MMTSA

In figure 5 MMTSA able to identify 104 mismatched cases out of 164 misclassified cases in positive sentiment assignment in IMDB movie Reviews. Moreover, the negative sentiment assignment for 103 cases is correctly recognized by model out of 123 cases.

Table 17: Mismatched Cases of Existing Models and MMTSA of IMDB Movie Reviews

IMDB Product Reviews				
Well, I'm not the world's biggest Sondheim fan, so although I have the cast album and I've listened to it a few times I've never actually seen this show performed and I haven't seen the Tim Burton movie version either. ……				
Actual Label	**Vader**	**Text Blob**	**CoreNLP**	**MMTSA**
Positive	Positive	Positive	Negative	Positive
Televised in 1982, from a Los Angeles production, this is probably the finest example of a filmed stage musical you are likely to encounter……				
Positive	Negative	Positive	Positive	Negative
I remember when they made a big deal about this when it was coming out. They showed clips every week on WWF				
Negative	Positive	Positive	Neutral	Positive
... in search of the cheesiest "so bad it's good" movie, I've repeatedly laughed at the first fifteen minutes of various films….				
Negative	Positive	Negative	Negative	Negative

Table 17 depicts Vader and Text blob is struggling in identifying the negative in a sentence especially a long text and assigned it Positive.

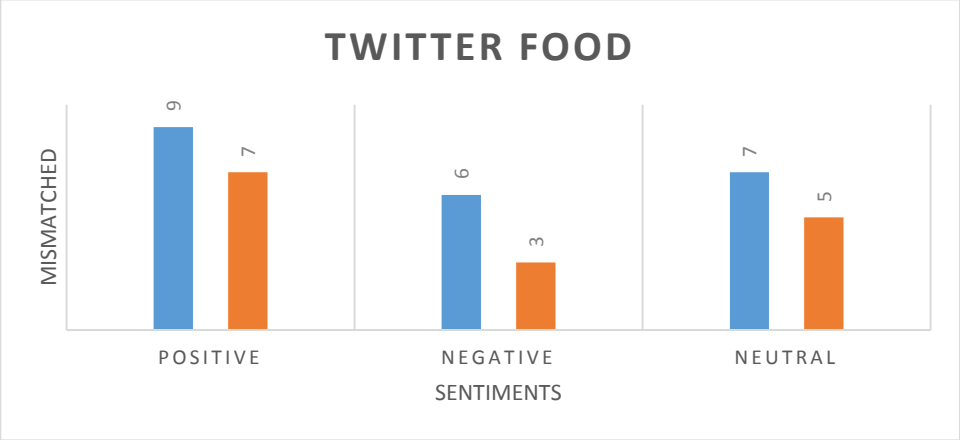

Figure 6:Mismatched of Twitter Food Cases Corrected by MMTSA

In figure 6 MMTSA able to identify 9 mismatched cases out of 16 misclassified cases in positive sentiment assignment in Twitter food dataset. Moreover, the negative sentiment assignment for 6 cases is correctly recognized by model out of 9 cases. Out of 12 mismatched cases in neutral sentiment category,7 classified correctly.

Table 18: Mismatched cases of Existing Models and MMTSA of Twitter Food

Food Twitter				
Big standard stout, no discernible vanilla or sweetness from the lactose - Drinking a Jet Black Heart				
Actual Label	Vader	Text Blob	CoreNLP	MMTSA
Negative	Neutral	Positive	Negative	Negative
Drinking a glass of water				
Neutral	Negative	Neutral	Neutral	Neutral
I've not had pizza in so long I think I'm actually getting withdrawal				
Positive	Neutral	Neutral	Negative	Negative

Table 18 depicts that MMTSA identifying neutral in a simple sentence and negative in a sentence. But if the sentence is a bit perplexed so MMTSA struggled in assigning sentiment to the text.

Table 19: Kappa Score for the Datasets

	Datasets	**CoreNLP**	**Vader**	**Text Blob**	**MMTSA**
Labels	IMDB Movie Review	0.3714	0.419	0.4	0.80
	Product Review	0.35	0.64	0.42	0.782
	Sentiment 140	0.26	0.68	0.516	0.700
	Twitter Food	0.16	0.23	0.21	0.58

Table 19 shows that the kappa score of MMTSA is significantly increased and in IMDB and Product Review datasets it reaches 80% and 78% respectively. Similarly, the kappa score of Vader is good for Product Review and Sentiment-140 dataset i.e. 64% and 68%

respectively. CoreNLP performs fair for IMDB and Product review datasets i.e. 37% and 35%. Text Blob has good accuracy on Sentiment-140 dataset i.e. 51.6%.

Table 20: Kappa Score Interpretation for the Datasets

	Datasets	**CoreNLP**	**Vader**	**Text Blob**	**Multi-Model**
Labels	IMDB Movie Review	Fair	Moderate	Fair	Very Good
	Product Review	Fair	Good	Moderate	Good
	Sentiment 140	Fair	Good	Moderate	Good
	Twitter Food	Slight Agreement	Fair	Fair	Moderate

Table 20 of kappa score interpretation depicts that the score of IMDB Movie reviews has a significant improvement and reaches to “Very good” while the Twitter Food dataset score comes under “Moderate” as initially, it is “Slight Agreement”.

