

Classifying #MeToo Hash-tagged Tweets by Semantics to Understand the Extent of Sexual
Harassment

by
Claire Elise Hubacek

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of the
requirements of the Sally McDonnell Barksdale Honors College.

Oxford
May 2018

Approved by

Advisor: Dr. Naeemul Hassan

Reader: Dr. Kirsten Dellinger

Reader: Dr. Dawn Wilkins

©2018 2018
Claire Elise Hubacek
ALL RIGHTS RESERVED

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Naeemul Hassan for the continued support and knowledge while guiding me throughout my thesis. I cannot overstate my gratitude for the degree of patience and understanding demonstrated with me throughout this process and the contribution Dr. Hassan brings to the computer science department.

In addition to my advisor, I would also like to thank my other readers: Dr. Kirsten Dellinger and Dr. Dawn Wilkins for their insight, questions, and flexibility.

My sincerest thanks also goes to the computer science department at the University of Mississippi, for accepting an art student with no background in computer science and developing me into the competent programmer I am now. Thank you to the Sally McDonnell Barksdale Honors College faculty and staff for the opportunities and challenges given to me these past five years.

Thank you to everyone who participated in this project and took the time to read my research on sexual harassment, my categorization rules, and assist me in manually categorizing a set of tweets: John Joseph Angel, Ainsley Ash, Tim Dolan, Jason Hale, Lily Hassan, Gracie Hubacek, Karen Hubacek, Krish Lamba, Will Lewis, Mallory Loe, Ethan Luckett, Jake Wooley, Lina Ye, Dr. John Wiginton, and Dr. Debra Young.

I would also like to thank the Title IX office and Department of Violence Prevention for the accommodations, help, and support given during my own struggles throughout this past year. Thank you for empowering me with the knowledge and support I needed to finish my last undergraduate year, and for inspiring me to take this direction with my thesis.

Lastly, I would like to thank my family and friends who have supported me throughout this thesis as well as my undergraduate career. Looking back on my transcript and resume as I prepare to graduate, I see nothing but the contributions of my support system and how my achievements would not have been possible alone. Without your patience and support, I would not have finished. In particular I would like to thank Dr. Kristin Davidson, Gracie Hubacek and Ethan Luckett for not only your help in all my academic challenges but also for your unending support as friends. I would like to extend a special thank you to my pet rabbits who have continued to inspire and motivate me every day since I have loved them: Ellie, Ezra, and Buddy.

ABSTRACT

Classifying #MeToo Hash-tagged Tweets by Semantics to Understand the Extent of Sexual Harassment
(Under the direction of Naeemul Hassan)

My thesis advisor Dr. Naeemul Hassan has collected approximately half a million tweets that contain the hashtag #MeToo and completed an initial analysis regarding various aspects of this data. His work, as well as others', could be expanded if the tweets were categorized according to their semantics and if analyses could be performed on isolated sets of this data. This thesis contains a program that can process these tweets and categorize them by the type of sexual harassment expressed, if applicable. By using this tool to filter out extraneous tweets to isolate the ones that express a personal experience, researchers would be privy to more useful analysis of the data by working exclusively with tweets that express a personal experience with sexual harassment while knowing the type of harassment described.

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES	vii
LIST OF ABBREVIATIONS	viii
INTRODUCTION	1
PREPARATORY WORK	2
CLASSIFICATION METHODOLOGY	6
IMPLEMENTATION	13
RESULTS	16
SUPPORTING MATERIALS	19
CONCLUSION	22
BIBLIOGRAPHY	23

LIST OF FIGURES

4.1	Supervised Machine Learning Architecture	14
4.2	Testing and Integration Flow	15

LIST OF TABLES

2.1	Categorization Header	5
3.1	Tweet Examples - Irrelevant	6
3.2	Tweet Examples - Stance	7
3.3	Patronizing Tweet Examples	8
3.4	Unwanted Sexual Attention Tweet Examples	9
3.5	Predatory Tweet Examples	11
3.6	Not Enough Context Tweet Examples	12
5.1	Positive and Negative Classifier Terms	16
5.2	SVM Accuracy - Related	16
5.3	SVM Accuracy - Stance	16
5.4	SVM Accuracy - Harassment Category	17
5.5	Naive-Bayes Accuracy - Related	17
5.6	Naive-Bayes Accuracy - Stance	17
5.7	Naive-Bayes Accuracy - Harassment Category	17
5.8	Category Reduction	18

LIST OF ABBREVIATIONS

CHAPTER 1

INTRODUCTION

1.1 Problem

“If all women who have been sexually harassed or assaulted wrote ‘Me too.’ as a status, we might give people a sense of the magnitude of the problem.” - Alyssa Milano

Beginning in October 2017, victims of sexual assault or harassment began using the hashtag “#MeToo” to illustrate the magnitude and prevalence of sexual assault and sexual harassment. With the growth of the movement, the hashtag #MeToo is used widely on social media from people who are in support of the movement but not victims themselves, from people who are antagonists or critics of the movement, in general discussion and news coverage as well as for its original purpose of victims communicating a personal experience. As the study of sexual assault and harassment grows in relevance and popularity, the #MeToo movement exists as an unprecedented platform to be vocal about personal experiences regarding sexual assault and harassment. This thesis explores this platform and endeavors to use it to draw new conclusions about the demographic groups who experience these problems.

1.2 Scope

This thesis only covers a basic classification implementation and also provides an online interface to allow others to use the classification tool without compiling and processing on their local machine. However, research indicates many other possible avenues through which this program might be applied. These possibilities and applications are discussed within the Supporting Materials chapter.

1.3 Overview

1.3.1 Natural Language Processing (NLP)

blah

1.3.2 Supervised Machine Learning

blah

1.3.3 Classification Model

blah

CHAPTER 2

PREPARATORY WORK

This chapter addresses the research and work conducted before the development process. It describes sexual harassment as it has developed and recognized both legally and psychologically as well as a discussion of the literature regarding sexual harassment categorization. This was fundamental to designing and applying a strong, comprehensive categorization topology to the dataset. This chapter also discusses the data used in the program, how it was collected, and how it was manually categorized for the supervised machine learning process.

2.1 Legal History of Sexual Harassment in the United States

Defining different types of sexual harassment is a task that remains incomplete despite steady, incremental progress because “sexual harassment” didn’t exist as a criminal offense until it was first supported through case law. Initially, the Title VII of the Civil Rights Act of 1964 made it illegal to discriminate employment opportunities based upon gender (and Title IX later upholding the same philosophies within education). Throughout the 1970’s, several women used Title VII provisions to sue their employers for coercing (or attempting to coerce) them into sexual acts, and their successes within the courts enshrined sexual harassment as a criminal offense under these provisions. Since these cases, appropriately defining sexual harassment continues to be refined through improvements in research, laws, and court decisions today. As a controversial topic with many nuances and varied perspectives, proper categorization of sexual harassment necessitates a comprehensive legal understanding alongside a psychological one.

