STUDENTS' PERCEPTION OF FACULTY THROUGH TWITTER DATA

Leela Saiteja Vatrapu

A thesis report submitted in partial fulfillment of
the requirements for the degree of
Master of Science

Department of Computer Science

Central Michigan University
Mount Pleasant, Michigan
February 2017

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Dr. Jesse Eickholt of the Computer Science Department at Central Michigan University. The door to Prof. Eickholt office was always open 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 Dr. Lisa Gandy, and Dr. Gongzhu Hu for their valuable directions when this project began, reviewed the draft of the document, and many contributions during my work. Throughout the process, these faculty members drew on their experiences of working with many graduate students as professors, dissertation advisors, and dissertation committee members.

ABSTRACT

STUDENT'S PERCEPTION OF FACULTY THROUGH TWITTER DATA

by Leela Saiteja Vatrapu

The main objective of this thesis is to examine the perceptions of students regarding faculty. Several scripts were developed along with programs to answer questions on the specific topic using data mining techniques. The dataset is a large random sample of tweets that have been collected over a 9-month period and represent ~1% of tweets created during that time frame. We are working on a specific topic of students' views about their professors. In this work, many questions such as what students are tweeting about their professors and what are the complaints they are expressing about their instructors were examined. For this purpose, a classifier was constructed to identify a Twitter user as student using historical data and then their tweets were analyzed.

To characterize the public perception of the faculty, a large classifier was built using historical twitter data to categorize a twitter user as a student or non-student. The tweets of those users classified as students were manually processed to see the complaints the students were making about the professors. We categorized the types of complaints to get the perceptions of the students. Perceptions and misconceptions of students about faculty affect the policy makers and it would be beneficial to know the extent to which the misconceptions reach in the Twittersphere. Analysis was also performed to determine the commonality of keywords used to characterize faculty behavior.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

What do people think about faculty? Anecdotally, stories of the professor who puts minimal effort into a course, an aloof professor out of touch with “reality” and everyday life, or the professor that the neighbors always find at home during the work week seem to be common place. Are these anecdotal stories in-line with what students’ view or perceive? How do student perceptions align with the larger narrative and perception of faculty and faculty actions? These are important questions given the public’s general concern for the cost of higher education and decreasing levels of state funding for higher education.

Studies into these types of questions have been conducted before. As an example, past studies concerning professor ratings have been conducted on gender bias (i.e. male or female teachers) (Feldman, 1993; Basow, Phelan, & Capotosto, 2006). Their studies described how male students describe their female professors and female students about their male professors and when they are rating their professors the studies report that they are not gender biased. The professor rating is based on their personal views and the behavior of the professor not on the gender.

With the rise of the internet and mobile devices, social media has also evolved into a cheap, easy and effective data provider through user-generated content that is published and shared. By analyzing textual content, which represents the thoughts and communication between users, it is possible to pull out and make general inferences about public wants and views regarding a variety of topics. As users push content about their everyday lives, for some this inevitably includes information about faculty. This information can come through websites, blogs, microblogs (e.g., Twitter) or other sources of web-based forums for students (e.g., RateMyProfessor).

Twitter users tweet about a large variety of subjects including students discussing their professors. These tweets are their views and perceptions of their professors. A number studies have been made to research the students' views and perceptions about their professors. All of them are mainly survey based. Some have used many sources like RateMyProfessor.com and Facebook to evaluate the students' views (Hewitt & Forte, 2006; Kowai-Bell, et al., 2011). They determine the sentiment of the public views and perceptions of the faculty. Consider, for example, a study about the impacts on the students' impressions of professors through Ratemyprofessor.com and how it directly affects students' perception of control over the course (Kowai-Bell et al., 2011). Their study showed that there is an impact from comments. They said that the students who read positive comments rather than the negative comments are more likely to report an increase in perceived control, look forward to the class and recommend it to a friend.

Our main aim is to look at people's views and analyze the topics they are discussing about their professors and draw their perception of them. For this purpose, we have taken the twitter as our source because the people are using it to post their views and feelings. The data we use is the random data from twitter so it is not biased. From this we are mining the tweets that are relevant to our study, i.e. students' tweets about professors. We are looking at the interesting topics about the faculty they are discussing by using the hashtags and the words the users are using to discuss about their instructor.

