Teaching Case (Instructor Notes)

An Experiential Learning Project using Sentiment Analysis of Twitter Posts

Joel Asay
asayj@xavier.edu

Elaine Crable
crable@xavier.edu

Mark Sena
sena@xavier.edu

Business Anaytics and Information Systems Department
Xavier University
Cincinnati, OH 45207 USA

Hook

What's the word on the tweet? This project enables students to collect tweets about topics of their choice and use sentiment analysis to measure the current perspectives from the Twittersphere about those topics.

Abstract

In this teaching case, we describe an experiential learning project that allows students to perform sentiment analysis on a set of tweets (posts made on the social media platform, Twitter) by collecting and analyzing posts that include key words selected by the students. Sentiment analysis refers to the process of identifying and categorizing opinions expressed in a piece of text. The project requires students to make edits to an R script, execute the script to save a collection of tweets that contain specific keywords, then open the file and paste the results into a macro-enabled Excel file that is provided. Students then edit the dataset to cleanse the data and write a report to interpret the findings. The assignment requires only a cursory knowledge of programming and Excel. We assign the project to students taking an introductory information systems course but the project could be suitable for courses in business analytics, marketing, social media, computer science, and other subjects.

Keywords: experiential learning, analytics, sentiment analysis, twitter

1. INTRODUCTION

Students enjoy coursework that is experiential, engaging, and relevant to their interests. Kolb and Kolb (2005) are known for their theory of experiential learning which includes learning as a whole process. In this study, we describe a project that allows students to perform sentiment analysis on a set of tweets (posts made on the social media platform, Twitter) by collecting and analyzing posts that include key words selected by the students. The project enables students to perform real-time analytics while providing exposure to a programming language without having to learn in-depth coding skills. This project allows students to grasp new concepts, to be

creative and reflective resulting in a much deeper level of learning due to the experiential format of the assignment.

2. OVERVIEW OF SENTIMENT ANALYSIS

Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information (Medium.com, 2021). Sentiment analysis models have been developed and tested to estimate the extent to which a string of text is positive or negative and to quantify the emotions that are expressed. Sentiment analyses have been applied to customer reviews, survey responses, social media, and other data sources across various fields of study. For example, one study demonstrated how the analysis of Twitter sentiment was closely correlated to a Gallup poll of public opinion (O'Connor, Balasubramanyan, and Routledge, 2010). Another study showed how the moods depicted in tweets can predict stock market trends (Bollen, Mao, and Zeng, 2011).

In this project, three different methods of analyzing sentiment are provided to students. The AFinn model uses a list of English words rated on a scale between -5 (negative) and +5 (positive). The model was developed by Finn Årup Nielsen between 2009 and 2011 (Nielsen, 2011). The Bing index (Liu & Hu, 2004) model assigns selected words as either negative (-1) or positive (+1). Applied to Twitter entries, each word in a tweet string is tabulated to determine the net positive or negative score. The NRC model (Mohammad & Turney, 2014) was coordinated by the National Research Council of Canada. Its model categorizes English words in alignment with eight emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. When applied to Twitter entries, each tweet's word count for each emotion is tabulated. The emotions are then compared to a total to determine the relative percentage of each emotion that is found in the set of extracted tweets.

3. PROJECT DESCRIPTION

Pre-Project Requirements

To search and extract the tweets that contain the selected keywords into a dataset, students need to register for a (or use a provided) development account on Twitter that allows access to an open API. Students must also either register for a cloud based R-studio account or students can download R and R Studio to their computer or use applications installed on lab or classroom computers. We give students a short introduction to R, including an example that edits the script used in the assignment. Similar open-source scripts are available for use in Python if instructors prefer using Python instead of R. Our students have prior instruction using Excel prior to starting the assignment. The assignment timeline can be expected to last approximately one week, given the possible delay in approval for the Twitter development account and the time to give students a brief introduction to programming in R.