The first Supreme Court case ruling in favor of the victim alleging sexual harassment was in the 1974 case *Barnes v. Train*, in which the harasser was found at fault for firing a female employee for refusing his sexual advances although the term “sexual harassment” was not used yet at this time. A few years later, the Supreme Court upheld that this type of behavior from employers was a violation of Title VII, and subsequently the Equal Employment Opportunity Commission (EEOC) refined the rules to explicitly cover this type of sexual harassment. The first case to grant relief for sexual harassment explicitly under Title VII provisions was *Williams v. Saxbe* (1976). Because of the legal distinction and precedents set, the plaintiff in the landmark case *Meritor Savings Bank v. Vinson* (1986) effectively established *quid pro quo* behaviors as a form of sexual harassment where the plaintiff won a case against her employer who attempted unsuccessfully to coerce her into sexual acts. This case also recognized that verbal remarks and questions of a *quid pro quo* nature were in violation of Title VII even if the victim suffered no tangible consequences because the language and attempt itself created a “hostile work environment” [Pierce, 1989]. Now, legally recognized instances of sexual harassment continue to fall under the domain of the the Equal Employment Opportunity Commission (EEOC) and Title VII, and the accepted definitions categorize sexual harassment as either being *quid pro quo* acts or behaviors that contribute to a “hostile work environment” [York and Brookhouse, 1988]. Through a myriad of rule changes, public statements, and Supreme Court decisions, the Department of Education’s Office for Civil Rights has uphold these same principles as they pertain to education because sex-motivated violence or harassment to a student creates a hostile environment.

What behaviors qualify as sexual harassment as a criminal offense are continuously revisited through new court decisions. Under U.S. law, *quid pro quo* sexual harassment, also referred to as sexual bribery, includes attempted and actual pursuits of a sexual nature against a person in a professional or academic environment when tangible benefits could be given or denied to the victim. For a long time, researchers overestimated the frequency of this type of harassment compared to

other forms because sexually explicit cases have historically been preferred over sex-related work [Fitzgerald et al., 1995].

Actions that contribute to a "hostile work environment" are the much more common category of sexual harassment. Less explicitly sexual behaviors, such as offhand remarks, teasing or banter, and personal questions are not explicitly banned as acts of sexual harassment. However, if these acts occur with a sufficient degree of severity or frequency, they would then be considered as elements that create a hostile environment in work or school and therefore qualify as sexual harassment. When evaluating behaviors that do not take place in an educational or professional setting, the actions cannot be evaluated under Title VII or Title IX regulations. These instances of sexual harassment that don't occur in a professional setting must occur with such a degree of frequency or severity that they can be considered (with context) as falling under harassment, stalking, cyberstalking, sexual assault, or other criminal laws. This technicality makes proper evaluation of sexual harassment complex because the behavior in question, while it might be considered sexual harassment in a professional setting, does not have an equal opportunity regulation to dictate it as such when the action occurs between peers.

2.2 Literature Review

As the courts first began to address more and more claims of alleged sexual harassment, the need for a proper categorization grew in order to appropriately assess the degree and severity of the actions. From both a legal and social research perspective, the lack of consistency among categorization definitions caused problems when trying to compare and use instances or research from one context as a guide when evaluating another. The first widely used standardization was designed by Frank J. Till in 1980. In his work, Till defines five major categories that consist of generalized sexist remarks or behavior, inappropriate and offensive but essentially sanction-free sexual advances, solicitation, coercion, and sexual crimes [Till, 1980]. This categorization was the most frequently used throughout the 1980's, although most organizations generally either rephrased the categories or consolidated them into three categories.

In 1992, James E. Gruber, a legal consultant, made a significant contribution to categorizing sexual harassment by defining three overarching categories and 3-4 subtypes for each one, for a total of 11 distinct, exhaustive, and mutually exclusive categories. Gruber's first category "verbal requests" includes sexual bribery, sexual advances, relational advances, and subtle pressures/advances. These are all behaviors that are said directly to the victim with the intent of a sexual or an increased personal or social relationship goal even if the relationship being sought isn't obviously sexual. Subtle pressures and advances generally refers to statements or questions that can be ambiguous and not directed towards the victim but are still inappropriate, such as accidentally "thinking out loud" or double entendres.

The second category, "verbal comments", includes personal remarks, subjective objectification, and sexual categorical remarks. These categories generally refer to statements of a nonsolitary nature directed either to a woman (ex. teasing or jokes), about a woman, or about women in general. Sexist verbal remarks, rumors, inappropriate comments regarding someone's body or perceived level of attractiveness, and bystander harassment are also classified here.

The third category, "nonverbal displays", includes sexual assault, sexual touching, sexual posturing, and possession/display of inappropriate sexual materials. Actual or attempted coercion is sexual assault regardless. Pinching, grabbing, groping are examples of sexualized touching. Sexual posturing includes violations of personal space and attempts to make physical contact, including behaviors like making an obscene sexual gesture with one's hands. The possession or display of sexual materials includes personal items regarding sexuality (tampons, pads, birth control, sex toys, etc.) in addition to pornographic materials or sexist materials.

Altogether, there are 11 distinct types of sexual harassment, and these categories are both mutually exclusive and reflective of the EEOC's guidelines [Gruber, 1992]. Gruber originally wrote these categories based upon reviews of sexual harassment that only included female victims and male harassers, but their continued application today shows that these definitions are not gender exclusive.

These mutually exclusive categories establish a fluid progression in how certain behaviors can contribute to a hostile work environment, and Gruber’s work continues to influence many more recent approaches to this topic. In a 2005 review of past, present, and future directions of improving gender and minority diversity in professional environments, Murrell and James refer to Gruber’s work as a cornerstone in developing a comprehensive legal definition of sexual harassment [Murrell and James, 2002]. A study performed in 2005 to evaluate the effect of an obscene television show on individuals’ perception of what constitutes sexual harassment used Gruber’s categorization in their participant surveys in order to do so [Ferguson et al., 2005]. Despite the significance of his contribution, there is a neglected space of defining sexual harassment beyond the scope of Title VII and Title IX that Gruber’s categories do not accommodate entirely. Consequently, many researchers continue to consolidate these categories. and in doing so, claim a lesser degree of specificity to permit the nuances of peer to peer harassment to be appropriately placed.

A common issue with classifying sexual harassment lies in the varied perceptions of what is and isn’t sexual harassment among observers [Studzińska, 2015]. The victim’s perception is integral not just for legal categorization but also to understand the degree and severity of the harm inflicted on those who are exposed to the behavior. Some concerns with relying too heavily on the victim’s perception include fear that the victim might be biased and in some rare cases, deliberately dishonest. However, psychological research into coping with sexual harassment suggest that this is not necessarily the case. A study by Aparna Pathak on sexual harassment and coping behaviors synthesized many different research publications over the past few decades in her work. This synthesis notes that a 1997 study found that experiencing sexual harassment, whether or not the victim is aware of it, will still have negative outcomes on the victim [Schneider et al., 1997] in terms of health and distress. Additional research into false claims shows that the prevalence of false allegations of sexual assault comprise between 2% and 10% of cases, which is the same statistical prevalence as false reporting of non-sexual crimes [?]. These types of research supports the research philosophy of giving the victim the benefit for the doubt in some cases.

Additionally, a lack of research exists in sexual harassment as it pertains to instances that don’t occur within title VII or Title IX, particularly on the streets. Pathak’s review notes that (as of 2015, the time this was written) a study performed in 2000 is one of the only known, peer-reviewed attempt at documenting the extent of unwanted sexual attention from strangers. MacMillan et al. found that over 80% of women experienced unwanted sexual attention (ex. catcalls) and just under 30% of women experienced direct confrontation of a sexual nature from strangers [MacMillan et al., 2000]. Overall, these caveats indicate that while Gruber’s categorization might be the most appropriate tool for evaluating sexual harassment in a professional capacity because of it’s adherence to EEOC guidelines, it does not necessarily scale towards including social environments outside of a workplace or school.

