May 2016

Assessing gender inequality from large scale online student reviews

Souradeep Sinha
Syracuse University

ABSTRACT

Career growth in academia is often dependent on student reviews of university professors. A growing concern is how evaluation of teaching has been affected by gender biases throughout the reviewing process. However, pinpointing the exact causes and consequential effects of this form of gender inequality has been a hard task.

Current work focusses on university-wide student reviewing system, that depends on objective responses on a Likert scale to measure various aspects of an instructor's quality. Through our work, we access online student review data which are not limited by geographies, universities, or disciplines.

Thereafter, we come up with a systematic approach to assess the various ways in which gender inequality is apparent from the student reviews. We also suggest a possible way in which bias related to the gender of a professor could be detected from both objective numerical measures and subjective opinions in reviews. Finally, we assess a logistic regression learning algorithm to find the most important factors that can help in identifying gender inequality.

Assessing gender inequality from large scale online student reviews

by

Souradeep Sinha

B. Tech., West Bengal University of Technology, 2014

Dissertation

Submitted in partial fulfillment of the requirements

for the Degree of Master of Science

in Computer Science

Syracuse University

May 2016

Acknowledgments

I would like to thank my advisor Dr. Reza Zafarani of the Department of EECS at Syracuse University. The door to Dr. Zafarani's office was always open, sometimes even several times in a day, whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right the direction whenever he thought I needed it.

I would also like to thank the experts who were involved in the Defense Committee:

- Dr. Carlos Caicedo, Director, Center for Convergence and Emerging Network Technologies (CCENT)
- Dr. Kishan Mehrotra, Chair, Department of Electrical Engineering and Computer Science
- Dr. Edmund Yu, Associate Professor, Department of Electrical Engineering and Computer Science

Without their participation and input, the Defense could not have been successfully conducted.

This thesis would not be complete without mentioning my laboratory mates Shengmin, Majed and Kandarp, who haven't just been friends, but have rooted for me during tough times. And the same goes to my friends at SCPS and ETS, who were the best support system that one can get away from family!

Finally, I must express my very profound gratitude to my parents Arunangshu and Mukul Sinha, my elder brother Kaustav Das, and my friend for a long time Aditi for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Author

Souradeep Sinha

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Perhaps the importance of achieving gender equality is well represented by the fact that the United Nations have made it one of their top organizational priorities[1]. The present is a critical moment in time, considering the disparate state of genders in today's society. While the equilibrium has been attained in some quarters of the society, many others still remain to achieve parity. According to various UN reports, no country has yet achieved complete equality, and sections of the world face an alarmingly high rate of gender inequality in the form of violence and lack of safety, health access, education and income disparity. Gender inequality marginalizes rights by virtue of a person's gender, thus depraving the chance to a fair economic, cultural and social environment. This form of inequality is often fuelled by discriminatory attitudes, social norms and cultural beliefs. Being a long standing problem, it has limited opportunities to realize one's own potential, especially among women. These disadvantages are often fuelled by lack of access to essential services, thus restricting growth. To combat this issue, it is highly essential to investigate the key underlying factors such as gender roles and gender bias that retard the achievement of gender inequality.

The internet, with its vast expanse of outreach and ability to engage the populations across political, social and geographical boundaries may serve as a common tool to assess as well as to curb gender related issues. Even today, texts serve as the major medium of communication between any two parties. This enables us to investigate a history

[1]Planet 50-50 by 2030: Step It Up for Gender Equality

of easily available communication, and may open avenues to detecting several forms of gender discrimination. While sensitive data is often protected, there are other public sources of data in the form of open social media transcripts, weblogs, public comments, discussion forums and threads and product/service reviews. A great platform to mine opinion, sentiment and other macro level reactionary information from the populace, the internet can serve as a great starting point in the pursuit of understanding various forms of gender biases. Also, the currency of information flow on internet platforms ensures that contemporary facts are easily identified. This makes the scrutiny of more recent issues relatively easier.

A very recent study aimed at identifying the reasons behind lack of women in scholarly pursuits revealed that gender biases have a key role in putting women at a tough spot in their pursuit of a career in research (Boring, 2015). The study also detailed at how gender biased student reviews hold back female faculty members from achieving their full potential. Our work is inspired from this study, and harnesses the power of the internet to expand the scope of available data to be put under observation.

Our goal in this study is to identify and report the key areas of the student review process where gender inequality is starkly evident, and formulate an empirical process by which gender biases can be detected, if indeed they are present. While it is hard to track every possible source of student review websites with vastly differing opinions about multiple faculty members across several universities, we fixate on one such popular website that records most of the important details in their reviews. Coincidentally, the website also curates the largest database amongst all of its competitors. This enables us to expand the horizons of study across different geographies, disciplines and allows us to follow the differing trends over time.

In this thesis, we follow a systematic approach to detecting the presence of gender bias. We first identify a list of hundred questions with regard to gender inequality, that can be further probed with the availability of apropos data. Thereon, we focus on a single issue - teaching evaluations - and collect relevant data from online sources. Once the raw data is cleaned and preprocessed, we use it to scope the presence of gender inequality. Finally, we proceed to formulating a hypothesis to validate the existing gender bias among student reviews. Our key contributions are:

1. Recognize the various ways in which the society is affected gender inequality, and wrap them into a hundred relevant questions.

2. Collect information about university level instructors, identify their gender and gather student review data from RateMyProfessors[2].

3. Perform data analyses to report some interesting artefacts of the reviews that hint at gender inequality.

4. Formulate a hypothesis on the presence of gender bias, and train a learning algorithm to test and validate the hypothesis.

In the next chapter, we review previous research related to gender inequality in the fields of economics, engineering, recruitment, software development and university level teaching. In Chapter 3 we describe our data and perform preliminary analysis and report some interesting facts and visualizations illustrating gender bias. Chapter 4 demonstrates the experimental set up, and training a learner algorithm that validates the presence of gender inequality, and identifies the key features that are most important in detecting gender bias. Chapter 5 provides concluding remarks and discussion on how this work can be carried on further. Finally among appendices, Appendix A lists a hundred questions that has been recognized in respect to gender inequality. Appendix B provides a brief overview of the data collection and class labelling process that enabled this study.

[2] http://ratemyprofessor.com

Chapter 2

Related Work

Gender equality - a key human right, has been hard to achieve, and continues to be a global challenge. Yet, it was only in 1964, that using gender as a basis of discrimination was finally prohibited by the law in the United States.[1]. Subsequently, national and international organizations have been established to achieve the much required parity. A growing number of social scientists, policy makers, psychologists, educational researchers, public health specialists, STEM researchers and human resource managers have stepped in to investigate gender inequality. Needless to say, this problem deters socio-economical growth and affects the world population at large.

It was perhaps Hausmann and colleagues who provided the most geographically detailed analysis of gender based inequalities including data from 142 countries (Hausmann et al., 2014). Analysing gender gap in health, educational attainment, economic participation and political empowerment, they find that "no single country in the world has fully closed the gender gap". Societally, Glick and Fiske showed that gender inequality is derived from both hostile and benevolent sexism, and impacts the genders by providing a justification for gender roles (Glick and Fiske, 2001). Fuwa investigated the division of work in family life and found that the share of work puts women at a disadvantage in lesser egalitarian countries (Fuwa, 2004), and Jacobs found that women fare poorly in terms of outcomes of schooling and access to college experience (Jacobs, 1996). A study by Kenworthy and Malami concludes that the number of women involved in professional-type jobs are

[1] 110 Cong. Rec. 2577 (1964)

positively correlated with political representation of women nationally (Kenworthy and Malami, 1999). Nevertheless, women have been preferred second to men in labor queues (Reskin and Roos, 2009). From the perspective of crimes and violence, sexual violence increases as educational and occupational statuses for women go lower (Yodanis, 2004); and their homes being the most common place of an attack, with their partner being the most likely offender, (Bruynooghe et al., 2000).

To provide further context to our work, we elaborate on (1) how gender inequality is assessed in **economics**; (2) the effects of gender bias in college education with focus on **women in engineering schools**; (3) how gender bias impacts **recruitment and hiring**; and (4) the inequality that is associated with women's contributions in **open source software**. Then, (5) we shed light on **teaching**, and the problems associated with student reviews. Finally, we discuss (6) how gender bias manifests itself in **student reviews** of professors at a distinguished French university.

2.1 Economics and capability theory

Robeyns reasons that gender inequality can be conceptualized and assessed with relevant capabilities using *Amartya Sen's capability approach* (Robeyns, 2003). Capabilities are people's abilities to carry out certain functions. Examples include, but are not limited to being well fed, community participation, being healthy, caring for others, and working in the labor market. She uses fourteen such capabilities (see Table 2.1) to assess gender inequality. While defending the choice of capabilities, she found that each capability is affected by gender inequality in varying proportions. While there were no markers of gender biases in the chances of being born in Western societies, women fared better than men in terms of life expectancy at birth. In the case of mental well being, women suffered more than men, with anxiety and depression being more common among women than

Table 2.1: Capabilities

Capability	**Description**
Life and physical health	Being able to be physically healthy and enjoy life of normal length
Mental well-being	Being able to be mentally healthy
Bodily integrity and safety	Being able to be protected from violence of any sort
Social relations	Being able to be part of social networks and to give and receive social support
Political empowerment	Being able to participate in and have a fair share of influence on political decision-making
Education and knowledge	Being able to be educated and to use and produce knowledge
Domestic work and nonmarket care	Being able to raise children and to take care of others
Paid work and other projects	Being able to work in the labor market or undertake projects, including artistic ones
Shelter and environment	Being able to be sheltered and to live in a safe and pleasant environment
Mobility	Being able to be mobile
Leisure activities	Being able to engage in leisure activities
Time-autonomy	Being able to exercise autonomy in allocating one's time
Respect	Being able to be respected and treated with dignity
Religion	Being able to choose to live or not to live according to a religion

men (Doyal, 2000). Women were also associated with more frequent and severe sexual and physical violence. Women experienced most domestic violent attacks with their partner being most probable attacker (Bruynooghe et al., 2000). As the frequency of reporting such crimes vary, results still remain inconclusive about gender inequality. Owing to more extensive networks in political and economic circles, men have better economic and public life. Women, on the other hand involve themselves with informal networks of friends and family (Fuhrer et al., 1999, Munch et al., 1997). While both genders have equal access to formal education, gendered social norms still continue to challenge young women in accruing knowledge and pursuing degrees. Parents have been found to favor their sons over their daughters when it comes down to providing encouragement to do well at higher education. Teachers are found to be more attentive of male students as compared to females and women are often alienated from traditionally 'male' courses such as science and technology (Connell, 1989, Warrington and Younger, 2000). Politically, females hold fewer offices across all countries. Female politicians often face masculine culture in politics including power tussles, aggressive tonalities, continuous interruptions between discussions, and so on (Robeyns, 2003). Domestically, women participate more in 'non-market' care for children, elderly dependants and the sick (Robeyns, 2003). Lelli and other researchers assess housing and neighborhood conditions which show no trace of gender inequality (Lelli et al., 2001). Men have been found to spend significantly more time towards leisure activities than women, though they are known to have similar amounts of free time. This can be explained by the interruptions in leisure that women have to take, to tend to unpaid work or childcare (Bittman and Wajcman, 2000, Glorieux et al., 2001). Women also face more time-pressure than men in the case of families where both the husband and wife are employed, in which women are found to handle household responsibilities that cannot be postponed (Phipps and Burton, 1995). In the case of being respected, women are systematically devalued by their gendered societies (Robeyns, 2003). Even for religion and freedom to practice, most religions make statements favoring men and the andro-centric ones do not identify women as religious leaders (Robeyns, 2003). Due to the nature and importance of these real-world capabilities, it is difficult to quantify them and measure gender inequality.

2.2 Women in engineering

McLoughlin's study on gender biases amongst undergraduate engineering education demonstrates 'spotlighting', which refers to women being singled out in ways that make them feel uncomfortable (McLoughlin, 2005). Spotlighting manifests itself in three ways, based on the intent of the 'spotlighter':

1. *Type I* spotlighting or overt sexism, that discriminates women with the intent to harm them;
2. *Type II* spotlighting or tacit sexism, *i.e.* with no intent to harm or to help women;
3. *Type III* spotlighting, which is a relatively newer observation, and arises from the intent to help women.

Type I is well towards recession. Even though some women have reported being subjected to inappropriate and offensive behavior, a majority of them felt that they did not encounter such obnoxious situations. *Type II* occurs when societally established gender roles are defied by women's aspirations, such as engineering students. Women have often felt like an 'outsider' and excluded in male dominated engineering schools. *Type III* emerged from the women targeted endeavors that aimed at curbing *Type I* and *Type II*. It is derived from the patronizing thought that women feel less adequate and capable. Another direct derivative from this type would be the hostility and criticism it generates from male students. Males can question the 'privileged' status of females, and form misconceptions and pre-judged notions about their opposite sexed colleagues.

2.3 Recruitment and hiring

It is widely acknowledged that gender biases are prevalent in the recruitment and hiring processes, and has lasting effects on the gender gap in employment, and subsequently on the pay gap. Forms of discrimination throughout recruitment can be hard to document, and therefore, hardly investigated. Often discriminatory practices arise due to several informal factors like professional networking, internal hiring, employee referrals, and so on. Petersen and Togstad use proprietary data involving job applications and subsequent job offers from 866 applicants for 15 open positions in a Norwegian bank, and provide a statistical analysis of gender inequality in the hiring process with interesting results (Petersen and Togstad, 2006). Counter intuitively, logistic analyses of managerial, professional and secretarial level jobs show that there is no evident disadvantage to women. While the applicants pool consists of less than 50% women across all but secretarial positions, the gender composition of applicants who get offers favors women. No evidence suggesting women receiving fewer offers than they would under equal circumstances was found. However, this picture deviates slightly when family-life is considered. The results show opposite trends for men and women getting an offer. While offers are extended at a higher rate to single women than single men, being married increases offers for men over women. Though having children results in higher likelihood of receiving offers for both genders, but ultimately it is the men who get an advantage in being offered a job, when being part of a family.

2.4 Open source contributions

Terrell and collaborators study gender bias by examining acceptance rates of contributions from software developers of both genders in an open source community (Terrell et al., 2016). They inspect historical Github data to assess how often pull requests from women are accepted. Investigating a large dataset of over 4 million users, they initially hypothesize that it is less likely that a woman's contribution is accepted. But, this hypothesis proves to be false, in fact, and men face a higher rejection likelihood. Considering both internal and external project contributors to a software project, acceptance rates are higher among women. Examining success rates over a history of pull requests, women have higher acceptance rates in the case of their first pull request being accepted. They maintain these higher rates till the sixty fourth such pull request made being accepted. Analyzing the need based problem or an 'issue', they find that the pull requests made by women are less likely to cater to an immediate need; however, this fact fails to explain the higher acceptance rates. Apart from these observations, they also find that not only do women participate in larger changes to source code (notwithstanding the risks of more buggy code), their acceptance rates trump those of men in the case of top programming languages. To test type III bias - a concept we discussed before (McLoughlin, 2005), which entails bias by singling out a woman with an intention to help - they observed the mean acceptance rates of contributors who give a clear indication of their gender versus those who do not. This helped shed light on the gender based discrimination. They found that women external contributors who identified their genders faced lower acceptance rates as compared to women who did not identify, or to contributors who were internal to a project. While being associated to a project may not affect acceptance rates, the discriminatory stance is more stark. The above results showed that while women may be more competent than men in the open source arena, being labelled by their gender leads to them losing 'favors' from the community, raising doubts that only merit alone defines success in open source communities.

2.5 Problems of student evaluation

Stark and Freishtat probes into the reliability of student ratings in the form of SET (student evaluations of teaching) to measure teaching effectiveness (Stark and Freishtat, 2014). SET scores act as a primary gauge for the faculty and the department alike, when considering important decisions, including promotions and tenure. Given the ordinal nature of scoring system, general statistical aggregations such as averaging, is prone to several forms of errors. The authors also argue that while SET is aimed at teaching effectiveness, it does not necessarily succeed; instead it ends up measuring "8what students say" based on their own perceptions of teaching. Moreover, SET have been found to show different kinds of trends. SET scores were found to be highly correlated with students' grade expectations (Marsh and Cooper, 1981), students' reaction to 30 seconds of silent video of the instructor is enough to predict how the instructor is going to perform on his/her SET (Ambady and Rosenthal, 1993), and Worthington showed that answers to some relevant questions are influenced by factors that are irrelevant to learning (Worthington, 2002). Due to the open and optional nature of SET surveys, the low response rates impact the reliability of the scores. However, while authors feel that ratings to objective questions do not make a fair assessment of the instructors, they recommend paying special attention to the comments collected from students to have a more subjective assessment of the teacher's performance.

2.6 Gender roles and student evaluations

A different study by Andersen and Miller examines how gender bias detrimentally affects student evaluations of teaching with respect to students' expectations (Andersen and Miller, 1997). Historical studies prove to bear ambiguous conclusions in how male and female professors are rated. But, student expectations of how their instructor conformed to gender roles had a consistent effect on the instructor's evaluation. In fact, ratings were generally positive if the instructor lived up to those expectations, and were adversely impacted if they deviated from those stereotypical perceptions (Kierstead et al., 1988). ***Gender*** and ***discipline of study*** were found to be the two most significant factors behind these expectations. By virtue of these stereotypes, male professors were encouraged to be more authoritative, adversarial than women, in their style of teaching. This negatively affected women professors, who faced a two sided blade. On one hand, they lost credibility as a professor, if they assumed the role of a gender approved nurturant. On the other, assuming a professional, competent and knowledgeable role, meant that they failed to be compatible with their associated stereotypes. An analysis showed that female faculty members were rewarded as compared to males for 'supportive' and 'nurturing' behavior, but were punished, for 'objective', 'authoritarian' role-inconsistent behavior. From a single semester study involving 9005 students at a particular university, it was found that female professors had lesser scores in 'global evaluations' and lower competency ratings than males, even when variables like GPA, expected grade, students' sex, discipline and course size were controlled for (Sidanius and Crane, 1989). Because women are assumed to be 'supportive listeners', students do not report women to be more accessible than their men counterparts, even when they admit that the female faculty have spent more time assisting them (Sandier, 1991). When studied in a laboratory with read out descriptions given to students, Kierstead and colleagues find no difference between social and non-social males, however female instructors received lower ratings if they were not friendly outside the classroom (Kierstead et al., 1988). In case of equally friendly male and female professors using 'slide lectures', male professors received higher ratings. Langbein observes that if the expected grades were low, female professors' evaluations were lower in comparison to males (Langbein, 1994). Overall, several studies have concluded that displaying a mixture of masculine and feminine characteristics helps in achieving higher ratings from student evaluations (Basow, 1995, Freeman, 1994, Martin, 1984). A recent study by Statham and researchers has examined gender differences in teaching styles from samples of student evaluations of widely varied disciplines at a large university. Considering departments with varying 'male dominance', studies have found that women focus on the student, whereas men, on themselves as the locus of learning. Adhering to gender roles - for females, acknowledging input and personalizing with students putting focus on interactions; while with males, teaching 'as an expert' while presenting materials, admonishing and interrupting students - led to higher evaluations (Statham et al., 1991).

