

# A demographic and sentiment analysis of e-cigarette messages on Twitter

Elaine Cristina Resende and Aron Culotta  
Department of Computer Science  
Illinois Institute of Technology  
Chicago, IL  
eresende@hawk.iit.edu, culotta@cs.iit.edu

## ABSTRACT

Social media platforms, such as Twitter plays a big role in the society nowadays and they have become an important source for public health surveillance and others applications. Electronic cigarette is an outstanding topic in the health area nowadays, and one important question that arises is to know how people are dealing with this new idea of "smoking free", in other words, how they are concerned about e-cigs? In order to answer questions like that we need to know what people are saying about e-cigs, if it is positive or negative and how this sentiment changes over the time. In this project, we have been working with Twitter data and we have tried to find out the sentiment of the users related to electronic cigarettes. Through our analysis we have noticed interesting facts that can contribute to the marketing and health campaigns, and also that most of sentiment is negative (marketing or criticism).

## Categories and Subject Descriptors

I.5.4 [Pattern Recognition]: Applications—Text processing; H.4 [Information Systems Applications]: Miscellaneous

## Keywords

Web mining, social media, public health

## 1. INTRODUCTION

Understanding evolving behaviors and attitudes related to alcohol, tobacco, and other drugs is a critical public health goal. Typically, such topics are investigated by survey research; however, surveys can be costly and time-consuming, making them ill-suited to rapidly changing environments. Furthermore, survey response rates have fallen precipitously in recent years (Kohut et al. 2012), leading researchers to seek non-traditional data sources. A promising new methodology is social media analysis, which analyzes online content expressing health-related behaviors and opinions to provide

insights into population-level trends. Online content offers several potential advantages over traditional data: one can in real-time measure how behaviors and attitudes change in response to rare events such as legal changes, new products, and marketing campaigns, and the open-ended nature of the content can provide more diverse data. Such approaches have been used to study depression [2], insomnia [8], and other health topics [3].

In this paper we present a descriptive study of attitudes towards electronic cigarettes (e-cigs) expressed on Twitter.com. E-cigs provide a nicotine-containing aerosol with different flavors, glycol and other ingredients, that users smoke by heating up a solution [6]. E-cig adoption has increased rapidly in the United States recently [9, 4]; by one estimate, usage by high school students tripled in 2014 [10]. Despite this rapid growth, there is still considerable debate over the health impact of e-cigs [17, 14, 1], also reflected in consumer surveys [15], leading to uncertainty as to how they should be regulated.

We analyze nearly one million tweets posted from October 2012 to September 2013 that contain keywords related to e-cigarettes. We use a supervised classification algorithm to annotate each message by sentiment (positive, negative, or neutral). Furthermore, using the first name of each user, we derive estimates of age and gender to further stratify results. Our main findings are as follows:

- Overall volume of e-cig tweets grew five-fold in our year sample, from 30K tweets in October 2012 to 150K in September 2013.
- Of tweets expressing sentiment, 65% are classified as positive sentiment (tweets either advocating for e-cigs or indicating that the user has tried e-cigs). This value ranges by month from 61% in March 2013 to 74% in July 2013. We find that external events (e.g., regulations) can have a large impact on this estimate.
- We find positive sentiment to be slightly higher for males than females (63% vs. 61%), and highest for users estimated to be 18-24 years old (67%).

We additionally perform a substantive analysis to provide a more fine-grained insight into the different ways people discuss e-cigs and how that is related to sentiment and demographics. For example, in November-January we find a spike in messages from users wishing for e-cigs as a Christmas present to help them stop smoking. We also find that many tweets from young female users mention smoking e-cigs with their parents or buying e-cigs for their parents;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

whereas male users are more likely to ridicule those using e-cigs.

The remainder of the paper is organized as follows: Section 2 reviews related work; Section 3 describes the Twitter data and our method of analysis; Section 4 describes our main results; Section 6 discusses implications, limitations, and future work; and Section 7 concludes.

## 2. PRIOR WORK

Myslin et al. [12] manually classified 4K tobacco-related tweets along 30 dimensions, including sentiment, theme, and genre, finding that tweets about e-cigs and hookahs tended to have more positive sentiment. Here, we focus specifically on e-cigs, expanding upon this initial work with a much larger set of tweets (1M) over a longer time span, and further stratifying by gender and age.

Huang et al. [7] analyzed 74K tweets related to e-cigs and, using a supervised classifier, found that 90% were commercial tweets, and about 10% mentioned smoking cessation. We build on this prior work, using their classifier to first filter out commercial tweets.

Other work has found a prevalence of smoking cessation accounts on Twitter [16], and has attempted to track the progress of cessation attempts by following Twitter accounts of identified users [11].