Shortly after Gruber published his categorization, Fitzgerald et al. developed a simpler, consolidated categorization architecture in order to consequently develop a better questionnaire for measuring sexual harassment. Their categorization was comprised of three types: unwanted sexual attention and gender harassment (hostile work environment) and sexual coercion (*quid pro quo*). This categorization was an attempt to distinguish between sexual harassment “as a legal concept and a psychological construct” in order to better accommodate how a victim might perceive or label a behavior. Accommodating this “gray space” in victim perception allows for more consistency among responses, which guided their research goal of developing a more scalable questionnaire for surveying the frequency of sexual harassment [Fitzgerald et al., 1995]. In a 2008 publication, Chamberlain et al. deviated from Gruber’s 11 types for similar reasons. Upon the basis that these legal-driven approaches “underscore diversity” and “suggest substantial variation with regard to intent and severity” [Chamberlain et al., 2008] from the victims’ perception. For their purposes, Chamberlain et al. uses the following sexual harassment categories: patronizing (sexist but non-sexual remarks, gestures, or condescension), taunting (sexual gestures, physical displays, and overly personal comments and queries), and predatory (encompassing sexual solicitation, sexual promises or threats, touching, and forced contact).

2.3 Data

A collection of 10,000 tweets were originally assembled into a single excel sheet with the tweet’s unique ID and contents in two columns. The next three column headers contained the checks of relevancy, stance, and type of sexual harassment. If the correct classification could not be determined from the text available, the index was left blank.

		Classification Labels		
		Relevant 1. Related 2. Unrelated	Stance 1. Support 2. Against 3. Neutral	Harassment Category 1. Patronizing 2. Unwanted Sexual Attention 3. Predatory 4. Not Enough Context
id_str	text			
1	‘... example ...’	1	1	3
2	‘... example ...’	2		
3	‘... example ...’	1	2	

Table 2.1: Categorization Header

To create a reliable training set of data, a sufficiently large amount of tweets were manually categorized by human readers. These voluntary participants were given a comprehensive explanation of sexual harassment both legally and psychologically, examples of properly categorized tweets, category definitions, as well as the categorization rules.

Participants were each provided 500 unlabeled tweets and asked to categorize them according to the rules provided. One male and one female of a similar age and level of education shared each set of 500 tweets and categorized them without having access to any other responses beyond the examples provided. Because research suggests that males and females assess sexual harassment differently, this process was performed to improve the integrity and consistency of the training set.

5,000 tweets were categorized total, with each set of 500 being categorized by a male and female. Cumulatively, 10,000 categorizations were made. Only the tweets that were categorized identically by both the male and female participant were considered in the training and testing sets of data.

CHAPTER 3

CLASSIFICATION METHODOLOGY

For every tweet, the program will make either one, two, or three determinations depending on the level of relevancy, context, and detail. It will first determine whether or not the tweet using #MeToo is relevant to the movement. If it is relevant, it will then determine the stance of the tweet regarding the movement. If there is enough context, the last check is to determine the type of sexual harassment or assault being described.

3.1 Determining Relevancy

The majority of the tweets using the hashtag #MeToo are relevant to the movement in some way. However, sometimes they are not. Examples of irrelevant tweets are ones that are written by bots and no meaningful determination can be made, the tweets are unintelligible, the entirety of the tweet's context can only be determined through following a URL or image, tweets that use the hashtag #MeToo for the purpose of winning a giveaway or for increased visibility in a promotion, or tweets that are obviously misrepresenting or misusing the hashtag. Relevant tweets were assigned the number 1 and irrelevant tweets were assigned a the number 2. Table 3.1 illustrates examples of irrelevant tweets.

ID	Original Tweet Text	Relevant	Stance	Harassment Category
1	https://soundcloud.com/zay-hippy #askNiall #Bel-lator185 #MeToo #BadTimesToTellAJoke #Fri-dayFeeling #RaiderNation #LouCity #bitcoin #WWEBuenosAires	2		
2	#MeToo You too can achieve salvation by doing worship as per our Holly scriptures. To know more watch SADHANA TV 7:40 pm pic.twitter.com/4wwhbEFZjM	2		
3	@Der.Peemann check this crazy track out #hiphop you like it? "Yes really"? #MeToo WHAT?	2		

Table 3.1: Tweet Examples - Irrelevant

3.2 Determining Stance

When categorizing tweets within the excel sheet, each tweet that is relevant to the #MeToo movement with enough context to ascertain a stance was assigned a 1, 2 or 3 accordingly:

1. Support
2. Against
3. Neutral

Tweets without enough context to determine the stance were left void. Tweets that are expressing a personal experience with sexual harassment or assault are considered supportive tweets. Tweets that are expressing a supportive sentiment but not claiming victimhood are also considered supportive tweets. The distinction between tweets that are supportive of #MeToo in solidarity and

supportive of #MeToo because of a personal experience are made in the third check of the program. Tweets that are critical or against the movement are labeled as thus, and tweets asking a sincere question or making an unaffiliated remark regarding the movement are considered neutral. Table 3.2 shows examples of the different stance categorizations.

ID	Original Tweet Text	Relevant	Stance	Harassment Category
1	Calling In – Not Calling Out – Men (#METOO BUT NOW WHAT?) Good men wondering what to do, this guide is for you.	1		
2	The #MeToo Photo Going Viral on Instagram https://buff.ly/2ysRfle pic.twitter.com/uyFONo00vA	1		
3	Okay, first off, with all the #metoo stuff going around, what exactly are you classifying as sexual assault?	1	3	
4	the whole MeToo thing seems pointless tbh, like literally all 3.whatever billion women on this earth have experienced degrees of harrassment	1	2	
5	Just another trend started by idiots to get attention. If someone abused you,you should’ve slapped that cunt.Not cry on social media #MeToo	1	2	
6	This #metoo thing has me nearly in tears. It’s not that I didn’t know, but it’s something else to confront the enormity of the problem.	1	1	
7	#MeToo	1	1	4

Table 3.2: Tweet Examples - Stance

3.3 Classifying types of sexual harassment

Following the school of thought belonging to those researchers who consolidated Gruber’s categories is ultimately the strongest approach to this particular problem, as it maintains consistency among the diversity in victims’ perceptions. With regards to the integrity of the classification, Twitter’s character limit prohibits users from providing an adequate context to classify every tweet with confidence, and the personal bias, ignorance, or language choice of each author could result in improper categorization if the algorithm were to attempt to place each tweet within one of Gruber’s eleven very specific categories. Ultimately, three broad categories have been defined that consolidate Gruber’s eleven types into each one: *patronizing*, *unwanted sexual attention*, and *predatory*.

