

Sentiment Analysis Study with an Emphasis on the Integration of Different Statement Polarities and the Evaluation of the Resulting Sentiments

By

Shawn Braddy

December, 2015

Director of Thesis: Dr. Mark Hills

Major Department: Computer Science

Social media has become an integral part of today's society and has continued to grow tremendously throughout the world. People communicate through social media constantly and social media has become a pivotal place for people to gather information related to things that they are interested in. It is also a place for people to express their views to friends, family and followers with just one click of a button immediately. Naturally in society there are a wide range of views encompassing an unlimited amount of subjects. Some of these views can be deemed negative, positive, neutral or maybe even other.

My research will attempt to identify the sentimental impact that different views in the form of social interactions online, can have on the people who are watching, listening and reading them. Furthermore, I will research how the integration of neutral and positive statements can stimulate positive and productive conversation amongst a community on a small scale as well as a large scale. Analysis of this will be quantified by changes in the overall reaction of the various communities time following intervention. The main contribution of this thesis is to show a correlation between statements with positive or neutral sentiment and a trend towards positive or neutral responses after moderation. Events of large discussion or interest will be evaluated to ensure a significant amount of data is available for analysis. In order to integrate the social aspect

into the equation, software will be explored and evaluated or built and then utilized to connect participants and extract the appropriate metrics.

Sentiment Analysis Study with an Emphasis on the Integration of Different Statement Polarities
and the Evaluation of the Resulting Sentiments

A Thesis

Presented To the Faculty of the Department of Computer Science

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Software Engineering

by

Shawn Braddy

December, 2015

© Shawn Braddy, 2015

Sentiment Analysis Study with an Emphasis on the Integration of Different Statement Polarities
and the Evaluation of the Resulting Sentiments

by

Shawn Braddy

APPROVED BY:

DIRECTOR OF
THESIS: _____

Mark Hills, PhD

COMMITTEE MEMBER: _____

Nasseh Tabrizi, PhD

COMMITTEE MEMBER: _____

Qin Ding, PhD

CHAIR OF THE DEPARTMENT
OF COMPUTER SCIENCE: _____

Venkat N. Gudivada, PhD

DEAN OF THE
GRADUATE SCHOOL: _____

Paul J. Gemperline, PhD

DEDICATION

To list the number of unique individuals that I would like to dedicate this thesis to, would likely take up the next several pages. Together my family, friend and mentors have pushed me to limits that I never imagined I would reach. Because of them ideas flow continuously through my head and because of them I push to be the best person that I can be towards everyone regardless of academia. It's a lot easier to read this dedication then to read this paper, but I encourage you to do so and provide me with feedback so I can improve for the people who believe in me.

ACKNOWLEDGMENTS

Dr. Nasseh Tabrizi, paved a road for me when I stumbled into the world of Computer Science and Software Engineering. It was he that taught me that to be a Software Engineer requires a plethora more than just knowing how to program. For this and many other things I thank him. Dr. Mark Hills's knowledge in this field is beyond a regular number², and I pray that I retain 40% of the information he so tirelessly embedded with me for over a year now. Lastly, I would like to acknowledge Dr. Qin Ding who truly understands the balance between learning and limits. Her efforts are equivalent in mentorship and academia to a select few people on earth. It is an honor to have these individuals on my acknowledgements page.

Table of Contents

DEDICATION	iv
ACKNOWLEDGMENTS	v
Introduction.....	1
Related Work	4
Background.....	6
Social Media and Sentiment.....	6
Sentiment Tools For Use	13
Politics, Government and Social Sentiment.....	15
Participant Selection	19
Initial Communication.....	19
Platform Construction.....	21
Choosing Topics.....	26
Analysis of Comments.....	27
Initialing Study.....	31
Results	32
Statement 1:.....	32
Statement 2:.....	36
Statement 3:.....	40
Statement 4:.....	44
Statement 5:.....	47
Statement 6:.....	50
Statement 7:.....	53
Statement 8:.....	56
Summary	59
Conclusion	61
References.....	62
Appendix – IRB Exempt Certification Form.....	64

Introduction

Opinions can be found everywhere you turn. Everyone has some type of opinion about an issue and many times it is safe to say that person is either for or against something. With the rise of social media over the past few decades it has become increasingly easier for people to find out things that are going on and express their opinions about the situation. There are wide varieties of tools that are used for this exact reason alone and social media is a booming business right now with millions of dollars made on a daily basis just from having a stream of people express their minds all day every day. It is important to define what sentiment is in order to understand this topic. Sentiment is an attitude toward something or the thought or feeling intended to be conveyed by words, acts, or gestures as distinguished from the words, acts, or gestures themselves. (Sentiment 2015) In other words, think about how you feel when you are in the process of expressing your feelings towards anything. It is likely that if someone were to ask you how you were feeling at the time, you are going to be able to give them an explanation of that feeling. This in itself is what the term sentiment refers too and the pivotal piece to what will be examined throughout this thesis. To be clear, the contributions of this thesis are as follows:

- The overall sentiment of a person can be changed towards a more positive or neutral sentiment through moderation using positive or neutral statements.
- This change can be tracked by analyzing a series of responses before and after moderation and quantifying the results.
- Analysis will show a measurable, predictable change in the sentiment of social media discussion by introducing feedback of the desired sentiment.

In order to fully understand the contributions that I will make to this field, I am going to delve into a few areas of sentiment analysis as a whole. This will provide a backbone into why this topic is so important in today's society in a variety of different areas including, politics, business, community development, education and other areas. In order to do this, I will take a look at some studies that have been used in the past in the form of social interaction, systems or analysis of data to determine the sentiment of people. This will help you understand the sheer magnitude of data that is being utilized currently in several different areas worldwide. I will also go in depth into how the ability to determine sentiment and adjust to it is being used in real companies to grow, avoid potential problems and of course find ways to monetize based on how people are feeling.

Another topic of importance is how exactly sentiment can be determined without specifically asking someone how they felt or what they meant by what they said or did. It is significant to note that in the context text of determining how someone felt throughout the majority of this paper, I will be referring to sentiment expressed throughout written text, either through social interactions, reviews in a variety of areas, or data and text mining scenarios. There is a lot of work that goes into evaluating the sentiment of what someone has typed, and the percentages of accuracy within these results are pretty eye opening. It really allows us to have an accurate representation of what someone may be feeling about a subject, at the point in time when they were expressing their thoughts. I will look at what tools are out there in this area and discuss how the tools come up with these results with so much accuracy.

All of this begins to lead up to my study and my contribution to this field. I am particularly concerned with how integrating neutral or positive sentiment statements can change the sentiment of people who are communicating on a particular subject. When looking into this,

it is necessary to understand why this would be important overall. In other words, why changing the sentimental mindset of someone can be important in business, socially, politically and a variety of other areas. I will provide insight into this area and point out the benefits of this using real world examples.

This brings me to the bulk of my contribution and how I will really evaluate how the integration of neutral and positive sentimental statements can cause a change in the resulting communication. I set up a real time social system that will allow participants to communicate with one another on a variety of current topics. Throughout this study, I will utilize the results to see the effects of integration in real time, to actually see if this changes the way people feel about a certain topic. I will use multiple tools that are used within the field of sentiment analysis to read the sentiment of user's responses and also tools to analyze the results. This will allow me to see if the polarity of the responses has changed significantly due to the integration of neutral or positive statements.

Finally, I will discuss how my findings can be used in the real world today along with what parts of society can benefit from this type of work. I will also provide examples of how this topic can be expanded on and improved in the future. My ultimate goal is to be able to strategically communicate with people online to directly help people to understand different sides to an argument and come to a positive resolution. This could be of benefit anywhere from a conflict arising between two countries, to bouts of racism due to police brutality, to a misguided ad put out by a company.

Related Work

There have been studies done in the past that in one way or the other relate to this thesis topic, although none directly. A study in 2013 set out to determine if there were any differences in social media sentiment based off of gender, age and country. (Pettit 2013) They examined millions of social media verbatims and scored them based on sentiment. Ultimately they found that women, older people and Austrailians show more positive sentiment and Americans show more negative sentiment. (Pettit 2013) In 2014 “FaceBook revealed that it had manipulated the news feeds of over half a million randomly selected users to change the number of positive and negative posts they saw. It was part of a psychological study to examine how emotions can be spread on social media.”(Goel 2014) The researchers selected 689,003 random users and found that moods were contagious. Positive posts led to more positive posts, while negative posts led to more negative posts. (Goel 2014) Both of these studies show a correlation between a person and the sentimental value of their communication online. The work here is different in that it evaluates the concept of introducing positive or neutral statements to real world conversations to see if this can lead the conversation in a more positive direction.

Another area of related work is sentiment analysis in social media. A 2012 study set to determine if it was possible to track online popular sentiment from twitter hashtags. “Considering hashtags original proposal and their current usage, we may hypothesize that they are also being used to express, group, and propagate people's sentiment with regard to some topics and events (e.g., “I watched amovie #IloveIt”), and may contribute to detect and track sentiment.” (Barbosa, 2012) The goal of this study was to get a better understanding on how hashtags are used on twitter to express sentiment. After randomly picking and analyzing 500 tweets, they found that in just 3% of the messages the hashtags contradict the message content. Ultimately, they concluded

that hashtags carry sentiments and there is a potential for using this to portray the sentiment of a population. (Barbosa, 2012)

Background

Social Media and Sentiment

The first type of sentiment analysis that I want to begin looking at is social media sentiment. There are many different social media platforms out there for people to use such as Facebook, Twitter, WhatsApp, Google +, WeChat and many more. “The unprecedented popularity of social platforms such as Facebook, Twitter, MySpace, and others has resulted in an unparalleled increase of public textual exchanges that remain relatively unexplored, especially in terms of their emotional content.” (Paltoglou, Gobron, Skowron, Thelwall, & Thalmann, 2010) Not only are there plenty of platforms, people have been using these platforms all across the world and at a very high rate. In fact, according to wearesocial.net, as of January 2015 there were 2.078 billion active social media accounts worldwide, with 1.685 billion of these also being active mobile social accounts.(Kemp,S) This is just a tremendous number and it is only increasing on a daily basis. Since 2014 alone there was an increase for the most active social networks per country of 12% in one year. (Kemp,S)

We are also seeing more and more time spent on the internet and social media. According to globalwebindex.net which annually polls 170,000 participants, the daily average of time spent on the internet has risen from 1.61 hours a day to 1.72 hours a day with social networking accounting for a whopping 28% of that time. (Mander, J.)

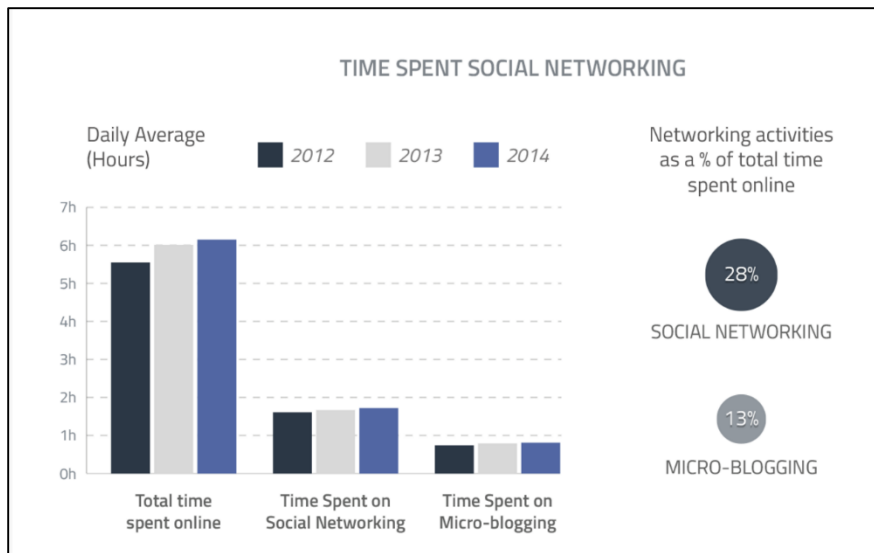


Figure 1

via globalwebindex.net

In figure 1 above, you can clearly see the networking activities trend going up over from 2012-2014 and the amount of time that is spent on average going increasing. There is no reason to believe that this trend will stop with the number of people around the world who have access to the internet rising. Internet growth globally has close to 3 billion and mobile use of mobile social media is on the rise with 77% of social network users accessing via mobile devices. (Kemp,S) This is significant when paired with the fact that more 3.6 billion unique users around the world now have access to cell phones. (Kemp,S)

With this much time spent on the internet and a large portion of that time spent on social media, it is understandable that a treasure trove of information is going to be put out there for analysis via social media. A lot of information comes in the form of communication by individuals to the world and it is this information that we then peruse though to see if can discern some type of sentiment during this communication process.