Very recently, Godea et al. [5] analyzed 106K tweets containing e-cig related terms collected over a two month period. They built a sentiment classifier using a hand-engineered lexicon of over 600 terms, achieving a positive/negative F1-score of 55-56%.

Compared to this prior work, our main contributions consist of (1) an analysis of a much larger sample of e-cig tweets than has been done previously, consisting of nearly one million tweets written over one year; (2) a sentiment classifier tuned for precision that identifies positive tweets with 96% precision and negative tweets with 70% precision; (3) an analysis of temporal trends in sentiment and demographics.

## 3. METHODS

In this section we describe the data collected and the methods used for classifying tweets by sentiment and for estimating user age and gender.

### 3.1 Data Collection

We use the data collection process described in Huang et al. [7]. Using the full Twitter Firehouse, tweets were collected from October 2012 to September 2013, using a set of keywords identified by expert consensus (*e-cigarette*, *ecigarette*, *e-cig*, *ecig*). Additional tweets were identified that matched the query: (*cig* OR *cigarette*) AND (*electronic* OR *blu* OR *njoy*) (the latter two terms referring to the top-selling e-cig brands in the U.S.). This resulted in 4,639,885 tweets. Using the classifier of Huang et al. [7], we retained only those tweets identified as “organic,” defined as “those reflecting individual opinions or experiences or linked to non-promotional content.” This left 992,633 tweets.

Aiming to analyze the sentiment of users we need to work just with organic tweets, so we have selected just organic tweets which represent (21.39%) of the data. Figure ?? shows the number of tweets (y axis) by month (x axis). The green line refers to organic tweets and blue line to all tweets. We can see that there is a big spike in March and also an

Figure 1: Precision-recall curves derived by classifying the labeled data using 10-fold cross-validation.

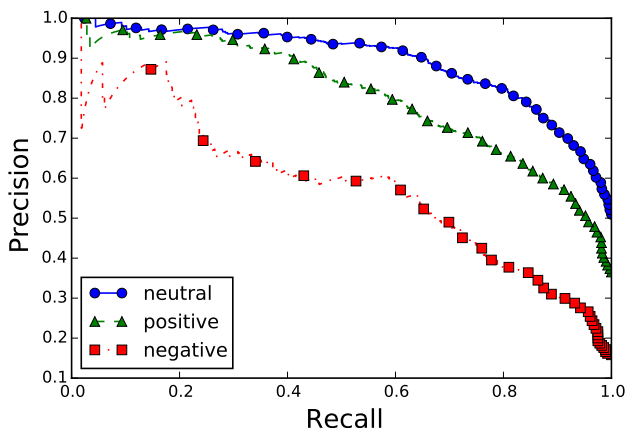


Table 2: Cross-validation classification accuracy

	Prec	Rec	F1	N
negative	0.60	0.55	0.57	279
neutral	0.78	0.84	0.81	1014
positive	0.75	0.68	0.72	707
avg	0.74	0.75	0.74	2000

increase in the months of May and September.

A random sample of 2000 collected tweets, from the set of organic tweets, were manually labelled as positive, negative and neutral. Each tweet was put in one class according to the following:

- tweets from merchandising companies, news about electronic cigarettes, and tweets belonging to other languages were classified as neutral class, they did not express any sentiment;
- tweets from users who expressed buying or use desire, or from users who tweeted about their e-cigs were considered positive class;
- and tweets from users who were complaining (or citing) about other people vaping were labelled as negative.

Table 1 describes some examples of our training data with tweets, their classes and also the distribution of each class. We can see that 52.45% of the tweets were labeled as neutral, 13.75% as negative and 33.80% as positive.

For this analysis we have joint negative and neutral tweets because we are more interested in the positive, in other words, telling that you use e-cigs is more important than just criticism or marketing campaigns. So, at this point we have 676 (33.8%) tweets classified as positive and 1324 (66.2%) as negative. The coded set of tweets was used to train a machine learning classifier using Logistic Regression algorithm, and the built model was applied to the full set of organic tweets.

### 3.2 Classification

A Logistic Regression (LG) classifier was built, trained it with coded tweets, and tested it with the organic set. Section

Table 1: Training Data Summary

Class	#Tweets	Tweet Sample
Positive	707	I bought a Ecig today Electric cigarettes are better than regular cigarettes!!
Negative	279	#IHatePeopleThat smoke ecigs e-cigs are bad, mmkay?
Neutral	1014	what is an electronic cigarette? A homeless guy just asked me if he could bum an e-cigarette

Table 3: Top Coefficients of Classifier

<b>negative</b>	you, smoking, smoking an, he, fuck, people, smokes, an, faggot, are, class, smoke, stupid, in, look, shit, her, pussy, sorry, one
<b>neutral</b>	URL, e-cigarettes, de, la, 99, retail, URL retail, ni, e-cigarette, markten, store, by, cigarette, of, dallas, smokers, 9999, @vaper_trail, electronic, may
<b>positive</b>	my, i, vaping, #vaping, #ecig, my ecig, me, we, #vape, got, e-cig, my e-cig, #euecigban, my e, vape, i'm, good, this, i need, to black

3.2.1 and 3.2.2 are going to show how the methodology of training and testing we followed.