Editing the R Script

One student accounts are approved, students then open and edit an R script that installs the rtweet open source package and library. The required edits include editing search terms, maximum number of records requested, variable names, and file names. Note that the script in Figure 1 represents a simple usage of the r-tweet application. Many additional arguments and options can be use with additional code to filter or refine the extracted set of tweets. (Kearney, 2019)

In our assignment, students select keywords that represent a particular brand, product, person, or topic then select two additional related keywords to make comparisons. For example, students could compare a specific style of Nike shoe with competing brands offered by Adidas or Under Armour. Students could also select keywords to compare sentiment towards politicians, celebrities, sports teams, or other topics.

Using Excel for Sentiment Analyses

After successful execution of the code, files will be generated that store the collection of tweets in csv (comma separate values) formatted files that could then be opened in Excel or read by a programming language to score the tweets in accordance with the sentiment models. The resulting csv file contains the text of each tweet in separate rows, along with the ninety other columns of data, including the screen name, date and time, location, number of retweets, and various other attributes. For this assignment, students can filter the csv file manually or by creating a table. Figure 2 shows a partial screen shot showing extracted tweets in the csv file with a table inserted to allow filtering of records.

```
# To collect tweets, pass a search criteria to the
# search_tweets() function and assign the output
# as a new R object.

# Define a new data object based on a Twitter
# search. The "n=200" augment is used to
# specify the number of tweets requested.
rtweet_xavier<-search_tweets("xavier
University", n = 200)

# Many other arguments exist, including
restricting language to only English...
rtweet_xavier_en <- search_tweets("Xavier
University", n = 200, lang = 'en')

# ...or you can opt to not include retweets:
rtweet_xavier_en_nort <- search_tweets("Xavier
University", n=200,include_rts = FALSE)

# You can learn about the many other possible
arguments by viewing the help:
help("search_tweets")
```

Figure 1: Sample R-Tweet Script

screen_name	text
mha_online	Online MSN in Nursing Administration Programs Ranked b
ShayTAllenESQ	Support a fantastic event on July 24th! Proceeds go towar
xuecosoc	With this, we welcome you to Xavier University Economics
gageywagey15	@DTV89 I once went to a party at Xavier. Can confirm, that
WyattEarp1969	@daraghson_ Welcome to the Xavier family, young man.
WindsorMBB	Xavier is a transfer from Dalhousie University. There he w
XULAalumni	The Alumni would like to formally introduce Mr. Phillip Ac
juellieana	xavier university see u next year
unitassateneo	The Xavier University United Arts and Sciences and Studer
MindanaoGSD	THAT question continues to loiter in the minds of Xavier U
XU_JMAofficial	The Junior Marketing Association of Xavier University is op
PaigeThomas_30	SO excited to announce my verbal commitment to St.Xavi
ericcrawford	CONGRATS to St. Xavier product Trey Sweeney, of Eastern
andypic17	The Cleveland Indians are selecting Naji Marshall, LF, Xavi
IAMKPSmith	Xavier University is coming up on #CollegeBowl.

Figure 2: Excel CSV File with Table

In our project, once students finish filtering, they copy and paste the data from the "text" column into a separate macro-enabled Excel file with embedded Visual Basic code that performs the sentiment analyses and summarizes the results. Figure 3 shows a partial screen shot of the sentiment analysis Excel file. The image shows the summary Afinn and Bing scores for the tweets and the first four (of eight) of the emotion scores from the NRC model. In order to compare the sentiment analysis results of multiple search terms, the processes shown above would need to be repeated or code from the R-script could be added to produce multiple csv files along with the need to copy and paste the text results into additional worksheets.

Average Tweet Score	1.15	0.96				
Total Subject Score	26.34	22.00	5%	26%	0%	5%
Text	Afinn	Bing	Anger	Anticipation	Disgust	Fear
Happy 4th of July, Musketeers!						
Wishing you a fun and safe holiday weekend.	2.67	3	0	3	0	0
Let's all give a big Xavier welcome to @PrezHanycz as she celebrates three big milestones today!						
Canada Day	1.75	1	0	0	0	0
Xavier is strongly encouraging our community to receive a #COVID19 vaccine this summer.<U+2063>						
Students who update their records this summer will be entered into a special	2.5	2	0	2	0	0

Figure 3: Excel File with Sentiment Analysis Results

Although we do not teach students the Visual Basic code that computes the sentiment scores for each model, students are shown how to enable the Developer menu in Excel and shown how to open and view the code. Note that the macro-enabled Excel file was developed in-house by one of this study's authors for teaching purposes. Sentiment analysis code can also be deployed in other programming languages but may be more complex for students and instructors to implement. Figure 4 shows a small portion of the code used to compute the scores summarized in the worksheet.