3.3.1 Category 1: Patronizing

The *patronizing* category aggregates the following categories from Gruber’s work:

- Relational Advances
- Possession/display of sexual materials
- Subjective objectification
- Sexual Categorical Remarks

The *patronizing* category is designed to address behaviors that commonly fall into the “gray space” of sexual harassment. Comprehensively, this category is comprised of generally sexist remarks, gender-motivated harassment that is not necessarily pursuing a personal relationship with the recipient, nonverbal displays of harassment that are sexual in nature, and nonsexual behaviors and remarks that the victim interprets as being sexual. Because many victims experience sexual

harassment in contexts that are not applicable to Title VII or Title IX where sexual harassment is upheld, there is a lot of controversy among perception of these behaviors when they occur in a social or non-professional context. In general, if a neutral, unaffiliated third party were to witness the behavior and reasonably deem the behavior as being nonsexual, yet the percipient still perceives it as thus, that type of harassment would be classified as *patronizing* behavior; additionally, this requires that the remark or behavior does not have an obvious sexual goal with the victim. Table 3.3 shows examples of the categorization of tweets describing patronizing behavior.

ID	Original Tweet Text	Relevant	Stance	Harassment Category
1	Simply walking to class in normal, baggy, purposely-unattractive clothes still somehow warranted catcalls and unsolicited comments. #metoo	1	1	1
2	#MeToo When I was 17 my boss screamed @me in front of a store full of customers what's ur problem? R u on ur period or something?	1	1	1
3	#MeToo Never have i been treated with so much disdain and lack of respect due to my sex than in the past 3 years working in a record store	1	1	1
4	because having big boobs means "all the boys will like you" #MeToo	1	1	1
5	Men, don't use derogating\belittling\demeaning words towards women or call men female words - you make it seem women are worth less #MeToo	1	1	1

Table 3.3: Patronizing Tweet Examples

Behaviors that are subject to controversy, such as crossing the boundary between flirting and harassment, are likely to be categorized as *patronizing* because they are not consistently and objectively interpreted as having the intention of pursuing a sexual relationship with the victim. More examples of patronizing sexual harassment include but are not limited to: sexist comments, obscene gestures or drawings about the victim, catcalling or ambiguously sexual behaviors and comments, teasing, banter, jokes, inappropriate comments regarding the victim's body (ex. weight, level of attractiveness, etc.), and other behaviors that would legally contribute to a hostile work environment but cannot be classified as thus when reviewing peer-to-peer harassment.

Some behaviors are ambiguous, and are ultimately evaluated by the effect on the victim. Relational advances, such as repeated contact of a nonsexual nature, could reasonably make a victim fear that the advances in question might be sexually motivated and therefore cause them to interpret the behavior as harmful. In some cases, it is revealed in hindsight that the offender had a sexually motivated goal in regards to the victim, but it could not be objectively determined at the time the behavior was exhibited. While these behaviors ought to be categorized as *unwanted sexual attention*, an unaffiliated third party might witness the behavior without context and objectively think otherwise; therefore, these behaviors should be classified as *patronizing* behaviors instead as it maintains the integrity of the claim and the accusation.

Gruber's category of "sexual categorical remarks" (possession or display of sexual materials and other sexist behaviors and comments) belongs in this category because of the discrepancy that exists between issues that occur in professional versus casual environments. Subjective objectification, which includes remarks made about a victim whether or not she is present (ex. rumors), is also evaluated as being within the patronizing category for the same purpose of accommodating the professional versus casual environment discrepancy. When evaluating contexts that do not apply to Title VII or title IX, these behaviors are not always present to a degree of severity or frequency that they could constitute a different criminal offense (such as stalking, cyberstalking, harassment, etc.). However, the same behavior would easily qualify as sexual harassment in a professional setting

because of the behavior’s contribution to a hostile work environment. Consistency among this legal discrepancy is resolved by classifying these types of behaviors that are not explicitly sexual as *patronizing* behaviors instead.

3.3.2 Category 2: Unwanted Sexual Attention

The *unwanted sexual attention* category aggregates the following categories from Gruber’s work:

- Sexual advances
- Subtle pressures/advances
- Personal remarks
- Sexual posturing

The category of *unwanted sexual attention* includes any behavior, language, questions, or comments of an explicitly sexual nature. Explicitly means that an unaffiliated, objective third party would also find the nature of the comment to be sexual. These comments are an easy classification to make when they take place in a professional environment because they directly follow traditional legal classifications when evaluating a hostile work environment. When evaluating social and casual environments, the behaviors become more complex to categorize because the harasser legally has the room to act within a reasonable degree of respectful, personal interest and flirtation with the victim. Table 3.4 illustrates examples of tweets describing this type of sexual harassment.

ID	Original Tweet Text	Relevant	Stance	Harassment Category
1	“Sleep in my bed I’ll take the floor” “You’re cold, let me in with you” “forget your boyfriend” “Come on I could make you so happy” #MeToo	1	1	2
2	Three diff. men,same day, same line “My wife & I are getting a divorce so...” Military women have experienced harassment since day 1 #MeToo	1	1	2
3	Was 18. New apt, new unlisted ph #. Applied 4 job. Mgr starts sexual, obscene calls 2 me. I called the business. Calls stopped. #MeToo	1	1	2
4	#MeToo I personally witnessed gay men shocked to the core when mildly harassed on the streets. I almost laughed. This is women’s reality.	1	1	2
5	#MeToo . On my bus just now. Men twice my size sitting at the back of a bus proclaiming what they’d do to my body given the chance.	1	1	2

Table 3.4: Unwanted Sexual Attention Tweet Examples

Regarding instances where the harasser claims to be merely flirting in such an environment, the comments or questions made would fall under this category of *unwanted sexual attention* if the victim has communicated their lack of interest or if the comments or questions were egregiously sexual in nature. If the alleged harasser has a reasonable defense for claiming their statements were innocent and a neutral observer would agree, the incident would instead be categorized within the previous category as *patronizing* instead.

In general, any form of non-consensual physical contact is *predatory* with the exception of socially and culturally acceptable forms of physical contact, such as shaking hands or using hugs as a greeting. However, the victim in some cases still interprets these forms of contact as harassment. In some cases this is because the socially acceptable contact is coming from someone who has previously behaved

inappropriately or another valid reason that could make the victim uncomfortable. To categorize these behaviors accurately, the standard of evaluation again is considering the opinion of what an unaffiliated, objective third party would interpret had they walked in to witness the behavior. If the neutral party would reasonably interpret the physical contact as socially appropriate yet the victim maintains that the contact was sexually motivated, the harasser’s behavior is classified as *unwanted sexual attention*. If the neutral third party would reasonably interpret the behavior as strange, unusual, or inappropriate, it would be categorized as *predatory* instead.

Bystander harassment is also considered *unwanted sexual attention* within this thesis. According to Gruber’s classification, a victim of bystander harassment (an individual witnessing harassment that happens to another person) would be considered within his category of sexual categorical remarks, which in this thesis is aggregated into patronizing behavior. However, for this thesis’s applied purpose of categorization, it is more appropriate to classify bystander harassment as a form of *unwanted sexual attention*. Gruber’s topology was developed regarding legally upheld forms of sexual harassment and consequently does not consider some impacts of sexual assault as opposed to sexual harassment. Gruber’s description of bystander harassment does not address witnessing sexual assault, possibly because such an instance would be considered under different criminal boundaries. The likelihood of witnessing sexual assault is extremely small in general, but even more so with regards to professional environments. Because this thesis needs to consider victims who witnessed assault happening to a peer, bystander harassment is classified as *unwanted sexual attention* as opposed to *patronizing*. This is the only significant philosophical or logical deviation from Gruber’s classification rules.