### 3.2.1 Training

A Logistic Regression classifier was used in Python language, which is responsible to predict the labels of all organic tweets; the set of labelled tweets were used as our training set. In order to have a better accuracy for the classifier, we used a grid search to check the best values for "Penalty" and "C", which are parameters of LG. Running the grid search we found out that Penalty = L2 and C = 2.6 is the best configuration for our LG model. Figure ?? shows the accuracy versus C parameter, and we can see that the accuracy is higher when C is equal to 2.6.

Our best classifier configuration has the accuracy of 81.8%. The confusion matrix of the model is shown in the Figure ??, and Table 2 shows precision, recall and F-score values for both classes. As we can notice the percentage of retrieved tweets that are relevant is 90% what is seen as a good value for our assignment.

### 3.2.2 Testing

Our testing set is defined as all organic tweets (900k tweets). The LG classifier was applied to them, as result we had 27% classified as positive and 73% as negative. What is comparable to our manual labeling which was 33.8% positive and 66.2% negative.

In order to check the quality of our LG model, 200 tweets were randomly selected and labeled. LG model was again applied to this same new set of tweets. The labels were used as true labels and the predicted labels (output from LG model) were compared to the true labels to get the accuracy. As result, the model had 79% of accuracy, so we conclude that our model is good for predicting the sentiment of tweets for this problem.

The top and bottom words of the classes are shown in the Table 3. Top coefficients refer to negative class and bottom coefficients show words related to positive class. Through Table 3 we can see that LG model learned pretty well because words in the Top Coefficients, such as THIS\_IS\_A\_URL, RETAIL and STORES are related to marketing tweets, also HE, YOU, PEOPLE and SHE refer to other people, what represents tweets criticizing people who use e-cigs, thus being negative

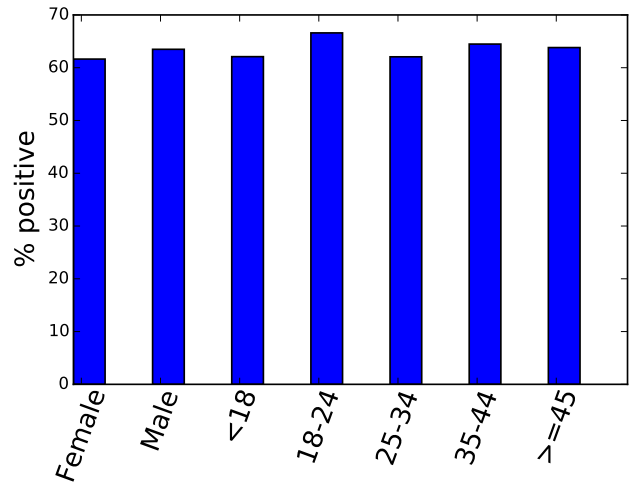


Figure 3: Percent of tweets from each group classified as positive sentiment.

class. The bottom coefficients MY, I, VAPING, VAPE and BUY refer to users who tweet about getting or using e-cigs.

Using three-class classifier: predicted label distribution on raw tweets: [(neutral, 537474), (positive, 329919), (negative, 125240)]

## 4. RESULTS

### 5. ANALYSIS OF THE SENTIMENT DURING A YEAR (OCT/2012-SEP/2013)

We analyzed the amount of tweets by month according to their sentiment (negative or positive) toward e-cigs. In the Figure ?? we have considered all organic tweets. In the Figures ?? and ?? it is considered just users who tweeted once. Figure ?? one tweet per user is considered, and in the Figure ?? does not consider re-tweets.

In those Figures red represents positive tweets, green represents negative and blue all organic tweets. Figure ?? shows spikes in April, June and September for all sentiment lines,

Figure 2: Tweets by month and sentiment.