There are many different ways to spin this data and get back things that are useful. For instance, a company called Biz360 built a project called IdolStats, and the purpose of idol stats

was to predict the weekly results of American idol voting by analyzing the sentiment of towards contests via social media. This information was combined with the volume of social media mentions to predict the results with accuracy. (Ogniva, M) Another example of utilizing social sentiment comes from a 2010 study where 3 million tweets from twitter was used to predict box-office revenues of movies prior to their release. (Asur & Huberman, 2010) They were able to utilize tweets a week in advanced and predict the revenue for that particular weekend. By analyzing the change in polarity over the course of the week, they were able to see a significant resemblance in revenue made that week. “The movie with the enormous increase in positive sentiment after release is The Blind Side (5.02 to 9.65). The movie had a lukewarm opening weekend sales (34M) but then boomed in the next week (40.1M), owing largely to positive sentiment. The movie New Moon had the opposite effect. It released in the same weekend as Blind Side and had a great first weekend but its polarity reduced (6.29 to 5), as did its box-office revenue (142M to 42M) in the following week. (Asur & Huberman, 2010)

Business and Social Sentiment

What customers and potential customer think about the services and products of a company is very important. In fact, it could very well be the difference between a successful company that has loyal customers and a company that is struggling to make a sale. For this reason alone, it is vital that a business understands what a customer feels about their products and services. These days, it is normal for a company to have customers all across the country and even across the globe at that. Communication becomes vital for businesses with a reach this

large because for one, obviously you may not be able to actually see all of your customers to directly communicate with them face to face and discern how they are feeling about things. This is where sentiment analysis can play a major role. What if you as a company, can figure out how people are feeling about your products and services, without actually having to talk to them directly. You could potential deal with issues in real time before it becomes a major problem for other customers. Businesses can also benefit by saving time and money on traditional avenues that they used to bring in information about consumers perspectives.

“Seeking consumers’ opinions has traditionally been a difficult and time-consuming task, which usually involves hiring expensive consulting firms to conduct surveys on representative groups of customers. The results of these surveys are restricted to a very limited sample of customers and need to be extrapolated from the sample to the population. Moreover, these surveys are only valid for a short time period, since customers’ opinions change fast nowadays, and frequently due to things that are completely out of the companies’ control.”

(Plaza & Carrillo de Albornoz 2012)

This speaks directly to the ability for businesses to utilize sentiment analysis online to get more information cheaper.

It is also clear that consumers are getting valuable data about things products and services beforehand and providing feedback afterwards. Two surveys were done with more than 2000 American adults to gather information about how much research they do to obtain and provide information on products that they are interested in. The results found had some interesting figures:

- 60% of Americans have done online research on a product at least once.
 - 15% of all Americans do so on a normal day.
 - Among readers of online reviews of restaurants, hotels, and various services. (e.g., travel agencies or doctors), between 73% and 87% report that reviews had a significant influence on their purchase.
 - Consumers report being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item (the variance stems from what type of item or service is considered).
 - 32% have provided a rating on a product, service, or person via an online ratings system, and 30% have posted an online comment or review regarding a product or service.
- (Pang & Lee, 2008)

This information clearly shows that customers are engaging in online activity when it comes to researching products and they are also providing feedback about their experiences with products.

Sentiment analysis can also help companies deal with problems in real time. Because many social media tools allow for public access, it makes it a lot easier for someone to search and see if their business is being mentioned via these social networks. Consumers can also directly “mention” businesses via some platforms to get their attention directly and express how they are feeling. Many companies even have dedicated employees whose job literally is to manage and respond to situations via social media.

Real time sentiment analysis becomes even more powerful when the communication is good amongst different departments and problems can be dealt with immediately. For example, a large airline company monitored tweets being sent concerning their flights and how customers

were feeling about various things such as delays, upgrades, new planes entertain etc. This information was then fed into a customer support platform for the airline. It was determined on the flight that a customer tweeted negatively about losing their luggage before boarding a connecting flight. The airline was made aware of the customers tweet, and directed an employee to wait for the customer to land at their destination and offer them a free first class upgrade on their return flight. They also tracked the customer's bag and provided him with information on exactly where he could pick it up when he got off his flight. The customer responded to this great customer service with another tweet expressing how happy he was with the response. (Nuseir, R) This is a clear example of how being able to understand the sentiment of a customer's tweet allowed a company to make the customer happy and potentially show other people how they are able to handle problems appropriately.

Businesses are also relying on social media more and more to directly market to potential customers on social media. Salesforce surveyed 5000 companies and 70% of these companies plan to increase social media advertising. 66% of these companies believed that social media was core to their business and 66% had a dedicated social media team. (Heine 2015)

Targeting Sentiment in Business

At this point it is understood that with one tweet a company can get information out there to their followers and get the word out about what they are offering or selling. Possibly even more important than actually getting information out about a product or a service is finding out what response the customers have based on the product or service. "In their study, Jansen et al. (2009) have found that 19% of a random sample of tweets contained mentions of a brand or product and that an automated classification was able to extract statistically significant

differences of customer sentiment (i.e., the attitude of a writer towards a brand).” (Tumasjan, Sprenger, Sandner, & Welppe 2010) This shows that people are directly mentioning companies and products when they have something on their mind. A key job task for a business’s social media expert is figuring out exactly what a customer is feeling about the company at any given time. As mentioned before, this is very important because in real time it can allow a company to get ahead of the competition, adjust some marketing campaign on the fly or resolve an issue. Clearly businesses are investing in these types of individuals who know how to figure out social strategies. Searching the platform LinkedIn for Social Media Specialist yields over 5,000 jobs all over the world. These companies include well know companies such as Jaguar Land Rover, CarMax, NBC Universal, Nordstrom, Deloitte, Ceaser’s Entertainment and even Universities such as Arizona State University. When looking through the job descriptions of these opportunities, you see a lot of verbiage that indicates understanding sentimental values. Examples include:

- “Provide recommendations to optimize content & channel activation to improve performance and the consumer experience” via Land Rover
- “Utilize strong analytical ability to evaluate the end-to-end consumer experience across responsible areas to deliver strong ROI, maximize performance & provide recommendations for optimization” via Land Rover
- “Reviewing and moderating conversation around billing, devices, and troubleshooting flows across all channels” via T-Mobile

Each one of these examples have the underlying criteria of understanding your customer base and reacting appropriately. This points directly to figuring out the sentiment value of what

customers are saying and doing online about you, your products are things that relate to your services.

So we know that through social media, a customer can immediately express their opinion to all of their followers and let them know exactly how they feel about a certain product. The question then becomes, how can you take this information as an organization and decipher it quickly as it is coming in to figure out how people are actually feeling. How can you get overall figures on what a person is feeling? There are tools out there that attempt to address this problem directly and give feedback into the sentiment of customers based on social media responses.

Sentiment Tools For Use

If you simply do a google search for what tools are currently out there that deal with sentiment analysis, you will get a lot of results. For example, I did a search and over 2.5 million results come up at the time of this research. Many of them allow you to link in to your own social media devices and track your followers to get a better understanding of how your followers are reacting to what you are saying. Later in the thesis I discuss the tool I used to assist with my study and why. If you are a company and you are looking to get a more in-depth about what others are saying about your company and feeling about your company, likely you are going to have to pay some type of fee to get this information. Ultimately if you purchase the services of one of these companies, there is going to be algorithms that claim to be able to decipher social media posts and decide on an individual post basis, what a person is actually feeling.

In order to understand what is out there, let's take a look at a few examples.

Salesforce.com indicates that it analyzes conversations from over 650 million different sources to find out what customers are saying about your products, competitors and brand. You can also utilize a feature to monitor the sentiment of consumers so that you can stay ahead of potential problems and evaluate new products and services. Directly from the tool, you have the ability to connect a variety of social platforms and respond from there as well. SocialMention.com allows you to search for a word or words and provides feedback from a variety of areas about the word you searched for. For example, searching for Walmart displays a sentiment ration of 31:1 and lets you know that the platform determined that the word had 61 positive sentiment mentions, 100 neutral mentions and 2 negative mentions. Of course, this is only between a certain timeframe which can also be adjusted. There are many more tools out there that companies use to analyze sentiment and manage social media at the same time. These tools are only growing and getting smarter as algorithms are tuned to understand more and more data is made available via the internet.

Other tools out there include tools that directly deal with sentiment analysis of text in that you do not have to connect a social media account, you can simply input your text and it will be analyzed for sentiment as well as a variety of other attributes. These tools include Semantria, SentiStrength and LIWC (Linguistic Inquiry and Word Count). Semantria's website describes this tool as being a highly customizable sentiment analysis, categorization & named entity extraction tool. They have an downloadable extension for excel that you can use to input data and get back sentimental results. SentiStrength describes itself as providing the ability to estimate the strength of positive and negative sentiment in short texts, even for informal language. It has human-level accuracy for short social web texts in English. LIWC works similarly in that it is

able to detect sentiment strength along with a variety of other word associations and display this information in a well organized and easily readable format. A drawback to these tools would be that they are not 100% accurate and it can be difficult to detect certain parts of speech such as sarcasm. LIWC website notes that it does not understand irony sarcasm or metaphor's. Of course this can lead to misreading the sentiment of some phrases due to the prescence of these attributes.

Politics, Government and Social Sentiment

Politics is also a tremendous area where sentiment analysis can be useful. This does not only refer to the election of officials, but other things such as feelings throughout the community and what things are important to your community. Understanding how your community is feeling about certain things can be vital when it concerns the wellbeing of a community as well as the response that elected officials give towards any variety of situations. We have certainly seen examples over the past few years where the responses of the government did not sit well with the people and ended up going to a bad place that no one really predicted. In many of these situations, the problem could have been handled differently if community officials really got a good understanding of what the people were actually feeling and dealt with it accordingly. Social media allows the possibility for this because if there is a situation that is going on, people are going to go to social media to express how they feel. Many times this is the only avenue that the average person may have to express their opinions. A 2006 survey of over 2500 American adults that were campaign internet users who gathered information about the 2006 elections online produced the following results:

- 28% said that a major reason for these online activities was to get perspectives from within their community, and 34% said that a major reason was to get perspectives from outside their community.
 - 27% had looked online for the endorsements or ratings of external organizations.
 - 28% said that most of the sites they use share their point of view, but 29% said that most of the sites they use challenge their point of view, indicating that many people are not simply looking for validations of their pre-existing opinions.
- 8% posted their own political commentary online. (Pang & Lee, 2008)

This data shows that even 9 years ago, people were utilizing the internet to gather information related to their community and politics. Obviously since then, people have even greater access to the internet and more information than ever is shared via the internet. In 2008, many people saw the election of Barack Obama as a very well put together strategy from his social media team as he was able to utilize social media outlets to get his ideas out to potential voters quicker than ever.

“Following Barack Obama's 2008 Presidential campaign, the world saw the crowning of the Internet as the king of all political media. Online activity indicators such as number of fans on Facebook, followers on Twitter, and likes on YouTube have been seen as indicators of a galvanized base, which ultimately contributed to Obama's victory.” (Mejova, Srinivasan & Boynton. 2013)

There have been a variety of studies done that analyzed sentimental values of tweets and used it to predict the outcomes of elections. “Wang et al. (2012) proposed a real-time sentiment analysis system for political tweets which was based on the U.S. presidential election of 2012.

They collected over 36 million tweets and collected the sentiment annotations using Amazon Mechanical Turk. Using a Naive Bayes model with unigram features, their system achieved 59% accuracy on the four category classifications.” (Bakliwal, 2013) In 2009, “One of the most influential papers is that of Tumasjan et al. (2010) who focused on the 2009 German federal election and investigated whether Twitter can be used to predict election outcomes. Over one hundred thousand tweets dating from August 13 to September 19, 2009 containing the names of the six parties represented in the German parliament were collected. LIWC 2007 (Pennebaker et al., 2007) was then used to extract sentiment from the tweets.” (Bakliwal, 2013) Tumasjan et al. concluded that the number of tweets/mentions of a party is directly proportional to the probability of winning the elections. (Bakliwal, 2013) These are just a few cases where sentiment analysis techniques were tailored to predict political outcomes.

Real time Social Sentiment Study

In order to directly determine and effectively analyze how people react amongst one another socially to different news events, I have created a small social platform that will allow users to interact with one another and react to the posted comment or news article. The overall purpose is to determine how neutral intervention throughout a conversation can influence the overall sentiment of the users during the course of the conversation. Below I will outline the various different areas of this study, along with describing what analytical measures will be

utilized. Each comment or article used will have a specific purpose which will be outlined individually. Below is an outline of the specific steps and layout of the study.