Overall, the category of unwanted sexual attention is comprised of any behavior that is explicitly sexual when the victim has expressed their lack of interest, bystander harassment, any explicitly sexual behavior that occurred within a workplace or academic environment (which falls within Title VII and Title XI regulations), any language or interaction that is beyond what is reasonably accepted as flirting, and forms of conventionally accepted physical contact that the victim has claimed to be sexual.

3.3.3 Category 3: Predatory

The *predatory* category aggregates the following categories from Gruber’s work:

- Sexual bribery
- Sexual assault
- Sexual touching

Predatory behavior includes all forms of non-consensual physical contact, excluding the commonly accepted forms of contact addressed within the previous category. *Predatory* behaviors include attempted or successful rape, sexual assault and battery, and *quid pro quo* arrangements. Table 3.5 demonstrates examples of tweets that fall under the *predatory* category.

ID	Original Tweet Text	Relevant	Stance	Harassment Category
1	Or the time I woke up to a friend assaulting me when I was drunk but never said anything because he made me think it was my fault? #MeToo	1	1	3
2	Latest incident breasts grabbed from behind at a concert. Man laughed at me when I spun around and hit! #MeToo thank you @Alyssa_Milano	1	1	3
3	I have had my ass slapped at two different jobs by two different men. #metoo	1	1	3
4	I was 14. Mom protected him, shamed me, kept me from help. Took me into my 20's to figure out I wasn't to blame. You aren't either. #MeToo	1	1	3
5	Denied an A in HS Chemistry at age 16 for refusing to go into the storage room with my pedophile Chemistry teacher. #metoo #manyexamples	1	1	3

Table 3.5: Predatory Tweet Examples

Some nuances regarding authority exist within the *predatory* behavior category. If the harasser is exhibiting behavior generally classified as *unwanted sexual attention* but exists in a position of authority over the victim (ex. boss or teacher), the behavior considered *predatory* due to the nature of the relationship between the harasser and victim, similar to the way statutory rape is regarded. Sending sexually explicit photos to a victim without the victim’s permission or consent is also *predatory* behavior. While not every instance of a victim being drugged results in actual or attempted assault or rape, being drugged should be considered *predatory* if the victim expresses the fear of assault or rape, particularly if bystander intervention is the reason the attack was unsuccessful. Any and all forms of sexual comments, remarks, questions, and contacts between an adult and a minor is considered within this category, even if the behavior would be considered *patronizing* or *unwanted sexual attention* if the two parties were of legal age; however, if the behavior is in fact *patronizing* or *unwanted sexual attention* and it happens to a minor from another minor, it is not inherently *predatory*.

3.3.4 Category 4: Not Enough Context

The majority of tweets that use #MeToo to express an experience do not describe it with a compelling level of detail. Many users do not describe their experience at all and merely attest their victimhood by writing “#MeToo” by itself. Anything that is clearly claiming to have experienced sexual assault or sexual harassment but does not have a significant enough level of detail to determine the type of harassment falls within this category. Tweets that were classified as having a supportive stance but did not claim victimhood are left blank in this evaluation. Examples of tweets that claim victimhood but without enough detail for a categorization are shown in Table 3.5.

ID	Original Tweet Text	Relevant	Stance	Harassment Category
1	#MeToo	1	1	4
2	I felt so ashamed when I understand what has happen. I thought it was my fault bc I never said anything. It's NOT #MeToo	1	1	4
3	#MeToo I could horrify you with the details, trust me, I see them everyday, but I prefer to show you how strong I've become in spite of it.	1	1	4
4	Talking to him still makes me feel furious and i want to throw up but im trying to forget. I need help #MeToo	1	1	4
5	Mine was the son of my mom's best friend. 20 years later, I still don't know what happened. #MeToo	1	1	4

Table 3.6: Not Enough Context Tweet Examples

Assembling Gruber’s 11 types of sexual harassment alongside into the three categories *patronizing*, *unwanted sexual attention*, and *predatory* successfully accommodates the ambiguous and controversial areas created by each Twitter user’s interpretation of the behavior that he or she experienced. As these users’ individual accounts of their experiences have not necessarily been reviewed (such as by a human resources department, the police, etc.), this is the best approach to this problem. It also avoids claiming a degree of accuracy that cannot be determined with confidence. The *not enough context* label is only for tweets where victimhood is clear; otherwise it is left blank.

3.4 Categorization Rules

Some rules exist to accommodate the parameters of the algorithm. This list of categorization instructions covers all categorization decisions that are not obvious or intuitive.

1. Remove tweets that are not written in English from the dataset.
2. Ignore URLs entirely, even if the rest of the context could be retrieved at that destination. Do not parse and consider words within the URL.
3. Do not consider words that are contained within username mentions beginning with the @ symbol.
4. Consider words that are contained within other hashtags.
5. Context that cannot be confidently determined because of sarcasm should be left blank.
6. Advocacy on behalf of a friend is considered to be only supportive and not a personal experience if the author did not personally witness the assault or harassment.
7. If the tweet only says “#MeToo”, the author should be given the benefit of the doubt and assume that they have indeed experienced a form of sexual harassment or assault. The appropriate label is *not enough context*.
8. Tweets that imply a personal experience but do not explicitly claim one with language are categorized as having a supportive stance but left blank when determining the type of harassment. The only exceptions are tweets that contain only “#MeToo.”
9. If the categorization is difficult to determine, the decision should be made by giving the benefit of the doubt to the author of the tweet. This decision is based on the research suggesting that victims of sexual harassment experience the negative physical consequences (anxiety, insomnia, etc.) regardless of whether or not the victim can accurately identify a behavior as being sexual harassment or not.

CHAPTER 4

IMPLEMENTATION

4.1 User Requirements

The objective of this program is to classify tweets using the #MeToo hashtag, and the core goal is to output a categorization that is as accurate as possible. The classification index of harassment and the ground truth dataset built will help the program ascertain three levels of classification to each tweet. This thesis also contains an evaluate of the classifier model. The last requirement is to build an online UI such that anyone can upload their dataset and, if formatted properly, receive the categorization results of the classifier.

4.2 Development, Design, and Deployment

This project is written with the Python programming language, chosen for its flexibility from the high quantity of libraries available to assist in natural language processing and for supervised machine learning. The pandas library was used to handle the datasets through DataFrame objects, NLTK was used for its text pre-processing features, and spaCy for more powerful prep-processing features. The scikit-learn library was used for classification algorithms and . Jupityr will be used to handle the preprocessing and exporting of the dataset. Matplotlib has been used be used for basic visualizations using the completed dataset.

Dr. Hassan has already collected the dataset, which amasses to approximately half a million tweets of data. Currently, I have manually labeled 1,000 of the tweets according to the proper categorization with the intention to raise that number up to 10,000 manually labeled tweets for the purposes of supervised machine learning. To verify the integrity of my categorization definitions, multiple volunteers of both genders are categorizing the dataset alongside me. The project is hosted on my personal GitHub account under the username “claireballoon”.

While the program can categorize a local dataset simply by running it in Python3 and outputting a file with proper categorization, further utility is desired from this project. The immediate step after completion of the categorization is to create a simple online interface for researchers to use to categorize their data without having to install Python and categorize them locally. Accomplish our core goals necessitates only the installation of Python3 and the appropriate libraries for those wishing to categorize their own datasets of tweets. As Jupityr can export to PDF or LaTeX, no further environment specifications are needed for researchers looking to use the results of this categorization in their data at this time.