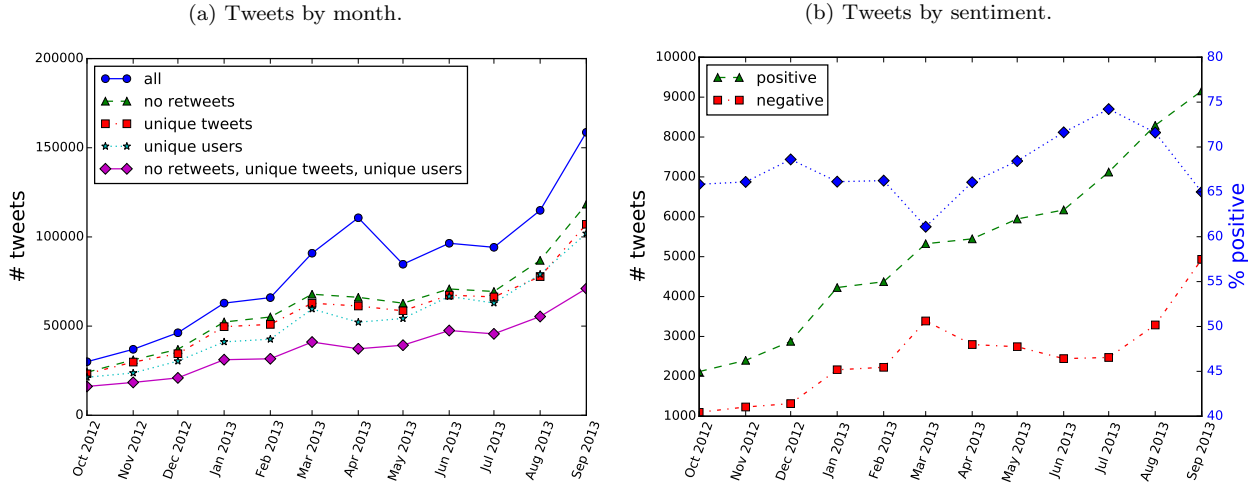
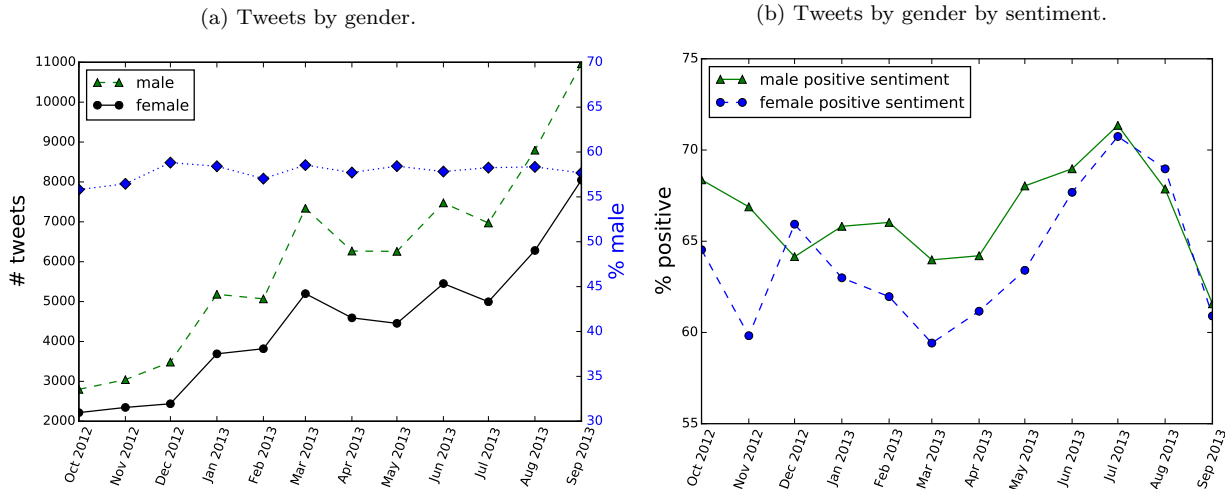


Figure 4: Tweets by gender and sentiment.



but Figures ?? and ?? have moved the first left spike from April to March, what tell us that the number of users was higher in March than April. Figure ?? and Figure ?? have similar shape (same for Figures ?? and ??), they differ just in the number of tweets in the y axis.

Aiming to try to find out why we have those spikes we analyzed n-grams, which are a sequence of words in the tweet, and also the 10 top words for each month during the year, shown in the Table ???. We can see that the words in the table are very correlated with the tweets posted in the respective month. For example, November has the word THANKSGIVING, CHRISTMAS appears in December, VALENTINE in February, and so on. These words can help us to understand the spikes, and what are the main topics for positive and negative tweets. Some of the spikes are due to some events that occurred during those months. Besides the significance of each word, the Table ?? also shows the percentage of tweets that each word is present, what can describe the importance of them to the entire data. Following we have a brief analysis description of the spikes shown in the graphs.

### March

Analysing Figure ?? which shows one tweet per user, March has 44788 negative and 17656 positive tweets. If we look in the Table ??, the month of March shows the word ONEW, what refers to a Korean singer vaping. A lot of Twitter users tweeted about him, and most of tweets were negative. We also have the words VATICAN and POPE which represent the month of the choice of the pope, and these are also classified most negative. VAPERWARE is a store and most of the tweets are classified as negative. SWSW is a music, film and interactive festival happening in March and most tweets are negative classified. 300 tweets were saying: BLACKFRIDAY #BLACKFRIDAY IS TOO GOOD FOR ME BECAUSE I USED ELECTRONIC CIGARETTE WHICH I MISSED A LOT and they were classified as positive tweets.