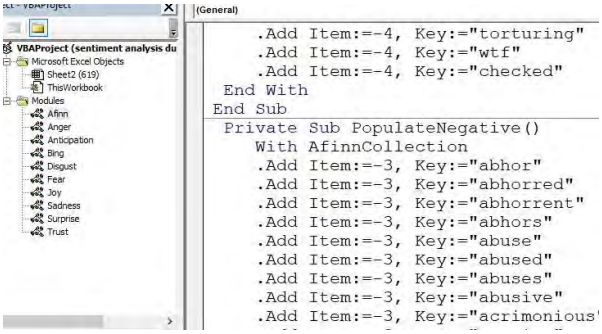

Figure 4: Excel Visual Basic Code to Compute Sentiment Analysis Scores

Limitations of Twitter for Sentiment Analysis

Students should be informed that sentiment analyses simply examine the words contained in posts and compares them against a set of words and their respective scores in accordance with each model. The process can be imperfect and can inaccurately score individual posts depending on the context of the use of the words contained in the post. Students should try to be specific in their search terms and be aware that some terms may be used in differing contexts. For example, if students chose "Google" as a brand to evaluate, tweets using google as a verb would also be captured. Since the assignment uses tweets, students should be aware that re-tweets and tweets generated by bots can cause records of the same tweet to be represented multiple times in the dataset which may impact the accuracy of

the overall sentiment scores. Twitter is also used as a marketing platform and not all posts represent opinions on a topic. For example, there may be a series of posts referencing the Nike shoe brand that are simply links to eBay product offerings. The extracted tweets might also include posts that are not in English or that contain images or other media that cannot be interpreted by the sentiment analysis code. Depending on the purpose of the sentiment analyses (a real-life consulting project, graduate thesis, etc.), time-consuming filtering may be needed to cleanse the data. As an alternative to this assignment, students could collect data from other sources (such as Yelp, TripAdvisor, Amazon, etc.) that contain text that can be scraped using programming language scripts or manually copied and pasted into the Excel file.

Final Report and Project Deliverables

We feel it is important for students to write a short report that summarizes their findings. While students may not fully comprehend the R code that extracts the data nor the VBA code that scores each record using the three sentiment analysis models, students should convey a big-picture understanding of the process and be able to interpret their findings along with any limitations or problematic records in their datasets. We require students to limit their report to a one-page document that summarizes and compares the average sentiment scores for each search term, compares significant emotion percentage differences between the terms, and reflects on the tweets collected and the reasons for the resulting scores. In terms of deliverables, students are required to turn in their R code, their excel workbook with separate worksheets for each search term, and the written report. We typically allow students to work in groups of three on the assignment, offer peer evaluation and provide students with a grading rubric that accompanies the project instructions.

4. CONCLUSIONS

This experiential sentiment exercise was assigned to undergraduate business students taking their first information systems course. Upon conclusion of the assignment, it is clear to us that a hands-on learning exercise such as this can provide a superior learning experience for students that transcends the typical activities found in a beginning IS or Computer Science course. As Kolb and Kolb (2005: 43-44) repeatedly discuss in their research, "Learning is a holistic process of adaptation. It is not just the result of cognition but involves the integrated functioning of the total person – thinking, feeling, perceiving and behaving." This activity successfully meets each of these experiential requirements.

On a separate Word document, summarize your findings on the three search terms in one page or less. Be sure to comment on:

- Average tweet score for Afinn and Bing sentiments. Is it positive or negative? Judging by some of the tweets and what you know of the terms, why do you think this might be the case?
- How do the emotions compare between search terms? Are they consistent or do they have different profiles?