Twitter

Twitter, a microblogging service started in 2006, has over 500 million users as of May 2015. Twitter users tweet on any topic within a limit of 140-characters and follow others to view their tweets. With that enormous increase in users, we are finding new ways to mine Twitter information about what people are thinking about products or other people. According to Pew Research Center's report (2014), about 74% of the online adult users are using social Networking sites. They also mentioned that 46% of adult users post their original photos and videos on online (McArthur & Bostedo-Conway, 2012). Early studies of Twitter for classroom use had unexpected results such as decreased student shyness, increased writing skills, and enhanced discussions. The Pew International and American Life (2014) project reported that 59% of students aged 12-17 shared artistic content through their posts.

For this work, we have gone through several research papers that deal with the sentiment analysis using Twitter and also perceptions of students and other people. McArthur and Bostedo-Conway (2012) in their work "Exploring the Relationship between Student Instructor Interaction on Twitter and Student Perceptions of Teacher Behaviors" discussed a survey they conducted with a sample of 144 students and 3 instructors. Each instructor assessed four courses and the instructor tweeted multiple times a week. They surveyed on teacher immediacy, teacher content relevance, teacher credibility and student-instructor interaction on Twitter. In processing the survey data and they found some interesting results about the student perceptions of teacher non-verbal behaviors in the class about the teacher smiles, gestures and movements. This has been perceived as a measure of closeness between student and instructor.

A work by Marín and Tur, (2014) entitled "Student Teachers' Attitude towards Twitter for Educational Aims" described an experiment with 100 student teachers to find out the students' perceptions. They tried to find out the students' perceptions on Twitter for educational aims. From this survey they concluded that most students show positive attitude after the use of Twitter in their learning. They suggested for some different educational designs that enhance the use of social media in education and the attitudes of student teachers toward technology in education.

Hewitt and Forte (2016), discussed a survey they conducted between the students and faculty at a mid-sized public research university to investigate if interaction on Facebook was influencing student perceptions' of faculty. About 79% of the enrolled students in the survey were on Facebook and 36% of them were friends with their professors. Their study showed that there is no impact of the communication on Facebook in rating their professor (Hewitt & Forte, 2006).

In a recent work over ideal qualities in an educator, students and faculty were surveyed concerning 214 pre-service teachers from a department of science and technology (Tunca, Şahin, Oğuz, & Bahar Güner, 2015). Respondents commented on their teachers with respect to professional roles and responsibilities, professional values, personal characteristics, professional ethic principles and social responsibilities. The results of their analysis concluded that when they rate their professor they consider some of the qualities like tolerance, patience, creativity, being far in evaluation, being enthusiastic, etc.

To date, most studies regarding Twitter and student perceptions of faculty have been survey based (i.e.., a survey was conducted by selecting a limited number of students and instructors). There are limitations to such approaches. First, the respondents are aware that the survey is going on. This may affect their responses. Second, they may be restricted in their form of response and may not be able to use any informal language. Twitter data contains free response by students and there is no expectation that the responses will be analyzed. This makes Twitter an attractive source for student opinions, although Twitter data does present some of its own challenges.

Challenges with Twitter Data

Given the short nature of each individual tweet, informal language or abbreviations are often used. This creates what can be considered noisy data as the meaning of the tweet can be unclear. The lack of context for a tweet can make it difficult to make inference about the user. In social network and media, there is also huge numbers of spam users (Benevenuto, Magno, Rodrigues, & Almeida, 2010). In their study, they manually prepared a labeled dataset by looking at the content of tweets of the users who tweet spam. Based on the tweets of these users they tried to predict other spam users. They took many attributes into consideration to detect the spammers and using these as a base they predicted the spam users from the large dataset of about 58 million users.

For the purposes of this study, it was important to identify potential student users to then be able to filter perceptions of faculty on a student/non-student basis. To accomplish this task a user profile was created using all the tweets which had been collected for a user. Some specific keywords were used to filter the tweets, label some of them and then build a classifier to identify all probable students. By making use of this historical data, the limitation of the tweet length was overcome.

In short, the principle focus of this work is on mining the Twittersphere with open source tools. This thesis analyzes a large amount of tweets to find out what students are tweeting about their professors, the sentiment of these tweets, and the types of complaints contained in the tweets with regard to faculty. As Twitter does not provide an easy means to identify users as students, the first step was to develop a classifier that would predict if a given user was a student, based on his or her tweets. With a means to identify potential students, a collection of Twitter users were classified and the tweets from those identified as students were retained for further analysis. This analysis consisted of both automated and manual means of identifying salient topics and complaints.