This month had about 16.5% of users who posted positively according to Figure ??.

### April

Considering just positive tweets from Figure ?? and going

Figure 5: Tweets by age and sentiment.

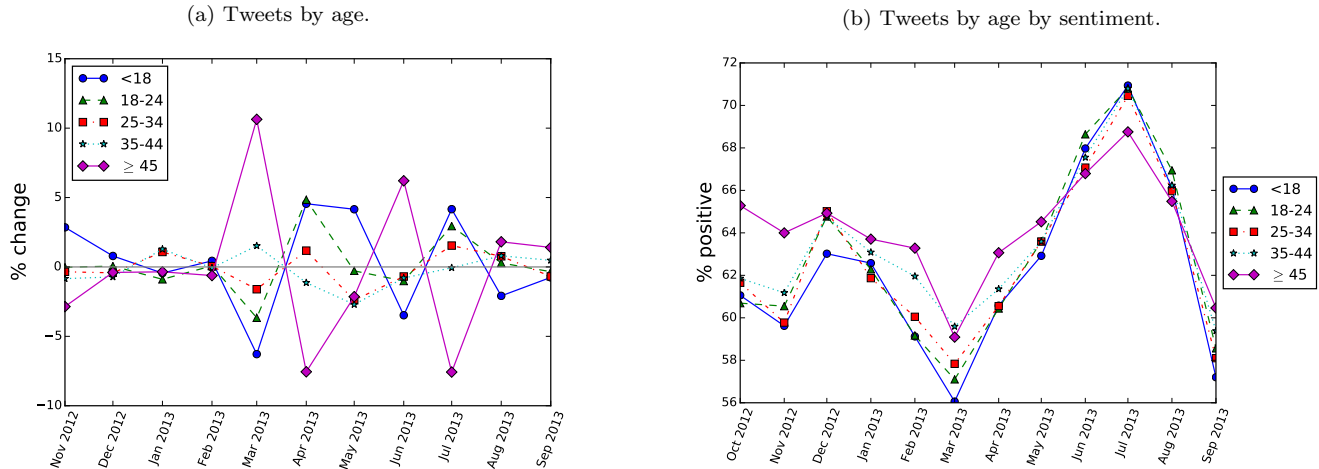
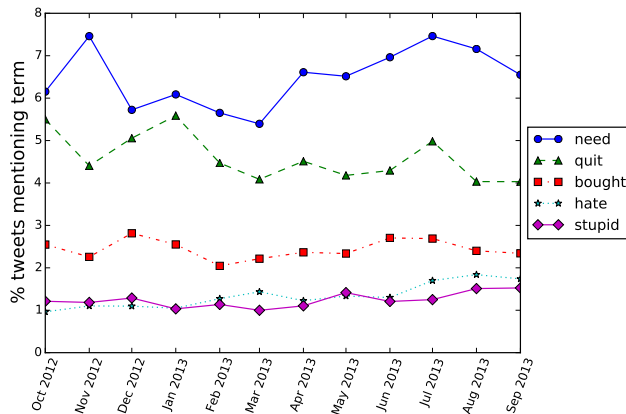


Figure 6: Percent of non-neutral tweets mentioning each term.



through the n-grams of the month of April we found in total 17548 tweets. After analyzing the n-grams, we took a further look in those tweets and we found more than 600 re-tweets of: " @SIMONCOWELL: I HAVE CUT DOWN SMOKING. A LOT. USING ELECTRIC CIGARETTES TO HELP. ARE THEY BAD FOR YOU?" and they were labeled as positive.

According to Figures ?? and ??, April was the month which people talked less in positive way toward e-cigs. In the Table ??, we can see that is because some senators in US made regulations to restrict the sale of e-cigs <sup>1</sup>, and also about Courtney TV commercial.<sup>2</sup>

Some other words from Table ?? are described below and most of the tweets containing them were classified as negative, what contributed for negativeness in this month.

#welcometomyschoolwhere PEOPLE SMOKE ELECTRONIC CIGS DURING CLASS

E-CIGARETTES **primarily** USED TO QUIT TOBACCO: STUDY

<sup>1</sup><http://www.ewspsf.com/latest-news/health-and-fitness/42182-senators-call-on-fda-to-restrict-the-sale-distribution-and-marketing-of-e-cigarettes.html>

<sup>2</sup>[http://www.huffingtonpost.com/2013/04/03/courtney-love-njoy\\_n\\_3006660.html?ncid=edlinkusaolp00000003](http://www.huffingtonpost.com/2013/04/03/courtney-love-njoy_n_3006660.html?ncid=edlinkusaolp00000003)

[HTTP://T.CO/L2G1DksNJt](http://t.co/L2G1DksNJt) #EUECIGBAN

**Leachon:** E-CIGARETTES MAY CONTAIN CARCINOGENS, INGREDIENTS NOT APPROVED BY FDA

### September

This month is the third most negative with 15% of positive tweets overall. 28938 tweets are positive and 78341 are negative according to Figure ??.