Submit the following under the "R Twitter Project" assignment link:

- Your R code for extracting the data from Twitter using the rtweet package
- A sentiment analysis workbook with three sheets calculating sentiment (one for each search term)
- The final Word document summarizing your findings

Figure 5: Final Report and Project Deliverables

By the end of the exercise, we observed the following of our introductory students:

- Students initially perceived the assignment to be very intimidating. Most likely, this was due to their inexperience with programming and basic data analysis. Post-assignment, in almost all cases students indicated it was much easier than they imagined. • Students were surprised to find that results from analyses were easily swayed by when the data was collected and/or by the keywords used.
- Students found that sometimes the results didn't make sense because sentiment analysis is not perfect.
- At first, many students struggled to come up with things to compare because they were not familiar with how sentiment analysis tools work. After performing the tasks, they were able to perceive the value in the tools and could imagine easily applying to a variety of situations.
- The majority of students indicated that this was a fun and practical activity. They

were impressed with the simplicity and yet the power of this analytical method.

Experiential learning is a significant trend in higher education. It boosts students' involvement with the learning process. We believe employing this novel sentiment analysis exercise as a participatory experience enhanced both the students' ability to absorb the concepts as well as their retention of the knowledge to a significant degree.

5. REFERENCES

Bollen, J., Mao, H., & Zeng, X. (2010). Twitter mood predicts the stock market. *Journal of Computational Science,* 2(1), 1-8.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Kearney, M. (2019). rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software,* 4(42), 1829.

Kolb, A., & Kolb, D. (2005). Learning styles and learning spaces: Enhancing experiental learning in higher education. *Academy of Management Learning and Education,* 4(2), 193-212.

Mohammad, S., & Turney, P. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence,* 29 (3), 436-465.

Nielsen, F. Å. (2011) A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages,* 93-98.

O'Connor, B., Balasubramanyan, R., & Routledge, B. (2010). From Tweets to Polls: Linking text sentiment to public opinion time series. *Proceedings of the International AAAI Conference on Weblogs and Social Media,* 122-12.

APPENDICIES

```
# To collect tweets, pass a search criteria to the
# search_tweets() function and assign the output
# as a new R object.

# Define a new data object based on a Twitter
# search. The "n=200" augment is used to
# specify the number of tweets requested.
rtweet_xavier<-search_tweets("xavier University", n = 200)

# Many other arguments exist, including restricting language to only English...
rtweet_xavier_en <- search_tweets("Xavier University", n = 200, lang = 'en')

# ...or you can opt to not include retweets:
rtweet_xavier_en_nort <- search_tweets("Xavier University", n=200,include_rts = FALSE)

# You can learn about the many other possible arguments by viewing the help:
help("search_tweets")
```

Figure 1: Sample R-Tweet Script

screen_name	text
mha_online	Online MSN in Nursing Administration Programs Ranked b
ShayTAllenESQ	Support a fantastic event on July 24th! Proceeds go towar
xuecosoc	With this, we welcome you to Xavier University Economics
gageywagey15	@DTV89 I once went to a party at Xavier. Can confirm, that
WyattEarp1969	@daraghson_ Welcome to the Xavier family, young man.
WindsorMBB	Xavier is a transfer from Dalhousie University. There he w
XULAalumni	The Alumni would like to formally introduce Mr. Phillip Ac
juellieana	xavier university see u next year
unitassateneo	The Xavier University United Arts and Sciences and Studer
MindanaoGSD	THAT question continues to loiter in the minds of Xavier U
XU_JMAofficial	The Junior Marketing Association of Xavier University is op
PaigeThomas_30	SO excited to announce my verbal commitment to St.Xavi
ericcrawford	CONGRATS to St. Xavier product Trey Sweeney, of Eastern
andypic17	The Cleveland Indians are selecting Naji Marshall, LF, Xavi
IAMKPSmith	Xavier University is coming up on #CollegeBowl.