In other words, this work develops Python and Scala scripts to handle a large dataset of Twitter data. In this way, we are trying to capture the raw perceptions of people and students and analyze how they report on instructors in the real world. As a starting point, we used some keywords to filter the Twitter data and to perform stemming on the most frequent words. Students are identified by pulling out all the tweets for a particular user and building a profile. From the profile, student users can be identified and their tweets analyzed to see what each user is saying about the faculty in their tweets.

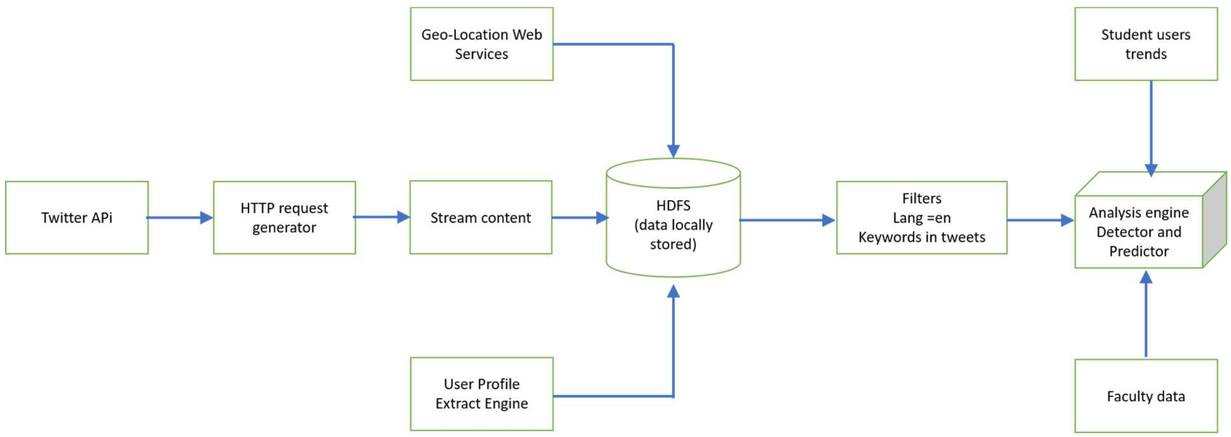

Figure 1. Workflow

CHAPTER II

METHODOLOGY

In this work, large amounts of Twitter data is mined to characterize the perceptions of faculty with open source tools. There are two main objectives: first, to characterize the topics currently taking place in twitter with respect to faculty and second to focus on student complaints regarding faculty. To this end, a classifier was developed that can be used to identify and pull out the student users based on their tweets. A novel aspect of this work is using historical twitter data. In other words, a user is not classified based on one tweet but on a collection of his or her tweets (i.e., a profile was created for each user). Keyword combinations were used to identify a seed set of Twitter users who self-identified as students. This data was used to create a classifier to identify a large set of Twitter users (i.e., more than 1.5 million from approximately 10 million) that mentioned faculty or coursework. These classifications were made on word usage and not self-identification. Various tools were used to handle the very large amount of data and also to process it. Of particular use was Hadoop, Python, and Scala. With the probable students identified, tweets were randomly selected and analyzed to identify categories of complaints against faculty. Once the categories were established, then common keywords for each category were identified and relative frequency of these keywords were used to characterize the prevalence of these complaints.

Data Collection, Format, and Filtering

To collect tweets, we used Apache Flume with the Twitter Application Programming Interface (API). The Twitter data related to faculty was collected over several months during the fall of 2015 and the historical Twitter data was being collected for a period of approximately nine months during 2014 and 2015. The total dataset was approximately 4.7 TB in size. The data we collected is in native JSON format as Twitter API provides. The JSON object contains 58 fields out of which we are considering three fields i.e. user description, user id and text. This process was done making use of custom Python and Scala scripts.

Due to the size of the dataset, a distributed framework was needed to store and process the data. Hadoop was used for this purpose. Hadoop is a software suite capable of distributing processing of large datasets over clusters of computers using simple programming models. It can be scaled from a single server to thousands of machines. All the Twitter data collected is stored in the Hadoop Distributed File System (HDFS) and Hadoop MapReduce or Spark Jobs are used to process the tweets. Hadoop MapReduce is used for parallel processing of large data sets, which is a YARN-based system. Hadoop YARN is a job scheduling and cluster resource management framework (which can support other frameworks such as Spark). The Hadoop cluster employed for this work contains four machines, 60 cores, 64 GB memory, and 28 TB of storage. With this cluster, it was possible to process several TBs of Twitter data in a few hours.