Table 2.2: Chances of male and female teachers to obtain *good* and *excellent* scores

Student Performance	**Chances of a male teacher**		**Chances of a female teacher**	
	Male student	**Female student**	**Male student**	**Female student**
Good	89%	85%	84%	83%
Average	84%	80%	78%	77%
Poor[2]	23%	16%	15%	14%

One of the more recent works by Boring investigates gender biases in student evaluations at a French university to find the key reason behind women coming second to men in academic careers (Boring, 2015). While research productivity and teaching effectiveness both impact academics, the former matters the most in career advancement decisions in top institutions. Research has shown that time allocation to the teaching-research balance differ by gender, with women spending more time on teaching compared to men. She puts forth and defends her argument that it is because of the presence of gender bias in teaching evaluations, women end up focusing more on teaching in order to stay on the same SET score levels as men. She also finds that students value time-consuming dimensions of teaching such as course preparation and feedback in women, whereas lesser time-consuming dimensions like class leadership and animation skills in men. Male students

[2] only *excellent* scores were reported for the students who performed poorly

were also found to give higher scores to male teachers in terms of overall satisfaction. In fact, male students were 30% more likely to rate a male teacher as *excellent* compared to a female teacher. When actual performance of students in their final exams were considered, they performed equally well, irrespective of the gender of their instructor. Thus, suggesting no explanation to the disparity of male students towards female teachers, other than a possible gender bias leading to an inequality among the genders. A trend of differential scores was noted where students rated teachers according to gender based stereotypes within the teaching dimensions. The four teaching dimensions tested were (1) preparation and organization of classes, (2) quality of class material, (3) clarity of course assessment criteria, and (4) usefulness of feedback. Though female teachers obtain more favorable ratings for the much more time-consuming first three of the above dimensions, male students still gave a small premium to male teachers. Male teachers score favorable ratings from both male and female students in dimensions like quality of animation and class leadership. Students also found men teachers more knowledgeable, even if objective measures proved that they learn the same from both genders. Students were awarded two grades for each coursework - a continuous assessment grade for a seminar style of classes awarded by the seminar teacher, and a final exam grade which was corrected anonymously. Students also were to complete SETs before they took their final exams, but at a point where they knew their continuous assessment grades. Not surprisingly, SETs were found to correlate with the continuous assessment and not the final grade, which also suggested that SETs merely represented student's perception of teaching effectiveness, and not the actual effectiveness. Considering the performances of students and how their grade expectations affected the SET scores (see Table 2.2), women teachers were consistently penalized with lower chances of receiving an *excellent* score, independent of the gender of the student, or whether the student received a good, average or poor grade. Another important result was how diversity of gender in a teaching population affected students.

By using regression analyses, she shows that using a combination of one male and two female in a triplet team of teachers scores the highest coefficients, suggesting that both male and female students are particularly satisfied with the singular male teacher, yet are harsher on the two female teachers. Using these facts, it becomes obvious that students rate teachers according to gender stereotypes.

Most of the previous work that discuss the impact of gender bias on the teachers, instructors and professors are restricted to analyzing the objective measures of student evaluations. Our work focuses on data generated from an online review website, where students voluntarily provide feedback about a professor whose class they may have taken in the past. To our knowledge, this is the first time that a systematic study of online student has been performed to an order of millions of reviews about hundreds of thousands of professors at American universities and beyond. In addition to objective measures, we also analyze textual characteristics of review content. This approach lets us to have a better realization of the reviewer's subjective comments, thus enabling us to mine their opinions depending on the professor's gender. We also demonstrate an approach that suggests the possible gender bias influencing the way professors are reviewed.

Chapter 3

Measurements

Gender inequality is a complex problem impacting societies. It can go unnoticed, and is often concealed by the grayer shades of gender roles. To systematically study gender inequality, we aimed to identify different areas and ways in which this problem can manifest itself. In order to do this, we studied available literature, blogs and articles among other sources, and identified a hundred key questions listed in Appendix A. These questions, if answered quantitatively using the right data, can help us understand the dynamics of gender inequality better. Encouraged by recent work on gender bias in student reviews, we focused on the following question: "Are male and female professors rated differently by students on online platforms?" (Boring, 2015, Powell, 2016). Treating this question as a starting point, we gathered relevant data from an online student review website, RateMyProfessor. Using this data, we performed an initial analysis of gender bias with a few more in-depth questions, which we shall discuss later on in the chapter.

3.1 Data

Before we study the effects of a professor's gender on the review that he/she earns, we provide some insight on our source of data - Rate My Professors, hereon abbreviated as RMP. According to the RMP 'About' page[1], the site is "the most highest trafficked site for quickly researching professors". While there are other websites that serve a purpose of

[1] http://www.ratemyprofessors.com/About.jsp

collecting student reviews such as Uloop[2], Koofers[3] and MyEdu[4], we chose RMP because it has the largest collection of student reviews, collected publicly and anonymously.

Table 3.1: Description of Professors

Feature	Type	Note
Professor ID	Nominal	Hexadecimal indices
Gender	Nominal	$gender \in \{M, F, U\}$
Details		
City	Nominal	$\mid cities \mid = 2128$
State	Nominal	$\mid states \mid = 107$
University / College	Nominal	$\mid universities \mid = 4181$
Discipline	Nominal	$\mid disciplines \mid = 2104$
Ratings		
Overall Quality	Numeric	Real value in $[1, 5]$
Clarity	Numeric	Real value in $[1, 5]$
Easiness	Numeric	Real value in $[1, 5]$
Helpfulness	Numeric	Real value in $[1, 5]$
Average Grade	Ordinal	Value in set $\{A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-, F, N/A\}$
Top 20 tags		
Tag Name	Nominal	See Figure 3.1 for the tags, maximum number of tags = 20
Tag Score	Numeric	$tag\_score \in \mathbb{N}$
All Reviews		See Table

Table 3.1 shows the data collected of each professor. We collected information for 921,859 professors on RMP. The professors were from 2,128 cities (US and other countries) across 4,181 universities and 2,104 disciplines. RMP reports the average $Clarity$, $Easiness$ and $Helpfulness$ scores assigned by students. RMP uses $Clarity$ and $Helpfulness$ to internally calculate the $OverallQuality$ score, which as the name suggests, serves the purpose of overall quality rating. The $AverageGrade$ is calculated by aggregating the grades that the reviewer indicates that he/she has received. The top 20 tags received by the professor are also reported, with corresponding $TagScore$ representing the number of

[2] http://www.uloop.com/professors/
[3] https://www.koofers.com/rate-professors
[4] https://www.myedu.com/professor-recommendations/

occurrence for each particular tag.

Table 3.2: Description of Reviews

Feature	**Type**	**Note**
Comment ID	Nominal	Hexadecimal indices
Date	Date	*mm/dd/yyyy*
Ratings		
Clarity	Ordinal	$clar \in \mathbb{N} \wedge clar \in [1, 5]$
Easiness	Ordinal	$ease \in \mathbb{N} \wedge ease \in [1, 5]$
Helpfulness	Ordinal	$help \in \mathbb{N} \wedge help \in [1, 5]$
Interest Level[5]	Ordinal	$interest \in \mathbb{N} \wedge interest \in [1, 5]$
Course Related		
Course Code	Nominal	ID Strings
Attendance Required?[5]	Binary	$isAttendance \in \{0, 1\}$
Grade Received[5]	Ordinal	$grade \in \{A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-, F, N/A\}$
Textbook Usage	Ordinal	$textbook \in \mathbb{N} \wedge textbook \in [1, 5]$
Online Class?	Binary	$isOnline \in \{0, 1\}$
Taken for Credit?	Binary	$isCredit \in \{0, 1\}$
Other		
Tags	List	$0 \leq\mid tags \mid\leq 3$
Comment	String	$length_{\max}(comment) = 350$

Table 3.2 describes a typical review. Each professor is associated with reviews from possible students. Overall, we collected a total of 15,497,536 reviews. While submitting a review, each reviewer has to provide details like the course code, whether class was taken for credit, and Likert ratings on characteristics of the professor such as *clarity*, *easiness* and *helpfulness*. Optionally, the reviewer could also select up to three *tags* which best described the professor, report their prior *interestlevel* in taking the course, the *grade* he/she received for the course, his/her *major*, and so on. A mandatory *comment* is required for a review submission, and is limited to 350 characters. The review serves as a subjective assessment of the faculty member, and can include thoughts, criticism, or commendation. RMP follows a strict policy of removing comments that are improper or inappropriate.[6]. Reviews can be in English or French. Reviewers can also mark

[5] Optional, N/A values ignored

[6] http://www.ratemyprofessors.com/TermsOfUse_us.jsp#guidelines

isAttendance (binary), where students can report whether the attendance of the class was mandatory, *isOnline* which specifies if it was an online course, *isCredit* which indicates whether the class was taken for credit and a *date* field which captures the datestamp of the review.

In our dataset, we successfully labelled[7] a subset of 915,334 professors and 15,467,632 reviews as Male or Female. We use this dataset in our experiments.

3.2 Initial Analyses

3.2.1 What are the tags received by male and female professors?

When submitting a review for a professor in RMP, the student is encouraged, but not forced to select three tags out of a list of twenty predefined tags[8] that best defines the professor to the student. Figure 3.1 and Figure 3.2 show the distribution of assigned tags to the male and female professors across all reviews, and the distribution of top three tags assigned to male and female professors. Since tags serve as an aggregate of the perception of the professors in general, we shall consider bars in Figure 3.1 as how students judge and attribute certain characteristics to professors in general. On the other hand, top three tags represent the major opinion that reviewers carry towards professors. So, for Figure 3.2, we shall assume the bars represent characteristics that are inherent to professors.

Observation 1: Students find female professors less respectable.

Male professors are tagged as *Respected by students* on 182,249 reviews, 23.52% more than female instructors. Females also get 44.73% less tags when considering the top tags

[7]Using second person pronouns, discussed further in Appendix C

[8]Y axis on Figure 3.1

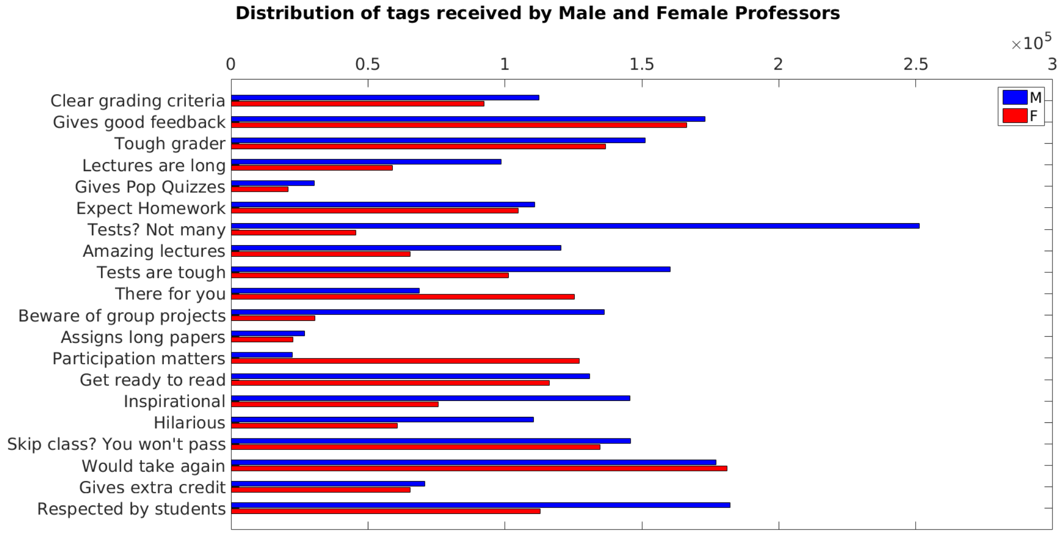


Fig. 3.1: Distribution of all tags received

received. So, not only do more reviewers think that females are not as respected as males, but females also fail to command respect from students when compared to males.

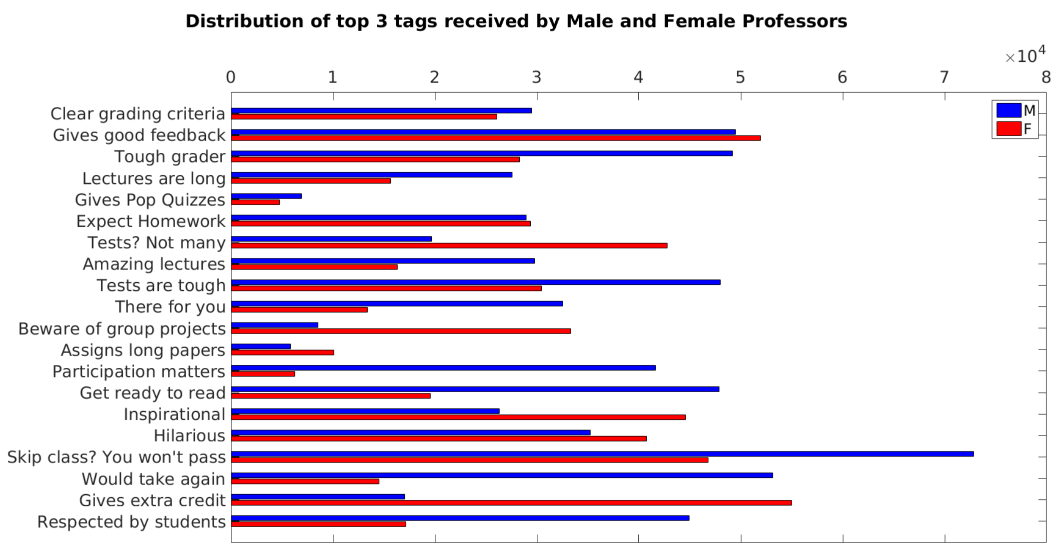


Fig. 3.2: Distribution of top 3 tags received

Observation 2: Male professors give better lectures

As for the tag *Amazing lectures*, male professors perform better than women by 29.78% more reviews. This is also backed up by the fact that 29.28% more male professors received this tag as one of their top three tags.

Observation 3: Female professors are thought to be less inspirational and less hilarious, but that is not necessarily true.

Male professors outnumber female professors by 31.56% and 28.98% reviews in cases of the tags *Inspirational* and *Hilarious*, respectively. However, females do far better than males - in fact 25.78% and 7.25% more females receive the very same tags as their top three tags received.

Observation 4: Female professors are perceived as participation oriented and more supportive, but that is not really the case.

Unsurprisingly, students perceive female professors to assume the role of a 'nurturant' as discussed by Andersen (Andersen and Miller, 1997). This is evident from the high margins of *Participation matters* and *There for you* tags received as compared to males. However, it is the males who do better by 73.97% and 41.76% for the same tags when considering the top three tags received.

Observation 5: Females give more extra credits, but a lot of students fail to notice it.

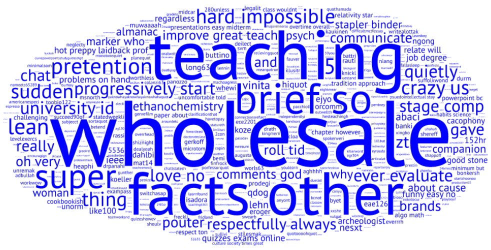
Contrary to one of the claims made by Boring (Boring, 2015), female professors receive the *Gives extra credit* tag as one of their top three 52.73% more times than males, but students fail to recognize that. In fact, there is no significant difference between the two genders when considering the *Gives extra credit* tag over all reviews.

3.2.2 What kind of comments do male and female professors garner?

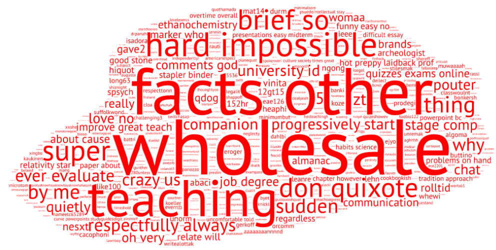
Each review is accompanied with a textual comment from the reviewer, which serves as a subjective platform to express thoughts about the professor being commented on. In order to mine information from about 15.5 million reviews, we first divided these comments into two classes - one for each gender. Then we filtered out commonly occurring English stop words using a popular corpus[9]. We stemmed the remaining words to reduce them to their word stems using Porter Stemmer.[10] To find the important word stems in both classes of comments, we fed these stems into the Gensim TF-IDF model.[11] This process yields the most important stemmed words from each of the classes. Each important word is assigned a weight to signify its importance within the class. To visualize the importance of words among both genders, we fed these word stems, along with their importance weights into a Word Cloud generator.[12] Figure 3.3 represents the words in comments received by male and female professors in blue and red respectively.

From these images, we can clearly see that even though TF-IDF modeling extracts various words with differing weights for male and female comments, many of those words are shared, albeit with different weights. So, to get an idea of what keywords actually represent both genders, we provide another word cloud, that only includes unique words from comments of either gender.