Table 21: Food Mood Results

	Healthy Tweets		**Unhealthy Tweets**		**Mixed Tweets**	
	# Tweets	% Tweets	# Tweets	% Tweets	# Tweets	% Tweets
Positive	25,268	32.6%	7,917	10.2%	5,853	7.5%
Negative	5,685	7.3%	31,364	40.5%	1,310	1.69%
Total	30,953	39.99%	39,281	50.75%	7,163	9.25%

Table 21 shows that 40% of overall Tweets are healthy out of which 32.6% are positive sentiment.51% are Unhealthy out of which 40% are with negative sentiment. However, a small ratio of tweets i.e. 9.25% are mixed (having the same number of healthy and unhealthy food items in the tweet) out of which 7.5% had positive sentiment.

Table 22: Example Tweets

	Example Tweets
Healthy Positive Food Tweets	"This Thai spinach, brown rice is wonderfully aromatic and delicious." "A ginger, lemon, orange and grapefruit juice is the best thing whenever I get cold symptoms."
Unhealthy Positive Food Tweets	"I love eating chocolate cake and ice cream after a show.
Healthy Negative Food Tweets	"I am sick of eating broccoli. I also hate spinach."
Unhealthy Negative Food Tweets	"I lovvvvvvve the Halloween cookies. I just wanna know why they made them so damn small"
Mixed Positive Food Tweets	"chicken enchiladas w cheesy Chipotle sauce with rice beans damn good"
Mixed Negative Food Tweets	"I am tired of eating eggs, sausage, and veggie casserole every morning for their first meal"

Table 22 depicts the example tweets reflecting users' sentiments toward food. Some of the tweets conveying how users are emphasizing on positive/negative words to show like or dislike towards food items.

Table 23: Word2Vec Embeddings

Base	**Word2Vec**
Healthy food	aubergine grits kaleslaw feta cabrales quinoa butternut habanero miso marinara egg halloumi broth prawn pilaf sweetcorn baguette lobster cilantro courgette shea sesame ricotta jambalaya naan peppermint dijon ceviche rhubarb callaloo chowder swiss cheese mozzarella cottage cheese chives
Unhealthy food	coffee Nutella Gin Salami Chorizo Calamari Poptarts Tikka Carnitas Chilaquiles Couscous Squid Mayo Tater

Using Apache Spark word2vec, in Table 23 we calculated the synonyms using window size 5 to increase the size of the food lexicon. Here, we can extract 38 Healthy food synonyms and 14 Unhealthy food synonyms from the lexicon.