To aid in exploration and understanding of the data, several filters were applied. We developed Python scripts to filter the tweets with the keywords like professor, student, prof, school, college, university, staff, instructor, etc. In this way, we filtered all the tweets to retain only those with these key words. We also tried to pull out which could pertain to a particular course (i.e., the tweets regarding course numbers which are combination of both alphabets and numbers). To do that we used a regular expression of 2-4 alphabets and 3-4 numeric values to pull out those tweets along with the keywords. When looking into the filtered data, there were irrelevant tweets in them. To remove such tweets, we have done stemming and used the high frequency words and pairs. This process will be discussed in detail in subsequent subsections.

Classification of Twitter Users as Student/Non-student

In order to build a classifier, labeled data is needed (i.e., data with categories). For this work, the labels that are needed are ‘student’ and ‘non-student’. Twitter users are not required to report student status; thus, it is difficult to obtain this information with certainty. As a result, heuristics were developed to label a Twitter user based on self-reported statements (e.g., a Twitter users makes use of the phrase ‘my prof’). The user’s profile was also checked for self-identification which could identify him or her as a student. For this, the student description was taken as the source and keywords like student, stdnt, etc. were used. Both manual and automated approaches were developed for these processes. We have manually identified 80+ users when we went through 10,000 tweets. When we ran the Python script over the test dataset of 200 GB we identified around 195,000 users. The users’ ids for these probable students were saved.

In order to ensure the widest possible reach of this analysis, historical data was used to create a profile for each Twitter user. The rationale being that word usage could characterize a user and identify additional student users which had not self-identified but nonetheless said something about faculty. Thus, to characterize a Twitter user, a number of tweets were used to build a user profile. These tweets came from a collection of some 1.2 billion tweets that had been collected over a 9-month period from the Twitter ‘gardenhose’ (i.e., a %1 sample of all tweets released in real-time via the Twitter API). Script were developed to pull out all the tweets making use of a list of user ids and by pulling out all the tweets of the tweets for those users.

Once grouped by user, the content of the tweets was filtered and stemmed. First, the tweets were filtered to only retain those which were reported to be in English. For the filtering, some characters and words were removed as many of the tweets contained URLs and common words such as ‘I’, ‘am’, ‘I’m’, ‘can’, ‘hi’, and other commonly used conjunctions. Emoticons were also removed and irrelevant to our study. When we were done with removing all the stop words and URLs, the result was then stemmed. This process reduces a word down to its root meaning (i.e., stemming means reducing modulated words into their stem or root word). In this way, we stemmed the root words and then paired them with all the stem words of that tweet. Later we made the word-pair count to see which words were more often used with our keywords. We ignored the common English words like a, the, are, about, from, for, etc. We also considered the other words that in general are used often in all tweets (i.e., not only those of interest for this study). From these lists of commonly used stems, the top 6,000 were used to create "bags". These bags represent stems commonly used by both students and non-students and their presence or absence in a user's profile characterizes that user.

college	967796	good	42261	friends	16985
course	881751	faculty	42257	already	16835
~~a~~	~~712121~~	~~b~~	~~41886~~	business	16356
~~for~~	~~514206~~	~~because~~	~~41622~~	home	15909
~~and~~	~~460363~~	action	40743	during	15547
~~at~~	~~261477~~	back	39564	again	15324
~~have~~	~~191305~~	class	38636	community	15281
~~be~~	~~187478~~	best	38155	check	15233
~~but~~	~~158944~~	~~her~~	~~38062~~	chemistry	15200
~~amp~~	~~141110~~	considering	36726	career	15099
~~are~~	~~130068~~	feelings	36010	does	15084
~~from~~	~~127512~~	easiest	35796	famnwhatd	14996
~~dont~~	~~111560~~	great	35332	another	14872
~~all~~	~~110461~~	expressing	35323	actually	14817
~~do~~	~~101119~~	aries	35309	biology	14736
~~about~~	~~96971~~	basketball	34367	grocery	14687

Figure 2.Sample of Word Count on Faculty Dataset

Feature vectors were created for every user. This is done by using the profiles created for both student and non-students and looking at word stem usage. A total of 6000 features, including 3000 features from the students and 3000 features from the non-student users were used to create the feature vector. The feature vectors were then fed into a machine learning tool to initially train a classifier and then to apply the classifier to evaluate all users. A Python script was developed to create the feature vector as a series of 0's or 1's for a user. If a particular word was present in the user profile, then its position in the feature list was set to 1 or else 0. Once provided the user profiles for all users as the input to this script, it created the feature vector for every user. Once the feature vector was created for all of the student users and a roughly equal size sample from the non-student user pool, we created a labeled dataset with the last place of the feature vector. If the last one is 1 then the user is a student and if it is 0 then he is non-student. In this way, a labeled dataset was constructed for training our model.