[9]NLTK Stopwords Corpus: www.nltk.org/nltk_data
[10]http://www.nltk.org/howto/stem.html
[11]https://radimrehurek.com/gensim/models/tfidfmodel.html
[12]Tagul: www.tagul.com

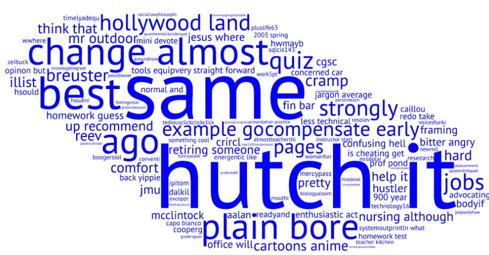


(a) Male comments

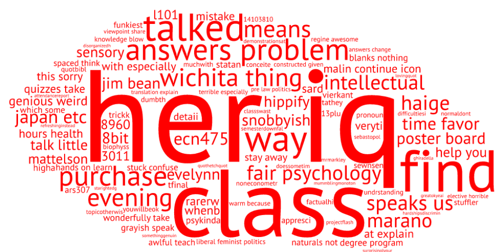


(b) Female comments

Fig. 3.3: Cleaned words in comments



(a) Male comments



(b) Female comments

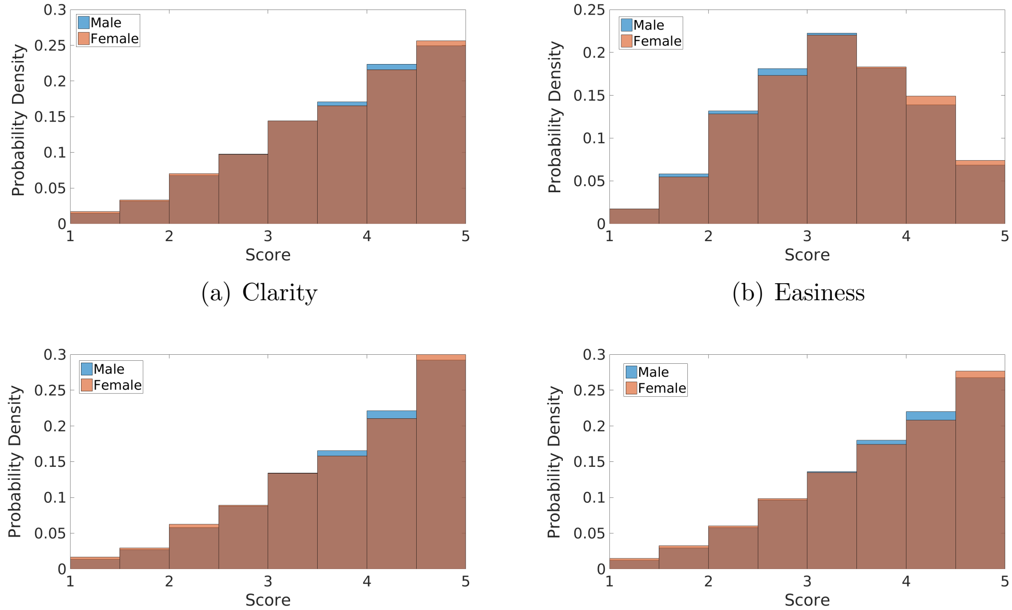
Fig. 3.4: Cleaned unique words by gender in comments

3.2.3 How are professors from both genders rated across all reviews?

With each submitted review, the reviewer has to rate the professor on three abilities, *clarity*, *helpfulness* and *easiness*. For each professor, RMP aggregates[13] these abilities, and reports the average. An additional feature, *overall quality*, is also reported and is calculated internally by the website. However, only *clarity* and *helpfulness* scores are used to calculate *overall quality*. Figure 3.5 shows the histograms of how professors are rated on the above abilities and overall quality. Table 3.3 also provides the probabilities of receiving different range of scores for *clarity*, *helpfulness*, *easiness* and *overall quality*. The table also shows differences in these probabilities observed across genders.[14]

[13] http://www.ratemyprofessors.com/help.jsp#tally

[14] Negative difference signifies that probability density is in favor of men



(a) Clarity (b) Easiness

(c) Helpfulness (d) Overall quality

Fig. 3.5: Histograms of rating measures by gender

Observation 1: Males are clearer and more helpful than women

Except for the range of 4.5 - 5 where females outperform males by slight margins, they consistently receive less ratings in the *good* range of 3 - 4.5 and receive more ratings in the *bad* range of 1 - 3.

Observation 2: Women are rated as easier professors compared to males

The trend seen in previous observation reverses as women get rated more in the range of 3.5 - 5, and men get ratings in the range of 1 - 3.5 more often than women.

Observation 3: Women perform poorly in terms of overall quality

Considering *overall quality* scores by both genders, males receive more ratings in the range of 3 - 4.5 range and women receive more of ratings in the 1 - 3 range. Women, do outdo men in the 4.5 - 5 range, but by a small margin.

Table 3.3: Comparison of male and female professor ratings

Rating	Clarity (%)			Easiness (%)		
	Male	Female	Difference	Male	Female	Difference
1 - 1.5	1.487	1.722	+0.235	1.758	1.736	-0.022
1.5 - 2	3.173	3.322	+0.147	5.824	5.466	-0.358
2 - 2.5	6.761	7.022	+0.261	13.16	12.81	-0.35
2.5 - 3	9.731	9.755	+0.024	18.11	17.33	-0.78
3 - 3.5	14.44	14.44	0	22.26	22.02	-0.24
3.5 - 4	17.1	16.53	-0.57	18.17	18.31	+0.14
4 - 4.5	22.36	21.6	-0.76	13.88	14.92	+1.04
4.5 - 5	24.93	25.63	+0.7	6.838	7.406	+0.568
Rating	Helpfulness (%)			Overall quality (%)		
	Male	Female	Difference	Male	Female	Difference
1 - 1.5	1.663	1.35	-0.313	1.224	1.476	+0.252
1.5 - 2	2.7	2.956	+0.256	2.935	3.259	+0.324
2 - 2.5	5.811	6.241	+0.43	5.801	6.035	+0.234
2.5 - 3	8.811	8.91	+0.099	9.666	9.839	+0.173
3 - 3.5	13.39	13.43	+0.04	13.6	13.48	-0.12
3.5 - 4	16.55	15.82	-0.73	18.02	17.42	-0.6
4 - 4.5	22.14	21.05	-1.09	22.01	20.82	-1.19
4.5 - 5	29.22	29.97	+0.05	26.73	27.67	+0.94

With these results, we may interpret that even though women do carry an advantage of being highly probable of being rated in the topmost rung of 4.5 - 5 for all different rating measures, they also run the risk of being subjected to lower scores of 1 - 2.5 in most of the rating measures. Men on the other hand are safe from the low brackets, and make themselves up in the mid to high ranges of 3 - 4.5 score.

3.2.4 Can gender disparity across varying disciplines be observed from the reviews?

The lack of representation of women as university level instructors has been reported as a cause of rising gender inequity by various reports (Rosser and Taylor, 2009, West and Curtis, 2006). While the reports mention representation gaps in terms of wages, we consider the size of the discipline and *OverallQuality* score obtained by professors depending on their gender as prime metrics.

We define *Disparity* which measures the relative *OverallQuality* score received by a female professor as compared to a male for a specific discipline. We also define *Imbalance*, which compares the representation of women to men for a given discipline.

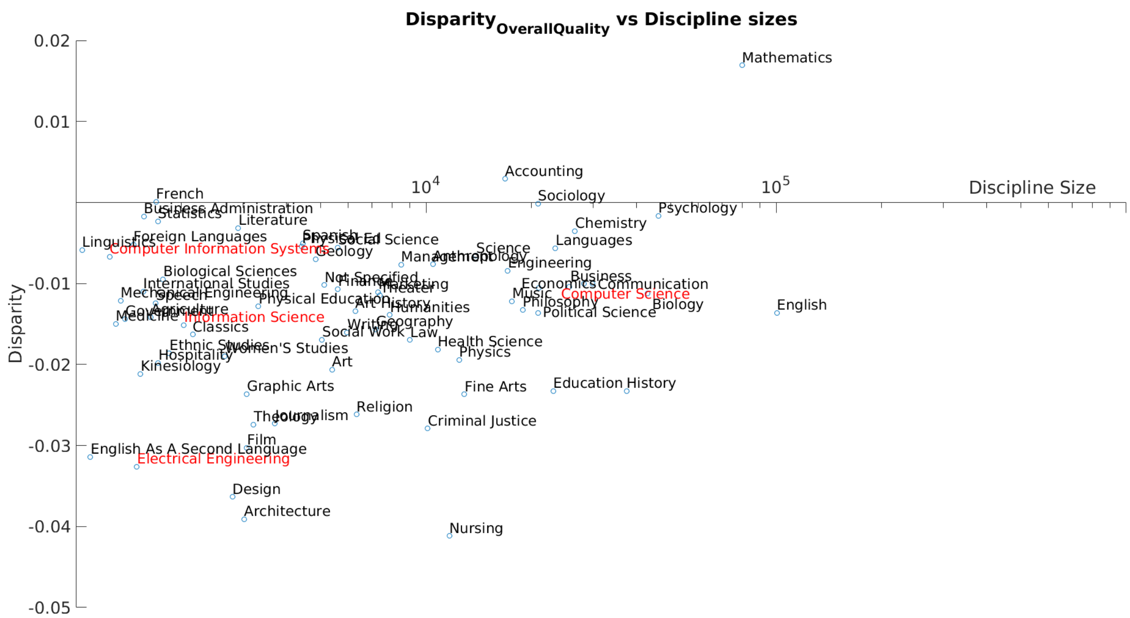


Fig. 3.6: Gender disparity against Discipline Sizes

For each discipline,

$$Disparity = \frac{(\overline{f}-\overline{m})}{(\overline{f}-\overline{m})} \text{ and } Disparity \in [-0.67, 0.67],$$

$$\text{where } \overline{f} = average(OverallQuality_{female}), \overline{m} = average(OverallQuality_{male})$$

$$Imbalance = \frac{|females|}{|males|} \text{ and } Imbalance \in [0, \infty]$$

Figures 3.6 and 3.7 gives the scatter plot of disparity against discipline sizes and against discipline imbalance respectively. Figure 3.8 produces a magnified view of the imbalance

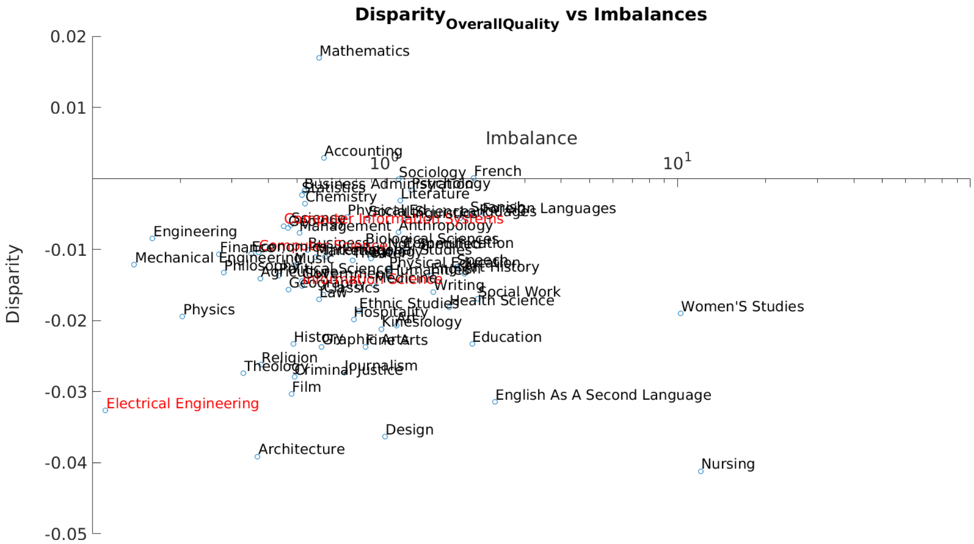


Fig. 3.7: Gender disparity against Discipline Imbalance

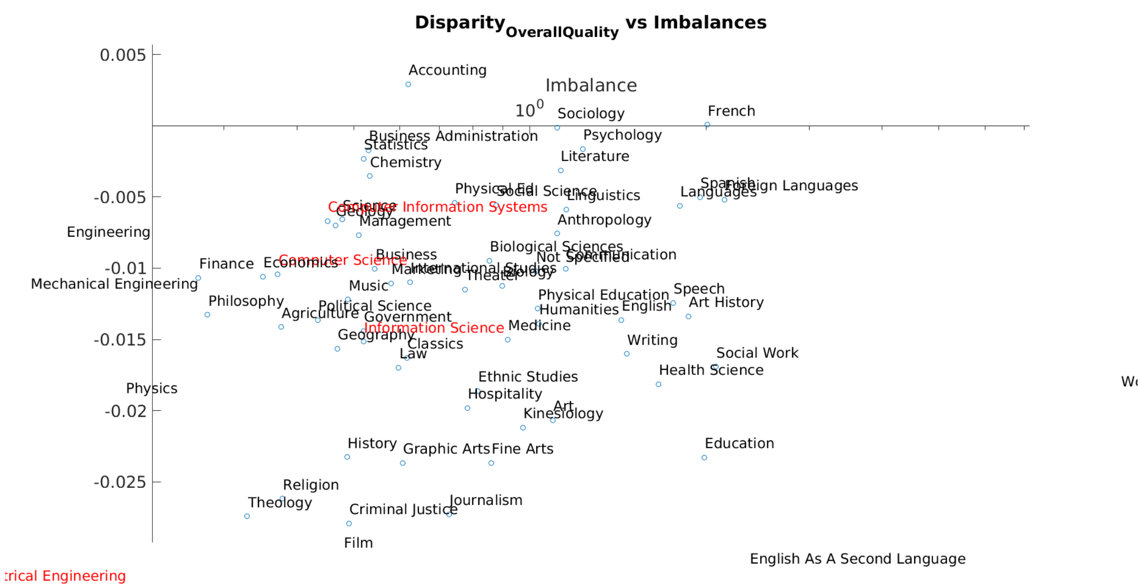


Fig. 3.8: Gender disparity against Discipline Imbalance

scatter plot from the most densely crowded region. Each point on the scatter represents a specific discipline. As we go down on the *Disparity* axis, the relative *OverallQuality* score for women goes down in comparison to men. Moving left on both $DisciplineSize$ and $Imbalance$ axes suggests a decreasing proportion of females for a discipline.

Observation 1: Women face disparate overall ratings for majority of disciplines

Observing Figure 3.6, it is clear that irrespective of the size of a discipline, most of the disciplines fall in the area of $Disparity \leq -0.01$. As a matter of fact, only French, Accounting and Sociology clear the $Disparity > 0$ mark.

Observation 2: Women in Electrical Engineering, Design, Architecture, English as a Second Language and Nursing are worst affected by disparate rating scores

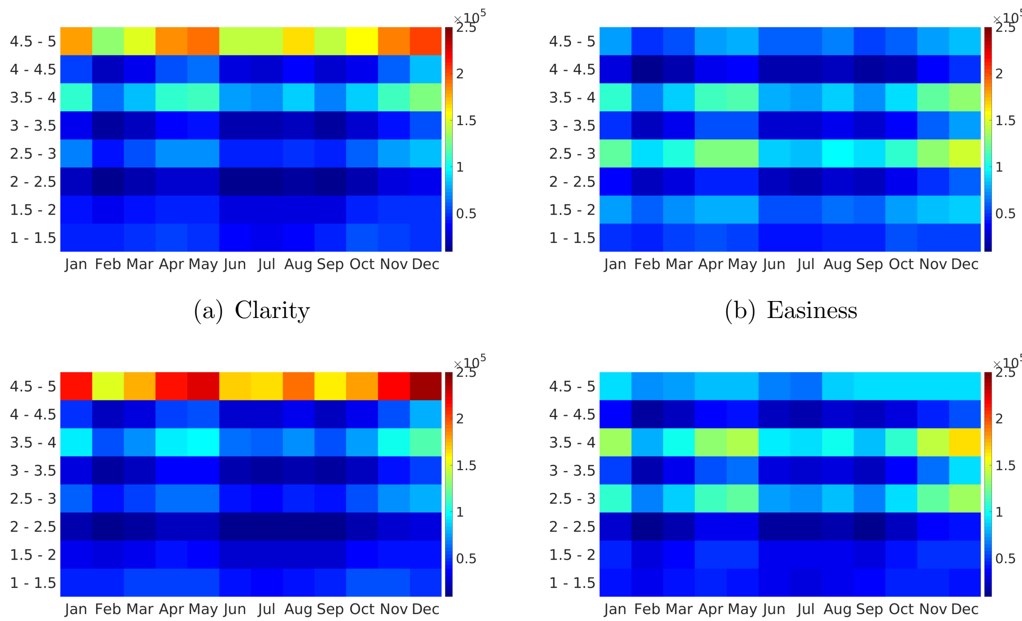
It is clear from Figures 3.6 and 3.7 that the aforementioned disciplines lie in the $-0.04 \leq Disparity \leq -0.03$ area. For Nursing and English as a Second Language, the $Disparity$ comes as a surprise as the $Imbalance$ is high and in favor of women for these disciplines.

Observation 3: Most of the disciplines are balanced in terms of representation of women faculty members

Following the dense region of scatter points around the $Imbalance = 1$ line in Figure 3.8, it can be said that most disciplines have achieved a balance in the number of male to female instructors. It is also worthy to note that the disciplines that achieve the highest parity in scores lie close to this line, thus signifying the importance of maintaining this balance to achieve equality.

3.2.5 Does time of the year affect how professors are rated?

In most universities across the US and other countries, the semester examinations are taken during similar time frames. According to Stark and Freishtat (2014), anger and satisfaction related to coursework often play a pivotal role in evaluating a pedagogue. To measure this effect, we look at how the rating trends change by the months. Figures 3.9 and 3.10 show heat maps that indicate density of professors with average ratings from 1 to 5 (divided into equally sized bins of width 0.5 rating points on the Y axes) against months of the year (X axes).