4.4.2 Topic Discovery Results

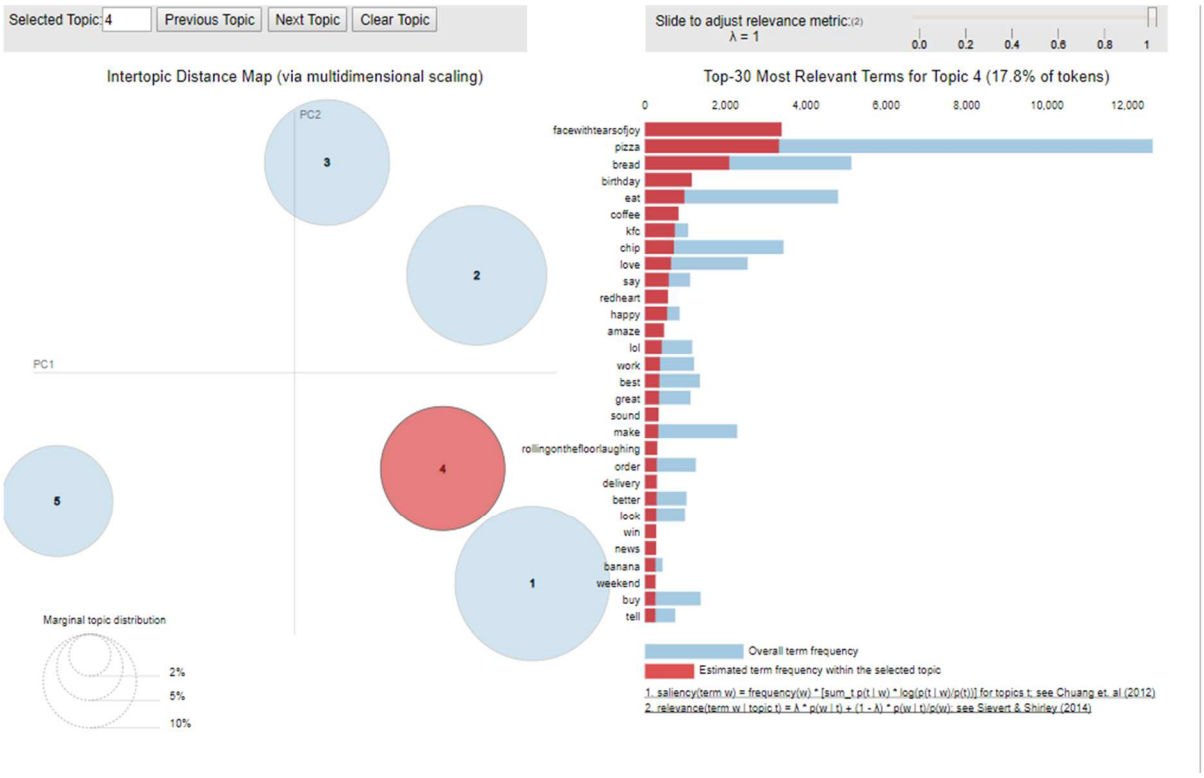

Figure 7:Twitter Food Topics with 5 Topics

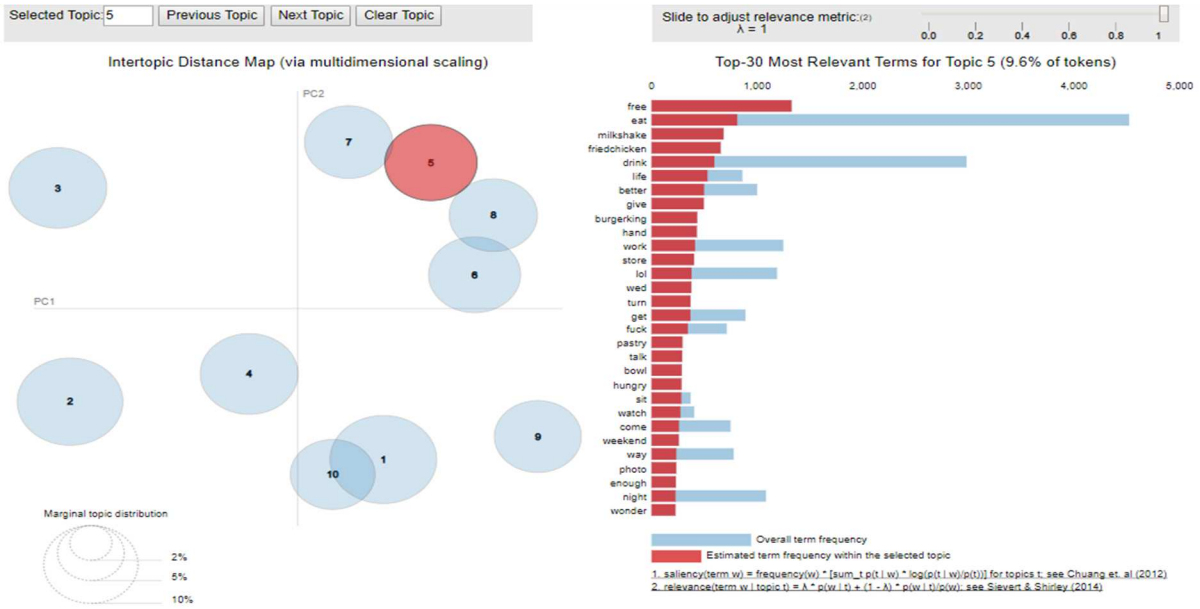

Figure 8:Twitter Food Topics with 10 Topics

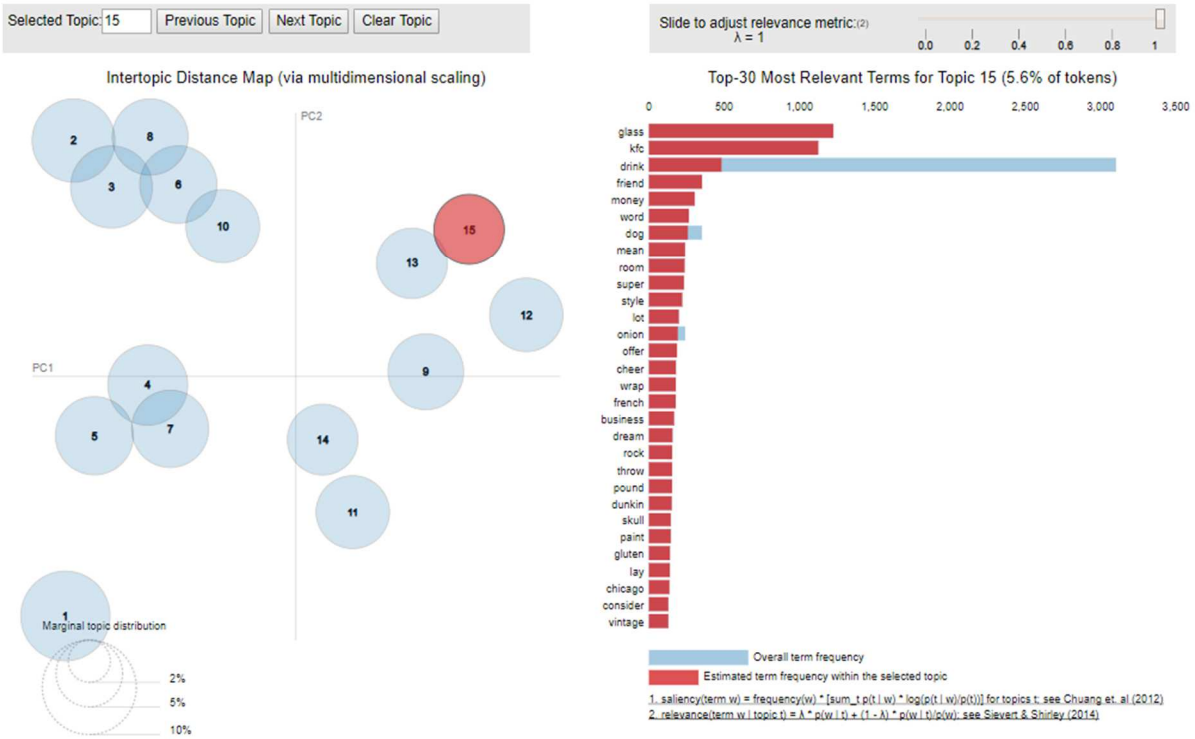

Figure 9:Twitter Food Topics with 15 Topics

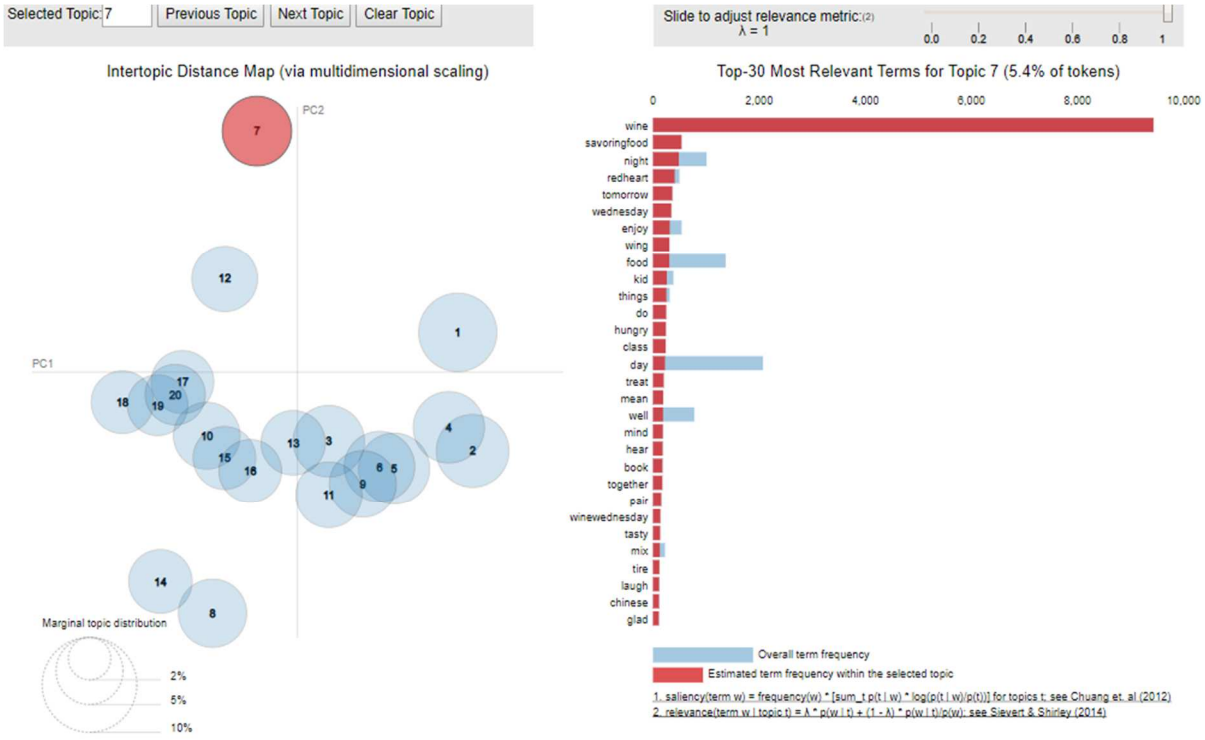

Figure 10:Twitter Food Topics with 20 Topics

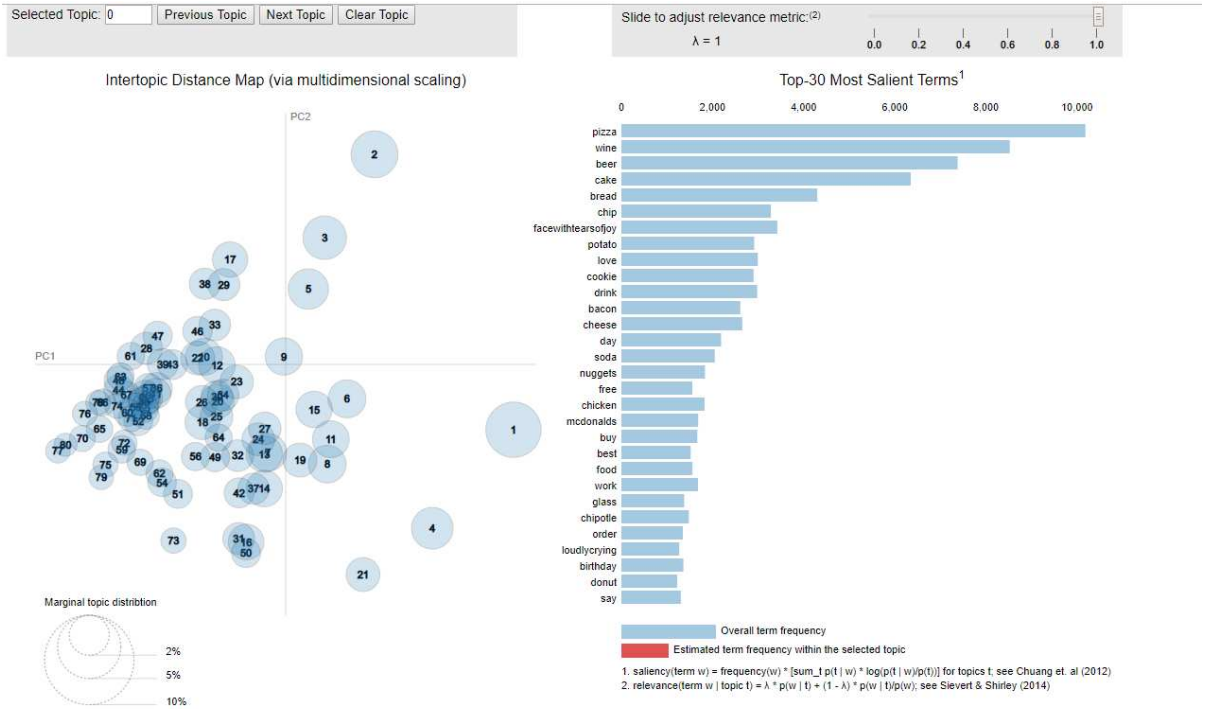

Figure 11:Twitter Food Topics with 80 Topics

Figure 7, 8, 9 and 10 show the Twitter food topics with 5, 10, 15 and 20 topics respectively. The value of k for those