A decision tree is a tree like structure which includes different root nodes and leaf nodes to illustrate every possible result. They have been used in data mining to simplify complex challenges and evaluate business decisions. A Decision tree is also used in machine learning (Safavian & Landgrebe, 1990). Traditionally the path from the root to leaf nodes represents classification rules. Given sets of labeled data, rules decision trees of specified depths and breadths can be generated to best classify the training data using the provided features.

Random forests, also known as random decision forests, is a machine learning technique to perform classification, regression, and other tasks (Pal, 2005). The random forest is used to find the nearest neighbor in data mining and machine learning and capable of performing classification. A random forest consists of an arbitrary number of simple trees, which are used to determine the final outcome. Random forests can be used as an algorithm in predictive models. Data assumptions with this algorithm are very low, so data preparation is less challenging which saves time for the user.

The data set for evaluating our model is prepared by creating the feature list of the data we require. Once the feature lists are re-generated with a label we use it as input for our model. First, we split the dataset in to three subsets: training, validation, and test. Eighty percent of the data is used for training our model and 10% each for cross-validation and testing. The test dataset is used to produce a final, unbiased evaluation of the expected accuracy of a model. The validation dataset is used for adjusting the parameters of the machine learning algorithm to the particular problem at hand and selecting the optimal values.

Tuning the model (i.e., decision tree) is one of the most important tasks when developing a classifier. For this purpose, we are using the hyperparameters to tune our model. This is done by choosing different values and building models and assessing the quality of the results. The parameters that control the tree's decisions are maximum depth, maximum bins, and impurity measure. Maximum depth limits the number of levels in the decision tree, and it is the maximum numbered of decisions that a classifier will make to classify the data. The decision rules are responsible for the decision tree algorithm to try at every level and are known as bins. The larger the number of bins, there is a chance of finding a more optimal solution for complex classification tasks but also increases the chance of overfitting. There are two commonly used measures of impurity: Gini impurity and entropy. Gini impurity is directly related to the accuracy of the random-guess classifier. Entropy captures how much uncertainty the collection of target values in the subset contains. We have tried various values to build the models and the results are analyzed to select the best combination of them.

Tools Employed

To aid with managing the large amounts of data in this project, an automated pipeline was created to curate the user profile and perform classification. MapReduce and Spark jobs were created to curate the user profile. After creating the profiles, they were filtered using shell scripts and then processing continued making use of the Hadoop and Spark platforms. Using the MLlib for Spark, it was possible to create and apply a classifier using the machine learning techniques like decision tree and random forest in a distributed fashion.

Hadoop is a framework that allows for the distributed processing of large data sets across clusters of computers using easy programming models ("Apache Hadoop," 2017). It is designed to scale from a single server to thousands of machines, with parallel computation and storage built in Instead of place confidence in hardware to deliver high-availability, the library itself is meant to find and handle failures at the applying layer, therefore delivering a highly-available service on prime of a cluster of computers, each of which can be vulnerable to failures. Hadoop includes four modules: Hadoop Common, Hadoop Distributed File System (HDFS), Hadoop YARN, and Hadoop MapReduce.

The origins of Hadoop came from the Google File System paper which was published in October 2003 ("Apache Hadoop," 2017). Hadoop version 0.1.0 was released in April 2006. Hadoop can be deployed in a traditional onsite datacenter and in the Cloud. Currently cloud based options are being offered by Microsoft, Amazon, IBM, Google, and Oracle. Yahoo! Search Webmap is a Hadoop application that runs on a Linux Cluster with more than 10,000 cores and produces data that was used in every Yahoo! Web search query. In 2010, Facebook claimed that they had the largest Hadoop cluster with 21 PB of storage, later upgraded to 100 PB in June 2012.

While Hadoop laid the ground work for large scale data processing, it has its limitations when it comes to processing large amounts of data in a repetitive manner. Apache Spark was designed to overcome these shortcomings and is a fast and general-purpose cluster automatic data processing system (Ryza, Laserson, Owen, & Wills, 2015). It provides high-level APIs in Java, Scala, Python and R, and an optimized engine that supports general execution graphs. It additionally supports an upscale set of higher-level tools such as Spark SQL for SQL and structured processing, MLlib for machine learning, GraphX for graph processing, and Spark Streaming.