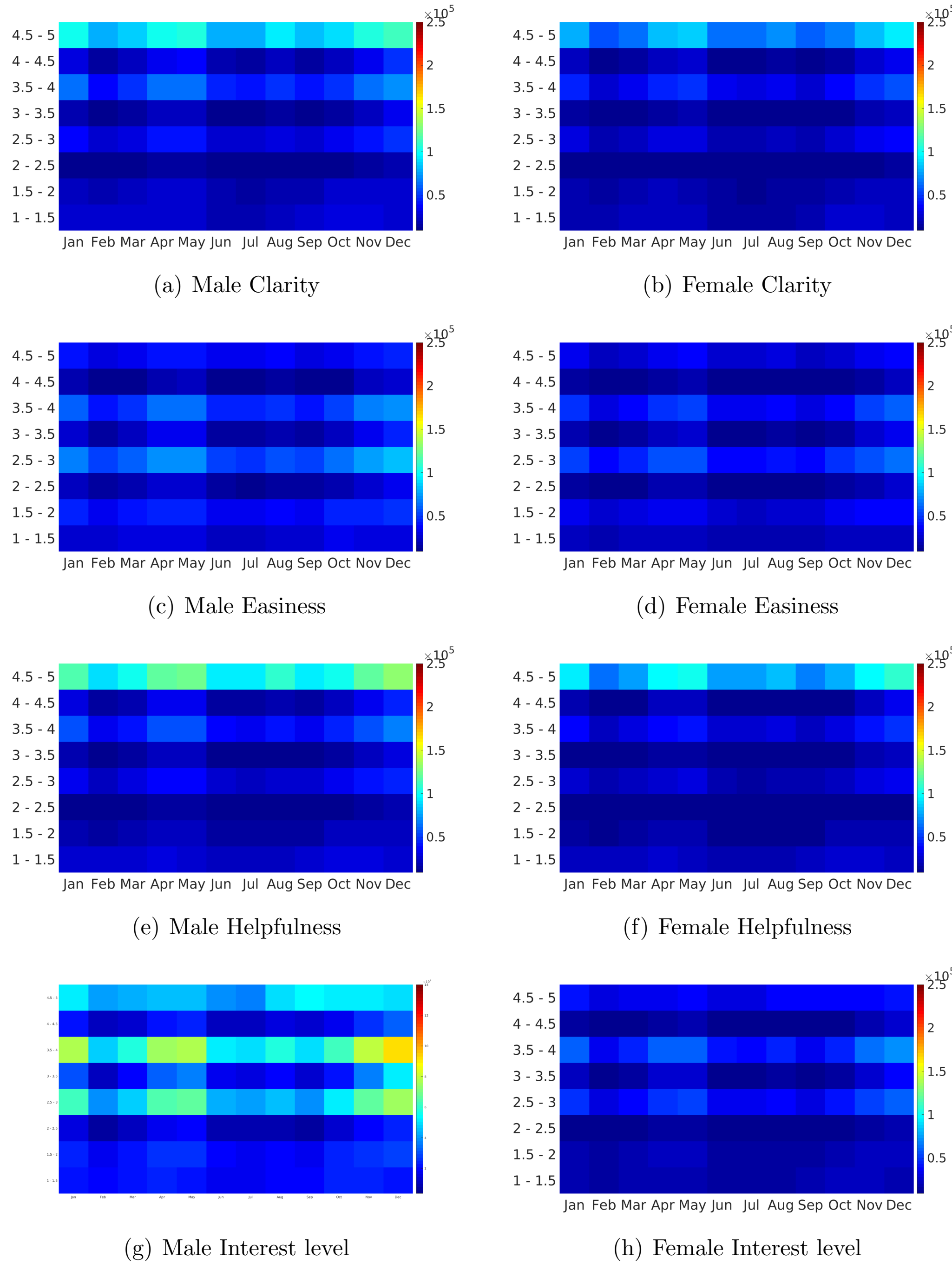
(a) Clarity

(b) Easiness

(c) Helpfulness

(d) Interest level

Fig. 3.9: Heatmaps measuring average rating density by month for all professors



(a) Male Clarity

(b) Female Clarity

(c) Male Easiness

(d) Female Easiness

(e) Male Helpfulness

(f) Female Helpfulness

(g) Male Interest level

(h) Female Interest level

Fig. 3.10: Heatmaps measuring average rating density by month for both genders

Observation 1: Most reviewers rate professors generously for clarity and helpfulness.

Clearly from 3.9(a) and 3.9(c), the bin with ratings in [4.5, 5] has the most number of reviews, though the number of reviews peak during the April - May and November -

January frames. However, such is not the case in 3.9(b) and 3.9(d), where the most reviews receive rating in the bins of [2.5, 3) and [3.5, 4), even though the number of reviews peak for the same time frame as before.

Observation 2: Generosity in clarity and helpfulness do not change by the gender of the professor

The bands of [4.5,5] in Figures 3.10(a), 3.10(b), 3.10(e) and 3.10(f) continue to receive most heat over all the months, and the loss of generosity is observed for *helpfulness* and *interest level* as the bands of [2.5, 3) and [3.5, 4) continue receiving more heat than other bands.

Observation 3: Male professors generate more interest than females

Observing Figures 3.10(g) and 3.10(h), it is evident that over all months, males receive more reviews, and proportionally receive more reviews in the band of [4.5, 5) as compared to women.

Observation 4: April-May and November-January receive most of the reviews

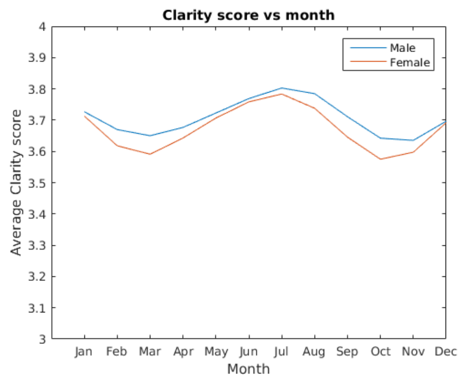
Figure 3.11 shows us number of comments and average ratings and interest levels by month. The number of reviews received peak two times during the year, once during April and May, and the other during December. Incidentally, for most of the universities in the US and other countries, these months coincide with the end of the semester, or right after the semester examinations are conducted. Because, the male and female lines on Figure 3.11(e) follow a similar trend, no evidence of a singular gender being more targeted by reviewers was found. However, all measures of the professor's abilities as well as reported average interest level of the reviewer in attending the professor's classes show different trends. Surprisingly, these trends have two troughs, which almost coincide with the peaks in number of reviews submitted. As a matter of fact, the months of July and August have the highest averages, but lowest number of reviews. This observation supports the claim made by Stark and Freishtat (2014) that indeed, people who are highly satisfied with a professor do not show much interest in writing reviews and it's the least satisfied who are most motivated to provide feedback in the form of a review.

From these observations, it can be concluded that a surge in the number of reviews is generally accompanied with dissatisfaction on the part of the reviewer, and results in the female professors receiving the lower scoring reviews.

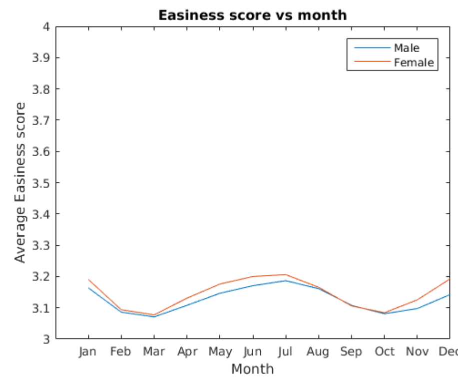
To further investigate the professors' performance, we look at their best and worst months in terms of ratings received in Figure 3.12. The best or worst month for a professor is calculated by taking the average of rating (for the above four measures) received over all months of the year, and then picking out the months with highest and lowest averages. The months with no review received is ignored from the above calculation. The red and blue lines signify the density of a male and female professor of having a particular month as their best or worst month. For a particular month, this density is calculated by taking the number of professors with their best/worst month in that month over the total number of reviews received during that month.

3.2.6 How does grades received by reviewers affect professors of both genders?

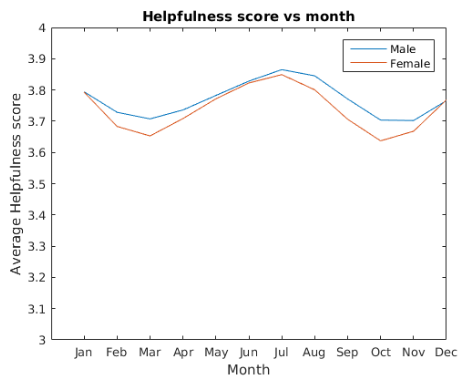
RMP allows reviewer to optionally mention the grade received in the class that they are reviewing for. To understand the effects of grade expectations on the ratings received, we repeated the experiment of investigating average scores in Question 5, but only with reviews that mentioned the grades that they received. To reduce the sparseness of data



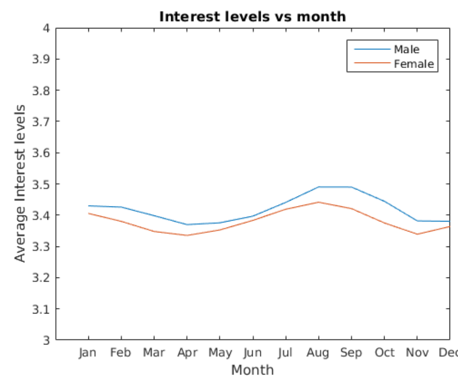
(a) Average Clarity Score



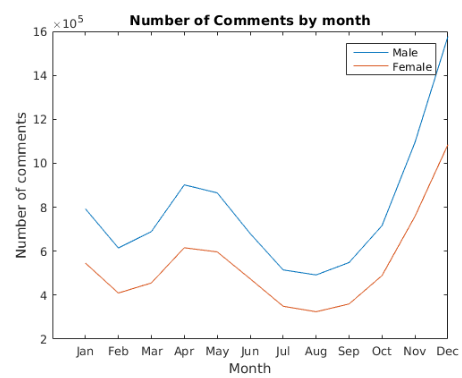
(b) Average Easiness Score



(c) Average Helpfulness Score



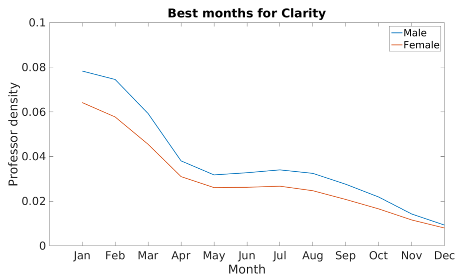
(d) Average Interest Level



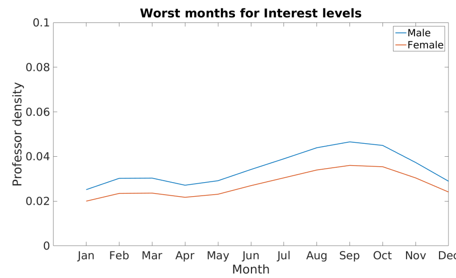
(e) Number of reviews

Fig. 3.11: Review details by months

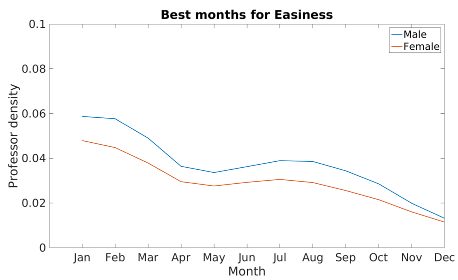
over a number of grades, we identified *A+*, *A*, *A-* and *B+* as good grades; and *C*, *C-*, *D+*, *D*, *D-*, and *F* as bad grades. Figures 3.13 and 3.14 show the distribution of averages



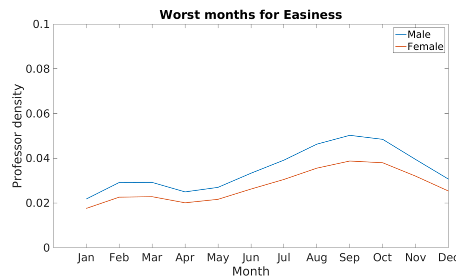
(a) Clarity - Best months



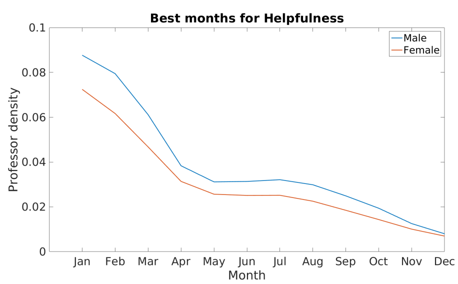
(b) Clarity - Worst months



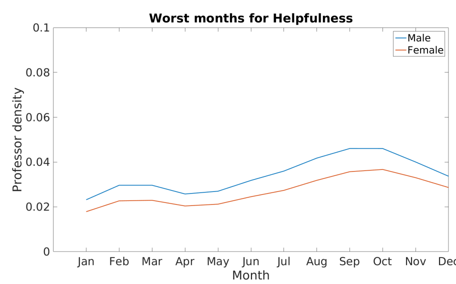
(c) Easiness - Best months



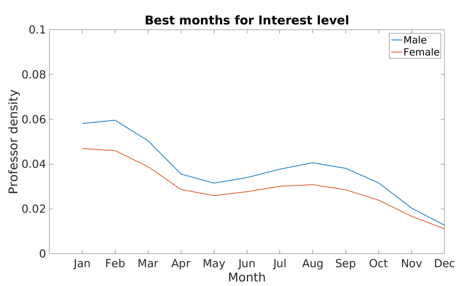
(d) Easiness - Worst months



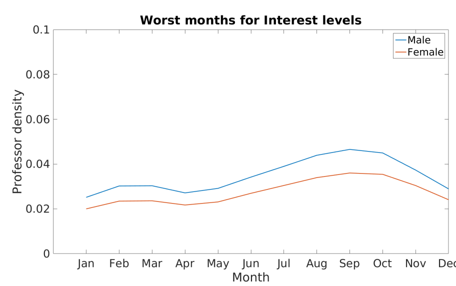
(e) Helpfulness - Best months



(f) Helpfulness - Worst months



(g) Interest Levels - Best months



(h) Interest Levels - Worst months

Fig. 3.12: Best and worst months for professors

for good and bad reviews respectively.

(a) Average Clarity Score (b) Average Easiness Score

(c) Average Helpfulness Score (d) Average Interest Level

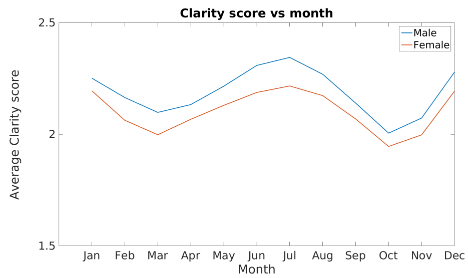
Fig. 3.13: Review details for reviews with good grades

Observation 1: Ratings, on the whole are indeed affected by the grades received by the students.

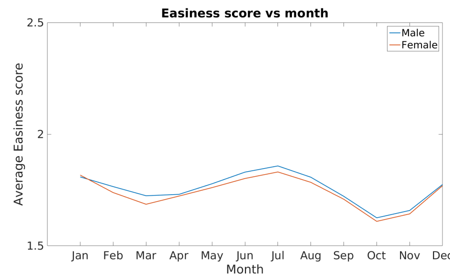
This is evident from the fact that the average of all scores range from [3.5, 4.5] in the case of reviews with good grades; and goes down to the range of [1.5, 2.6] for the ones with bad grades.

Observation 2: Reviewers who received good grades, found classes less interesting when taught by female professors.

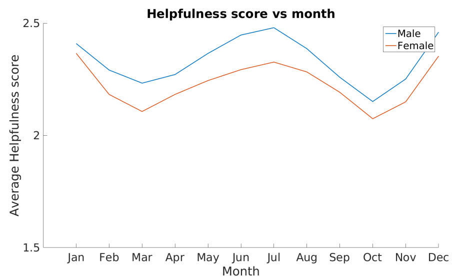
While they do not discern by gender in terms of *clarity*, *easiness* and *helpfulness* as the average line intertwine over the months, almost in every month, they show a visible difference in *interest level* in favor of male professors.



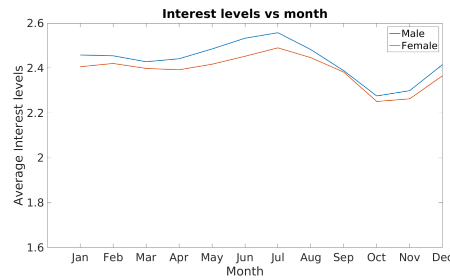
(a) Average Clarity Score



(b) Average Easiness Score



(c) Average Helpfulness Score



(d) Average Interest Level

Fig. 3.14: Review details for reviews with bad grades

Observation 3: Women are rated as unclear and unhelpful when the reviewer received bad grades

Despite both genders being rated poorly by reviewers who received bad grades, women professors find themselves receiving poorer ratings as compared to men, especially on measures of *clarity* and *helpfulness*. However, both genders are rated similarly easy (or not so easy), with the male and female average lines almost coinciding with each other.

Chapter 4

Investigating Gender Bias

By definition, gender bias is unequal treatment in opportunity and expectations due to attitudes based on the gender of a person, or a group of people. Providing opportunities and keeping expectations lead to discrimination. Gender roles arise from societal expectations, and it follows with gender based discrimination in the way opportunities are extended to both genders. In this chapter, we follow the various manners in which gender bias can be detected from the student reviews of professors in RMP. In order to follow how expectations and discrimination together lead to inequality based on the gender of professor, we first formally describe the formulation of gender bias. Then, give a detailed information on how we construct the dataset from the available data. We move on to detailing the setup of our experimental trials, discuss about avoiding learning bias, and selecting the best candidate for testing our hypothesis on gender bias. Finally, we provide a detailed analysis of the features that are most related to the hypothesis.