topics is randomly selected. Here some of the topics are overlapped. To find out the optimal number of topics we calculated model perplexity and topic coherence that provide a suitable measure to decide the “goodness” of the given topic model. But topic coherence gives more insights.

Figure 12 shows the perplexity calculated for 100 topics for the model. It is a metric to apprehend uncertainty in the model predicting topics for text. Lower the entropy lower in the perplexity of the topics. A model trained on good text and is being evaluated on fine test data, assign a higher probability, so the model has lower perplexity. In figure 12 topic 80 has a lower perplexity.

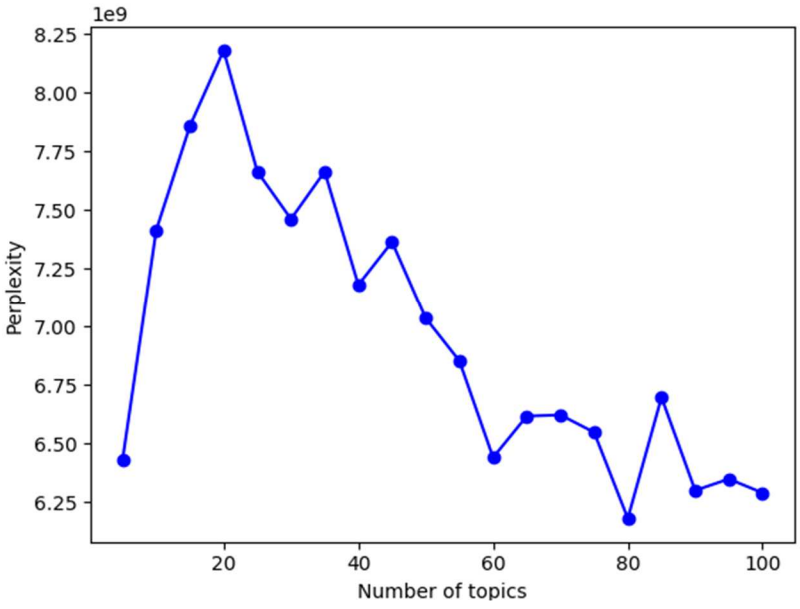

Figure 12: Perplexity score for Topic selection

In figure 11 we can see that 80 topics visualization is highly complex and most of the topics are overlapped. It is difficult to analyze the topics and find the relevant terms in it.

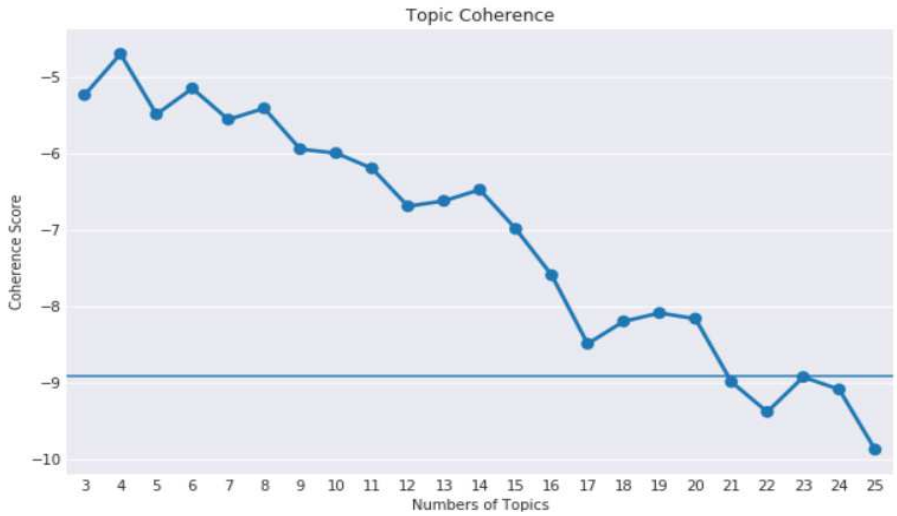

Figure 13:Coherence Score for Topic selection

Figure 13 shows the result of topic coherence using U-Mass. The relevance of the topic is decreased with the decrease in coherence score.