The word **whisp** refers to "RT YABOYLILB: \*GETS IN BOOSTER SEAT\* \*PUTS SEATBELT ON\* \*PUTS IN THE NEWEST KIDZBOP CD\* \*TAKES DRAG OF E-CIG\* \*LOWERS GLASSES\* \*WHISP" and it was classified as positive.

**worldvapingday** is a hashtag about <http://www.world-vaping-day.com/>, which is the day of vaping, and most of them were negative. Two main tweets appeared in the data:

1. THURSDAY IS #WORLDVAPINGDAY! TIME TO TELL THE WORLD HOW #ECIGS HAVE CHANGED YOUR LIFE!!!. It was classified as negative.
2. THURS. IS #WORLDVAPINGDAY! CONSERVATIVES NEED TO KNOW THE FDA WILL TRY TO BAN #ECIGS AGAIN IN OCT. STOP LIBERALS,SAVE LIVES. It was classified as positive.

PROPERLONGBOARD VAPING DALLAS PARTY GIFT E-JUICE & **Nude** SELFIES [HTTP://T.CO/mSKYYKfzZW](http://t.co/mSKYYKfzZW) explains the appearance of the word nude, and they were negative tweets.

The word **attorneys** represents tweets saying about e-cigarette regulations. E.g.: VEGASNEWSNOW ATTORNEYS GENERAL CALL ON FDA TO REGULATE E-CIGARETTES [URL] #VEGAS. Most classified as negative.

**Vapefest** occurs during September and most of tweets were negative. E.g.: CANT WAIT FOR #VAPEFEST NEXT WEEKEND. STOKED

**Teens** represents tweets talking about the grow of teens vaping, and they also represent negative class. E.g.: BAKO-COM STUDY SHOWS MORE TEENS TURNING TO ELECTRONIC CIGARETTES: #PALMSPRINGS #[URL]

Figure ?? shows, in the right y axis, the percentage of positive tweets over all organic with LG model built with 2

Table 4: Significant terms by category.

<b>Positive sentiment</b>	my, i, got, ecig, need, love, e-cig, me, i'm, cig, lost, mom, want, vaping, bought, dad, broke, so, just, get
<b>Negative sentiment</b>	class, smoking, an, you, in, kid, cool, electric, people, guy, look, smoke, he, you're, stupid, if, fuck, dude, cigarette, are
<b>Female</b>	cigarette, my, her, electric, mom, dad, smoking, me, amp, electronic, an, class, omg, #giveaway, so, x, his, boyfriend, i, onew
<b>Male</b>	vaping, green, e-cigs, bro, blu, pope, e-cigarettes, @blucigs, ban, #vape, @youtube, stephen, the, game, pussy, new, a, any, vapers, gay
<b>&lt;18</b>	electric, cig, class, an, cigarette, e, lol, just, mom, cool, those, guy, haha, was, hookah, mask, elaborate, dad, shit, my
<b>18-24</b>	electric, cigarette, those, lol, cig, class, her, fait, electronique, blu, hookah, got, smh, cutting, laugh, quand, feel, me, mom, commercial
<b>25-34</b>	an, cigarette, electric, cig, class, my, just, those, i, bought, guy, so, her, girl, his, mom, kid, dad, me, cigs
<b>35-44</b>	cigarette, #gotitfree, electric, those, blu, i, cigs, just, cig, bought, an, these, was, lol, electronic, luck, her, it, what, smoking
<b>≥ 45</b>	electric, thuis, markten, cigarette, alle, van, hq, those, blu, via, e-smoking, e-cigarette, elaborate, @lord_sugar, mask, just, cigs, mistic, reviews, was
<b>2012-10</b>	electric, cigarette, electronic, cigarettes, green, free, alternative, trial, class, those, obtain, an, during, enjoys, liked, looks, obama, photo, stephen, lol
<b>2012-11</b>	election, cigarette, electric, mask, elaborate, electronic, kristen, traditional, device, rob, halo, na, brand, knee, touching, kit, cigarettes, lol, was, class
<b>2012-12</b>	christmas, electric, elaborate, cigarette, mask, electronic, @overlymanlymann, na, ko, addiction, advertising, ako, ny, ng, cigarettes, hq, allowed, calif, report, those
<b>2013-01</b>	prevent, electric, accessories, regulating, rolling, sale, cigarette, banning, electronic, na, fda, continue, ng, ko, launch, tv, those, haha, cigarettes, ni
<b>2013-02</b>	amg, miracle, menace, boat, racing, electric, drive, cigarette, elaborate, mask, prevent, class, decision, 99/99, crave, continues, electronic, including, opinion, @e_swisher
<b>2013-03</b>	onew, green, utah, cigarette, gallagher, noel, muse, pope, drummer, caught, moves, criticises, electronic, electric, #gotitfree, risks, smoke, mistic, tax, an
<b>2013-04</b>	courtney, f-bomb, drops, ad, love, @simoncowell, electric, njoy, #gotitfree, inside, bring, commercial, cigarette, blu, class, an, lol, those, i, watch
<b>2013-05</b>	@lord_sugar, france, la, electronique, les, governo, logic, places, electric, ecig, une, my, an, public, c'est, electrique, lol, il, mais, cigarette
<b>2013-06</b>	e-smoking, package, restrictions, incredible, britain, rise, medicines, include, kits, medicine, website, parker, face, markten, experience, easy, perfect, alle, thuis, reynolds
<b>2013-07</b>	markten, van, alle, thuis, vuse, vaporizer, #uk, e-cig, ook, my, op, hookah, ecig, tweets, being, i, ik, regulation, blu, me
<b>2013-08</b>	promise, markten, companies, benowitz, neal, thuis, echo, van, cigs, heyday, alle, york, e, stadium, literally, merthyr, professor, never, bloomberg's, ook
<b>2013-09</b>	doubles, among, patches, students, teens, cdc, e-cigarettes, use, kids, effective, flames, 9-foot, survey, shows, markten, middle, u.s, teach, e-cigarette, study