- I will introduce a total of 8 topics via a customized website that I designed specifically for this purpose. The topics will be posted by a user named moderator.
- Each textual statement will be pre-analyzed to find out the sentiment value of the statement.
- Each user will be asked to respond at least once to the initial topic, and at least once to the comment of another user. They can also respond as many times as they like to other comments and this is encouraged.
- Every response will be analyzed for sentimental value to see if the sentiment is positive, negative or neutral and by what percentage.
- The Moderator will make one additional statement later after the initial discussion on the topic. This statement will be pre-analyzed for sentiment and will have a sentimental value of positive or neutral. The participants will be asked to respond to this statement.
- Each response to the positive or neutral moderator comment will be analyzed for sentimental value to see if the sentiment is positive, negative or neutral and by what percentage.
- The analytical results of all comments will be quantified to evaluate change in sentimental value throughout the course of the conversation.

Participant Selection

The participants are all volunteers who were selected by the researcher. Criteria for the participants were that they had the ability to access the social platform at all times in order to respond to the pending statement in a timely manner. For this reason the platform was built to cohesively work on mobile, desktop and tablet machines. Later I will provide further details on the building of this platform. In order to allow for a diverse range of sentiments and reactions, participants were specifically picked to be of varying races, ages and gender. The ages of the participants ranged from age 19-37. There were a total of 16 participants with 9 males and 7 females. The races of the participants were 6 Caucasians, 5 African Americans, 3 Indians, 1 mixed with black and white and 1 of Korean decent. Participates were also notified that their responses would be anonymous along with their username. The participants were informed via email of their usernames and passwords to log into the system. They were required to verify they were able to access the system before the beginning of the study and also provided with instructions on what they would need to do on a daily basis.

Initial Communication

In order to make sure everyone understood how they would be doing throughout the study, an initial email was sent out to all of the volunteers prior to the first post. These instructions were as follows:

Hello,

Thank you for participating in my thesis research. The overall topic of my thesis deals with social sentiment analysis and I will be utilizing the results of this research to analyze responses to a variety of different statements. Your responses to statements will be anonymous and your participation is 100% voluntary. With that being said, please express your opinions or feelings towards the given statement without holding anything back. Simply express what you feel.

Below I will outline the process and if you have any questions please feel free to reach out to me for clarification.

- I will introduce a total of 8 topics via a customized website that I designed specifically for this purpose. The topics will be posted by a user named moderator.
- Please respond at least once to the initial topic, and at least once to the comment of another user. You can certainly respond more than once and this is encouraged. The resulting conversation is what drives this research.
- The Moderator will make one additional statement on the same topic after the initial discussion. Please respond at least once to this statement as well. It will relate in some way to the initial topic. Again, respond at least once to the statement and also to a comment of another user.
- Feel free to agree or disagree with the moderator or someone else's comment if you wish to do so.

You should have received an email or will be receiving an email shortly requesting you to set your password to access the website. You will have to login in order to participate. Thank you for your participation and I look forward to working with you.

Regards,

Shawn Braddy

Communication Goals

The goal of the communication was to provide an outline of when posts should be expected and also to express that they should communicate with one another freely throughout the entire process. It was very important to try to simulate a natural conversation via a social platform in order to conduct research effectively. I also tried not to place emphasis on the fact that posts would be analyzed for sentiment purposes to limit the urge to write something just to sound a certain way. I really wanted to make sure the volunteers felt comfortable expressing their views. This is the reason why I chose to make everything anonymous and to emphasize this fact. In this way, people could react to the posts without wondering if someone would think

something of them personally. In social media today, many of the interactions that take place are from people that do not know one another at all. Of course, you also have those situations where a person will create an account for the sole purpose of replying to a post without being identified. I thought it was important overall to simulate this process.

Platform Construction

In order to conduct my study and make sure it feels like an actual social platform. I hosted the website on an independent server so I could have access to the database directly in order to allow me to quickly access and export the results in an organized fashion for examination. The database used was a MySQL, Apache and is running Linux. A WordPress backend was implemented on the server and I then customized this to remove unnecessary verbiage and indicate that this was the platform for a thesis social interaction study.

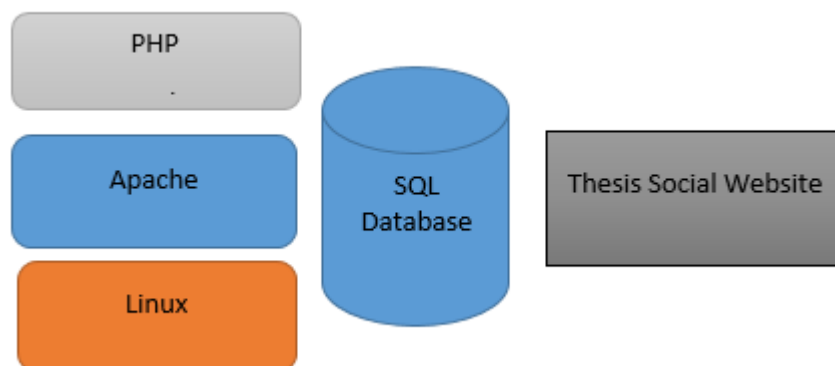


Figure 1.2

It was important for the participant to be able to access the platform from wherever they were at, so the website was customized to work well with on mobile, tablet as well as PC platforms. I also tested this to ensure that everything displayed and worked like it should. Performance issues could unnecessarily frustrate the participants and I did not want this to be an issue at all. The framework that this was built on has been thoroughly tested and utilized which gave me confidence that it would work well in this situation.

Another important aspect was the feeling that you are interacting with someone in a social way and not simply responding to statements. A lot of time was put in the design and implementation of this platform to ensure that this was the case. Each user has a customizable username that they have the ability to edit, with the criteria that the nicknames do not indicate who they are, or identify them in anyway. They were informed that the nickname should maintain anonymity. This allows the participant to customize their personality in whatever way that they like. They also have the ability to change the theme of their profile page to different colors to further enhance the feeling of their page. Another tool that they could use if they like, is the ability to add a status and update it whenever they liked.

Interacting with one another live is also a feature that is important on any social website or web domain. Facebook, Google and many others allow you to actually see who is online, and this helps creates the feeling that you are actually interacting with someone. For this reason, it was vital that you be able to see who was online at any given time so that you could directly interact with them. On the left hand toolbar of the website, the participants can see the avatar of anyone that is online and can also hover over the avatar to take a look at the nickname of the person. The social aspects were created using a mix of wordpress plugins and PHP customization. Another feature that was implemented to make this feel like a social media

platform was the ability to use the @ symbol to directly respond to someone else. This would let participants know directly if they were being responded to an addressed. You see similar things like this via the Twitter platform as well as Facebook. In order to get an understanding of the look and feel of the platform to better imagine the interaction that was ongoing, I have included screenshots as well as descriptions below:

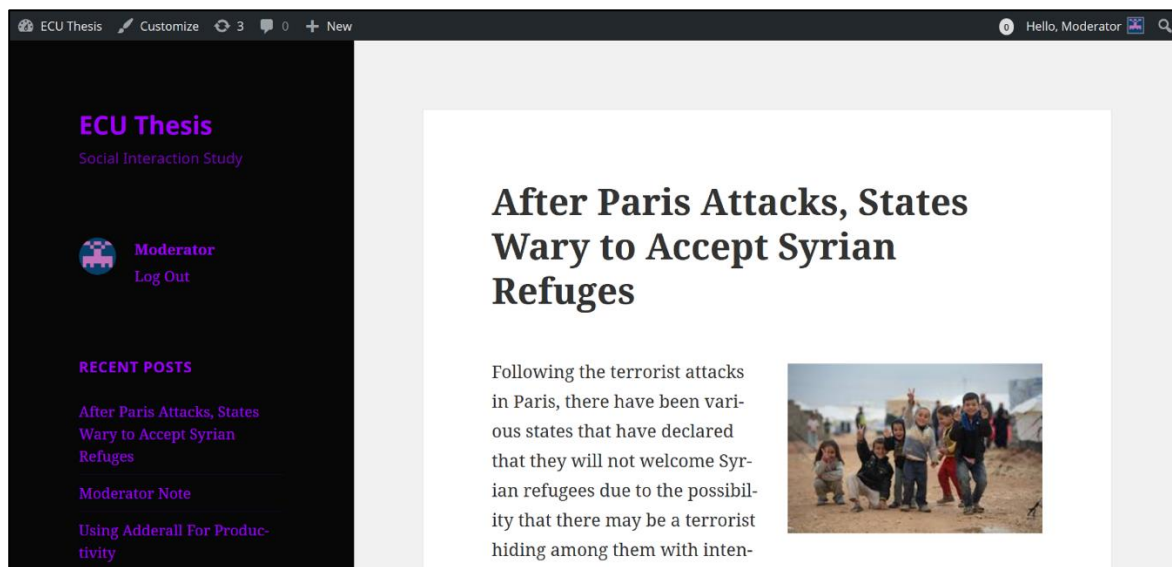


figure 2 - Homepage

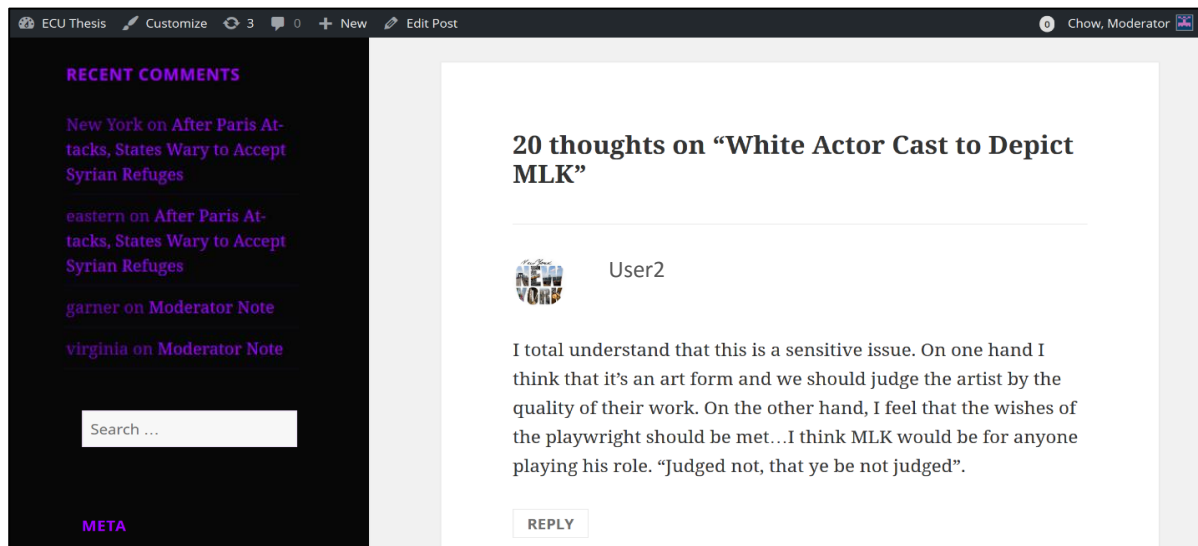


figure 3 - User Comment View

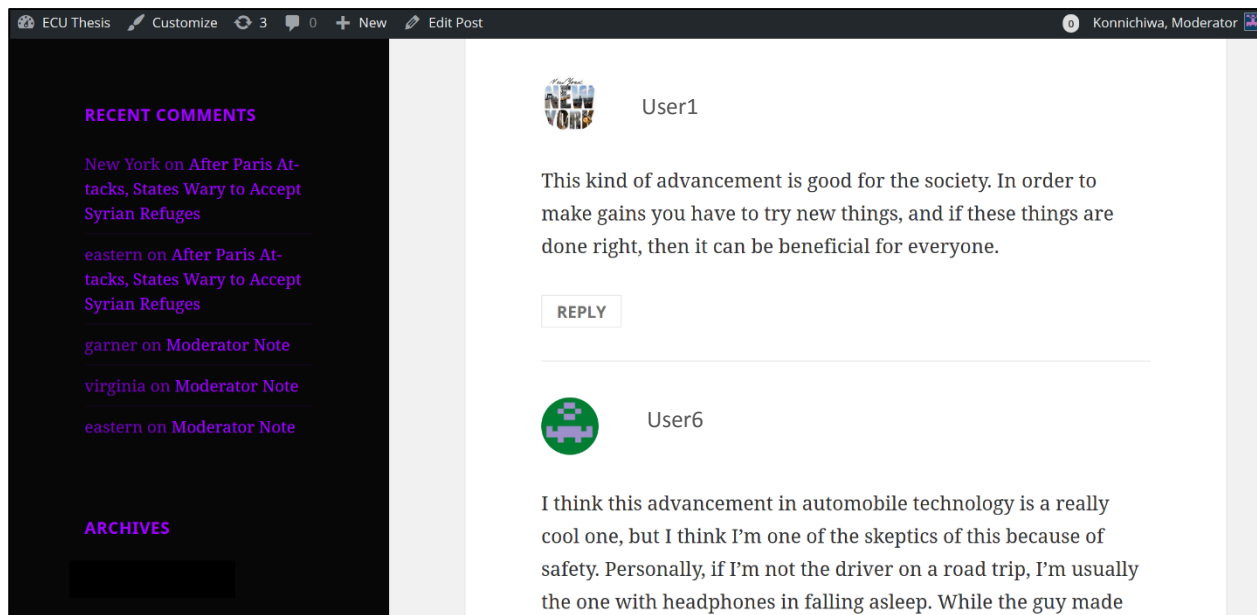


figure 4 – Multiple reply view

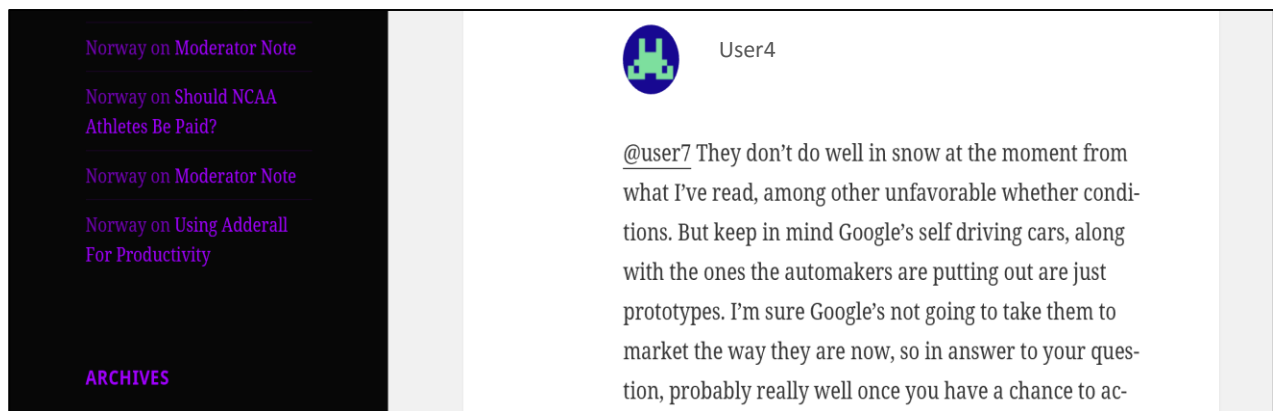


figure 6 – sample mention view

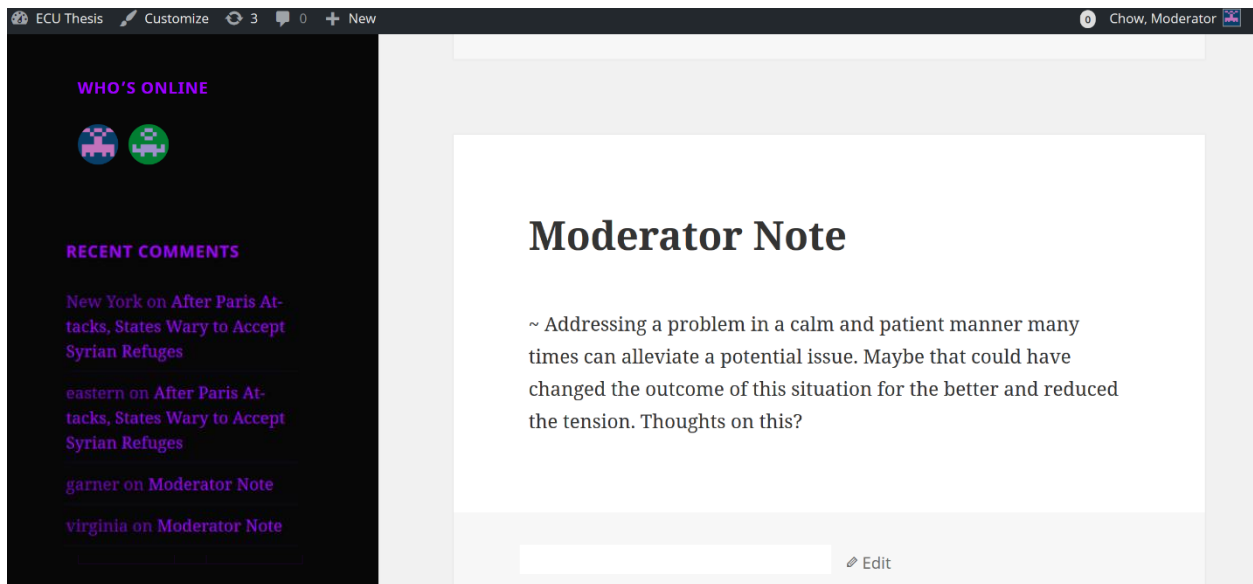


figure 7 – moderator note

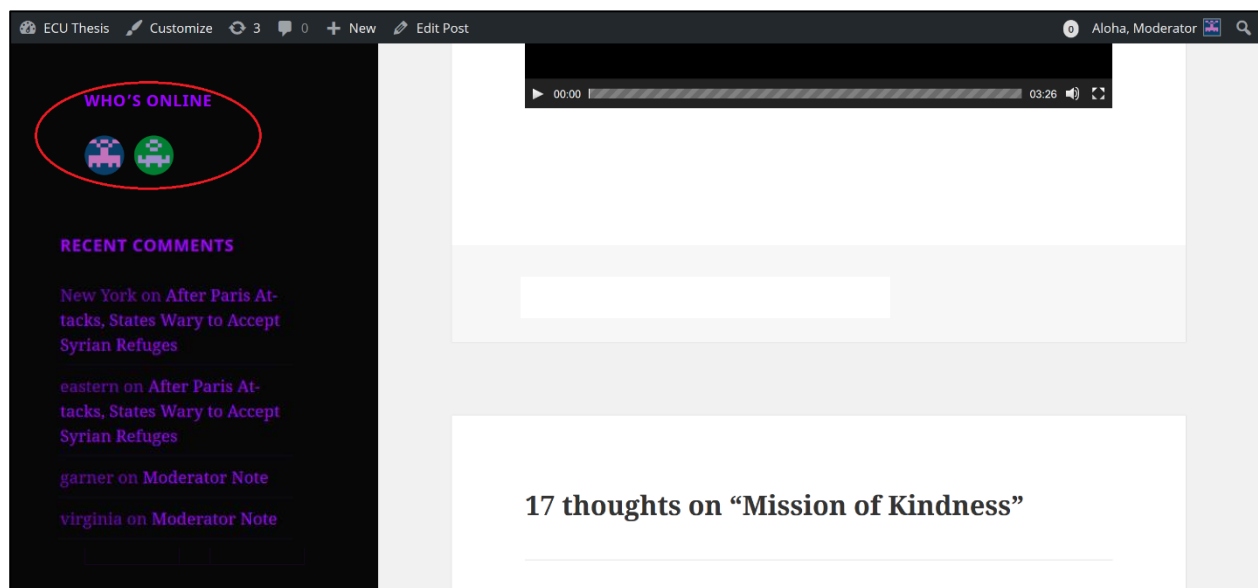


figure 8 – Users Online

Choosing Topics

As previously stated, a different topic of discussion was chosen for discussion every time. The topic selection was thoroughly vetted to really drive a conversation, and also pre-analyzed in a variety of ways. The title for each topic was also run through a sentiment analysis tool to get the sentiment of the title. For example, the second post's topic read "Altercation between police officers and students at The University of Alabama". This was carefully selected in an attempt to add no bias for the viewer when they read the title. Running this title through the sentiment tool provided a neutral sentiment value with an overall score of 53.4%. I changed the wording of the title to ensure that responses to this article will begin after looking at a neutral statement. Later, I will go into each topic, why they were chosen, how they were analyzed and expected results vs. results received. Changing the sentiment of the statement and the reaction afterwards was something that I wanted to make sure to keep track of.

Moving forward in the study, it was also important for me to have different types of topics. If you log onto twitter, Facebook, or just see what is trending around the world and creating conversation, you are going to see a wide range of different types of media. For example, doing a current Google Trends search, which is a tool that lets you see what trends are currently popular and people are talking about currently, brings up a wide variety of topics including, the republican debate, a fire in Oklahoma, a comedian Aziz Ansari, Bloomingdales, User3 Yankees etc. Looking at current trends on twitter brings you a variety of hashtags such as #Mizzou (In reference to protests on the University of Missouri's campus), #BlackOnCampus, Victoria's Secret, #WhatAFeeling (song by Beyonce), and #Biafra (referring to protests in Nigeria calling for the release of Nnamdi Kanu). The trend here is that people discuss a wide

variety of things and those things could be positive or negative. The aim is to simulate this randomness as much as possible.

Analysis of Comments

The overall goal in moderating the topics ultimately is to see how sentiment changes along the course of a topic of conversation once a moderator inputs a more neutral or positive comment. The idea is that when people are looking at things negatively, their feelings can possibly be swayed by reading and thinking about a positive point in the conversation. In order to do this, I begin by introducing a topic and analyzing the responses from everyone. Later on in the conversation, I introduce a statement via the moderator that is either positive, neutral or on the verge of being neutral. I then evaluate the responses afterwards to really see if this plays a part in changing the attitudes of the volunteers. I will also be evaluating responses between different users to see what happens when a positive or neutral comment follows a negative comment. Is there any correlation between the sentiment of the positive responding comment and the sentiment of the following comment? The list of topics include the following topics as well as moderator notes for each topic:

- Protests, Tension and Racism claim's drive University of Missouri President to Resign
- Altercation between police officers and students at The University of Alabama
- White Actor Cast to Depict MLK
- Mission of Kindness
- Self Driving Cars
- Using Adderall For Productivity
- After Paris Attacks, States Wary to Accept Syrian Refugees
- Should NCAA Athletes Be Paid

Social Sentiment Tools Used

In order to analyze the data, I initially did research on a variety of open source tools available to allow me to analyze text of any size and relatively accurately predict the overall sentiment value of the text. The tool that I chose to go with for the first part of the study was an open sourced tool located at <http://sentiment.vivekn.com/>. What drew me to this tool was that it had the ability to examine single words as well as words combined together in relatively small sequences. This was important because during the first phase of this study, there was criteria set on the amount of words that someone could type in response to a statement. I did emphasize “short” responses to try to limit the length of the responses, but I did not specify a length due to the fact that I wanted to get full and thought through responses to get participants used to expressing their feelings and used to the system as well. The tool has the ability to take in a large amount of text as well. In addition, when receiving results back about the sentiment value, this tool also provides you with a percentage of the confidence level in which the value was a certain type.

It is important to discuss how the sentiment tool makes its decisions in order to clearly understand what in effect can be classified as, positive, negative or neutral. This tool is based on the exploration of different methods to improve the accuracy of the Naïve Bayes classifier for sentiment analysis. Ultimately it utilizes a supervised sentiment classification model based on the Naïve Bayes algorithm. (Narayanan Vivek, 2013) The Naïve Bayes classification is based on Bayes’ theorem of posterior probability. It assumes class-conditional independence-that the effect of an attribute value on a given class is independent of the values of the other attributes. (Han Jiawei, 2011). In this case the classes are positive, negative or neutral.

Ultimately, naïve Bayesian classification looks at a training set of data and utilizes that data to predict what class the data may belong to. For example, in the scenario of this study, the Naïve Bayes classification would look at the data and determine whether a sentence is more likely to be positive, negative or neutral. It is also useful to take a look at how effective Bayesian classifiers are. “Various empirical studies of this classifier in comparison to decision tree and neural network classifiers have found it to be comparable in some domains. However, in practice this is not always the case, owing to inaccuracies in the assumptions made for its use, such as class-conditional independence, and the lack of available probability data.” (Han Jiawei, 2011) For example, earlier in the thesis I discussed how one of the evaluated tools was unable to understand sarcasm and this of course could lead to inaccuracies in predicting sentimental value.

The tool that I will be utilizing was trained using a dataset of movie reviews. It is a set of 25,000 highly polar movie reviews for training, and 25,000 for testing. Both the training and test sets have an equal number of positive and negative reviews. (Narayanan Vivek, 2013) This set and the algorithm itself derives from a supervised learning model in that the class label attributes are known and identified for the training data. It is important to note that this model being used has been modified utilizing the Naïve Bayes model along with linear training and testing time complexities to increase the accuracy of the results based on the type of information the algorithm is being applied to. This ultimately resulted in an accuracy of 88.80% on popular IMDB movie reviews dataset. (Narayanan Vivek, 2013) This is significantly important because it clearly shows that this can be used to relatively accurately predict sentiment of different given statements. The main reason I chose to utilize this tool to assist with this research domain is because it was trained using a large amount of data and the accuracy was still high in predicting the class attributes of positive or negative when testing sentiment of movies reviews.