4.3 Architecture

Supervised machine learning first requires the raw data to be properly formatted as needed as well as manually classified in order to form a reliable training set. The original, classified set of data is broken up into two sets of roughly 20% (test set) and 80% (train set) of the original. The train set is used alongside the classification rules and machine-learning algorithm being used to develop the model. The test set is used twice: once with the classification removed to which the model is applied, and once at the end with the classification kept in order to compare the accuracy.

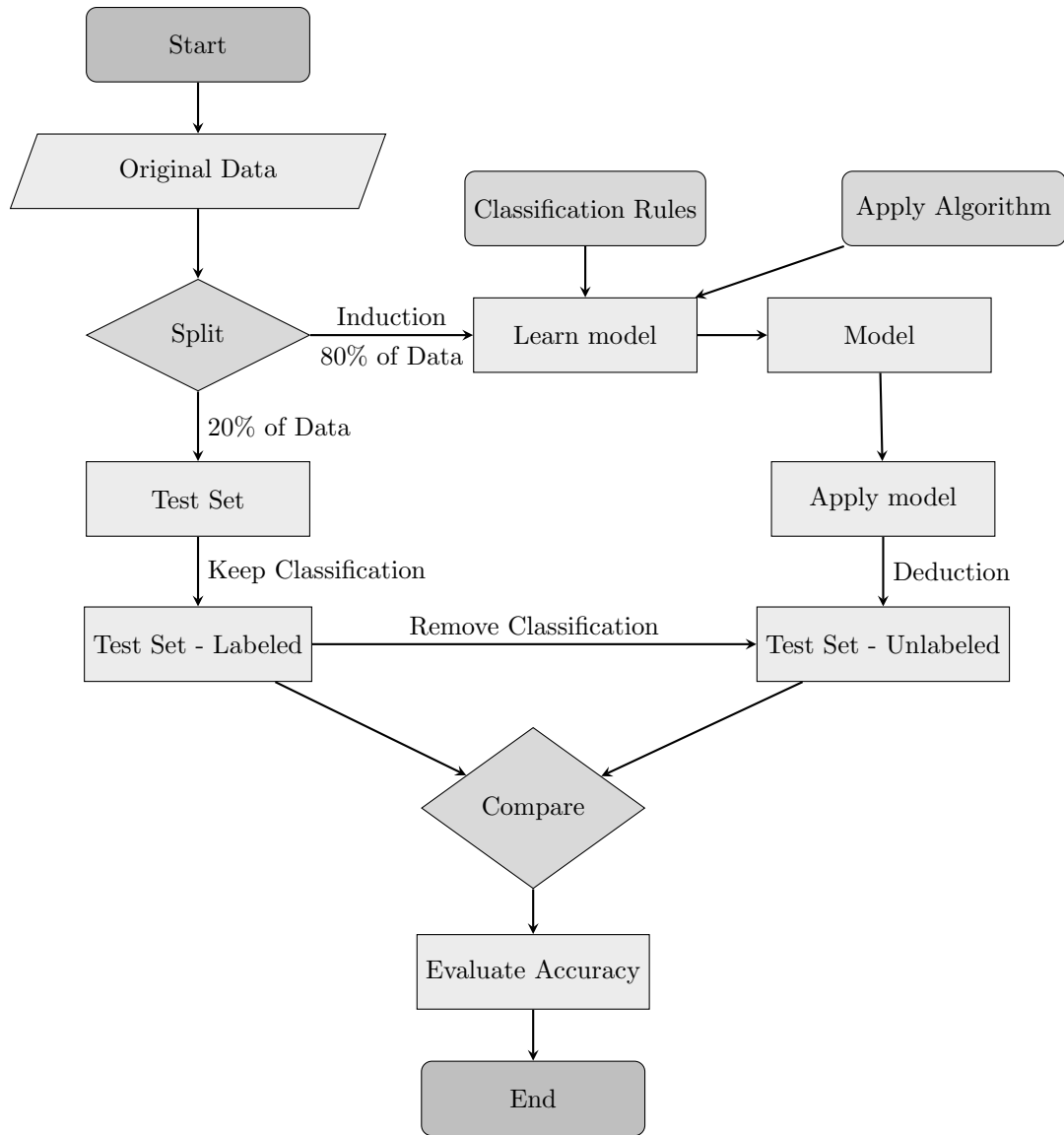


Figure 4.1: Supervised Machine Learning Architecture

4.4 User Interfaces

There are two ways to apply the model. The user can use the command line and a csv file with the dataset of tweets formatted properly (ID of the tweet and the text contained within the tweet). The program will then output a new file with the categorizations and possibly some visualizations. However, this requires the installation of Python3 as well as libraries, packages, and dependencies and can be an inefficient process. Therefore, an online UI has been built that allows the user to see the display of findings and to explore the tool as a testing feature.

4.5 Testing and Integration

The initial hand-tagged tweets will be used to verify whether or not the program has successfully categorized the tweets. By manually categorizing a large enough data set, it can be determined with a high degree of accuracy how successful the algorithm and project will be. Figure 4.2 shows the algorithmic flow in optimizing and testing the accuracy of a model.