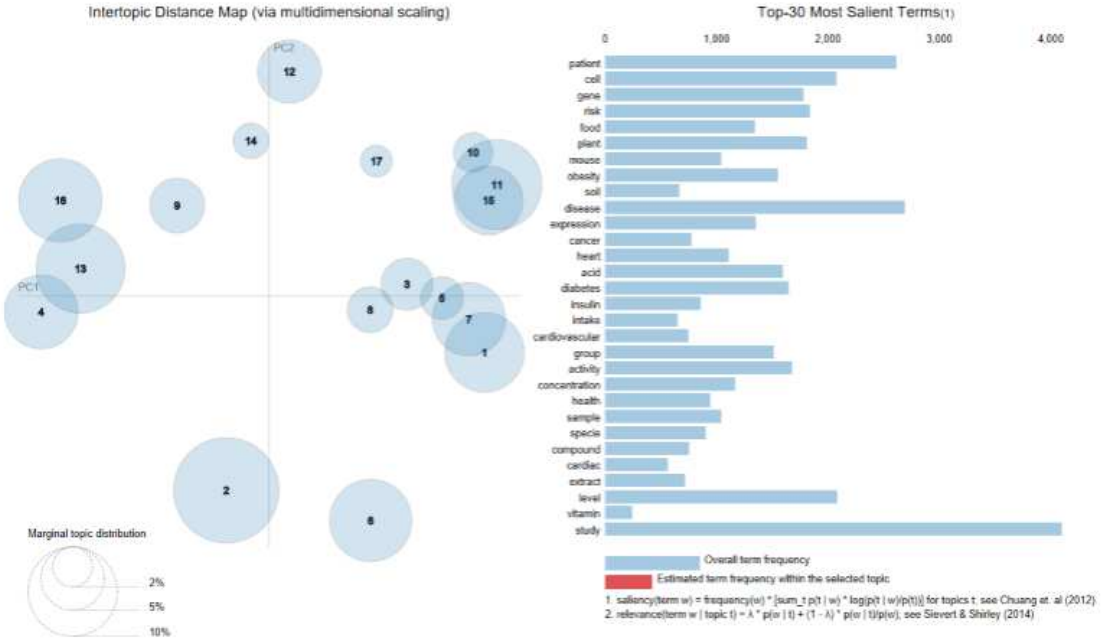

Figure 14:Twitter Food Topics

In Figure 14, each bubble is representing a topic. The size of the bubble shows the significance of the topic in a corpus. The topics which are similar appears closer, while less similar topics are farthest apart. Once the topic is selected the most relevant words and percentage of tokens of that topics appeared. Hovering over the word the topic size will be adjusted based on the representation of that words in the topics.

Table 24: Topics with Relevant Terms

Topic	Word with Relevance	Coherence score	Topic Name
Topic 2	cheese, slice, savoringfood, avocado, lunch, sandwich, breakfast, egg	-3.6935	Healthy Food Neutral Sentiment
Topic 5	Tomato, potato, sweet, recipe, soup, bean. potatoes, chipotle, rice, broccoli	-4.0078	Healthy Food Neutral Sentiment
Topic 10	sip, age, drink, Wednesday, pour, glass, bottle, soda, fine, wine	-4.9674	Unhealthy Food Neutral Sentiment
Topic 1	Dominos, kissblow, crust, deliver, order, hut, pepperoni, pineapple, roll, pizza	-5.290	Unhealthy Food Neutral Sentiment
Topic 6	craftbeer, smilingfacewithsmilingeyes, wink, beer, craft, brew, cold, celebrate, happy	-5.4506	Unhealthy Food Positive Sentiment
Topic 4	Facewithtearsofjoy, lmao, big, chip, fee, plaintain, salsa, store, chocolatecake	-5.5043	Unhealthy Food Positive Sentiment
Topic 15	cookies, macncheese, cup, smoothie, chocolate, strawberry, sour, drool, french-fries, blueberry	-5.663	Unhealthy Food Neutral Sentiment
Topic 16	loaf, love, duck, garlic, pudding, beamingfacewithsmilingeyes, parmesan, rollingonthefloorlaughing, cheesy, bread	-5.9627	Positive Unhealthy
Topic 13	tuna, pepper, kale, cry, chili, shrimp, hot, fry, pickle, heartaward	-6.443	Healthy Food Negative Sentiment
Topic 18	oreo, valentine, cheap, ice, book, hear, choose, pancake, forward, cream	-6.4662	Unhealthy Food Neutral Sentiment
Topic 7	king, skull, wed, taste, weary, sir, birthday, cute, pastry, cake	-6.6648	Unhealthy Food Neutral Sentiment

Topic 9	dad, mexican, guacamole, weight, restaurant, food, jimmyjohns, holy, lose, frown	-6.8406	Negative Healthy
Topic 19	winkingwithtongue, friedchicken, bucket, smilinghearteyes, kfc, eggplant, chicken, wings, nuggets, breast	-6.9125	Unhealthy Food Neutral Sentiment
Topic 17	peel, car, drop, jar, bed, juice, cranberry, orange, banana, hotdog	-7.8341	Healthy Food Neutral Sentiment
Topic 22	brown, coconut, oil, joke, water, skin, hair, olive, ketchup ,spice	-8.3578	Healthy Food Neutral Sentiment
Topic 20	little, onionrings, mushroom, pot, salmon, peas, boys, checker, plum, red	-8.7968	Healthy Food Neutral Sentiment
Topic 8	word, like, send, reason, carrot, know, smile, baby, krystal, tell	-9.0262	Mixed Food Neutral Sentiment
Topic 14	ass, vinegar, peach, onion, chicago, west ,wit, kiss, burgerking, mean	-9.0327	Healthy Food Neutral Sentiment
Topic 12	crumb, hes, fix, orangs, cupcake, dunkindonuts, swear, ginger, class, friend	-9.2008	Unhealthy Food Neutral Sentiment
Topic 11	grinningfacewithsweat, redheart, bacon, sausage, dough, delivery, scout, damn, cookie, card	-10.804	Unhealthy Food Neutral Sentiment
Topic 3	olives, loudcrying, three, tide, donut, mango, apples, donuts, rollingeyes, pod	-11.754	Healthy Food Neutral Sentiment
Topic 21	Kiwi, twohearts, million, pie, fight, naked, key, fake , news, meet	16.444	Healthy Food Neutral Sentiment

Table 24 represents the topics with relevant terms, the topic with high coherence score has good topic terms so we can identify the better topic description either the topic is Healthy food Positive Sentiment, Unhealthy food Negative Sentiment, Healthy food Negative Sentiment or Unhealthy food Positive Sentiment so on. As the topic score is getting lower, the description of a topic is becoming unidentified.