classes (positive and negative). We can see that the percentage of positivism is higher during October, December and February, and then it goes down during March and April. May starts going up again, but for the rest of the months no one goes upper than May. The left y axis describes the number of users, and in the x axis, the bars show the number of users who are posting positively by month, from October 2012 through September 2013 in blue color, and we can notice that this number is much less than negative, what tell us that most of tweets are propaganda or criticism about someone using e-cigs.

In order to understand more Figure ??, we have done an analysis considering the 3 classes explained in Section 3.1. Figure ?? plots the percentage of positive tweets over the year considering positive and negative tweets from LG model using 3 classes (positive, negative and neutral). As it is expected, the percentage of positivism increases. April

still is the month with less positive tweets, October is the most positive month overall.

## 5.1 Gender Inference

U.S. Census provides list of names from population. We use the 1990's census<sup>3</sup> to compute the proportion of males and females for positive and negative classes. We keep 75% of each names list, eliminate ambiguous names, at the end the total names are 226 males and 518 females. Each tweet gets then its label, if the first name of the user is present in the male list, it is labeled as male or the same for female list. Figure ?? shows the distribution of genders in our data. 74% overall tweets have their gender unknown because the users did not have their names in the census data and majority of tweets belonged to users of companies (369127 unique users are labeled as "unknown"). We can see that 28-30% of users,

<sup>3</sup><http://www2.census.gov/topics/genealogy/>

independently of gender, express positive sentiment towards electronic cigarettes, but the majority express negativeness toward them.

## 5.2 Age Inference

In a recent study Silver and McCanc [13], showed how the age of a person can be indicated by her/his name. The Social Security Administration (SSA) provides a list of baby names<sup>4</sup> who are born in each year, from 1880 to 2014. Through this list we can know gender and number of babies born with a name by year. Another resource we use to infer age is a cohort life table also provided by SSA<sup>5</sup>, that give us the estimate of how many people born in a given year are still alive.

We define the years in age brackets (under 18, 18-24, 25-34, 35-44 and 45+) and with the number of people in certain age and who are still alive we can compute the fraction of people who are still alive for each age bracket. For example, given a name we can see the fraction of people who are 18 years old, which is in the first bracket. Figure ?? shows an example for the name Elaine, the first column represents the brackets; "n" is the number of people who born during that period, "n\_alive" is the number of people born in that period who are still alive, and "fraction" shows the percentage of people in that bracket who are still alive. In this figure we can see that the most common age estimate for Elaine is 45+, followed by 35-44, 25-34, 18-24 and 18-.

The age inference analysis was done by month, considering both classes of our data set. The next figures show the age brackets estimates by month during the year data, plotting separately by sentiment. Figures ??, ??, and ?? plot the sentiment by age for each age bracket. Figures ??, ??, and ?? show the differences in the estimates of age.

## 6. DISCUSSION

### 6.1 Relation to prior survey results

"Likewise, Pearson et al. (2012) found no difference in men and women's ever-use of e-cigarettes (men: 12.6

Younger and older individuals appeared equally likely to have ever used e-cigarettes in the U.S., but younger individuals were more likely to ever-use them in other regions.