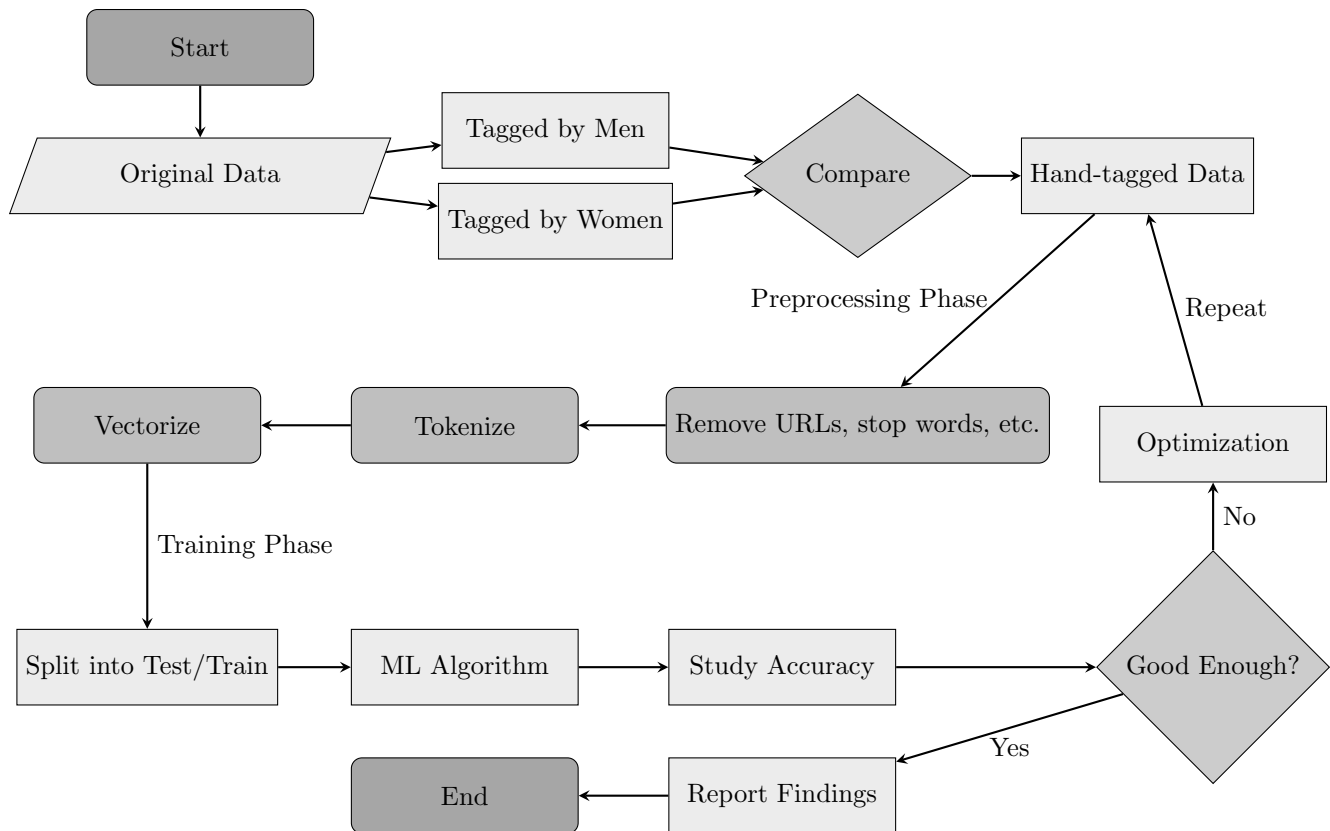


Figure 4.2: Testing and Integration Flow

The models are repeatedly tweaked and optimized until a satisfactory level of accuracy has been achieved. The pre-processing phase initially includes nothing, but after the overall pipeline has been established is continuously revisited and hyper-tuned. This includes the removal of stop words (ex.), removal of URLs, removal of username mentions beginning with the @ symbol, the removal of other punctuation, and more. The pre-processing phase also tokenizes the tweets into “tokens” that create the vocabulary. From there, the vocabulary is vectorized and each token is weighted with its value in the dictionary.

In the training phase, the dataset is split into a train and test set. A machine learning algorithm (SVM, Naive-Bayes, etc.) is applied to the training set with the new vocabulary established during the pre-processing phase, and the test set is used to evaluate the accuracy of that model. If the desired accuracy is not achieved, there are optimization options available, such as hyper parameter tuning.

CHAPTER 5

RESULTS

5.1 Accuracy

This section displays the accuracy obtained for each label and classification. Different classification algorithms were used to compare the relative accuracy, and the individual results of each of these are within their own subsections.

When evaluating the accuracy of a classifier, there are four possible results for the classification attempt: True Positives, True Negatives, False Positives, and False Negatives. These parameters are used when evaluating the accuracy and reliability of the classifier.

		Predicted Class	
		Class=Yes	Class=No
Actual Class	Class=Yes	True Positive (TP)	False Negative (FN)
	Class=No	False Positive (FP)	True Negative (TN)

Table 5.1: Positive and Negative Classifier Terms

Accuracy is defined as the sum of the True Positives and True Negatives over the number of classification possibilities ($\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$). Precision deviates slightly in that it describes the ratio of true results over the number of positive classification attempts made ($\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$). Recall is similar to precision, but describes the ratio of true positives over all of the positive observations in the class ($\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$). The F1-Score is the harmonic mean, or weighted average, of the precision and recall values and takes both false positives and false negatives into the evaluation ($\text{F1-Score} = 2 * (\text{recall} * \text{precision}) / (\text{recall} + \text{precision})$).

5.1.1 Support Vector Machine

The following tables detail the success in accuracy of the SVM classification algorithm.

Classification	Precision	Recall	F1-Score	Support
Relevant	0.00	0.00	0.00	
Irrelevant	0.00	0.00	0.00	
Average/Total	0.00	0.00	0.00	

Table 5.2: SVM Accuracy - Related

Classification	Precision	Recall	F1-Score	Support
Support	0.00	0.00	0.00	
Against	0.00	0.00	0.00	
Neutral	0.00	0.00	0.00	
Average/Total	0.00	0.00	0.00	

Table 5.3: SVM Accuracy - Stance

Classification	Precision	Recall	F1-Score	Support
Patronizing	0.00	0.00	0.00	
Unwanted Sexual Attention	0.00	0.00	0.00	
Predatory	0.00	0.00	0.00	
Not Enough Context	0.00	0.00	0.00	
Average/Total	0.00	0.00	0.00	

Table 5.4: SVM Accuracy - Harassment Category

5.1.2 Naive-Bayesian

Classification	Precision	Recall	F1-Score	Support
Relevant	0.00	0.00	0.00	
Irrelevant	0.00	0.00	0.00	
Average/Total	0.00	0.00	0.00	

Table 5.5: Naive-Bayes Accuracy - Related

Classification	Precision	Recall	F1-Score	Support
Support	0.00	0.00	0.00	
Against	0.00	0.00	0.00	
Neutral	0.00	0.00	0.00	
Average/Total	0.00	0.00	0.00	

Table 5.6: Naive-Bayes Accuracy - Stance

Classification	Precision	Recall	F1-Score	Support
Patronizing	0.00	0.00	0.00	
Unwanted Sexual Attention	0.00	0.00	0.00	
Predatory	0.00	0.00	0.00	
Not Enough Context	0.00	0.00	0.00	
Average/Total	0.00	0.00	0.00	

Table 5.7: Naive-Bayes Accuracy - Harassment Category

5.1.3 Accuracy Improvements

Each label had at least one category that occurred significantly more frequently than the other categories. When determining relevancy, X% of the original data was categorized as *relevant* while only X% of the data was categorized as *irrelevant*. When determining stance, Y% of the original data was categorized as being in *support* but only Y% of the data was categorized as being *against* and only Y% of the data categorized as being *neutral*. When classifying the type of sexual harassment, Z% of the original data was classified as being *predatory*, Z% of the data was classified as being *unwanted sexual attention*, Z% of the data was classified as being *predatory*, and Z% of the data was classified as *not enough context*. These were not proportional training sets. Consequently,

changes were made to improve performance, and this process included the arbitrary reduction of the predominant category from the training set in order to have comparable training sets for each category. Table 5.8 contains the results for improvements in efficiency for the reduction of each major category.

Percentage Eliminated	Remaining Quantity	Improved Precision	Percent Increase
Support			
10%	0.00	0.00	0.00
30%	0.00	0.00	0.00
40%	0.00	0.00	0.00
50%	0.00	0.00	0.00
60%	0.00	0.00	0.00
Relevant			
10%	0.00	0.00	0.00
30%	0.00	0.00	0.00
40%	0.00	0.00	0.00
50%	0.00	0.00	0.00
60%	0.00	0.00	0.00
Not Enough Context			
10%	0.00	0.00	0.00
30%	0.00	0.00	0.00
40%	0.00	0.00	0.00
50%	0.00	0.00	0.00
60%	0.00	0.00	0.00

Table 5.8: Category Reduction

CHAPTER 6

SUPPORTING MATERIALS

6.1 Mitigating Sexual Harassment

The context for this research topic is to explore possible avenues in which sexual harassment could be mitigated or reduced. This context was to be explored by finding new trends among the tweets themselves or by implementing the classifier online to assist in mitigating sexual harassment through technology. Research into sexual harassment and harassment online guided the development of a survey questionnaire that was incorporated into the direction of this project.