4.1 Formulating bias

If there was no bias, there would be no way in which the gender of a professor X could be guessed from the review that he/she receives.

Conversely, *If the gender of the professor(s) could be predicted with probability of more than random, then it would entail that a bias exists against one or the other gender.*

To test the above hypothesis, we classify the professors by their genders with the help of information present in their reviews. Before proceeding to the task of classification, we explore the feature set (See Table 4.1 for all features) that we use for predicting the target gender in section 4.2.

4.2 Dataset Construction

RMP provides a list of attributes that the student is expected to provide while rating a professor. Some of these are mandatory, however others are not. We now look into these attributes from a list of reviews for each professor, and further construct more related features.

4.2.1 Direct features

Features # 1 - 7 in Table 4.1 are borrowed directly from the RMP profile of a professor, and includes the aggregate rating scores received, the average interest level of reviewers in the courses offered by the professor, the professor's teaching discipline and the state where the professor's associated university is situated in. These features help us understand if gender bias can be observed from the public profile of a professor.

4.2.2 Sentiment features

We examine the phrase-wise sentiment composition to measure how changes in sentiment affect the process of gender prediction among professors (Kiritchenko and Mohammad, 2016). VADER (?), a sentiment analysis tool, calculates the weights of positive, neutral and negative sentiment present in a word. The highest sentiment weight in that word is considered as the dominant polarity (or simply as the polarity) of the word. For a professor, we look at all the comments, and construct word bigrams of the stopword-free

Table 4.1: List of features

Feature Name	Feature #	Feature Name	Feature #
Gender	Class	***Part of speech tags***	
Direct		Adverbs	31
Clarity	1	Nouns	32
Easiness	2	Adpositions	33
Helpfulness	3	Particles	34
Interest Level	4	Determinants	35
Overall Quality	5	Pronouns	36
State	6	Verbs	37
Discipline	7	Numbers	38
Sentiment		Conjunctions	39
(+, +)	8	Adjectives	40
(+, 0)	9	Punctuations	41
(+, -)	10	Others	42
(0, +)	11	***Word features***	
(0, 0)	12	Unique Words Count	43
(0, -)	13	Positive Words Count	44
(-, +)	14	Negative Words Count	45
(-, 0)	15	Positive Comments Count	46
(-, -)	16	Negative Comments Count	47
Crowd Bias		Ratio (44 : 45)	48
Clarity	17	Ratio (46 : 47)	49
Easiness	18	***Consistency***	
Helpfulness	19	Clarity	50
Positivity	20	Easiness	51
Negativity	21	Helpfulness	52
Interest Level	22	Interest Level	53
Text features			
Comments	23	***Word Vectors***	
Commas	24		
Quotes	25	Note: All word vectors follow after	
Semicolons	26	Feature # 53. Two length of word	
Colons	27	vectors have been used, which include	
Pauses [‘,’, ‘;’, ‘:’]	28	106 of the top important words, and	
Question Marks	29	205 of the top important words	
Exclamation Marks	30		

text. Then, we process these bigrams with VADER, and count the number of times the bigrams have opposing or matching polarities. Features # 8 - 16 measure the number of times bigrams with varied combination of polarities occur in the review comments, and lets us analyze its relation to gender bias.

4.2.3 Crowd bias

It has been previously found that 'Wisdom of Crowd' biases judgement of a reviewer based on previous users' ratings (Surowiecki, 2005). To test the effects of crowd bias, and how they subsequently contribute to gender bias, we construct Features # 17 to 22. To measure these factors, we sort the reviews of a professor in chronological order. Then, we calculate the Pearson's correlation coefficient of the n^{th} rating score against the average of $n-1$ scores. This is repeated for *clarity*, *easiness*, *helpfulness* and *interest level* scores received. The calculated coefficient is now the crowd bias measure of ratings. We also include the crowd bias measure in the sentiment of the comments received.

4.2.4 Text features

Linguistic evidence from text have been found to help in the prediction of gender (Kucukyilmaz et al., 2006). We adapt some of the text-based classification techniques as Features # 23 to 30, where we count the number of comments and the occurrence of punctuations like commas, semicolons, colons, question and exclamation marks in all comments.

4.2.5 Part of Speech

Writing styles, such as parts of speech (hereby referred to as POS) have been found to have significant effects on gender classification (Mukherjee and Liu, 2010). While most of previous work use POS to classify the author of text, we would like to analyze the effect of the same on the target of a review, in this case, a professor. We use the Stanford POS

Tagger, which generates the POS tags for every word in all the comments attributed to a given professor. Once generated, we count the 12 distinct tags from the Universal Part-of-Speech Tagset[1] in Features # 31 to 42, and examine the effect on gender prediction and bias.

4.2.6 Word features

To further understand the effect of text-based and sentiment-based features, we include Features # 43 to 49. While number of unique words (43) gives an idea about how the number of different words used helps in describing gender, frequency of positive and negative words/comments (44 - 47) gives some insight into the effect of sentiment on gender. The ratio of positive to negative words (48) and comments (49) determine if a gender is associated with relatively positive or negative sentiments over all comments in reviews.

4.2.7 Consistency

In case of ordinal rating scores, it has been argued in the past that scatter of scores determine the overall teaching effectiveness (Stark and Freishtat, 2014). In Features # 50 to 53, include the standard deviation of *clarity*, *easiness*, *helpfulness* and *interest level* rating scores to measure the effect of scatter on bias.

4.2.8 Word vectors

A recent project found out that word patterns can detail the gender of a reviewee (Schmidt, 2015). To investigate if specific words in reviews are related to gender bi-

[1] http://www.nltk.org/book/ch05.html

ases, we use Gensim TF-IDF model[2] to find out the 100 and 200 most important words that are present in comments involving male and female professors. From these sets, we remove the commonly occurring words, and this yields a list of 106 and 205 words respectively. We perform a word vectorization based on these words for all comments received by each professor, and include them as predictors in gender classification.

4.3 Experimental Setup

In the previous section, we discuss how the various features in the feature set were constructed, and a few of them were adapted from the profiles on RMP. In this portion, we experiment with different combinations of features, recorded as trials in Table 4.2. These combination of features may conform to one or a combination of many feature types discussed in the previous section.

Table 4.2: List of experimental trials

Trial #	**Features Used**	**Trial #**	**Features Used**
1	1 - 3	13	8 - 16, 44 - 47
2	1 - 3, 5	14	31 - 42
3	1 - 5	15	23 - 42
4	1 - 5, 7	16	23 - 37, 39 - 40
5	1 - 7	17	1 - 7, 23 - 47
6	17 - 19, 22	18	1 - 16, 23 - 47
7	20 - 21	19	1 - 16, 23 - 47, 106 word vectors
8	8 - 16, 20 - 21	20	1 - 16, 23 - 47, 205 word vectors
9	8, 10, 14, 16	21	1 - 16, 23 - 53, 205 word vectors
10	8, 10, 14, 16, 20 - 21	22	1 - 16, 23 - 49, 205 word vectors
11	23, 43 - 47	23	1 - 53, 205 word vectors
12	1 - 3, 23, 29 - 37, 39 - 40, 43		

- Trials 1 to 5 considers the general features only, and can help observe gender bias from the public information available from RMP.

[2] https://radimrehurek.com/gensim/models/tfidfmodel.html

- Trials 6 and 7 gives an idea of the effect of crowd bias on gender bias.
- Trials 8, 9 and 10 considers the effects of sentiment on both polar bigrams and crowd bias.
- Trial 11 uses text based features only. Trial 12 uses certain text-based and POS features, while trial 14 uses only POS tags.
- Trial 13 uses both polar bigram features, and some text based features.
- Trials 15 and 16 uses both text based and POS tag features.
- Trial 17 uses general, text and POS tags as features. Trial 18 adds polar bigram features to trial 17.
- Trial 19 adds a word vector length of 106 to trial 18, and trial 20 does the same but for a word vector length of 205.
- Trial 21 includes the standard deviation and some text based features (relative sentiment ratios only) to map the behavior due to scatter.
- Trial 22 uses all but the scatter features of trial 21.
- Trial 23 uses the entire dataset with word vectors of length 205.

We have deliberately considered mixing and matching different types of features, in order to understand and evaluate which features have the most impact on gender prediction. In some cases, a few redundant or irrelevant features (like Numbers, Punctuations and Other POS tags, and bigrams associated with neutral sentiment) were wilfully ignored.

4.4 Selection of Learning Model

To test our hypothesis, we start off by picking a learning model to perform classification by gender. We train a Naïve Bayesian Classifier over the trial runs discussed in the previous section and in Table 4.2. We report the respective accuracy and AUC scores as a percentage in Table 4.3.

Table 4.3: Results of Naïve Bayesian Classification

Naive Bayesian Classification					
Trial #	Accuracy	AUC	Trial #	Accuracy	AUC
1	58.1043	51.7	12	53.0792	54.6
2	57.9301	51.6	13	52.6141	53.1
3	57.7946	52.2	14	52.2377	53.6
4	62.8469	65.7	15	51.7506	54.6
5	62.9217	65.8	16	52.257	54.8
6	53.4291	51.0	17	56.8469	60.8
7	55.8588	50.3	18	54.9047	59.9
8	52.2125	51.9	19	56.1615	65.2
9	58.1224	52.7	20	57.3234	67.6
10	55.7643	51.4	21	57.3963	67.5
11	55.451	53.0	22	57.241	67.5

Although the AUC scores are higher than random, we wished to ensure that the classification was not being affected by learning bias. To ensure that our model is free of learning bias, we proceed to training a few other classification models.

Putting this theory to test, we evaluate our hypothesis using multiple learning algorithms, including a linear decision boundary based J.48 tree, and non linear classification algorithms like l_1-regularized l_2-loss SVM and l_1-regularized logistic regression. Given the high dimensionality of our dataset with a large number of features, it can be assumed that some features may not be as important as the others in learning a suitable model. Presence of word vectors and other features which are optionally responded to by the reviewers ensure that out dataset is sparsified. Hence, it is necessary to select a learning algorithm which is well suited to both handle sparse data, and provide weightage to features which clearly indicate feature importance. Considering these factors, we think l_1-regularized logistic regression would be a good candidate. The results of applying our choice of learners to different trials are reported in Table 4.4.

Clearly, l_1-regularized logistic regression (AUC: 82.1441%) and l_1-regularized l_2-loss SVM (AUC: 82.5141) outperform the the previously tested Naïve Bayesian Classifier (AUC: 67.5%) in Section 4.1 as well. This validates our speculation that a non-linear algorithm is better suited to handle the target dataset. Even among the two non-linear learners, l_1-regularized logistic regression performs better over most of the experiment trials. Naturally, we select this as the best candidate for performing further analysis of gender bias.

4.5 Analyzing Bias

The results of previous section gives us a clear indication about the performance of our candidate model. The high accuracy and AUC also validates our hypothesis that a bias related to gender undoubtedly exists. However, to obtain further insight into what specific factors contribute the most to this bias, we include all possible features of the dataset to train our candidate, and rank the features on the basis of their predictive power. This form of analysis allows us to confirm the most important factors that helps in prediction of gender, and consequently contribute to gender bias.

Quite expectedly, the model returns with an accuracy of 74.9768% and an AUC score of 82.1819%, which are even better to the best performance of the same model in Trial# 22 (See Table 4.4). LibSVM logistic regression allows us to access the weights that were assigned to features depending on their power of prediction. While all features combined do yield a better performance from the classifier, it is obvious that there are going to be

[3] Insufficient memory

Table 4.4: Results of experimental trials

Trial #	J48 Tree		l1 regularized l2 loss SVM		l1 regularized logistic regression	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
1	58.1264	50.0	58.1235	51.0924	58.1249	51.022
2	58.1264	50.0	58.1274	51.1199	58.1239	51.0422
3	58.1264	50.0	58.1164	52.4126	58.1218	52.363
4	62.8594	65.5	58.1247	53.1156	58.1228	53.0993
5	62.8602	65.5	58.1208	53.2223	58.1237	53.1908
6	-[3]	-	54.5787	49.6222	58.117	52.3618
7	-	-	53.9004	49.1787	53.8935	49.1782
8	-	-	58.2426	57.4292	57.9544	53.8952
9	62.8602	65.5	57.6824	53.0707	57.7788	53.0431
10	-	-	57.5265	52.8866	57.5589	52.9118
11	58.1264	50.0	58.1469	55.8746	58.1758	55.8091
12	58.1264	50.0	61.7238	64.3347	61.7068	62.3015
13	-	-	59.8682	60.5729	59.2511	59.1736
14	-	-	61.3449	62.6072	61.4744	62.5586
15	-	-	62.1214	63.8882	61.6601	63.9978
16	-	-	61.9736	63.8882	61.9536	63.8853
17	-	-	63.1187	67.7264	63.063	66.2326
18	-	-	63.8776	67.7264	63.6178	67.2293
19	-	-	75.1344	79.3292	72.3891	79.0202
20	-	-	75.1168	82.5049	78.8846	82.1361
21	-	-	75.1185	82.5365	74.9592	82.1753
22	-	-	75,1081	82.5141	74.9575	82.1441

a subset of features that may be more important than the others. Consequently, these features would also be the factors that contribute the most to gender bias. We use the odds-ratio for the analysis of feature ranks.

Because of the implicit feature selection capability of l_1-regularization, a few features are assigned a negligible weight, while the others are given more precedence. For a binary class problem such as ours, the model outputs the negative weights to signify the features which are important for predicting the other class. We report the positive and negative weights as Male and Female specific features. The top 20 features that are most important for predicting the gender of a professor are listed in Table 4.5.

Table 4.5: Top 20 ranked features(all features)

Rank	Feature/Feature #	Feature type	Weight	Gender preference
1	'sweet'	Word vector	0.7524	Female
2	'person'	Word vector	0.3642	Female
3	'wonderful'	Word vector	0.3615	Female
4	'cool'	Word vector	0.283	Male
5	'funny'	Word vector	0.2475	Male
6	52	Standard deviation	0.2149	Female
7	'loved'	Word vector	0.2072	Female
8	'love'	Word vector	0.1707	Female
9	'amazing'	Word vector	0.1538	Female
10	53	Standard deviation	0.1444	Male
11	3	Direct	0.1207	Male
12	'fun'	Word vector	0.1086	Female
13	'helpful'	Word vector	0.098	Female
14	'ever'	Word vector	0.0924	Male
15	'boring'	Word vector	0.0911	Male
16	4	Direct	0.0892	Male
17	'sense'	Word vector	0.0872	Male
18	'great'	Word vector	0.0868	Male
19	1	Direct	0.0868	Female
20	'worst'	Word vector	0.0808	Female

Among features adapted or derived from RMP rating scores, *helpfulness* and *interest level* have the best predicting power, and their standard deviations being ranked higher than the average scores. Most of these features (3 out of 4) among the top 20 features, the rating scores favor men. Interestingly, *interest level* scores are also the ones which are excluded from the calculation of *overall quality* of a professor.

It is very clear from the above Table 4.5 that while all features contribute in varying extent to detecting gender bias, it is indeed the word vectors that play the most important role. In fact, 8 of the top 10 and 14 of the top 20 features are word vectors that have helped in prediction of gender. Words such as 'sweet', 'wonderful', 'loved' and 'amazing' are preferred in the predicting female professors, males are described better with words like 'cool', 'funny' and 'boring'.

Judging by their importance, we further perform a classification using just the word vectors as features. The classifier returned an accuracy of 70.9371% and an AUC score of 77.005%. The top 20 weighted vectors are reported in Table 4.6.

Table 4.6: Top 20 ranked word vectors

Rank	Feature/ Feature #	Weight	Gender preference
1	'sweet'	0.738	Female
2	'wonderful'	0.3488	Female
3	'cool'	0.2973	Male
4	'funny'	0.2447	Male
5	'loved'	0.2107	Female
6	'love'	0.1459	Female
7	'amazing'	0.1328	Female
8	'sense'	0.1232	Male
9	'fun'	0.1064	Female
10	'great'	0.1026	Male
11	'helpful'	0.0956	Female
12	'boring'	0.0920	Male
13	'real'	0.0918	Male
14	'knowledgable'	0.0911	Male
15	'problems'	0.0820	Male
16	'loves'	0.0812	Male
17	'tough'	0.0765	Female
18	'knows'	0.0751	Male
19	'best'	0.0713	Male
20	'interested'	0.0682	Male

While it is interesting to note that the performance of the classifier falls a little in comparison to all features. However, the performance remains better than the best trial performance without using word vectors. This is a significant result as this entails the necessity of considering the text comments in student reviews, and validates a suggestion from an earlier study involving teaching evaluations (Stark and Freishtat, 2014). Thus, in the case of computational constraints, the classifier can be expected to perform satisfactorily in predicting gender, provided the most important features (word vectors and direct features like *helpfulness* and *interest level*) are used.

Chapter 5

Conclusions and Future Work

The obscurity of underlying causes that gives rise to gender inequality often veil implicating evidence of gender bias, and this is true irrespective of whether in physical life or in the virtual online presence. Therefore, it is often hard to instantly notice the bias with enough clarity. However, once the root elements that imply the divide are better understood, it becomes easier to notice the previously unobvious.

We faced a similar problem in our pursuit of deciphering the inequality. To revisit our course of action, we took a systematic path that included, but were not limited to:

1. Understanding gender inequality and it's various related forms. We followed up and expanded on Boring's work on gender bias in student evaluations (Boring, 2015). Her work explains the different ways in which women instructors face disparity in terms of rating scores as compared to men. She also explains the bias that is apparent from not just the gender of the instructor, rather the gener of the student reviewer adds in to the bias, where male and female students tend to rate male instructors higher than women. When considering the case of a triad of teaching instructors, the combination that worked the best in terms of gaining student satisfaction consisted of one male and two female instructors. But even for this case, the male outperformed the female instructors throughout all rating measures.

2. In order to understand the various ways in which gender inequality has affected the society at large, we followed up on published materials and internet media, and

identified a hundred questions. These questions can be followed up with appropriate data from online sources. Along the way, we recognized student reviews to be affected by the stigma of gender inequality, and focused our research goals towards the same.