MLlib is a machine learning library. It provides tools like ML algorithms, featurization, pipelines, persistence, and utilities ("MLlib: Main Guide - Spark 2.1.0 Documentation," n.d.). ML algorithms includes common machine learning algorithms such as classification, regression, clustering, and collaborative filtering. In featurization, tasks provided include feature extraction, transformation, dimensionality reduction, and selection. Pipelines are the tools for constructing, evaluating, and tuning ML applications and persistence is used for saving and load algorithms, models, and pipelines. In additional to these higher-level functions, utilities (i.e., underlying functionality used by and made available through MLlib) include linear algebra, statistics, data handling, etc. MLlib supports both the RDD-based API and the DataFrame-based API. Spark, via MLlib, supports decision trees for classification and for regression, using both categorical and continuous features. The implementations partition data by rows, allowing distributed training with millions or even billions of instances ("MLlib: Main Guide - Spark 2.1.0 Documentation," n.d.).

Manual Processing of Student Tweets

Manual inspection of all the tweets is not possible and as a result only a sample of the Twitter data was inspected manually. During this inspection, some of the student tweets were collected (i.e., 1500 tweets) that contained some sort of complaint about their professor. When processed, they were found to be discussing the professors' behavior, professional values, and personal characteristics. In this way, there are many categories of complaints regarding professors. Some of them are coming late to class, giving difficult exams, speaking on irrelevant topics for the course, or about simply expressing their feelings about professors regarding their classroom behavior.

As an example, when sifting through the tweets discussing the professor coming late to class, it was discovered that some students were tweeting about themselves as late to the class. So, first the keyword "late" was considered to see how many times students were using it. There were about 41,900 times "late" was repeated and as a result we used the following keywords to find their count in the probable student profiles, "late", "I am late", "I'm late", "professor late", "late to class". All these words were considered to see the usage of the students from their tweets

CHAPTER III

RESULTS

The principle results of this work are the classifier (i.e., methods to determine if a twitter user is a student or non-student from a collection of tweets) and exploration of tweets by users classified as students.

Classifier

Training and validation data was created for the classification task and several decision trees were trained and evaluated. For this purpose, many parameter values were used to build our classifier and look at the results to see the ideal combinations of the values. This is to say that many hyperparameter combinations were tried to tune our model and get the best results. The table shows the accuracy and area under curve for the different values for impurity, depth of the tree and number of bins. These were the hyperparameters used to tune our model. The results reported are for a validation set.

Table 1. Accuracy and Area Under Curve for Various Parameters

Impurity	Max. Depth	Max. no. of BINS	Accuracy	Area Under Curve
gini	1	100	0.784390817	0.757468131
entropy	1	100	0.784390817	0.757468131
gini	2	100	0.805111171	0.810055609
entropy	2	100	0.805111171	0.810055609
gini	4	100	0.834734273	0.833858234
entropy	4	100	0.834711678	0.830397804
gini	10	100	0.860222343	0.852136438
entropy	10	100	0.855545011	0.845363521
gini	20	100	0.864990058	0.858553727
entropy	20	100	0.862866052	0.855680353
gini	30	100	0.8577594	0.851612213
entropy	30	100	0.856335864	0.850150985

A random forest model was also constructed using the best combination of hyper parameters obtained on a decision tree. Using a random forest was a way to further improve the highest accuracy to build our model and train the classifier. A random forest with 40 trees using gini, maximum depth of 30 and 300 bins achieved an accuracy of 0.867 and an AUC of 0.864. In this way, we trained our model and predicted probable students from the entire faculty dataset. It approximately predicted about 1.2 million users as students. This represents a many fold increase in the number of student users who self-identified. Thus, the classifier was able to increase the reach of the analysis.

Application of Classifier and Results of Manual and Semi-Automated Analysis of Tweets

During manual processing of some 1500 tweets of the students, there were several students tweeting about their professors and many complaints about them. When we out listed these complaints, they were mainly concerning the professors' behaviors. The complaints were categorized based on the content. For each category, we have selected certain keywords to identify that category. These keywords are used to give an indication as to how many times these complaints were used by the students in their tweets. In this way, we came to know how students perceived faculty through the categories of student complaints about their professors. Here four types of complaints by the students are considered. They are:

First category of the complaints deals with the professor's punctuality and also student's tweeting themselves coming late to the class. The table shows the keywords used and their frequency in the profiles of the predicted students.