CHAPTER 5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this thesis, we presented the implementation of contextual sentiment framework with improved rules to increase the capabilities of existing models. The rules are focusing on mismatched cases identified using Machine-Human and Machine-Machine analysis. The mismatch cases cover ambiguity, sarcasm, emoji, positive, neutral and negative. These mismatched cases remain unidentified in existing sentiment models namely Vader, Text Blob and CoreNLP. The MMTSA framework shows good accuracy on a benchmark dataset.

Moreover, we discover hidden trends and sentimental topics from the Twitter dataset using LDA topic modeling technique. To find the optimal topic number we calculated different measures namely perplexity and topic coherence. We also find the topic with relevant terms to extract a good topic from the corpus.

For a case study, sentiment and topic discovery on Twitter food data are applied to identify the eating trends of social media users. We developed a food classification model to the categorization of topics as Healthy food Positive sentiment, Healthy food Negative sentiment, Healthy food Neutral sentiment, Unhealthy food Positive sentiment, Unhealthy food Negative sentiment, Unhealthy food Neutral sentiment, Mixed food Positive sentiment, Mixed food Negative sentiment and Mixed food Neutral sentiment.

5.2 Future Work

In the future, we are planning to develop an integrated multi-modality (text model + image model) for a comprehensive interpretation of tweet messages. The model will be

able to identify the eating items present in the images and evaluate the food sentiment of the image. We are willing to extend it to large-scale and long-term tweet sentiment analysis framework. In the future, the framework will be able to assign sentiment to a huge amount of data. Addition of more rules and lexical features in a model to improve the accuracy of the model.

BIBLIOGRAPHY

[1] Blei, David M., Andrew Y Ng, and Michael I Jordan. “Latent Dirichlet Allocation”. *Journal of Machine Learning Research*, vol.2, pp993–1022, Jan. 2003.

[2] Girolami, Mark, Kaban.” On an Equivalence between PLSI and LDA” in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003, pp. 433-434.

[3] Hu, Han, NhatHai Phan, James Geller, Huy Vo, Bhole Manasi, Xueqi Huang, Sophie Di Lorio, Thang Dinh, and Soon Ae Chun. "Deep Self-Taught Learning for Detecting Drug Abuse Risk Behavior in Tweets." presented at the *7th International Conference, CSoNet 2018, Shanghai, China* ,2018.

[4] Jangid, Hitkul, Shivangi Singhal, Rajiv Ratan Shah, and Roger Zimmermann. "Aspect-Based Financial Sentiment Analysis using Deep Learning." in *Proceedings of the Companion of the Web Conference 2018 on The Web Conference,*2018, pp. 1961-1966.

[5]Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global Vectors for Word Representation." in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)”*, 2014, pp. 1532-1543.

[6] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed Representations of Words and Phrases and their Compositionality." in *Advances in Neural Information Processing Systems*, 2013, pp. 3111-3119.

[7] Lund, Kevin, and Curt Burgess. "Producing High-dimensional Semantic Spaces from Lexical Co-occurrence." *Behavior Research Methods, Instruments, & Computers,*vol.28,pp 203-208,1996.

[8] Levy, Omer, and Yoav Goldberg. "Linguistic Regularities in Sparse and Explicit Word Representations." in *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*,2014, pp. 171-180.

[9] Wadhawan, Kahini. "Investigation of Word Representation Methods for Biomedical Domain." , Ph.D. dissertation, the University of Colorado at Boulder, USA, 2016.

[10] Socher, Richard, John Bauer, and Christopher D. Manning. "Parsing with Compositional Vector Grammars." in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, vol. 1,pp. 455-465, 2013.

[11] Socher, Richard, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank." in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*,2013, pp. 1631-1642.

[12] Yau YH, Potenza MN.(2014,Oct.). “Stress and Eating Behaviors.” *Minerva Endocrinologica* , vol.33, pp.255-76, 2013.

[13] Nguyen QC, Kath S, Meng HW, Li D, Smith KR, Van Derslice JA, Wen M, Li F. "Leveraging Geotagged Twitter Data to Examine Neighborhood Happiness, Diet and Physical Activity.” *Applied geography (Sevenoaks, England)*,vol. 73, pp.77-88, 2016.

[14] Eichstaedt, Johannes C., Hansen Andrew Schwartz, Margaret L. Kern, Gregory Park, Darwin R. Labarthe, Raina M. Merchant, Sneha Jha. "Psychological Language on Twitter Predicts County-level Heart Disease Mortality." Internet: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4433545/.

[15] “Center for Disease Control & Prevention”.Internet: https://www.cdc.gov/.

[16] Paul, Michael J., and Mark Dredze. "You are what you Tweet: Analyzing Twitter for Public Health." *in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*,2011, pp.265-272.