3. Collecting information-rich data about professors at the university level and their student reviews. Further wrangling allowed some preliminary analysis, and led to interesting facts that detailed the areas where gender inequality could be identified and observed. Based on the perceptions of the student reviewers, we found that students do not find women to be as respectable as men. They also feel that male instructors are better at being inspirational and hilarious, while knowledge about the professors' inherent quality proved this perception wrong. Looking at the rating scores, we find that male professors do better than females better on *clarity*, *helpfulness* and *easiness*, and also in the *overall scores* they have achieved. Considering disciplines of study, it was found that most of the disciplines show disparity in the average overall scores of female professors as compared to males.

4. Testing the hypothesis that the dataset is biased by gender, and validating the hypothesis with a l_1-regularized logistic regression learning model. The results showed a significant evidence of gender bias. When considering only word vectors, the model was learned with a good enough AUC, further realizing the importance of the words, and more importantly, the text content in a review, in detecting gender bias.

5. Ranking features depending on their predictive power. This assured that a reduced dataset with select few features can be used to achieve similar model performance.

While most of the closely related work look at primary Likert scale responses from students collected via university wide surveys to perform their analysis of gender inequality

(Andersen and Miller, 1997, Boring, 2015, Stark and Freishtat, 2014), our approach assumes a more holistic view, encompassing subjective responses in text comments which allows scope for sentiment and text mining. Our framework distinguishes itself by the following ways:

1. The scale of data used is larger than any previously known study in this area. Also, the data is collected from online sources, where student voluntarily submit reviews, as opposed to surveys where anonymousness is a key issue for various reasons. Moreover, the online platform allows us a wider reach in terms of geographical locations, universities and disciplines.

2. Not only do we consider numeric features, we also take into account linguistic features like parts of speech, sentiment measures, text features, and crowd bias features.

3. Word vectors, a key feature type ensure that we scrutinize the comments for important words pertaining to each of both genders. This also opens up the avenues for further lexical analyses.

To our knowledge, this is one of the first attempts at using data science algorithms to solve the global problem of gender inequality. While our approach may have been straightforward, variations of the same could be evaluated to understanding and assessing the problem, the rich nature of data collected and features such as *grade*, *isAttendance*, *tags* can play an important role in better understanding bias. It shall also be interesting to see how the reviews vary depending on whether the institution/university is a public or private. While the states of America do not show much difference of rating the professors based on their gender, looking at conservativeness of states depending on the geographical divisions o USA could open up newer results. A time series analysis of reviews based on their datestamps may uncover interesting results. Further to our good performance using text features, other forms of natural language processing algorithms may evaluate the linguistic features alongside the text-based ones.

Due to the anonymized reviews, it is difficult to evaluate the gender bias that arises from the reviewer being of same or opposite gender to the reviewee. It will be interesting to follow this work up with a method that allows labelling the reviewer's gender, and then tallying the results.

Appendices

Appendix A

Gender Inequality - 100 Questions

A.1 Reaction

1. How does Twitter feel about the Academy Awards Best Actor versus the Best Actress? (O'Neil, 2015)

2. What kind of comments do male and female 'YouTubers' garner? (Luscombe, 2014a,c)

3. How does Facebook/Twitter react to breaking news reports when reported by a male journalist as compared to a female journalist? (Sonderman, 2011)

4. What do YouTube users think of similarly popular male and female artists? Examples below

   Eminem (Not Afraid - Views: 710,276,180, Comments: 860,756) (Marshall, 2010)

   Lady Gaga (Bad Romance - Views: 649,444,263, Comments: 877,813) (Gaga, 2009)

5. What were reactions to the death of popular male and female celebrities by drug overdoses, when other demographic features are controlled for? (mercuryblues, 2014)

6. Are the reactions to appointment of CEOs gender biased? (Lee, 2003, Oyotode et al., 2015)

7. How does the society react to single fathers as compared to single mothers? (Kate, 2014)

8. What is the internet's stance on equal Wimbledon's prize money for Men's and Women's Singles winners, eight years after it was officially announced? (Ford, 2014)

9. How did the social media react to Microsoft CEO Satya Nadella's public suggestion to women to not worry about raises? (Barr, 2014)

10. JK Rowling assumed the pen name "Robert Galbraith" for The Cuckoo's Calling, and a few more crime fiction novels What were the reactions to either of the gendered alter egos of the same person? (Hugel, 2013)

11. What are the reactions to laws that promote equal salaries for men and women in similar positions? (Commission, 2010, Jowit, 2015)

12. How do audiences react to different genres of music from male and female musicians? (Mayberry, 2013)

13. What is the attitude towards representation of men and women in Renaissance, Modern and Victorian Literatures? (Fortin, 2014)

14. How does Twitter react when male and female characters from popular TV shows are killed off? (Konnikova, 2013)

A.2 Perception

15. Gender roles - myth or reality? (Barnett, 2004)

16. How are content in gender-targeted magazines (like Women's Health Magazine and Men's Health Magazine) received by their readers? (Ratcliff, 2014)

17. What are the chances of content being re-shared when it is posted by a male as compared to a female? (Bennett, 2012)

18. Would 'genderbending' popular fictional characters be equally popular? How would fans like their favorite comic book characters gender-swapped? (Huh and Williams, 2010)

19. What forms of gender inequality can be observed from trending hashtags on Twitter? (Rightler-McDaniels and Hendrickson, 2014)

20. Is there benevolent sexism involved in the portrayal of men and women in advertisements? (Sheehan, 2013, Tuchman, 1979)

21. How are the roles of breadwinner and homemaker associated with gender? (Cunningham, 2008, Gardner, 2003, Luscombe, 2014b)

22. Is smoking perceived differently by the society when done by men as compared to women? (Warren et al., 2006)

23. Women are bad drivers - how true is this stereotype? (Hamilton, 2015)

24. What are the effects of sexism in comedy and humor? (Ford and Ferguson, 2004)

25. Is there gender bias in dating websites? (Hwang, 2013)

26. How are male and female pivotal characters portrayed in fiction? (Mlawski, 2010, Mosse, 2013)

27. How strong are laws regarding sex-selective abortions? (Kalantry, 2014)

28. What roles are generally preferred for male and female characters in literature? (Konnikova, 2013)

29. Is there sexism among the LGBTQIA+ community? (Ward, 2000)

30. Does the society think women are more emotional than men? (Lalama, 2004)

31. Do men and women both have equal opportunities of assuming the nurturing role? (UNFPA, 2005)

32. What were the overall representation of male and female protagonists in popular TV shows? (Lowe, 2014)

33. In a lawless society, is a man or a woman more likely to get killed/victimized? (Kellermann and Mercy, 1992)

34. What can be said about the gender associations in an Implicit Association Test (IAT)? (Greenwald et al., 2009)

35. Do curfew hours change by gender? (Kyine, 2013)

36. How does gender bias affect medical diagnoses? (Young et al., 1996)

37. Does the man or the woman talk more in a conversation? (Tannen, 1991)

38. 'Going Dutch' for a dinner date - equality or lack of etiquette? (Gneezy et al., 2004, Talbot and Quayle, 2010)

39. How are inanimate objects assigned genders? (Found in Old English) (Curzan, 2003, Jackson, 2012)

40. How likely is subject A (gender blinded) to be classified as a 'dick' or a 'bitch' for being involved in different kinds of socially unacceptable behaviors? (Basile, 2011)

41. What is the likelihood of men or women being the victim in cases of domestic violence? (Chuirazzi, 2015, Sarrel and Masters, 1982)

42. How right or wrong are the perceptions of the society to treat a transgendered person as one of the classical genders? (Schilt and Westbrook, 2009)

43. Are men's rights movement very similar to radical feminism? (White, 2011)

44. How equal are the participation of women and men in men's rights movements and feminism respectively? (Deven, 2011, Pape, 2011)

45. Should men and women be segregated in professional sports? (Messner, 2002)

46. Should men and women be allowed equal participation in the defense forces? (ASSOCIATION, 2013)

47. Who lies more - Males or Females? (Brooke, 2015, Grose, 2015)

48. How are toys for children gender-coded and stereotyped? (Cherney et al., 2003)

49. How does gender based discrimination affect physical sports/activities like bodybuilding? (Lowe, 1998)

50. How do commercials correlate products with gender? (Meaning, 2013)

51. What form of gender inequality prevails in the adult entertainment industry? (Catania et al., 1990)

52. Should chivalry be expected of both the genders? (Andress, 2012)

A.3 Economics

53. How much is the gender gap in wages and salaries in the IT and Tech Industry? (Farrell, 2005, PayScale, 2012)

54. What are the attrition rates in different industries affected by gender bias? (Farrell, 2005, PayScale, 2012)

55. What is the impact of gender inequality on global economy? (McBain, 2014)

56. Is the distribution of pay gender-dependent for different professions? (Farrell, 2005)
57. How is crowdsourcing affected by gender? (Phillips, 2011)
58. How many hours of work does a woman need to put in to match a man's $1 of income? (Kochhar, 2013)
59. Who controls the household finances - Men or Women? (Bialik, 2011)
60. Who does the grocery shopping for the family? (Goodman, 2008, Grocer, 2013)
61. Are insurance/retirement plans same for men and women? / Is there a plan that is not gender driven? (Palmer, 2015, Writers, 2013)
62. How equal are the rights to property for both the genders globally? (Agarwal, 1994)
63. How does gender equality affect the GDP of nations? 76
64. "Reproductive health indicators used in the GII do not have equivalent indicators for males". What forms of inequalities arise from this unequal treatment of gender? (Report, 2015)
65. Twice the funds are spent on breast cancer research as compared to prostate cancer research Are there other healthcare research disparities when considering gender-related diseases? (Benatar, 2012)
66. How equal are the concepts of maternal and paternal leave in a professional environment? (Ravn and Wetterberg, 2009)
67. What are the implications of men and women earning differently in a household? (Torabi, 2014)

A.4 Career, Education and Research

68. Are male and female professors rated differently by students on online platforms like RateMyProfessors? (Powell, 2016)

69. Do the teachers in elementary/primary schools give preferential treatment (in terms of grades) to children on the basis of their gender? (Cornwell et al., 2013)

70. What is the impact of 'guy thing versus girl thing' notion when considering gender roles in academia and research participation? (Rice, 2012)

71. How are gender differences and informal professional networking correlated? (Waldstrøm et al., 2001)

72. Are women more likely to fit into a receptionist position than men? (Csanady, 2015)

73. How does gender gap affect the fields of Science and Mathematics? (Angier and Chang, 2005, Niederle and Vesterlund, 2010)

74. How much is the divide in the participation of men and women in hazardous occupations? (Hoskins, 2005, Knestaut, 1996, Mantle, 2015, Smith, 2013)

75. How likely is it for a woman to be associated with research with another man as compared to another woman? (Rice, 2015)

76. Is it more likely to find men in career pole positions than women? (Slaughter, 2012)

77. How does motherhood slow women down at their professions? (Paik et al., 2007, Rice, 2011)

78. How does being a family person affect men and women as professionals? (Parker, 2015)

79. If there is equal opportunity between genders, will there be equal outcomes? (Anonymous, 2015, Arneson, 2015, of Mauritius, 2014, Reisman, 1997)

80. How are both genders progressing in STEM fields? (Beede et al., 2011)

81. What is the likelihood of a woman being promoted at work as compared to a man? (Ibarra et al., 2010, Lebowitz, 2015)

82. What is the gender-wise distribution of levels of education in the world? (Lutz, 2014)

83. How do employees of both the genders perform under female supervisors? (Buchanan, 2014, Chrisler and Clapp, 2008, Newport and Wilke, 2013)

84. What measures do firms take to ensure equality in hiring when considering men and women? (Banning, 2015, Egan, 2015)

85. How does gender stereotypes put pressure on both the genders to 'prove a point' for the society? (Basow, 1992, Garst and Bodenhausen, 1997)

86. How are both the genders taken into account when designing leadership development workshops? (Ely et al., 2011)

87. Does gender inequality affect postdoctoral researchers? (Doornebal, 2014)

88. How do hiring rates differ for men and women during periods of job scarcity? (Access, 2014)

89. Who is more likely to be the next successful entrepreneur - a man or a woman? (Strohmeyer et al., 2010)

90. How does gender bias affect a musician? (McSweeney, 2013)

A.5 Policy and Decision Making

91. Who handles the authoritative/leadership roles better - Men or Women? (Center, 2015a, Garber, 2012)

92. Which gender makes better political executives? (Center, 2008, 2015b)

93. How does gender impact the political agendas and activism on social media? (Loiseau and Nowacka)

94. How different are punishments for criminals in case of male and female perpetrators? (Mossière and Dalby, 2008, Starr, 2015)

95. Are there gender stereotyped perceptions regarding the victim and perpetrator? (Houry et al., 2008)

96. Do we have equal disciplinary measures for males and females, ranging from the household to the multinational companies to the judicial system? (Shapiro, 2000)

97. How does the law protect both the sexes from cases of domestic violence? (Statistics, 2012)

98. How would expectations change from a female President, should the US elect a female as POTUS? (Fox, 2013)

99. What is the difference between economic aids provided to both the sexes in third world countries? (Gunn, 2012)

100. How does gender bias affect efficiency in non-domestic roles? (Silverstein and Sayre, 2009)

Appendix B

Data Collection and Labeling

An initial scan of public repositories did not yield any suitable data source that could be used for the purpose of our work. As a solution, we scraped the required information, performed preliminary clean up of the raw data, and labelled the professor's gender with the help of second person pronouns used in the comments.

B.1 Scraping

RMP web pages are built on JavaServer Pages (JSP) to support their dynamic framework. A blank search on their Search[1] page returned a paginated list of professors, each of whom are assigned an internally generated numeric ID[2]. We sifted through all the pages and noted down their IDs.

To collect information on the professors, we iterated through our list of IDs and appended to a web URL prefix, which redirected to a professor's profile page. We ran a source code scraper, and processed it with a Beautiful Soup 4[3], HTML parsing tool and stored the relevant information in a suitable database for future querying.

Reviews were collected using a similar strategy. A JavaScript request was constructed using a prefix URL containing the professor's profile ID. This returned a paginated list of reviews with required details in JSON format. We cleaned the raw JSON, stored it in

[1] http://www.ratemyprofessors.com/search.jsp

[2] This type of searching has since been disabled by RMP

[3] https://www.crummy.com/software/BeautifulSoup/bs4/doc/

our database and cross referenced review ID with the professor's ID to keep track of the one-many relationship.

B.2 Labeling

RMP does not enable reviewers to report their own or the reviewee instructor's gender. However, we were able to decode the gender of the professor by closely following the review comments.

Even though the review ratings do not require reporting gender, the comments tend to not be gender-proof, owing to gendered second person pronouns like 'he', 'him', 'his', 'she', 'her', among the others. They also sometimes contain words like 'guy', 'gentleman', 'lady' which clearly indicate the gender of the professor.

For each professor, we iterated through all their review comments, and mapped the counts of words mentioned above into Male or Female word buckets. Whichever bucket contained more words was considered as the gender label of the professor.

Appendix C
Screenshots

RateMyProfessors (or RMP) is the largest online destination for student reviews. Depending on the reviews received, RMP aggregates the ratings received by the professor on their *clarity*, *helpfulness* and *easiness*. A few other details such as *overall quality*, calculated from the *clarity* and *helpfulness* and reported for the professor. The top tags received over all reviews are also reported. A screenshot for a professor's rating summary is given in Figure C.1.

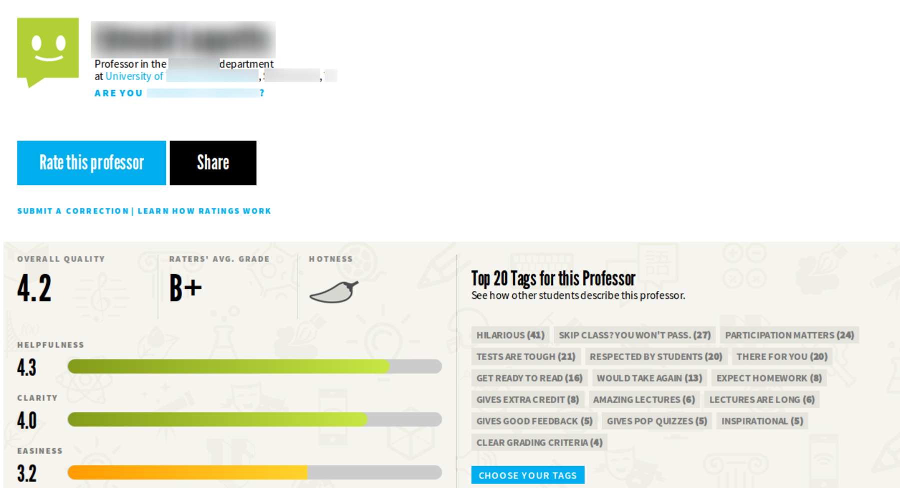


Fig. C.1: Professor Details on RMP

However, a professor's rating points consist of a collection of reviews received from students. In each review, a student can mention their own rating points for the professor on *clarity*, *helpfulness* and *easiness*. In addition to this, the students can choose upto three tags that are one word descriptions that the student believes best describes the professor.

A text comment follows each review, which serves as a more subjective measure of the professor's teaching. An example of reviews is given in Figure C.2.

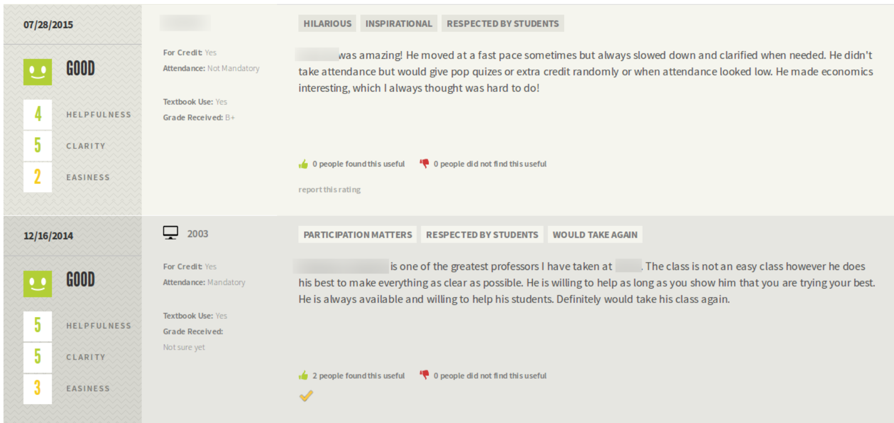


Fig. C.2: Sample reviews on RMP

Figures C.3 and C.4 on Pages 67 and 68 gives a detailed explanation of the review collection process, including the Likert scale rating system for *clarity*, *helpfulness* and *easiness*, the tags field, the comments field, the grades received by the student and a few other details.