In this Naïve Bayes model, the class attribute is given a class of positive or negative and it is assumed that the words are conditionally independent of one another. The formula that interprets the probability that the word belongs to some class is represented by the following expression

$$P(x_i | c) = \frac{\text{Count of } x_i \text{ in documents of class } c}{\text{Total no of words in documents of class } c}$$

It is important to discuss how a negative and positive word can be determined in sentiment analysis and also using the tool. Many sentiment analysis tools have a word bank which identifies what types of words are negative and positive. For example abnormal, abused, inaccurate, grouch, and lies would be considered negative while affection, awesome, beautiful, helpful and luckiest may be considered positive words. The tool that was ultimately used also deals with negation to help with accuracy. For example, the word the word “good” in the phrase “not good” will be contributing to positive sentiment rather than negative sentiment as the presence of “not” before it is not taken into account. (Narayanan Vivek, 2013) This problem was resolved by including an algorithm that uses a state variable to store the negation state. It transforms a word followed by a not or n’t into “not_” + word. Whenever the negation state variable is set, the words read are treated as “not_” + word. (Narayanan Vivek, 2013)

Initialing Study

Since this is a study that is a nascent idea, I am going into the study with an open mind and adjusting as I go along. Each statement that is presented will bring new insight into the study and hopefully will allow me to make better statements and get more out of the responses. My initial idea was to begin the study with a statement that is current and participants likely will have an opinion about. I also wanted to get everyone used to the structure of the study so they could be comfortable with the system and expressing the way they feel. Figure 8 below shows the structure of the study and in what order statements will be presented and analyzed.

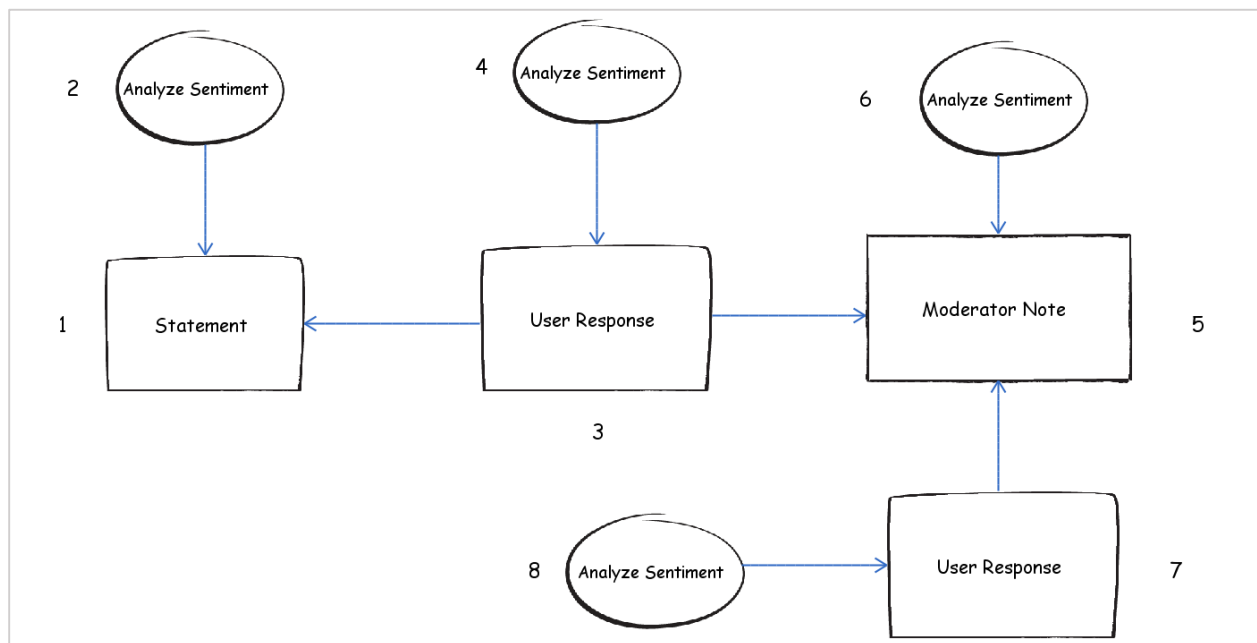


Figure 9

Results

Statement 1:

The title of the first statement was as follows:

Please look at the articles below which describe what led to the resignation of University of Missouri president. Once you have a grasp of the article, let us know what you think. Particularly if you felt he should have resigned, or stayed and handled the situation in a different way.

Why this was chosen:

I decided to begin with this particular statement for several reasons. The first reason was because it was something that had just happened recently and was all over the news and internet. The issue had been ongoing for several days and I knew that it was a topic that could possibly be highly opinionated as it dealt with race and also the resignation of a high official because of a race factor. I was interested in seeing the initial reactions within the system to this issue and to also see how people were communicating with one another in regards to the topic. I expected to get a relatively even mix of sentiments because I felt that some people would be happy with the resignation while others would feel that there could have been other solutions besides him resigning.

Statement sentiment analysis score:

Result – Positive

Confidence Level – 65.7%

Comment to evaluate

User – User 1

From my understanding of the articles, and what I have read in other articles, there have been several racially charged incidents in the months leading up to his call for resignation. Apparently 9 department heads from 9 different departments called for his resignation due to inaction.

The job of the university foremost is to provide an academic setting for students to learn. I think that anyone will agree that students should feel safe on campus and while attending campus events.

Racism is an issue that is actually black and white. Either you are pro-racism and racist language and actions, hate speech, or you against it. There is no room for neutral/middle ground when it comes to racism. In my opinion, failure to address the racially charged incidents that have occurred over the last few months, or not addressing them in a timely matter is the same as being supportive of them. This is clearly a case where “better late than never” does not apply.

When Wolfe finally did respond to the activists and protesters, he responded to their inquiries about systematic oppression and he responded with “I will give you an answer and I’m sure it will be a wrong answer, systematic oppression is because you don’t believe that you have the equal opportunity for success.” Many students took the “you don’t believe” portion to mean that it was their own fault that they were being oppressed and that Wolfe was blaming them for their failures.

It is my opinion that if Wolfe had taken the complaints and racist incidents more seriously and reacted sooner that his resignation would not have been needed.

Sentiment analysis score:

Result – Positive

Confidence Level – 75.7%

Moderator Note

Working together to calm a situation down before it gets out of hand is likely something that could save the president his job and help everyone come together a lot sooner.

Sentiment analysis score:

Result: Positive

Confidence Level: 84.8%

Sample Comment to evaluate

@user2 I'm all for team work and working together for a common goal, but in this issue with his position and job title, is group effort efficient? Ultimately, the blame falls on him, especially since the actions to stop the issues were not present from him. I personally don't feel that anyone else could have stepped in to save the situation. You get what out what you put in. In this case, nothing was put in by the president, so in the end, his only option was to resign.

Sentiment analysis score:

Result: Negative

Confidence Level: 99.6%

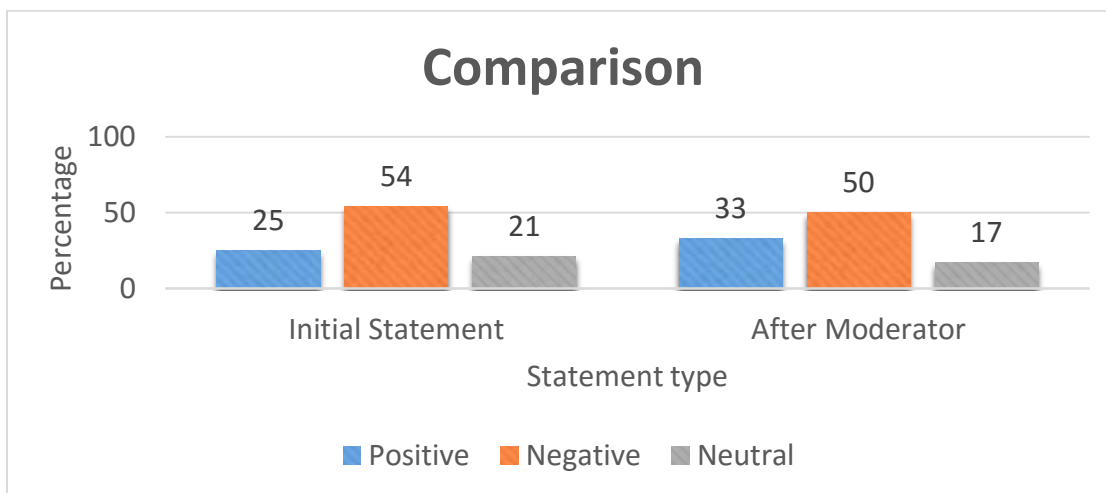
Analysis

Analyzing the comments of the participants after the first statement showed that there were varying levels of sentiment throughout the responses, For example, user @User3 stated “I think that there is a possibility for him to understand and make changes. I personally have come to grips on a situation in the past without having to get fired first. Yes it built up, but could it have been worked out...maybe”. This resulted in a neutral sentiment result with a confidence level of 55%. @User5 stated “I believe it was in the best interest of the University for Mr. Wolfe to resign. The events leading up to the protest should have been handled appropriately when they were first brought to Mr. Wolfe’s attention. As the president of a university you should always want to provide an environment for students to learn safely and fairly. By not addressing these issues immediately, the feeling of safety and equality in many students was compromised. By stepping down as president, Mr. Wolfe is allowing the healing and rebuilding process at the university to begin.” This resulted in a positive result. The overall majority of the initial responses resulted in mixed sentiments. 54% was negative 25% was positive and 21% was neutral. This was a good start to see if the moderator note would make any difference in the sentimental scores going forward. The moderator note that was later in the day put out, had a positive sentiment with a confidence level of 84.8%. The sentimental value of this statement and the responses afterwards produced overall 50% of the resulting responses being negative and

50% positive or neutral. In this case, the change in percentage for positive or neutral after moderation increased by 4%.

Initial Statement		
Sentiment	Count	Percentage
Positive	6	25%
Negative	13	54%
Neutral	5	21%
Total	24	100%

After Moderation		
Sentiment	Count	Percentage
Positive	2	33%
Negative	3	50%
Neutral	1	17%
Total	6	100%



Initial Statement 46 % Positive or Neutral

After Moderation 50% Positive or Neutral

Change +4%

Statement 2:

Please watch the video above and let us know what you think about this altercation between students and police officers. Later in the day, the moderator will make a statement about this situation, please respond to this as well.

It is important to note that this is actually a video and it is clear in the video that the result is a negative altercation. I have still included the sentiment value of the statement because it does discuss the altercation. The video is of students informing police officers that they need a warrant to enter their home and the result is the officers dragging a student out and tazing as well as arresting him.

Why this was chosen:

I wanted to change from just a direct statement and give the participants something to look at and evaluate instead. Again, this was a situation that had just recently happen in the past few days so it was a new video that some people may have heard about while others may have not. I felt that because of the nature of the video, there would be more negative response overall to this initial statement.

Statement sentiment analysis score:

Result – Negative

Confidence Level – 82.2%

Sample Comment to Evaluate

User – User2

This altercation is another example of police brutality that continues to go on. It is evident that the police were abusing their power. It even looks like this attack on the students was planned.

As soon as the police tried to pull the student by his shirt, instructing him to step outside, he looks to his left and gets frustrated and decided to go inside and start an altercation. That altercation led to more police officers coming in for back up and attacking the other students. A situation like this should have never gotten this far. The police were rude and unreasonable with the students. It amazed me on the use of several forms of police misconduct, such as intimidation, false arrest, and police corruption.

Sentiment analysis score:

Result – Negative

Confidence Level – 91.6%

Overall Comments Sentiment

Moderator Note

Addressing a problem in a calm and patient manner many times can alleviate a potential issue. Maybe that could have changed the outcome of this situation for the better and reduced the tension. Thoughts on this?

Sentiment analysis score:

Result: Neutral

Confidence Level: 61%

Sample Comment to evaluate

@User9 I agree. A clear and calm mind will yield an organized and thoughtful way of alleviating an issue.