[17] Madan, Anmol, et al.(2010) "Social Sensing: Obesity, Unhealthy Eating and Exercise in Face-to-Face Networks." in *Proceedings of the Conference of Wireless Health 2010, ACM,* pp. 104-110,2010.

[18] Scanfeld, Daniel, Vanessa Scanfeld, and Elaine L. Larson. "Dissemination of Health Information through Social Networks: Twitter and Antibiotics." *American Journal of Infection Control,* vol.38, p.182-188, April 2010.

[19] Poria, Soujanya, Iti Chaturvedi, Erik Cambria, and Amir Hussain. "Convolutional MKL based Multimodal Emotion Recognition and Sentiment Analysis." in *Data Mining (ICDM), 2016 IEEE 16th International Conference*, 2016, pp. 439-448.

[20] Go, Alec, Lei Huang, and Richa Bhayani. "Twitter Sentiment Analysis." Internet: https://nlp.stanford.edu/courses/cs224n/2009/fp/3.pdf .

[21] Dixon, Natalie, B. Jakić, Roderick Lagerweij, Mark Mooij, and Ekaterina Yudin. "Food Mood: Measuring Global Food Sentiment One Tweet at a Time."in *Proc. of Sixth International AAAI Conference on Weblogs and Social Media,* 2012.

[22] Lu, Ziyu, Nikos Mamoulis, Evaggelia Pitoura, and Panayiotis Tsaparas. "Sentiment-Based Topic Suggestion for Micro-Reviews." In *Association for the Advancement of Artificial Intelligence ICWSM*, 2016,pp. 231-240.

[23] Xiong, Shufeng, Kuiyi Wang, Donghong Ji, and Bingkun Wang."A Short Text Sentiment-Topic Model for Product Reviews." *Neurocomputing* , pp.94-102,2018.

[24] Lin, Chenghua, and Yulan He. "Joint Sentiment/Topic Model for Sentiment Analysis." In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 2009, pp. 375-384.

[25] Gilbert, CJ Hutto Eric. "Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text." In *Eighth International Conference on Weblogs and Social Media (ICWSM-14*),2014.

[26] "Stanford CoreNLP". Internet: https://stanfordnlp.github.io/CoreNLP/.

[27]"TextBlob : Simplified Text Processing". Internet : https://texblob.readthedocs.io

[28] Flegal, Katherine M., Margaret D. Carroll, Cynthia L. Ogden, and Clifford L. Johnson. "Prevalence and Trends in Obesity among US Adults, 1999-2000." *JAMA,* vol. 288, no. 14,pp: 1723-1727,October 2002.

[29] "USDA MyPlate. The USDA MyPlate (2015-20 dietary guidelines for Americans for children)", Internet: https://www.choosemyplate.gov/snapshot-2015-2020-dietary-guidelines-americans, 2018.

[30] "USDA. The USDA Standard on Food and Nutrition", Internet: https://www.fns.usda.gov/usda-standardized-recipe, 2018.

[31] "BusinessInsider. The 8 Unhealthiest Restaurant Meals in America", Internet: https://www.businessinsider.com/most-unhealthy-meals-in-america-2017-7, 2017.

[32] "Eatthis. The #1 Worst Menu Option at 41 Popular Restaurants", Internet: https://www.eatthis.com/restaurant-menu-worst-options/, 2017.

[33]" For Academics - Sentiment140 - A Twitter Sentiment Analysis Tool", Internet: http://help.sentiment140.com/for-students, 2018.

[34]" Amazon Review Data", Internet: http://jmcauley.ucsd.edu/data/amazon/, 2018.

[35] "Large Movie Review Dataset", Internet: http://ai.stanford.edu/~amaas/data/sentiment

[36] "What is Topic Coherence?", Internet: https://rare-technologies.com/what-is-topic-coherence.

[37] "Evaluation of Topic Modeling: Topic Coherence", Internet: https://datascienceplus.com/evaluation-of-topic-modeling-topic-coherence.

[38] "Learn to Find Topics in a Text Corpus", Internet: https://medium.com/@soorajsubrahmannian/extracting-hidden-topics-in-a-corpus-55b2214fc17d.

[39] “A.I Wiki” , Internet: https://skymind.ai/wiki/accuracy-precision-recall-f1

[40] “sklearn.metrics.precision_recall_fscore_support — scikit-learn 0.20.1” ,Internet: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.fbeta_score.html

[41] Anastasi, Anne, and Susana Urbina. *Psychological Testing*,Prentice-Hall,1997.

[42] Viera J.,Joanne M. Garrett.” Understanding Interobserver Agreement: The Kappa Statistic Anthony”. Internet: https://www.ncbi.nlm.nih.gov/pubmed/15883903

[43] Sooraj Subrahmannian ,”Learn to Find Topics in a Text Corpus” . Internet: https://medium.com/@soorajsubrahmannian/extracting-hidden-topics-in-a-corpus-55b2214fc17d

VITA

Sidrah completed her Bachelor's degree in Computer Science and Information Technology from NED University of Engineering & Technology, Karachi, Pakistan. She started her Masters in Computer Science at the University of Missouri-Kansas City (UMKC) in January 2016, with an emphasis on Data Sciences and graduates in December 2018. While she was studying at UMKC, she has worked as a Graduate Teaching assistant for Advance Software Engineering ,Big Data Analytics and Applications, and Python & Deep Learning Courses. She was also a Data Science Intern at IBM from September2018 to December 2018. Upon completion of her requirements for the Master's Program, she plans to work as a Data Scientist/Machine Learning Engineer.