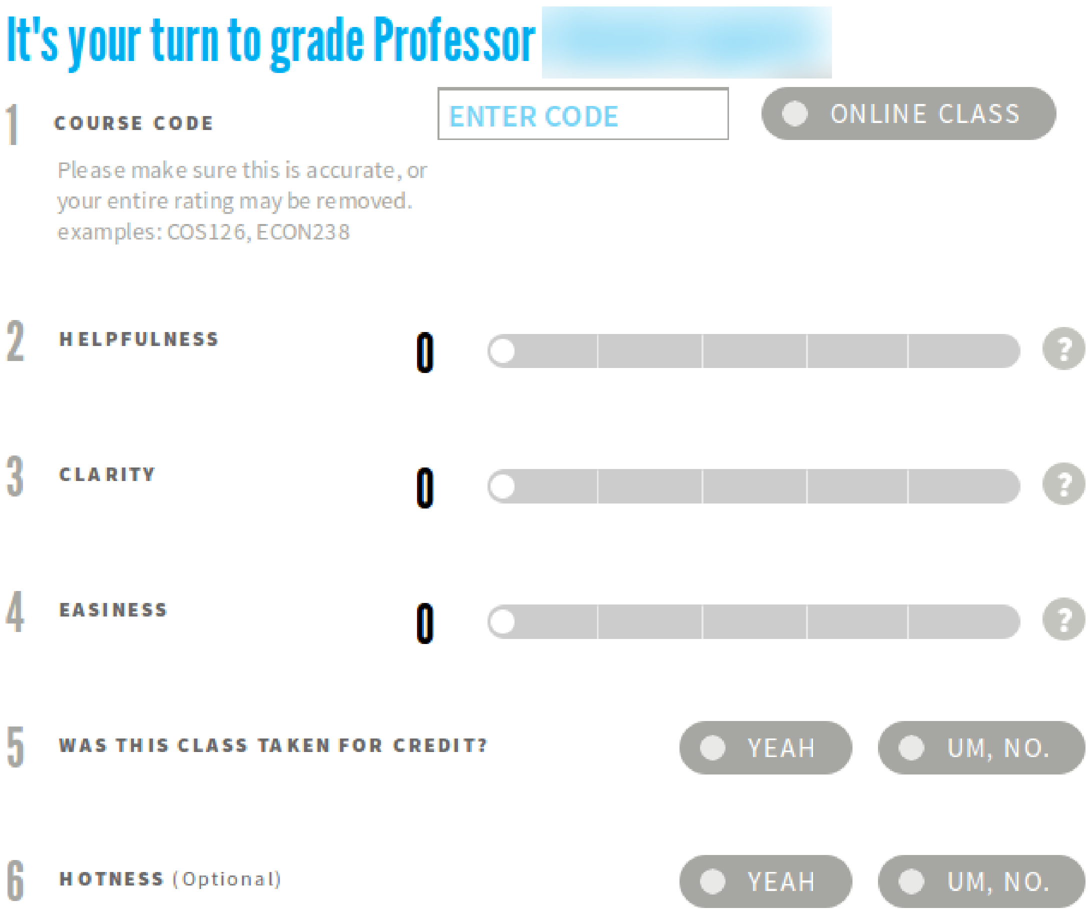


Fig. C.3: Review Collection page on RMP - Part 1

Fig. C.4: Review Collection page on RMP - Part 2

Bibliography

GGP Nesstar Online Access. Generations and gender survey lithuania wave 1 and wave 2. Technical report, Institut National DStudes Demographiques, 2014. URL http://ggpsurvey.ined.fr/documents/guide/GGPNesstarUG.pdf. 61

Bina Agarwal. Gender and command over property: A critical gap in economic analysis and policy in south asia. *World development*, 22(10):1455–1478, 1994. 59

Nalini Ambady and Robert Rosenthal. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of personality and social psychology*, 64(3):431, 1993. 11

Kristi Andersen and Elizabeth D. Miller. Gender and student evaluations of teaching. *PS: Political Science and Politics*, 30(2):216–219, 1997. ISSN 10490965, 15375935. URL http://www.jstor.org/stable/420499. 11, 21, 51

Justin Andress. Study shows female porn stars have higher quality of life than 'average' women, 2012. URL http://www.examiner.com/article/study-shows-female-porn-stars-have-higher-quality-of-life-than-average-women. 58

Natalie Angier and Kenneth Chang. Gray matter and sexes: A gray area scientifically. *New York Times*, 24, 2005. 60

Anonymous. If there is equal opportunity between genders, will there be equal outcomes?, 2015. 61

Richard Arneson. Equality of opportunity. http://plato.stanford.edu/archives/sum2015/entries/equal-opportunity/, 2015. 61

RESERVE OFFICERS ASSOCIATION. Selective service act registration of women, mar 2013. URL https://www.roa.org/sites/default/files/docs/13-03_Selective_Service_Registration_of_Women__exp_nc_20.pdf. 58

Taylor Banning. The surprising countries with more women in corporate leadership than the us—or even scandinavia. *Catalyst*, 2015. 61

Rosalind Chait Barnett. Women and multiple roles: Myths and reality. *Harvard review of psychiatry*, 12(3):158–164, 2004. 55

Valerie Barr. Much ado about salary: Grace hopper and satya nadella. *ACM SIGCAS Computers and Society*, 44(4):13–14, 2014. 55

Theresa Basile. Gender-neutral insults: Everyone's an asshole!, 2011. 57

Susan A Basow. *Gender: Stereotypes and roles* . Thomson Brooks/Cole Publishing Co, 1992. 61

Susan A Basow. Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87(4):656, 1995. 13

David N Beede, Tiffany A Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E Doms. Women in stem: A gender gap to innovation, 2011. 61

David Benatar. *The second sexism: discrimination against men and boys*. John Wiley & Sons, 2012. 59

Shea Bennett. On twitter, men are retweeted far more than women (and you're probably sexist, too). *SocialTimes*, 2012. URL http://www.adweek.com/socialtimes/twee-q-sexist-twitter/467654. 56

C. Bialik. Who makes the call at the mall, men or women?, 2011. URL http://www.wsj.com/articles/SB10001424052748703521304576278964279316994. 59

Michael Bittman and Judy Wajcman. The rush hour: The character of leisure time and gender equity. *Social forces*, 79(1):165–189, 2000. 7

Anne Boring. Gender biases in student evaluations of teachers. Documents de Travail de l'OFCE 2015-13, Observatoire Francais des Conjonctures Economiques (OFCE), 2015. URL http://EconPapers.repec.org/RePEc:fce:doctra:1513. 2, 13, 16, 21, 49, 51

C. Brooke. Women tell more fibs than men... honestly! four in five say they tell a lie on a daily basis, 2015. 58

Rosemarie Bruynooghe, Sigrid Noelanders, and Sybille Opdebeeck. Vreedzame samenleving nog niet in zicht. geweld als maatschappelijk probleem voor mannen en vrouwen. *CGSO-jaarboek. Gent: CGSO*, 2000. 5, 6

M. Buchanan. More women (and men) prefer male bosses to female bosses, 2014. URL http://www.avoiceformen.com/feminism/feminist-lies-feminism/more-women-and-men-prefer-male-bosses-to-female-bosses/. 61

Joseph A Catania, David R Gibson, Dale D Chitwood, and Thomas J Coates. Methodological problems in aids behavioral research: influences on measurement error and participation bias in studies of sexual behavior. *Psychological bulletin*, 108(3):339, 1990. 58

Pew Research Center. *Men or Women: Who's the Better Leader?* Women and Leadership. 2008. 62

Pew Research Center. Women and leadership. Technical report, Pew Research, 2015a. URL http://www.pewsocialtrends.org/2015/01/14/women-and-leadership/. 62

Pew Research Center. *What Makes a Good Leader, and Does Gender Matter?* Women and Leadership. Pew Research, 2015b. URL http://www.pewsocialtrends.org/2015/01/14/chapter-2-what-makes-a-good-leader-and-does-gender-matter/. 62

Isabelle D Cherney, Lisa Kelly-Vance, KATRINA GILL GLOVER, Amy Ruane, and BRIGETTE OLIVER RYALLS. The effects of stereotyped toys and gender on play assessment in children aged 18-47 months. *Educational Psychology*, 23(1):95–106, 2003. 58

Joan Chrisler and Sarah K Clapp. When the boss is a woman. *The psychology of women at work*, pages 39–67, 2008. 61

Sara Chuirazzi. Feminism: It's not about blaming men, it's about equality, 2015. 57

U.S.Equal Employment Opportunity Commission. Facts about equal pay and compensation discrimination, apr 2010. URL http://www.eeoc.gov/eeoc/publications/upload/fs-epa.pdf. 55

Robert W Connell. Cool guys, swots and wimps: the interplay of masculinity and education. *Oxford review of education*, 15(3):291–203, 1989. 7

Christopher Cornwell, David B Mustard, and Jessica Van Parys. Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1):236–264, 2013. 60

Ashley Csanady. 'female candidates are preferred' for receptionist gig: Toronto web firm's online posting slammed, 2015. URL http://news.nationalpost.com/toronto/female-candidates-are-preferred-for-receptionist-gig-toronto-web-firms-online-posti 60

Mick Cunningham. Changing attitudes toward the male breadwinner, female homemaker family model: Influences of women's employment and education over the lifecourse. *Social Forces*, 87(1):299–323, 2008. 56

Anne Curzan. *Gender shifts in the history of English*. Cambridge University Press, 2003. 57

Mandy Deven. Is their room for men in the women's movement?, 2011. URL http://www.herizons.ca/node/255. 58

A. Doornebal. Designing a career ladder for women postdocs in science. Technical report, Elsevier, sep 2014. URL //www.elsevier.com/connect/designing-a-career-ladder-for-women-postdocs-in-science. 61

Lesley Doyal. Gender equity in health: debates and dilemmas. *Social science & medicine*, 51(6):931–939, 2000. 6

Matt Egan. Still missing: Female business leaders. *CNNMoney, March*, 24, 2015. 61

Robin J Ely, Herminia Ibarra, and Deborah M Kolb. Taking gender into account: Theory and design for women's leadership development programs. *Academy of Management Learning & Education*, 10(3):474–493, 2011. 61

Warren Farrell. *Why men earn more: The startling truth behind the pay gap–and what women can do about it*. AMACOM Div American Mgmt Assn, 2005. 58, 59

Bonnie D Ford. Venus was 'at the forefront'. Technical report, ESPN W, 2014. 55

Thomas E Ford and Mark A Ferguson. Social consequences of disparagement humor: A prejudiced norm theory. *Personality and Social Psychology Review*, 8(1):79–94, 2004. 56

Elaine Fortin. Early nineteenth century attitudes toward women and their roles as represented by literature popular in worcester, massachusetts. *Teachushistory. org*, 2014. 55

L. Fox. Poll: Voters ready for a woman president, 2013. URL http://www.usnews.com/news/articles/2013/05/02/poll-voters-ready-for-a-woman-president. 62

Harvey R Freeman. Student evaluations of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. *Journal of Educational Psychology*, 86(4):627, 1994. 13

Rebecca Fuhrer, Stephen A Stansfeld, J Chemali, and Martin J Shipley. Gender, social relations and mental health: prospective findings from an occupational cohort (whitehall ii study). *Social science & medicine*, 48(1):77–87, 1999. 6

Makiko Fuwa. Macro-level gender inequality and the division of household labor in 22 countries. *American Sociological Review*, 69(6):751–767, 2004. 4

Lady Gaga. Bad romance, 2009. 54

M. Garber. Why we prefer masculine voices (even in women), 2012. 62

R. J. Gardner. Alpha women, beta men, 2003. URL http://nymag.com/nymetro/news/features/n_9495/. 56

Jennifer Garst and Galen V Bodenhausen. Advertising's effects on men's gender role attitudes. *Sex Roles*, 36(9-10):551–572, 1997. 61

Peter Glick and Susan T Fiske. An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56(2):109, 2001. 4

Ignace Glorieux, Suzana Koelet, and Maarten Moens. Vlamingen in tienduizend en tachtig minuten: een tijdsbudgetonderzoek. *Vlaanderen gepeild*, pages 157–184, 2001. 7

Uri Gneezy, Ernan Haruvy, and Hadas Yafe. The inefficiency of splitting the bill*. *The Economic Journal*, 114(495):265–280, 2004. 57

Jack Goodman. Who does the grocery shopping, and when do they do it. *The Time Use Institute*, 2008. 59

Anthony G Greenwald, T Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R Banaji. Understanding and using the implicit association test: Iii. meta-analysis of predictive validity. *Journal of personality and social psychology*, 97(1):17, 2009. 57

Progressive Grocer. Women dominate grocery shopping: Study. *Progressive Grocer*, apr 2013. URL http://www.progressivegrocer.com/women-dominate-grocery-shopping-study?nopaging=1. 59

Anouchka Grose. Why do women lie more than men? because we're nicer, 2015. 58

Dwyer Gunn. The challenge of closing the gender gap in developing countries, 2012. URL http://www.theatlantic.com/sexes/archive/2012/12/the-challenge-of-closing-the-gender-gap-in-developing-countries/266617/. 62

David L Hamilton. *Cognitive processes in stereotyping and intergroup behavior*. Psychology Press, 2015. 56

Ricardo Hausmann, Laura D. Tyson, Yasmina Bekhouche, and Saadia Zahidi. *Gender Gap Report 2014*. World Economic Forum, 2014. 4

Anne B Hoskins. Occupational injuries, illnesses, and fatalities among women. *Monthly Lab. Rev.*, 128:31, 2005. 60

Debra Houry, Karin V Rhodes, Robin S Kemball, Lorie Click, Catherine Cerulli, Louise Anne McNutt, and Nadine J Kaslow. Differences in female and male victims and perpetrators of partner violence with respect to web scores. *Journal of interpersonal violence*, 23(8):1041–1055, 2008. 62

Melissa Hugel. Why did j.k. rowling use a male pen name for her crime novel? *Arts.Mic*, jul 2013. 55

Searle Huh and Dmitri Williams. *Dude looks like a lady: Gender swapping in an online game*. Springer, 2010. 56

Wei-Chin Hwang. Who are people willing to date? ethnic and gender patterns in online dating. *Race and Social Problems*, 5(1):28–40, 2013. 56

Herminia Ibarra, Nancy M Carter, Christine Silva, et al. Why men still get more promotions than women, 2010. 61

S. B. Jackson. Masculine or feminine? (and why it matters). *Psychology Today*, 21, sep 2012. URL https://www.psychologytoday.com/blog/culture-conscious/201209/masculine-or-feminine-and-why-it-matters. 57

Jerry A Jacobs. Gender inequality and higher education. *Annual review of sociology*, pages 153–185, 1996. 4

Juilette Jowit. Women will get equal pay . . . in 118 years, 2015. 55

Sital Kalantry. Replacing myths with facts: Sex-selective abortion laws in the united states. *Cornell Legal Studies Research Paper*, 14(34), 2014. 56

Ellen Kate. Why we martyr single dads, but demonize single moms (and what to do about it), 2014. URL http://everydayfeminism.com/2014/08/single-parent-double-standard/. 55

Arthur L Kellermann and James A Mercy. Men, women, and murder: gender-specific differences in rates of fatal violence and victimization. *Journal of Trauma and Acute Care Surgery*, 33(1):1–5, 1992. 57

Lane Kenworthy and Melissa Malami. Gender inequality in political representation: A worldwide comparative analysis. *Social Forces*, 78(1):235–268, 1999. 5

Diane Kierstead, Patti D'Agostino, and Heidi Dill. Sex role stereotyping of college professors: Bias in students' ratings of instructors. *Journal of Educational Psychology*, 80 (3):342, 1988. 12

Svetlana Kiritchenko and Saif M Mohammad. Sentiment composition of words with opposing polarities. In *Proceedings of The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL), San Diego, California*, 2016. 38

A Knestaut. Fewer women than men die of work-related injuries. *Compensation and Working Conditions Online*, 1(1), 1996. 60

Rajesh Kochhar. How pew research measured the gender pay gap. Technical report, Pew Research Center, 2013. URL http://www.pewresearch.org/fact-tank/2013/12/11/how-pew-research-measured-the-gender-pay-gap/. 59

M. Konnikova. *Do Readers Judge Female Characters More Harshly Than Male Characters?* The Atlantic, 2013. 55, 56

Tayfun Kucukyilmaz, B Barla Cambazoglu, Cevdet Aykanat, and Fazli Can. Chat mining for gender prediction. In *Advances in Information Systems*, pages 274–283. Springer, 2006. 40

Daw Kyine. Are women asking for trouble if they go out alone late at night?, 2013. URL http://www.mmtimes.com/index.php/special-features/167-the-modern-woman/7099-are-women-asking-for-trouble-if-they-go-out-alone-late-at-night.html. 57

Christina Lalama. Are women more emotional than men? *Making*, page 39, 2004. 57

Laura I Langbein. The validity of student evaluations of teaching. *PS: Political Science & Politics*, 27(03):545–553, 1994. 12

Shana Lebowitz. A new study from lean in and mckinsey finds exactly how much more likely men are to get promoted than women, 2015. URL http://www.businessinsider.com/women-are-less-likely-to-get-promoted-2015-10. 61

Peggy M Lee. She'-e-os: Gender effects and stock price reactions to the announcements of top executive appointments. In *Academy of Management Proceedings*, volume 2003, pages GG1–GG6. Academy of Management, 2003. 54