Table 2. Keywords for Complaints Category I

keyword	freq. (prob-stud)
late	41900
I am late	13
I'm late	353
professor late	34
late to class	290

Another category of student complaints concerns their tests, assignments, and grades. Some students are tweeting about the assignments that were to be submitted. Students tweeting about the fairness of the professor while giving tests and assignments. They were discussing their professor’s way of evaluating tests and complaining about their grades they getting in the course. The table shows some of the keywords.

Table 3. Keywords for Complaints Category II

keyword	freq. (prob-stud)
test	48206
hard test	7
worst test	3
assignments	12355
grade	17680
exam	18958
bad grade	167
good grade	1131

Moving to the next category of the complaints, tweets were about professors canceling the classes. People were thinking about how the instructors cancel classes at the last moment. Students expressed feelings about their professors annulling the class. The table below shows the keywords and their occurrences.

Table 4. Keywords for Complaints Category III

keyword	freq. (prob-stud)
cancel class	627
canceled	2402
no class today	23
cancelled class	532

The last category of complaints, vulgar words being used by the users. Nowadays students are showed rude nature through their tweets about the professors and used bad words while tweeting about their professors.

Other than these complaints, students were also tweeting about their communication with their professors outside the class. For this type of tweet, we have identified some keywords and retrieved the count from the predicted students.

Table 5. Keywords for Communication

keyword	**freq. (prob-stud)**
reply	2120
email	12906
Office hours	711
meeting	6296

CHAPTER IV

DISCUSSION

This study offers many avenues for discussion and future research on the feasibility of social networking, specifically Twitter. Here, we address many of these findings and uses the user tweets to identify students and clarify some of the interests they are discussing about the faculty by this thesis.

To identify the students from the raw data, we had one of the greatest challenges we faced during our work. In the previous studies, a specific number of students and teachers who are using social network were selected and the analysis was based on these students which could lead to biased results. Survey methods also have the challenge in that respondents know they or their responses are being observed. To overcome this, we collected Twitter data from random and selected users for over a 9-month period which is ~1% of the tweets that were produced in that time frame. This amount of data and the fact that users do not expect to be monitored give an arguably more unbiased result.

Given the performance of the constructed classifier, i.e. 86% of the multi-class predictions were correct and reasonable values for the AUC values for the binary classification tasks. It is possible to effectively access larger amounts of tweets in a manageable fashion and predict the student users. It is possible to peruse tweets by category more efficiently by varying the decision threshold. With the advantage of the classic tradeoff between precision and recall, we can obtain less numbers of confident predictions or many less confident predictions.

The surveys done so far have selected a limited number of students and professors, and their usage of social networking is observed and conclusions were made which do not necessarily mirror the more general public view. In our study, we have selected the real-time data from Twitter from all the people instead of randomly sampled individuals. We have identified the students from the data we had which arguably gives a more diverse result. If we consider the work of Mc Arthur et al., (2012) they performed a survey with 144 students and 3 instructors which gave a very limited scope to see the relationship between students and faculty. Whereas, we have predicted about 1.2 million students based on their tweets and gone through them to see the real-world student perceptions' of faculty. Until now people have evaluated relationships based on some predefined criteria (i.e. feedback from students). But in this study, we are manually inspecting the tweets from the students that our classifier predicted. This makes a broad way for people to know what students are actually thinking about and the faculty in their views without any prejudice. When manually reviewing the tweets, we found some similar criteria with the previous studies like the classroom behaviors of the faculty, etc. Beside these qualities, we also observed the personal views of the students based on their interaction with their faculty both inside and outside the class.

Overall, the studies that took place until now to identify student perceptions' of faculty were not fully convincing because of their biased results that are based on the surveys with very small numbers of students. The classifier we built predicted approximately 1.2 million users as students and then we manually went through 1500 tweets of these students to get more precision in our results.

In the future, when trying to identify the students and their perceptions, the entire sentence could be considered instead of the keywords in the tweets. By considering the entire tweet and applying the natural language processing or latent concept mining, it could show a greater effect in finding the students' perceptions. Also, of interest would be the associations which could be identified by "mentions" in tweets. By identifying groups of interacting Twitter users, the types of complaints within groups could be examined to see if likeminded individuals are simply echoing similar concerns or if there is a variety of concerns voiced within a group.

CHAPTER V

CONCLUSION

This study developed and implemented a means to characterize the users on Twitter as students or non-students and then examine students' perceptions of the faculty. Using tools such as Flume, Hadoop, Spark, and machine learning techniques, it is possible to collect and process large amounts of Twitter data to see what words, or phrases were dominating the online discussion. In constructing a classifier to predict the student status of a Twitter user, historical data (i.e., all of the tweets for a particular user) were used to build a user profile. In manually examining the Tweets, three salient categories of student complaints were identified and keywords for each category were determined. These keywords were used as a crude measure of the prevalence of each type of complaint.