## 7. CONCLUSION

As we could notice from our analysis the over all tweets increase over the time, but negative tweets (marketing, criticism) increases more than positive tweets. Most of tweets are negative and the positive tweets don't change much over the time, they remain around 16% and 23%. Most of the spikes that the Figures showed are due to a particular event during its respective month. So, we are sure that TV commercials and new campaigns are relevant and they help to make changes in the data.

## 8. ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grant #IIS-1526674. Any opinions, findings and conclusions or recommendations expressed in this material are the authors' and do not necessarily reflect those of the sponsor.

<sup>4</sup><http://www.ssa.gov/oact/babynames/>

<sup>5</sup>[http://www.ssa.gov/oact/NOTES/as120/LifeTables\\_Tbl7.html](http://www.ssa.gov/oact/NOTES/as120/LifeTables_Tbl7.html)

## References

- [1] A. Bhatnagar, L. P. Whitsel, K. M. Ribisl, C. Bullen, F. Chaloupka, M. R. Piano, R. M. Robertson, T. McAuley, D. Goff, and N. Benowitz. Electronic cigarettes a policy statement from the american heart association. *Circulation*, 130(16):1418–1436, 2014.
- [2] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. In *ICWSM*, 2013.
- [3] M. Dredze. How social media will change public health. *IEEE Intelligent Systems*, 27(4):81–84, 2012. ISSN 1541-1672. doi: 10.1109/MIS.2012.76.
- [4] C. for Disease Control, P. (CDC, et al. Notes from the field: electronic cigarette use among middle and high school students-united states, 2011-2012. *MMWR. Morbidity and mortality weekly report*, 62(35): 729, 2013.
- [5] A. K. Godea, C. Caragea, F. A. Bulgarov, and S. Ramisetty-Mikler. An analysis of twitter data on e-cigarette sentiments and promotion. In *Artificial Intelligence in Medicine - 15th Conference on Artificial Intelligence in Medicine, AIME 2015, Pavia, Italy, June 17-20, 2015. Proceedings*, pages 205–215, 2015.
- [6] R. Grana, N. Benowitz, and S. A. Glantz. E-cigarettes a scientific review. *Circulation*, 129(19):1972–1986, 2014.
- [7] J. Huang, R. Kornfield, G. Szczypka, and S. L. Emery. A cross-sectional examination of marketing of electronic cigarettes on twitter. *Tobacco control*, 23(suppl 3):26–30, 2014.
- [8] S. Jamison-Powell, C. Linehan, L. Daley, A. Garbett, and S. Lawson. "I can't get no sleep": Discussing #insomnia on Twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 1501a–1510, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208612. URL <http://doi.acm.org/10.1145/2207676.2208612>.
- [9] B. A. King, S. Alam, G. Promoff, R. Arrazola, and S. R. Dube. Awareness and ever-use of electronic cigarettes among us adults, 2010–2011. *Nicotine & Tobacco Research*, 15(9):1623–1627, 2013.
- [10] K. Leonard. E-cigarette use triples among teens: The cdc chief blames aggressive marketing for the devices' growing use among young people. <http://www.usnews.com/news/articles/2015/04/16/e-cigarette-use-triples-among-teens>, 2015. Accessed 2015-06-01.
- [11] E. L. Murnane and S. Counts. Unraveling abstinence and relapse: Smoking cessation reflected in social media. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1345–1354, New York, NY, USA, 2014. ACM.
- [12] M. Myslín, S.-H. Zhu, W. Chapman, and M. Conway. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *Journal of medical Internet research*, 15(8), 2013.

- [13] A. M. Nate Silver. How to tell someone's age when all you know is her name. <http://fivethirtyeight.com/features/how-to-tell-someones-age-when-all-you-know-is-her-name/>, 2014. Accessed 2015-06-01.
- [14] J. Pauly, Q. Li, and M. B. Barry. Tobacco-free electronic cigarettes and cigars deliver nicotine and generate concern. *Tobacco Control*, 16(5):357–357, 2007.
- [15] J. K. Pepper, S. L. Emery, K. M. Ribisl, C. M. Rini, and N. T. Brewer. How risky is it to use e-cigarettes? smokers' beliefs about their health risks from using novel and traditional tobacco products. *Journal of behavioral medicine*, 38(2):318–326, 2015.
- [16] J. J. Prochaska, C. Pechmann, R. Kim, and J. M. Leonhardt. Twitter= quitter? an analysis of twitter quit smoking social networks. *Tobacco control*, 21(4):447–449, 2012.
- [17] T. E. Sussan, S. Gajghate, R. K. Thimmulappa, J. Ma, J.-H. Kim, K. Sudini, N. Consolini, S. A. Cormier, S. Lomnicki, F. Hasan, et al. Exposure to electronic cigarettes impairs pulmonary anti-bacterial and anti-viral defenses in a mouse model. *PloS one*, 10(2):e0116861, 2015.

sigproc.bib is the name of the Bibliography in this case