Sara Lelli et al. *Factor Analysis vs. Fuzzy Sets theory: Assessing the influence of different techniques on sen's functioning approach*. Kath. Univ., Department Economie, Center for Economic Studies, 2001. 7

Estelle Loiseau and Keiko Nowacka. Can social media effectively include women's voices in decision-making processes? 62

Linda Lowe. Hollywood by the numbers: Men still largely outnumber women in movies and tv shows, 2014. URL http://parade.com/264408/linzlowe/hollywood-by-the-numbers-men-still-largely-outnumber-women-in-movies-and-tv-shows/. 57

Maria R Lowe. *Women of steel: Female bodybuilders and the struggle for self-definition.* NYU Press, 1998. 58

Belinda Luscombe. A dramatic reading of my youtube hate comments, January 17 2014a. URL http://healthland.time.com/2014/01/17/a-dramatic-reading-of-my-youtube-hate-comments/. 54

Belinda Luscombe. The real problem with women as the family breadwinner, 2014b. URL http://time.com/87890/fox-news-female-breadwinner/. 56

Belinda Luscombe. [interview], January 2014c. 54

Wolfgang Lutz. *Inequality in Human Capacity (by Education, Gender and Age)*. IIASA-Alpbach Group, aug 2014. URL http://www.alpbach.org/wp-content/uploads/2014/08/Conceptual_note_on_educational_inequalities.pdf. 61

Hollie Mantle. Men over twenty times as likely as women to die at work, mar 2015. URL http://www.avoiceformen.com/men/men-over-twenty-times-as-likely-as-women-to-die-at-work/. 60

Herbert W Marsh and Terri L Cooper. Prior subject interest, students' evaluations, and instructional effectiveness. *Multivariate Behavioral Research*, 16(1):83–104, 1981. 11

Malcolm Marshall. Eminem - not afraid, 2010. URL https://www.youtube.com/watch?v=j5-yKhDd64s. 54

Elaine Martin. Power and authority in the classroom: Sexist stereotypes in teaching evaluations. *Signs*, 9(3):482–492, 1984. 13

Lauren Mayberry. Chvrches' lauren mayberry: 'i will not accept online misogyny', 2013. URL http://www.theguardian.com/music/musicblog/2013/sep/30/chvrches-lauren-mayberry-online-misogyny. 55

S. McBain. Gender inequality is costing the global economy trillions of dollars a year, 2014. URL http://www.newstatesman.com/economics/2014/02/gender-inequality-costing-global-economy-trillions-dollars-year. 58

Lisa A McLoughlin. Spotlighting: Emergent gender bias in undergraduate engineering education. *Journal of Engineering Education*, 94(4):373, 2005. 8, 10

E. McSweeney. The power list: Why women aren't equals in new music leadership and innovation. *New Music Box*, apr 2013. URL http://www.newmusicbox.org/articles/the-power-list-why-women-arent-equals-in-new-music-leadership-and-innovation/. 61

How Gender Shapes Meaning. Gender and advertising. *Controversies in Contemporary Advertising*, page 89, 2013. 58

mercuryblues. A tale of two celebrity overdoses, 2014. 54

Michael A Messner. *Taking the field Women, men, and sports*. U of Minnesota Press, 2002. 58

Shana Mlawski. Are male characters more likable than female characters? Technical report, 28; //www.overthinkingit.com/2010/06/28/are-male-characters-more-likable-than-female-characters/, jun 2010. URL www.overthinkingit.com/2010/06/28/are-male-characters-more-likable-than-female-characters/. 56

K. Mosse. Where have all the brave girls gone?, 2013. URL http://www.theguardian.com/books/booksblog/2013/aug/24/where-have-all-brave-girls-gone-kate-mosse. 56

Annik Mossière and J Thomas Dalby. The influence of gender and age in mock juror decision-making. *Europe's Journal of Psychology*, 4(4), 2008. 62

Arjun Mukherjee and Bing Liu. Improving gender classification of blog authors. In *Proceedings of the 2010 conference on Empirical Methods in natural Language Processing*, pages 207–217. Association for Computational Linguistics, 2010. 40

Allison Munch, J Miller McPherson, and Lynn Smith-Lovin. Gender, children, and social contact: The effects of childrearing for men and women. *American Sociological Review*, pages 509–520, 1997. 6

Frank Newport and Joy Wilke. Americans still prefer a male boss. Technical report, Gallup Research, 2013. URL http://www.gallup.com/poll/165791/americans-prefer-male-boss.aspx?ref=image. 61

Muriel Niederle and Lise Vesterlund. Explaining the gender gap in math test scores: The role of competition. *The Journal of Economic Perspectives*, 24(2):129–144, 2010. 60

University of Mauritius. Difference between gender equality and gender equity, 2014. 61

Lorena O'Neil. Is red carpet coverage sexist? battle lines drawn at oscars over #askhermore, February 21 2015. URL http://www.hollywoodreporter.com/news/oscars-red-carpet-2015-battle-775698. 54

Renee Oyotode, Zubair Ali Raja, and Jorge O Brusa. The bond market responses to female ceos appointment. *Available at SSRN 2662155*, 2015. 54

In Paik, Shelley Correll, Stephen Benard, et al. Getting a job: Is there a motherhood penalty? Technical report, The Field Experiments Website, 2007. 60

Kate Palmer. Men are still charged more than women for car insurance, despite eu rule change, 2015. URL http://www.telegraph.co.uk/finance/personalfinance/insurance/motorinsurance/11521781/Men-are-still-charged-more-than-women-for-car-insurance-despite-EU-rule-change.html. 59

Pierette Pape. The ewl's seminar on men and feminism: new paths of reflection for the women's movement. Technical Report Autumn 2011, European Women's Voice, 2011. 58

Kim Parker. Women more than men adjust their careers for family life. Technical report, Pew Research Center, jan 2015. URL http://www.pewresearch.org/fact-tank/2015/10/01/women-more-than-men-adjust-their-careers-for-family-life/. 60

PayScale. *Do Men Really Earn More Than Women?* PayScale Human Capital, 2012. URL http://www.payscale.com/gender-lifetime-earnings-gap. 58

Trond Petersen and Thea Togstad. Getting the offer: Sex discrimination in hiring. *Research in Social Stratification and Mobility*, 24(3):239–257, 2006. 9

Anna Lena Phillips. Crowdsourcing gender equity, 2011. 59

Shelley A Phipps and Peter S Burton. Sharing within families: implications for the measurement of poverty among individuals in canada. *Canadian Journal of Economics*, pages 177–204, 1995. 7

Avery Powell. Rate my professors data shows students' gender bias against professors, 2016. URL http://college.usatoday.com/2015/02/16/rate-my-professors-data-shows-students-gender-bias-against-professors/. 16, 60

C. Ratcliff. Why are health magazines dumbing down content for women?, 2014. URL http://www.survivaloftherealist.com/dumbing-down-womens-health/. 55

Anna-Birte Ravn and Christina Carlsson Wetterberg. *Gender equality and welfare politics in Scandinavia: the limits of political ambition?* Policy Press, 2009. 59

David Reisman. Equality of outcome, 1997. 61

Human Development Report. What is the rationale for using indicators for health without equivalents for men? Technical report, UNDP, 2015. URL http://hdr.undp.org/en/content/what-rationale-using-indicators-health-without-equivalents-men. 59

Barbara F Reskin and Patricia A Roos. *Job queues, gender queues: Explaining women's inroads into male occupations*. Temple University Press, 2009. 5

Curt Rice. The motherhood penalty: It's not children that slow mothers down, dec 2011. URL http://curt-rice.com/2011/12/08/the-motherhood-penalty-its-not-children-that-slow-mothers-down/. 60

Curt Rice. Science: It's a girl thing! statement from members of the gender expert advisory group. Technical report, Science in Balance, 2012. 60

Curt Rice. Eve and evidence: what research tells us about gender equality, mar 2015. URL http://curt-rice.com/2015/03/08/eve-evidence-research-tells-us-gender-equality/. 60

Jodi L Rightler-McDaniels and Elizabeth M Hendrickson. Hoes and hashtags: constructions of gender and race in trending topics. *Social Semiotics*, 24(2):175–190, 2014. 56

Ingrid Robeyns. Sen's capability approach and gender inequality: Selecting relevant capabilities. *Feminist Economics*, 9(2-3):61–92, 2003. doi: 10.1080/1354570022000078024. URL http://dx.doi.org/10.1080/1354570022000078024. 5, 7

Sue V Rosser and Mark Zachary Taylor. Why are we still worried about women in science? *Academe*, 95(3):7–10, 2009. 25

Bernice Resnick Sandier. Women faculty at work in the classroom, or, why it still hurts to be a woman in labor. *Communication Education*, 40(1):6–15, 1991. 12

Philip M Sarrel and William H Masters. *Sexual molestation of men by women*, volume 11. Springer, 1982. 57

Kristen Schilt and Laurel Westbrook. Doing gender, doing heteronormativity "gender normals," transgender people, and the social maintenance of heterosexuality. *Gender & Society*, 23(4):440–464, 2009. 57

Ben Schmidt. Gendered language in teacher reviews. *Bookworm Project*, feb 2015. 41

Andrea Shapiro. Unequal before the law: Men, women and the death penalty. *Am. UJ Gender Soc. Pol'y & L.*, 8:427, 2000. 62

Kim Bartel Sheehan. *Controversies in contemporary advertising*. Sage Publications, 2013. 56

Jim Sidanius and Marie Crane. Job evaluation and gender: The case of university faculty. *Journal of Applied Social Psychology*, 19(2):174–197, 1989. 12

Michael J Silverstein and Kate Sayre. The female economy. *Harvard Business Review*, 87(9):46–53, 2009. 62

Anne-Marie Slaughter. Why women still can't have it all, aug 2012. 60

J. Smith. America's 10 deadliest jobs, 2013. URL http://www.forbes.com/sites/jacquelynsmith/2013/08/22/americas-10-deadliest-jobs-2/#4d06b9d85095. 60

Jeff Sonderman. Women journalists confront harassment, sexism when using social media. *Poynter.*, 2011. 54

Philip B Stark and Richard Freishtat. An evaluation of course evaluations. *Center for Teaching and Learning, University of California, Berkley*, 2014. 11, 29, 32, 41, 48, 51

Sonja B Starr. Estimating gender disparities in federal criminal cases. *American Law and Economics Review*, 17(1):127–159, 2015. 62

Anne Statham, Laurel Richardson, and Judith A Cook. *Gender and university teaching: A negotiated difference*. SUNY Press, 1991. 13

Domestic Violence Statistics. Domestic violence statistics, 2012. 62

Robert Strohmeyer, Vartuhí Tonoyan, and Michael Woywode. Why are women less likely to become entrepreneurs than men? a jack-of-all trades perspective of human capital (summary). *Frontiers of Entrepreneurship Research*, 30(8):7, 2010. 61

James Surowiecki. *The wisdom of crowds*. Anchor, 2005. 40

Kirsten Talbot and Michael Quayle. The perils of being a nice guy: Contextual variation in five young women's constructions of acceptable hegemonic and alternative masculinities. *Men and Masculinities*, 13(2):255–278, 2010. 57

Deborah Tannen. *You just don't understand: Women and men in conversation*. Virago London, 1991. 57

Josh Terrell, Justin Middleton, Clarissa Rainear, Emerson Murphy-Hill, Chris Parnin, and Andrew Kofink. Gender bias in open source: Pull request acceptance of women versus men. *PeerJ PrePrints*, 2016. URL https://peerj.com/preprints/1733v1. 9

Farnoosh Torabi. *When She Makes More: 10 Rules for Breadwinning Women*. Hudson Street Press, 2014. 59

Gaye Tuchman. *Women's depiction by the mass media*, volume 4. JSTOR, 1979. 56

UNFPA. Frequently asked questions about gender equality, 2005. URL http://www.unfpa.org/resources/frequently-asked-questions-about-gender-equality. 57

Christian Waldstrøm et al. Gender differences and informal networks in a scandinavian context, 2001. 60

Jane Ward. Queer sexism: Rethinking gay men and masculinity. *RESEARCH ON MEN AND MASCULINITIES SERIES*, 12:152–175, 2000. 56

CW Warren, NR Jones, MP Eriksen, S Asma, Global Tobacco Surveillance System (GTSS) collaborative group, et al. Patterns of global tobacco use in young people and implications for future chronic disease burden in adults. *The lancet*, 367(9512): 749–753, 2006. 56

Molly Warrington and Michael Younger. The other side of the gender gap. *Gender and education*, 12(4):493–508, 2000. 7

Martha S West and John W Curtis. *AAUP faculty gender equity indicators 2006*. American Association of University Professors Washington, DC, 2006. 25

Jared White. What's the difference between the men's rights movement and feminism?, 2011. URL http://www.avoiceformen.com/mens-rights/whats-the-difference/. 58

Andrew C Worthington. The impact of student perceptions and characteristics on teaching evaluations: a case study in finance education. *Assessment & Evaluation in Higher Education*, 27(1):49–64, 2002. 11

Staff Writers. *Equally Insured? What You Need to Know About Women and Health Insurance Coverage*. Insurance Quotes, 2013. URL http://www.insurancequotes.org/health-insurance/health-resources/coverage-for-women/. 59

Carrie L Yodanis. Gender inequality, violence against women, and fear a cross-national test of the feminist theory of violence against women. *Journal of interpersonal violence*, 19(6):655–675, 2004. 5

Terry Young, Rebecca Hutton, Laurel Finn, Safwan Badr, and Mari Palta. The gender bias in sleep apnea diagnosis: are women missed because they have different symptoms? *Archives of internal medicine*, 156(21):2445–2451, 1996. 57

Souradeep Sinha

Education

2014–2016 **MS in Computer Science** *Syracuse University*
GPA (out of 4) - 3.47
Relevant Coursework: Analytical Data Mining, Database Management Systems, Design and Analysis of Algorithms, Operations Research, Cryptography

2010–2014 **BTech in Information Technology** *West Bengal University of Technology*
GPA (out of 10) - 8.44

Master thesis

title *Assessing gender inequality from large scale online student reviews*

supervisor Dr. Reza Zafarani

description Using a systematic approach, we evaluate a logistic regression learning model to detect gender inequality from anonymous student reviews.

Professional Experience

Related Experience

2015–Present **Graduate Researcher - Data Science** *Data Lab, Syracuse University*
- Designed multi-threaded scrapers to collect 16 million external data points as JSON objects
- Cleaned, preprocessed, indexed and warehoused numerical and textual data into MongoDB
- Tested model validity using ROC curves with Weka Naive Bayes, J48 Decision Tree and LibSVM logistic regression classifiers (High dimensional learning models)
- Visualized novel concepts like Crowd Bias and Mood Swings using NLTK, NumPy, Scikit, POSTagger, VADER, Matplotlib APIs
- Mapped data artefacts using Python, PyMongo API and MATLAB toolkits

2013–2013 **Data Analysis Intern** *Quikr,* India
- Investigated online advertisement data to identify consumer trends and buying patterns
- Developed statistical and predictive models that helped increase customer engagement time on premium ads by 14%

2013–2014 **Database Lab Administrator** *Institute of Engineering and Management,* India
- Developed MySQL queries to effectively test core concepts of relational databases
- Designed course project with appropriate Visio schema diagrams and sample data to facilitate stored procedures, functions, triggers and error handling

4 Evesboro Road – Chesterfield, NJ 08515
+1 (315) 412 8844 • *ssinha04@syr.edu* • *in /in/souradeepsinha*
ssinhaonline

Additional Experience

2014–Present **Late Night Manager** *Student Centers and Programming Services,* Syracuse

2015–Present **Event Staff** *Events and Technical Services,* Syracuse

2010–2014 **Business Development Associate** *In Good Taste,* India

Projects

α-Tweet: Retweet prediction with Naive Bayes *Syracuse University*

- Designed Selenium scraper bots to collect tweet history for users and their followers
- Used NLTK to perform word tokenization and Gensim's TF-IDF filtering to clean data
- Modified Naive Bayes algorithm and Pearson's coefficient and identified negative correlation between tweet words and retweet possibility

Analysis of elasticity on AWS Cloud *Syracuse University*

- Used Botocore to combine EC2, S3 and DynamoDB resources to store vectorized clusters
- Designed shell scripts to enable command line autologins to multiple EC2 instances
- Automated the procedure of data upload, download and performance monitoring with simple Python Boto3 API calls

Comparative study on dimensionality reduction techniques *Syracuse University*

- Compared Random Projection and PCA to reduce a synthetic dataset of 60 features
- Employed Agglomerative Hierarchical Analysis with Conditional Entropy and Normalized Mutual Information for cluster validation

Improved accuracy: @sholiday genderPredictor *Syracuse University*

- Tweaked statistical variables to include a probabilistic distribution in Bayesian estimate
- Improved accuracy of prediction from 82% to 96%

Skills

{Languages}	Python, Java, MATLAB	{DB/rDBMS}	MySQL, MongoDB
{Big Data}	Hadoop, MapReduce, HiveQL	{AWS Stack}	EC2, S3, DynamoDB
{OS}	Linux, Unix	{Packages}	Weka, Pandas, Scipy, Gensim
{APIs}	PyMongo, Scikit, LibLinear, NLTK, Tweepy, Botocore, Matplotlib, POSTagger, VADER		

Leadership Roles

2015–Present **Chair** *Professional Development Committee* Syracuse University

2015–Present **Senator** *Computer Science, Graduate Student Organization* Syracuse University

2014–2015 **Fellow** *Leadership and Cultural Program* Syracuse University

2014–2015 **Graduate Ambassador** *College of Enginering and Comp Science* Syracuse University

Certifications

- The Data Scientist's Toolbox - Johns Hopkins University
- Introduction to Hadoop and Mapreduce - Cloudera
- Data Wrangling with MongoDB - MongoDB

4 Evesboro Road – Chesterfield, NJ 08515

+1 (315) 412 8844 • *ssinha04@syr.edu* • *in /in/souradeepsinha*

ssinhaonline