Sentiment analysis score:

Result: Positive

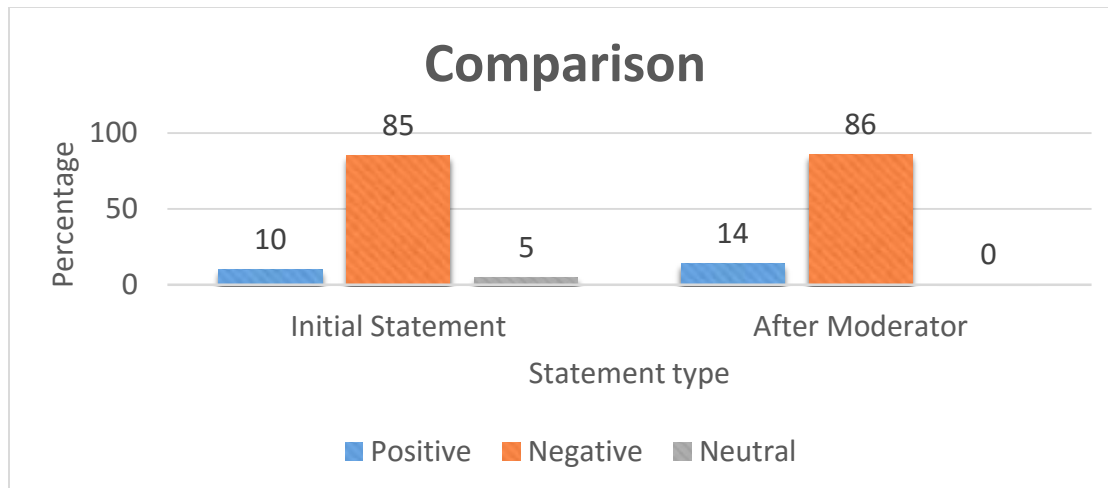
Confidence Level: 93.5%

Analysis

Looking at the video initially, I expected to get mostly negative responses to the video initially and this is exactly the result that I got after getting the sentiment scores from the tool. 17 responses were negative, 1 was neutral and 2 was positive, so the responses were 85% negative. With this being the second post, I was interested in seeing how this would vary once the neutral moderator comment was posted. The purpose of the moderator note was to encourage participants to really think about what would happen if the officers would have approach the situation in a calmer manner. I wanted to see if thinking about this would reduce the amount of negative comments. Overall the negative comments actually went up to 86% from 85%. This is actually a change in the opposite direction that I was expecting although it was not a large change.

Initial Statement		
Sentiment	Count	Percentage
Positive	2	10%
Negative	17	85%
Neutral	1	5%
Total	20	100%

After Moderation		
Sentiment	Count	Percentage
Positive	1	14%
Negative	6	86%
Neutral	0	0%
Total	7	100%



Initial Statement 15 % Positive or Neutral

After Moderation 14% Positive or Neutral

Change -1%

Statement 3:

At Kent State University, a white actor was cast to play the role of Martin Luther King Jr in an amateur production of the play. The original playwright caught wind of this and was upset at this depiction of the play. She penned an article for the website The Root about her disappointment in this choice. Below I have added a link to an article about this and above there is a link to the original playwrights response. Let us know what you think about this, specifically if you feel that a white actor playing the role of MLK is okay or not.

The sentiment score of this statement was taken before the word “below” because the rest of the statement was just direction on what should be evaluated.

Why this was chosen:

I felt that this topic was perfect for the participants to debate. The article discusses how the original playwright was upset with this choice, but the playwright of this particular depiction felt that he was just being an artist. I understood that there would be varying points of view as some would feel that this was an issue while others would feel differently. I expected overall sentiment to lean more towards neutral. I also made sure the sentiment value of the statement was high on the positive end.

Statement sentiment analysis score:

Result – Positive

Confidence Level – 93.5%

Sample Comment to Evaluate

User – User3

I total understand that this is a sensitive issue. On one hand I think that it's an art form and we should judge the artist by the quality of their work. On the other hand, I feel that the wishes of the playwright should be met...I think MLK would be for anyone playing his role. "Judged not, that ye be not judged".

Sentiment analysis score:

Result – Positive

Confidence Level – 99.9%

Moderator Note

The reality is that the audience is aware and not blindfolded, they can see the play clearly and likely know the story. If the actor remains African American, then everyone is satisfied with the play if it is good. Doesn't this change everything for the better?

Sentiment analysis score:

Result: Positive

Confidence Level: 71%

Sample Comment to evaluate

@User6 It would be interesting to see if the audience who actually watched the play thought the race of the actor affected the performance. Were they phased by the fact a white man played MLK? It probably would be easier for them to cast an African American and MLK was also indeed African American so it makes sense. However, I think everyone should be given a chance. The man who was chosen to play the role of MLK stated 'I was humbled. I wanted to honor this man and honor this text'. The man who was cast may have just happened to be a better actor than those who also auditioned for the part. It's the message of the play and the message MLK wanted to spread across the world that is the most important thing.

Sentiment analysis score:

Result: Positive

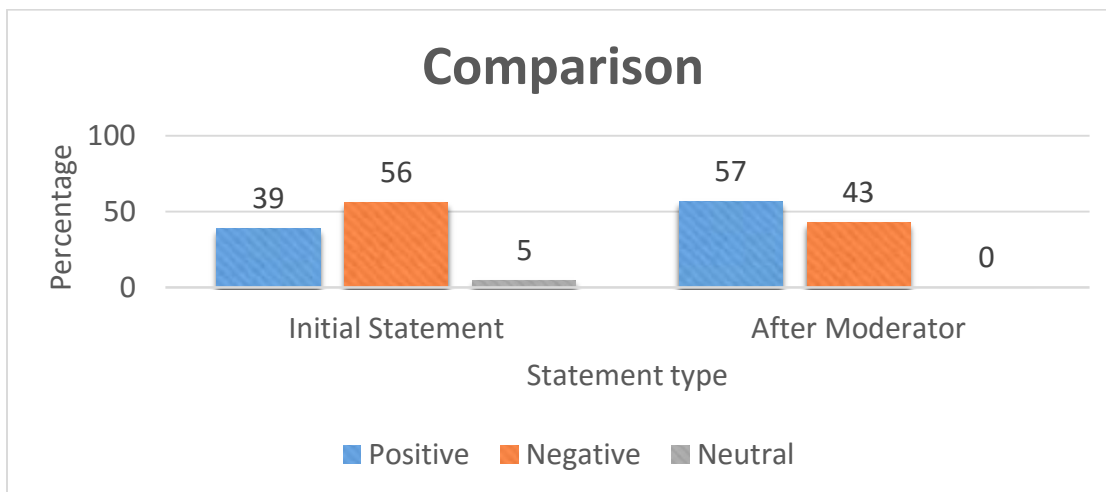
Confidence Level: 99%

Analysis

The results of this statement were initially what I expected in that 56% of the results of the initial statement was negative and the rest was positive or neutral. I expected that to be close to half and it was. People really seemed to be on either one side or the other in this case. The moderator note ended up having 75% positive which is certainly an increase from the initial value of 57% being positive. After moderation, an increase of 13% was observed in positive or neutral sentiment. This is really the results that I expected to see and falls in line with the idea that positive statements can reduce the resulting comments.

Initial Statement		
Sentiment	Count	Percentage
Positive	7	39%
Negative	10	56%
Neutral	1	5%
Total	18	100%

After Moderation		
Sentiment	Count	Percentage
Positive	4	57%
Negative	3	43%
Neutral	0	0%
Total	7	100%



Initial Statement 44 % Positive or Neutral

After Moderation 57% Positive or Neutral

Change +13%

Statement 4:

In honor of national kindness day, check out the video below and see how a #gobekind campaign helps out a homeless single father in Denver. Let us know your thoughts on this in the comments below. The sentiment score of this statement was taken before the word “below” because the rest of the statement was just direction on what should be evaluated.

It is important to note that this is a video and the sentiment score initially is based off of a general description of the video and not the actual video. It is clear to see that the video is done in a positive light as the single father is given money and lodging until he finds a place to stay.

Why this was chosen:

This topic was chosen because I wanted to have a topic that was ultimately positive and do the evaluation to see what the results would be after people watched the video and discussed it. It is important to see how some people may look at something in a positive light while others look at it negatively. Ultimately, the goal is to see if the negative sentiment values can be pushed more towards neutral.

Statement sentiment analysis score:

Result – Negative

Confidence Level – 62.2%

Sample Comment to Evaluate

User – User3

Call me a skeptic but this did video did not feel like a genuine act of kindness for several reasons

1. This video was sponsored by Trivago in collaboration with the donor. Giving away a \$1000 is a small sum to promote your company. A brilliant marketing strategy actually. Not to mention the money being made on Youtube through viewership.

2. A person's private life need not be made public to show off your generosity. If you want to donate, just donate.

3. It could be argued that making the video public encourages others to do the same but how many are in a position to donate enough to truly enforce a change in someone's life?

For the above reasons, I believe this was a publicity stunt rather than an act of generosity.

Sentiment analysis score:

Result – Negative

Confidence Level – 99%

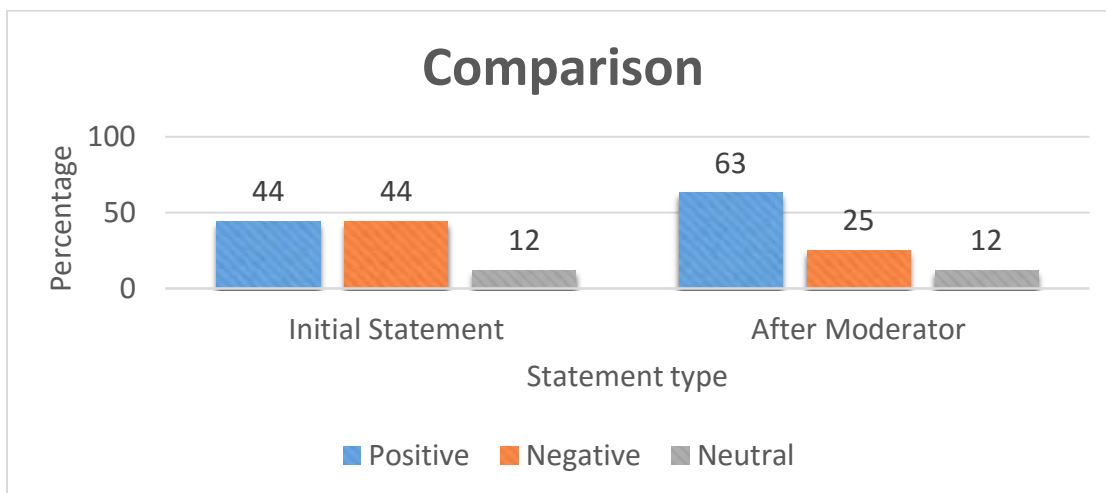
Analysis

The results of this statement varied in several ways. I chose to evaluate the comment above because the user had a different point of view then what I was expecting in that he saw the video as a marketing trick. The sentiment tool saw this response to be negative overall, which when

you read it that is certainly how it feels. Overall the positive and neutral responses prevailed here with an increase of 19%

Initial Statement		
Sentiment	Count	Percentage
Positive	7	44%
Negative	7	44%
Neutral	2	12%
Total	16	100%

After Moderation		
Sentiment	Count	Percentage
Positive	5	63%
Negative	2	25%
Neutral	1	12%
Total	8	100%



Initial Statement 56 % Positive or Neutral

After Moderation 75% Positive or Neutral

Change +19%

Statement 5:

Below is a video of a driver testing Tesla's self driving feature and a video of Stephen Colbert discussing the new feature. Some people are excited about these features on vehicles, while others worry about safety. Check out one, or both of the video's below and let us know how you feel about this very real technology.

It is important to note that this is a video and the sentiment score initially is based off of a general description of the video and not the actually video.

Why this was chosen:

This topic was chosen because I felt that there would be varying opinions in regards to whether or not the participants felt that it would be safe to ride in a car without hands. The video shows a demonstration of a Tesla driving by itself.

Statement sentiment analysis score:

Result – Negative

Confidence Level – 86.7%

Sample Comment to Evaluate

User – User7

This is great. Self driving cars are truly phenomenal. A true technological advance for mankind but I do see some problems. What happens at 4 ways stop sign? This is something that requires human interaction. Drivers interact with drivers of other cars to determine who goes. However I'm sure they will soon have cars communicating with each other road so each car knows of the

the cars around it and can coordinate turns and traffic patterns. Very cool stuff and I'm excited to see the future.

Sentiment analysis score:

Result – Positive

Confidence Level – 99%

Moderator Note

There was once a time when we couldn't fly and someone had to bend the rules to put airplanes in the air. Advancement many times means taking risks, which can lead to positive change.

Sentiment analysis score:

Result: Positive

Confidence Level: 82%

Sample Comment to evaluate

@User6 You have to take risks to advance. That's life. @User3 is right. As long as the advancements don't put people in danger.