Figure 2: Excel CSV File with Table

Average Tweet Score	1.15	0.96				
Total Subject Score	26.34	22.00	5%	26%	0%	5%

Text	Afinn	Bing	Anger	Anticipation	Disgust	Fear
Happy 4th of July, Musketeers! Wishing you a fun and safe holiday weekend.	2.67	3	0	3	0	0
Let's all give a big Xavier welcome to @PrezHanycz as she celebrates three big milestones today! Canada Dav	1.75	1	0	0	0	0
Xavier is strongly encouraging our community to receive a #COVID19 vaccine this summer.<U+2063> Students who update their records this summer will be entered into a special	2.5	2	0	2	0	0

Figure 3: Excel File with Sentiment Analysis Results

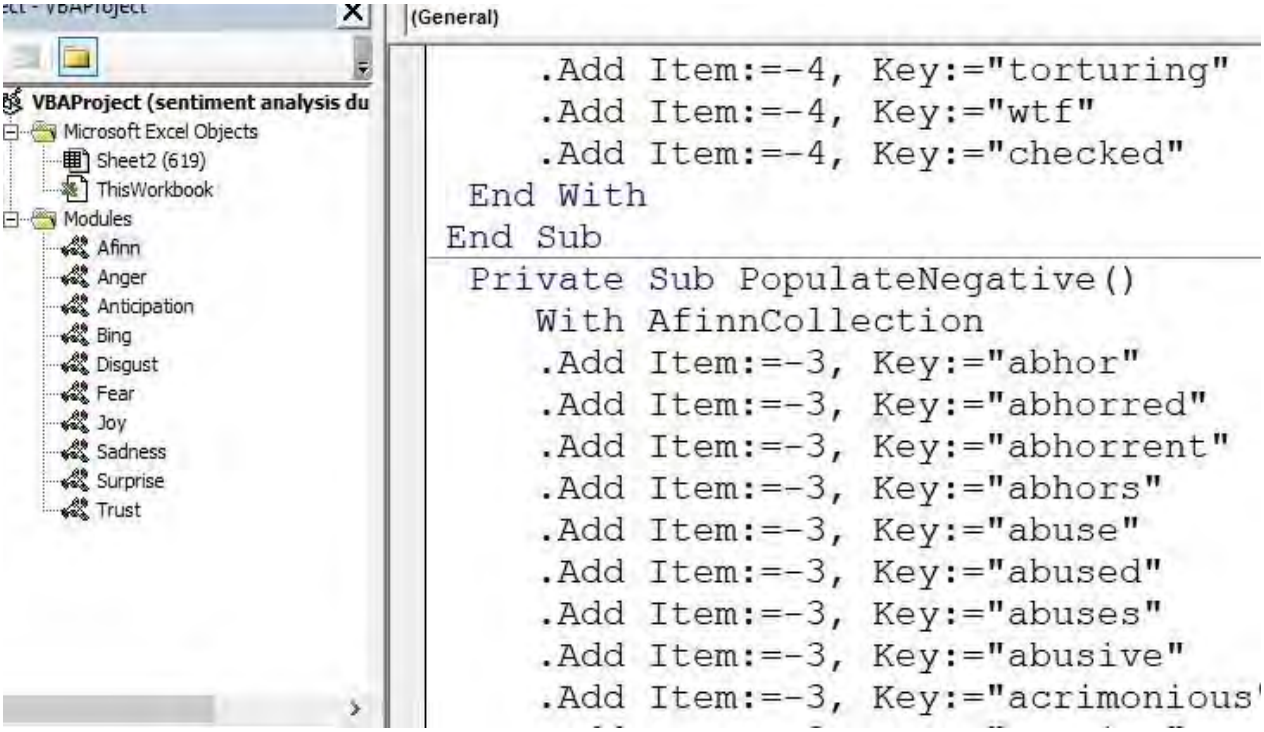

Figure 4: Excel Visual Basic Code to Compute Sentiment Analysis Scores

On a separate Word document, summarize your findings on the three search terms in one page or less. Be sure to comment on:

- Average tweet score for Afinn and Bing sentiments. Is it positive or negative? Judging by some of the tweets and what you know of the terms, why do you think this might be the case?
- How do the emotions compare between search terms? Are they consistent or do they have different profiles?

Submit the following under the "R Twitter Project" assignment link:

- Your R code for extracting the data from Twitter using the rtweet package
- A sentiment analysis workbook with three sheets calculating sentiment (one for each search term)
- The final Word document summarizing your findings

Figure 5: Final Report and Project Deliverables