6.1.1 Literature Review

Existing papers and their limitations.

6.1.2 Other Problems

Other issues exist that have not been studied in depth, and therefore a quantifiable impact or degree of severity of the issue has not been quantifiable measured. Regardless, some of these issues are worth nothing.

Woman, Action, and the Media (WAM!) created a team in 2015 to assist Twitter in evaluating harassment that took place within the website. While this tool does not look to analyze whether or not a tweet may be of a harassing nature, this report afforded insight into Twitter's system that could potentially have use in other contexts. Serious security flaws exist in the lack of maintenance invested on archiving and storing all tweets and messages posted by users, particularly those of a harassing nature, and this results in two majors problems. First, it eliminates the evidence that victims have to defend themselves; secondly, Twitter's own reporting system *only* accepts live versions of the tweet, further reducing victims' options.

Twitter cannot remove a tweet of a harassing nature if the evidence submitted is a screenshot—the tweet must still be available to be viewed live to be used as evidence when determining whether or not to take action against a Twitter user. More importantly, this report also reveals that when a harasser deletes his or her own tweet, or Twitter removes the offending tweet, that tweet is no longer available to law enforcement agencies [Matias et al., 2015].

Historically, harassing or stalking types of behavior occurred through phone calls, cards, or other more direct forms of inserting a presence in the victim's life. Now, it is quite common for harassment (both sexual harassment and not) to persist over social media. WAM!'s report indicated that the lack of persisting data and messages over social media platforms is a disservice to victims as it often leaves them without evidence to take to the police and without evidence to get their harasser removed.

Twitter is not the only social media platform that has this problem. Instagram users have the option of deleting their comments and direct messages as well as permanently hiding their presence from a user entirely, including their username. Instagram's public policy regarding their cooperation with law enforcement states that they will comply with valid subpoenas, court orders, or warrants under outlined circumstances. However, if a victim were to be harassed over Instagram and that user removed their messages and presence from the perspective of the victim, the victim would have no evidence to bring to law enforcement in order to file charges and proceed with a legal request for the data. The lack of persistence and lack of archiving disenfranchises victims of sexual and non-sexual harassment on social media.

6.2 Survey

In order to further develop an understanding of how to mitigate sexual harassment, a questionnaire was developed. The questionnaire was reviewed by IRB and found not to need approval, as it dealt with voluntary participants above the age of 18, no identifying information was collected, and responses were completely anonymous.

6.2.1 Questionnaire

Answer choices marked with an asterisk (*) indicates that the answer choice had a text box where the participant could type an answer.

- 1) What is your age?*
- 2) What is your gender identity?
 - Male
 - Female
 - Other*
 - Prefer not to respond
- 3) Select all of the following behaviors that you consider to be forms of sexual harassment and the degree of severity you consider the infringement to be.

Action	Mild	Moderate	Severe	Not Sexual Harassment
Staring or leering	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Whistling, catcalling, or winking at someone	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pinching or poking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sexist comments	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Inappropriate drawings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Messages on social media from strangers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Making lewd/sexual remarks about someone's looks or body	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Obscene gestures or sounds	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sending repeated messages, calls, or other forms of contact after the receiver expresses disinterest	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Asking overly personal questions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stalking or Harassment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Groping or other touching	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sending unsolicited photos of private body areas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (please specify)*	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- 4) What social network platform do you use the most frequently?
 - Facebook
 - Youtube
 - Instagram
 - Twitter
 - Reddit

- Tumblr
 - Other (please specify)*
- 5) Considering the social media network designated above, how would you rate this social network platform in terms of its ability **to prevent** sexual harassment from happening?
- It does not have any mechanism to prevent sexual harassment from happening
 - It has some mechanisms but these are not adequate
 - It has a good set of mechanisms to prevent sexual harassment from happening
 - I don't know
- 6) Considering the social media network designated above, how would you rate this social network platform in terms of convenience of **reporting sexual harassment**?
- It does not have any mechanism to report an instance of sexual harassment
 - It has some mechanisms but these are not adequate
 - It has a good set of mechanisms to report sexual harassment
 - I don't know
- 7) In your opinion, in what ways could this social network platform be designed to prevent sexual harassment from happening? Mark all that apply.
- Increasing the prevalence/priority advertisement which addresses sexual harassment
 - Stricter regulation of posts/content on social media sites
 - Adjustments to news feed algorithms to prioritize content regarding victims of sexual harassment
 - Automatically archiving or documenting instances of sexual harassment online for evidence
 - Connecting victims on social media with help (local authorities, other victims, or local resource options)
 - Please describe any other thoughts you have on ways technology could mitigate sexual harassment:*

6.3 Findings

No results at this time.

CHAPTER 7

CONCLUSION

None becuase I suck

BIBLIOGRAPHY

- [Chamberlain et al., 2008] Chamberlain, L., Crowley, M., Tope, D., and Hodson, R. (2008). Sexual harassment in organizational context. *Work and Occupations*, 35(3):262–295.
- [Ferguson et al., 2005] Ferguson, T., Berlin, J., Noles, E., Johnson, J., Reed, W., and Vincent Spicer, C. (2005). Variation in the application of the “promiscuous female” stereotype and the nature of the application domain: Influences on sexual harassment judgments after exposure to the jerry springer show. *Sex Roles*, 52:477–487.
- [Fitzgerald et al., 1995] Fitzgerald, L. F., Gelfand, M. J., and Drasgow, F. (1995). Measuring sexual harassment: Theoretical and psychometric advances. *Basic and Applied Social Psychology*, 17(4):425–445.
- [Gruber, 1992] Gruber, J. E. (1992). A typology of personal and environmental sexual harassment: Research and policy, implications for the 1990s. *Sex Roles*, 26:447–464.
- [MacMillan et al., 2000] MacMillan, R., Nierobisz, A., and Welsh, S. (2000). Experiencing the streets: Harassment and perceptions of safety among women. *Journal of Research in Crime and Delinquency*, 37(3):306–322.
- [Matias et al., 2015] Matias, J. N., Johnson, A., Boesel, W. E., Keegan, B., Friedman, J., and DeTar, C. (2015). Reporting, reviewing, and responding to harassment on twitter.
- [Murrell and James, 2002] Murrell, A. J. and James, E. H. (2002). Gender and diversity in organizations: Past, present, and future directions. *Sex Roles*, 45:243–257.
- [Pierce, 1989] Pierce, M. R. (1989). Sexual harassment and title vii – a better solution. *Boston College Law Review*, 30(4):1071–1101.
- [Schneider et al., 1997] Schneider, K. T., Swan, S., and Fitzgerald, L. F. (1997). Job-related and psychological effects of sexual harassment in the workplace; empirical evidence from two organizations. *Journal of Applied Psychology*, 82:401–415.
- [Studzińska, 2015] Studzińska, A. (2015). Gender differences in perception of sexual harassment. (2015TOU20052).
- [Till, 1980] Till, F. J. (1980). Sexual harassment: A report on the sexual harassment of students. pages 4–22.
- [York and Brookhouse, 1988] York, K. M. and Brookhouse, K. J. (1988). The legal history of work-related sexual harassment and implications for employers. *Employee Responsibilities and Rights Journal*, 1(3).