Sentiment analysis score:

Result: Neutral

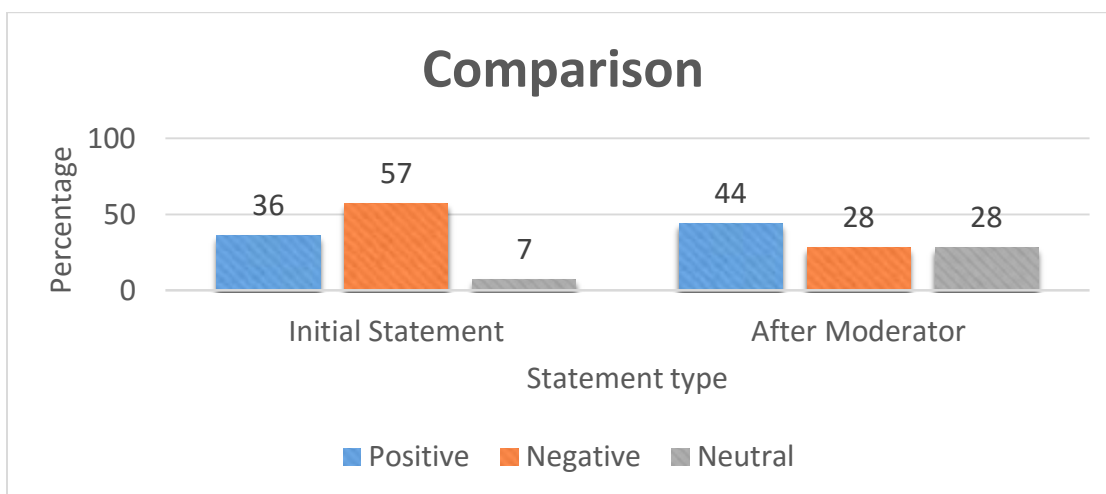
Confidence Level: 53.9%

Analysis

The results of this statement and the corresponding moderator note again trend in the direction of my expectations. The initial statement had 43% of the responses being either positive or neutral and after the moderator note, this percentage rose to 72%. The intentions of the moderator note was to utilize key words to get the user to lean towards taking risks that can lead to further advancement.

Initial Statement		
Sentiment	Count	Percentage
Positive	5	36%
Negative	8	57%
Neutral	1	7%
Total	14	100%

After Moderation		
Sentiment	Count	Percentage
Positive	3	44%
Negative	2	28%
Neutral	2	28%
Total	7	100%



Initial Statement 43 % Positive or Neutral

After Moderation 72% Positive or Neutral

Change +29%

Characters Limited

At this point in this point, I am halfway through the study overall and felt that it would be beneficial to change the structure of the study overall to evaluate how participants respond to a different structure. I chose to limit the amount of characters that you could type in response to a statement. By doing this, it would feel more like you were responding via twitter. I also felt like this would force participants to really think about what they were writing more. Participants were informed of this change in advanced and told that if they wanted to express more than the 200 character limit, they could simply add more comments.

Statement 6:

Below I have included a link to an opinion page that discusses the use of the prescription drug Adderall for reasons other than ADHD.

Why this was chosen:

I chose this topic because it has been a topic of debate for a long time now and many people have differences of opinions on this topic. The article has several different opinions on the page and it is a good avenue for you to get a wide range of opinions overall.

Sample Comment to Evaluate

User – User5

I believe that using the medicine as prescribed is beneficial, but using it when it is not prescribed to you and abusing it can be detrimental for your health.

Sentiment analysis score:

Result – Neutral

Confidence Level – 53%

Moderator Note

Drug abuse is a reality and unfortunately many times you do not see it coming. Ultimately everyone makes their own choices. Risk vs. reward is an entirely different beast when it comes to prescription drugs.

Sentiment analysis score:

Result: Positive

Confidence Level: 91%

Sample Comment to evaluate

@User2 I think this stems to the adolescence age when peer pressure is higher for children. Yes, drug abuse is a reality for all ages, but most often than not we try it because others are too.

Sentiment analysis score:

Result: Positive

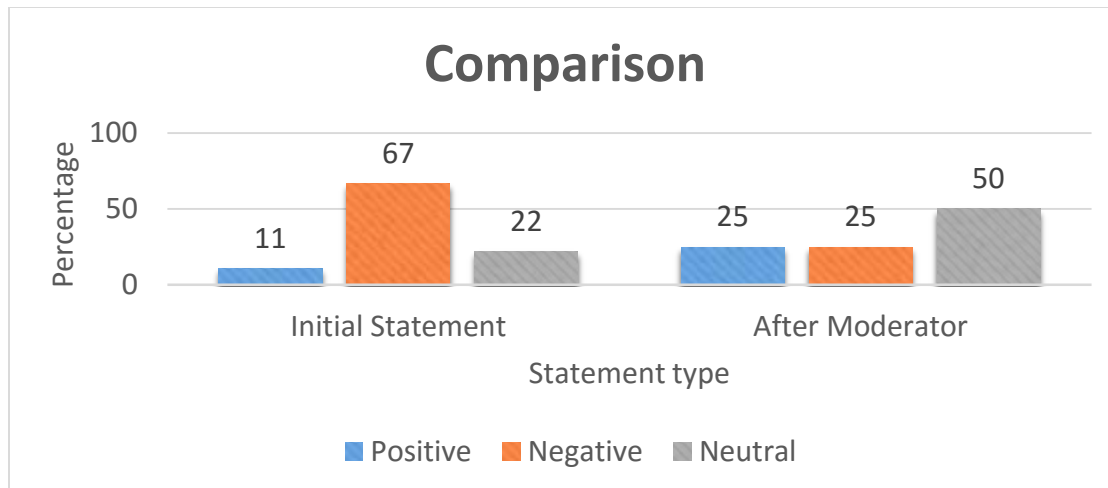
Confidence Level: 99%

Analysis

I wasn't really sure what the results of this statement would be when I chose it. I did expect mixed opinions but I did not know which way it would trend. After evaluation, I found that initially 33% of the statements were positive or neutral and after moderator note, the overall polarity went up to 72% positive or neutral with the majority of this falling in the neutral category. It is important to note that there were less responders to the moderator note, but the math is still done in correlation to all that answered each question individual. Again, this trends towards positive polarity following the moderator note.

Initial Statement		
Sentiment	Count	Percentage
Positive	2	11%
Negative	12	67%
Neutral	4	22%
Total	18	100%

After Moderation		
Sentiment	Count	Percentage
Positive	2	25%
Negative	2	25%
Neutral	4	50%
Total	8	100%



Initial Statement 33 % Positive or Neutral

After Moderation 75% Positive or Neutral

Change +42%

Statement 7:

Following the terrorist attacks in Paris, there have been various states that have declared that they will not welcome Syrian refugees due to the possibility that there may be a terrorist hiding among them with intentions to cause harm. Below are two recent articles on this this topic.

Please review one or both of these articles and let us know your thoughts on this topic.

Why this was chosen:

I chose this topic because this tragedy occurred the weekend before this statement and I knew that it would be fresh on the minds of many of the participants. I also knew that the debate over whether or not refuges should be allowed into America was a hot topic where many people had differences of opinions.

Sample Comment to Evaluate

User – User8

To date, not a single US refugee has committed a terrorist attack. The Boston Marathon bombers were asylum seekers. Europe will already be vetting the refugees before they're sent to the US anyway.

Sentiment analysis score:

Result – Neutral

Confidence Level – 58%

Moderator Note

In the wake of a tragedy such as the terrorist attack in Paris, it is normal for people to be worried. It is also important to realize that America is built on immigration at its core. Everyone does not wish ill intent, but it is important to keep out those who do.

Sentiment analysis score:

Result: Positive

Confidence Level: 99%

Sample Comment to evaluate

@User4 I feel it is normal to be on edge and weary of other possible attacks. Id rather keep everyone out to keep everyone safe versus let a few people in and risk lives.

Sentiment analysis score:

Result: Positive

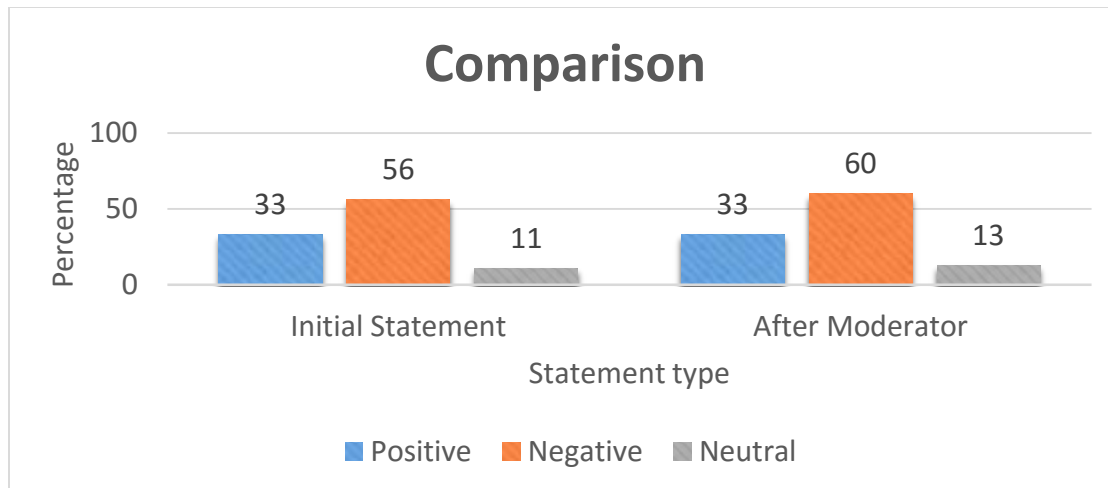
Confidence Level: 65%

Analysis

The results of this statement went in a different direction then the past statements had been going in. The initial statement had an overall positive or neutral polarity of 44% while the after moderation had a positive or neutral sentiment value overall of 40% This is definitely a deviation from the past few as the sentiment score is actual down by 4%, but it is expected to have some deviation from the norm.

Initial Statement		
Sentiment	Count	Percentage
Positive	3	33%
Negative	5	56%
Neutral	1	11%
Total	9	100%

After Moderation		
Sentiment	Count	Percentage
Positive	5	33%
Negative	9	60%
Neutral	1	7%
Total	15	100%



Initial Statement 44 % Positive or Neutral

After Moderation 40% Positive or Neutral

Change -4%

Statement 8:

Below I have included a link to an opinion page that provides a few viewpoints on whether or not NCAA athletes should be paid. Review some of the opinions on this topic and comment your thoughts on this subject. You can add multiple comments on the subject if needed.

Why this was chosen:

Student athletes getting paid is a strong topic, especially during the NCAA football season. There are a variety of opinions in this area overall, where some people think players should be paid while others feel the opposite. Normally it is difficult to convince one side that they should go the other way.

Sample Comment to Evaluate

User – User4

This is a double sided sword. NCAA makes billions of dollars annually off of these student athletes and penalizes them heavily for the slightest financial help they receive.

Sentiment analysis score:

Result – Neutral

Confidence Level – 61%

Moderator Note

An education provides tremendous value to athletes, many times beyond the notoriety that comes with playing sports. Universities and colleges have the responsibility of ensuring that an athlete has everything the he or she needs to be successful. If this is being done successfully, then it is naturally to say that everyone benefits.

Sentiment analysis score:

Result: Positive

Confidence Level: 99%

Sample Comment to evaluate

@User2 Education needs to be kept as a large focus for these athletes. If we were to pay college athletes, their education would quickly fall from being important to them and they would lose sight of it.

Sentiment analysis score:

Result: Neutral

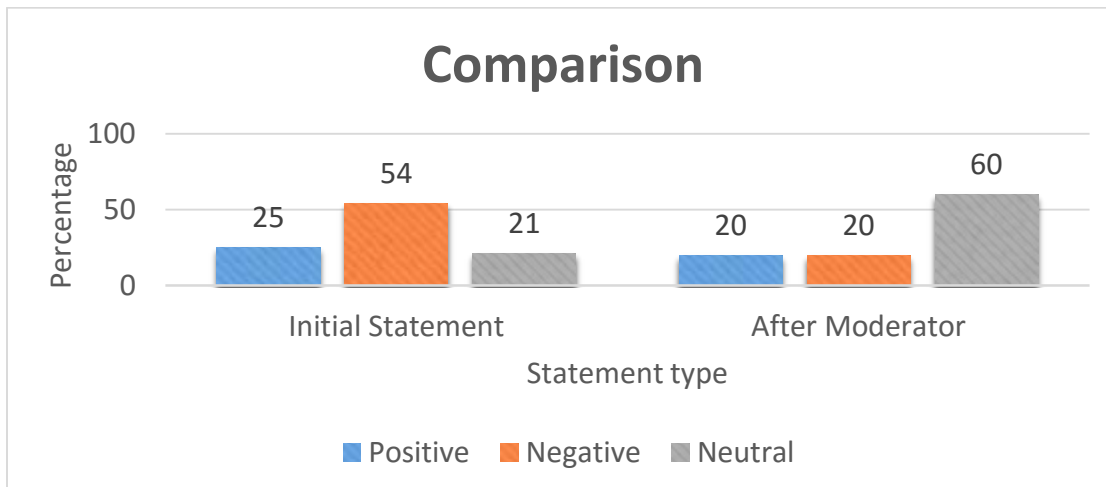
Confidence Level: 58%

Analysis

Analysis of the sentiment for this statement showed that the comments in reaction to the initial statement had a total of 46% positive or neutral sentiment, while 80% of the comments following the moderator comment was positive or neutral with the majority being neutral. This is right in line with what we have been seeing in the majority of the other statements.