REFERENCES

Basow, S. A., Phelan, J. E., & Capotosto, L. (2006). GENDER PATTERNS IN COLLEGE STUDENTS'CHOICES OF THEIR BEST AND WORST PROFESSORS. *Psychology of Women Quarterly*, *30*(1), 25–35.

Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)* (Vol. 6, p. 12). Retrieved from http://gmagno.net/papers/ceas2010_benevenuto_twitterspam.pdf

Bergsma, S., Dredze, M., Van Durme, B., Wilson, T., & Yarowsky, D. (2013). Broadly Improving User Classification via Communication-Based Name and Location Clustering on Twitter. In *HLT-NAACL* (pp. 1010-1019).

Feldman, K. A. (1993). College students' views of male and female college teachers: Part II—Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, *34*(2), 151–211.

Hewitt, A., & Forte, A. (2006). Crossing boundaries: Identity management and student/faculty relationships on the Facebook. *Poster Presented at CSCW, Banff, Alberta*, 1–2.

Kendra, R. L., Karki, S., Eickholt, J. L., & Gandy, L. (2015). Characterizing the Discussion of Antibiotics in the Twittersphere: What is the Bigger Picture? Journal of Medical Internet Research, 17(6), e154. http://doi.org/10.2196/jmir.4220

Kowai-Bell, N., Guadagno, R. E., Little, T., Preiss, N., & Hensley, R. (2011). Rate My Expectations: How online evaluations of professors impact students' perceived control. *Computers in Human Behavior*, *27*(5), 1862–1867. https://doi.org/10.1016/j.chb.2011.04.009

Lin, M.-F. G., Hoffman, E. S., & Borengasser, C. (2013). Is Social Media Too Social for Class? A Case Study of Twitter Use. TechTrends, 57(2), 39–45. http://doi.org/10.1007/s11528-013-0644-2

Machine Learning Library (MLlib) Programming Guide - *Spark 1.2.0 Documentation - https://spark.apache.org/docs/1.2.0/mllib-guide.html*

Marín, V. I., & Tur, G. (2014). Student Teachers' Attitude towards Twitter for Educational Aims. Open Praxis, 6(3). http://doi.org/10.5944/openpraxis.6.3.125

Martins, J., Goncalves, R., Branco, F., & Peixoto, C. (2015). Social networks sites adoption for education: A global perspective on the phenomenon through a literature review. In 2015 10th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1–7). http://doi.org/10.1109/CISTI.2015.7170490

McArthur, J. A., & Bostedo-Conway, K. (2012). Exploring the Relationship between Student-Instructor Interaction on Twitter and Student Perceptions of Teacher Behaviors. International Journal of Teaching and Learning in Higher Education, 24(3), 286–292.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217-222.

Pennacchiotti, M., & Popescu, A. M. (2011). A Machine Learning Approach to Twitter User Classification. *ICWSM*, *11*(1), 281-288.

Ryza, S., Laserson, U., Owen, S., & Wills, J. (2015). *Advanced analytics with Spark* (First edition). Beijing ; Sebastopol, CA: O'Reilly.

Safavian, S. R., & Landgrebe, D. (1990). A survey of decision tree classifier methodology.

Street, 1615 L., NW, Washington, S. 800, & Inquiries, D. 20036 202 419 4300 | M. 202 419 4349 | F. 202 419 4372 | M. (2013, December 27). Social Networking Fact Sheet. Retrieved from http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

Tunca, N., Şahin, S. A., Oğuz, A., & Bahar Güner, H. Ö. (2015). Qualities of Ideal Teacher Educators. *Turkish Online Journal of Qualitative Inquiry*, *6*(2), 122. https://doi.org/10.17569/tojqi.48192

Apache Hadoop. (2017, February 15). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Apache_Hadoop&oldid=765691738

Wright, N. (2010). Twittering in teacher education: reflecting on practicum experiences. Open Learning: The Journal of Open and Distance Learning, 25(3), 259–265. http://doi.org/10.1080/02680513.2010.512102

Yang, M.-C., & Rim, H.-C. (2014). Identifying interesting Twitter contents using topical analysis. Expert Systems with Applications, 41(9), 4330–4336. http://doi.org/10.1016/j.eswa.2013.12.051