Initial Statement		
Sentiment	Count	Percentage
Positive	3	25%
Negative	6	54%
Neutral	2	21%
Total	11	100%

After Moderation		
Sentiment	Count	Percentage
Positive	1	20%
Negative	1	20%
Neutral	3	60%
Total	5	100%



Initial Statement 46 % Positive or Neutral

After Moderation 80% Positive or Neutral

Change +34%

Summary

Overall a total of 200 comments were analyzed in response to 8 different statements. The majority of the time, after moderation the resulting comments trended towards a positive or neutral sentiment. 6 out of 8 times the or 75% of the time in the scenarios, after moderating a conversation with a neutral or positive statement, the overall results were positive. This is shown visually in figure 13.

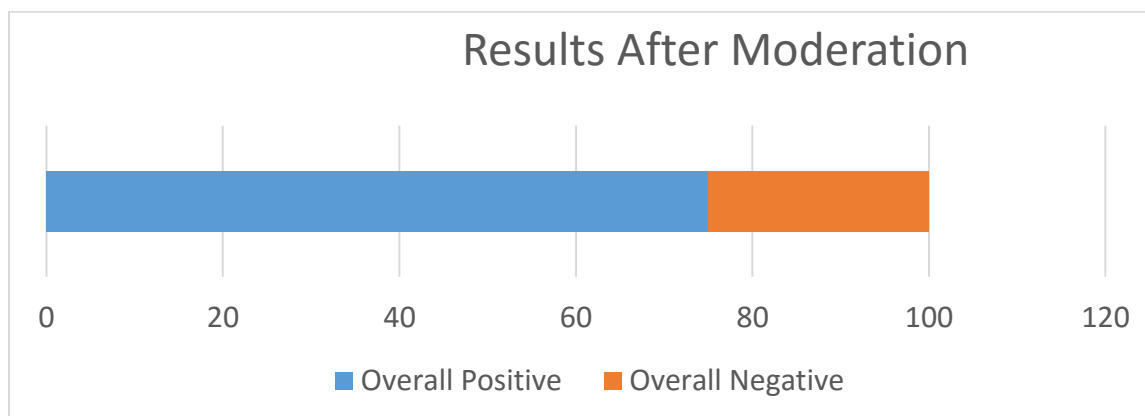


figure 13

The chart below (figure 14) shows the increase or decrease in positive/neutral polarity over the course of all of the statements. Overall you can visually see a positive trend in the line graph which shows that this factor consistently trended in a positive direction. There were some statements that had a trend towards negatives such as statement 2 and 7. The trendline also highlights the positive trend of the chart and sentiment values.

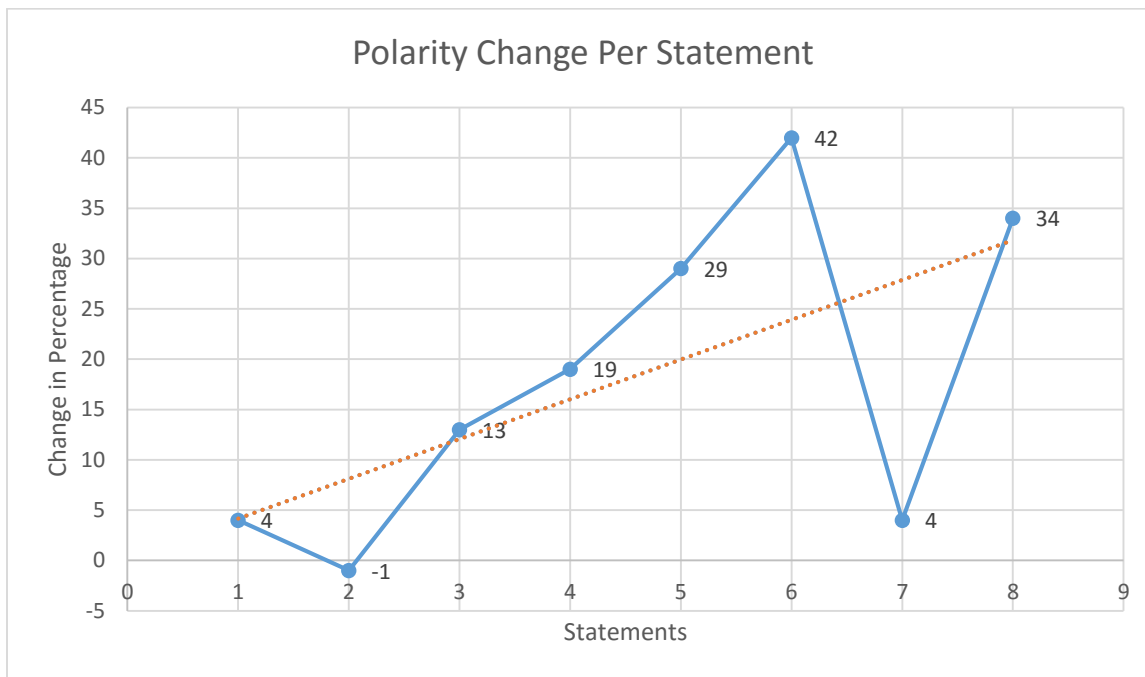


Figure 24

After reducing the character limit following statement 4, it is important to note that 2 out of 3 times the sentiment change was positive after moderation. This is similar to before the change in character limit where 3 out of the 4 times the resulting change was positive. So in this case, changing the character limit did not have a noticeable quantifiable affect on the overall change.

Conclusion

Sentiment analysis as a whole is a growing field and a field that has proven to be useful in a variety of areas. One of the major benefits is that it can be applied and adjusted in so many different ways to fit the need of the user. Throughout my thesis, I ultimately wanted to show that there is indeed a correlation between the polarities of certain statements. I specifically set out with the notion that sentiment values could be shifted to a more positive polarity with the integration of neutral or positive statements. My study attempted to target this area directly by hand picking out statements of neutral or positive polarity and using them in real life conversations to find out what the result would be. My analysis clearly showed that change in sentiment was more positive after moderation.

Going forward I believe that this is a good study to build off of and see if it can be expanded to include more user and more topics over a longer period of time. I feel that this could be of tremendous value in the world today for a variety of reasons. Imagine being able to get a community to think more positively by interjecting conversations online. This could potentially stop a riot from happening or turn a protest into understanding for all sides. It could also be very beneficial within business as companies can use this type of strategy to communicate with their customers and help them to understand their point of view and vice versa. Overall I hope my contributions will help further understanding of sentiment analysis and the effects sentiment can have on textual conversations.

References

- Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. doi:10.1016/j.apenergy.2013.03.027
- Bakliwal, A., Foster, J., Puil, J., O'Brien, R., Tounsi, L., & Hughes, M. (2013). Sentiment Analysis of Political Tweets: Towards an Accurate Classifier. *Language Analysis in Social Media*.
- Barbosa, G., Silva, I., Zaki, M., Meira, W., Prates, R., & Veloso, A. (2012). Characterizing the effectiveness of twitter hashtags to detect and track online population sentiment. *Proceedings of the 2012 ACM Annual Conference Extended Abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*.
- Goel, V. (2014, June 29). Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry. Retrieved December 3, 2015, from http://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html?_r=0
- Han Jiawei, K. M. (2011). Data Mining - Concepts and Techniques. San Francisco: Morgan Kaufman Publishers Inc.
- Heine, C. (2015, January 12). 70% of Marketers Will Spend More on Social Media Ads This Year. Retrieved November 25, 2015, from <http://www.adweek.com/news/technology/70-marketers-will-spend-more-social-media-ads-year-162291>
- Kemp, S. (2014, August 8). Global Social Media Users Pass 2 Billion. Retrieved November 9, 2015, from <http://wearesocial.net/blog/2014/08/global-social-media-users-pass-2-billion/>
- Kemp, S. (2015, January 21). Statistics. Retrieved November 20, 2015, from <http://wearesocial.net/tag/statistics/>
- Mander, J. (2015, January 26). Daily time spent on social networks rises to 1.72 hours. Retrieved November 20, 2015, from <http://www.globalwebindex.net/blog/daily-time-spent-on-social-networks-rises-to-1-72-hours>
- Pettit, A. (2013). Identifying the real differences of opinion in social media sentiment. *International Journal of Market Research Int. J. Market Res.*, 757-757.

Vivek Narayanan, Arora Ishan, Bhatia Arjun. "Fast and accurate sentiment classification using an enhanced Naive Bayes model". *Intelligent Data Engineering and Automated Learning IDEAL 2013 Lecture Notes in Computer Science Volume 8206, 2013, pp 194-201*

Nuseir, R. (2014, March 14). Strange Uses for Sentiment Analysis. Retrieved October 20, 2015, from <http://www.smartdatacollective.com/raminuseir/191466/strange-uses-sentiment-analysis>

Ogniva, M. (2010, April 19). How Companies Can Use Sentiment Analysis to Improve Their Business. Retrieved November 22, 2015, from <http://mashable.com/2010/04/19/sentiment-analysis/#vdMjLjL.NiqR>

Paltoglou, G., Gobron, S., Skowron, M., Thelwall, M., & Thalmann, D. (2010). Sentiment analysis of informal textual communication in cyberspace. *proceedings of ENGAGE*, 13-25.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 1-135.

Plaza, L., & Carrillo de Albornoz, J. (2012). Sentiment Analysis in Business Intelligence: A survey. In R. Colomo-Palacios, J. Varajão, & P. Soto-Acosta (Eds.) *Customer Relationship Management and the Social and Semantic Web: Enabling Clients Conexos*(pp. 231-252). Hershey, PA: Business Science Reference. doi:10.4018/978-1-61350-044-6.ch014

Sentiment. (n.d.). *Dictionary.com Unabridged*. Retrieved November 12, 2015, from Dictionary.com website: <http://dictionary.reference.com/browse/sentiment>

Tumasjan, A., Sprenger, T. O., Sandner, P. G. & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment.. In W. W. Cohen & S. Gosling (eds.), *ICWSM*, : The AAAI Press.

Yelena Mejova, Padmini Srinivasan, and Bob Boynton. 2013. GOP primary season on twitter: "popular" political sentiment in social media. In *Proceedings of the sixth ACM international conference on Web search and data mining (WSDM '13)*. ACM, User3, NY, USA, 517-526. DOI=<http://dx.doi.org/10.1145/2433396.2433463>

Appendix – IRB Exempt Certification Form



EAST CAROLINA UNIVERSITY

University & Medical Center Institutional Review Board Office

4N-70 Brody Medical Sciences Building · Mail Stop 682

600 Moye Boulevard · Greenville, NC 27834

Office 252-744-2914 · Fax 252-744-2284 · www.ecu.edu/irb

Notification of Exempt Certification

From: Social/Behavioral IRB
To: [Shawn Braddy](#)
CC: [Mark Hills](#)
Date: 12/22/2015
Re: [UMCIRB 15-002222](#)
Sentiment Analysis Study

I am pleased to inform you that your research submission has been certified as exempt on 12/21/2015. This study is eligible for Exempt Certification under category #2.

It is your responsibility to ensure that this research is conducted in the manner reported in your application and/or protocol, as well as being consistent with the ethical principles of the Belmont Report and your profession.

This research study does not require any additional interaction with the UMCIRB unless there are proposed changes to this study. Any change, prior to implementing that change, must be submitted to the UMCIRB for review and approval. The UMCIRB will determine if the change impacts the eligibility of the research for exempt status. If more substantive review is required, you will be notified within five business days.

The UMCIRB office will hold your exemption application for a period of five years from the date of this letter. If you wish to continue this protocol beyond this period, you will need to submit an Exemption Certification request at least 30 days before the end of the five year period.

The Chairperson (or designee) does not have a potential for conflict of interest on this study.