2013

AT&T vs Verizon: Mining Twitter for Customer Satisfaction towards North American Mobile Operators

Demitrios Pournarakis
Athens University of Economics and Business, pournadi@aueb.gr

Dionisios N. Sotiropoulos
Athens University of Economics and Business, dsotirop@gmail.com

Chris D. Kounavis
Athens University of Economics and Business, koun@aueb.gr

George M. Giaglis
Athens University of Economics and Business, giaglis@aueb.gr

AT&T VS VERIZON: MINING TWITTER FOR CUSTOMER SATISFACTION TOWARDS NORTH AMERICAN MOBILE OPERATORS.

Pournarakis, Demitrios, Athens University of Economics and Business, Elpidos 13, 10434, Athens, Greece, pournadi@aueb.gr

Sotiropoulos, Dionisios N., Athens University of Economics and Business, Elpidos 13, 10434, Athens, Greece, dsotirop@gmail.com

Kounavis, Chris D., Athens University of Economics and Business, Elpidos 13, 10434, Athens, Greece, koun@aueb.gr

Giaglis, George M., Athens University of Economics and Business, Elpidos 13, 10434, Athens, Greece, giaglis@aueb.gr

Abstract

The North American Telecommunications sector is one of the leading mobile broadband sectors worldwide, representing increasingly important revenue opportunities for mobile operators. Taking into consideration that the market is being saturated and revenue from new subscriptions is increasingly deteriorating, mobile carriers tend to focus on customer service and high levels of customer satisfaction in order to retain customers and maintain a low churn rate. In this context, it is a matter of critical importance to be able to measure the overall customer satisfaction level, by explicitly or implicitly mining the public opinion towards this end. In this paper, we argue that online social media can be exploited as a proxy to infer customer satisfaction through the utilization of automated, machine-learning based sentiment analysis techniques. Our work focuses on the two leading mobile broadband carriers located in the broader North American area, AT&T and Verizon, by analysing tweets fetched during a 15-day period within February 2013, to assess relative customer satisfaction degrees. The validity of our approach is justified through comparison against surveys conducted during 2012 from Forrester and Vocalabs in terms of customer satisfaction on the overall brand - usage experience.

Keywords: *Telecommunications, AT&T, Verizon, Twitter, Customer Satisfaction, Sentiment Analysis*

1 Introduction

1.1 Problem Definition

Taking into consideration that the market is being saturated and revenue from new subscriptions is increasingly deteriorating, mobile carriers tend to focus on customer service and high levels of customer satisfaction, in order to retain registered customers and maintain a low churn rate.

In this paper, we try to examine if mobile wireless carriers can benefit from performing sentiment analysis through social media networks in order to enhance and improve customer service, which will lead to increased customer satisfaction, thus keeping a low churn rate.

In order to examine this question we focus on two aspects:

- First we perform sentiment analysis on text mined from Twitter, of customers mentioning the two leading mobile wireless carriers in North America (AT&T and Verizon) and examine if customer satisfaction can be measured through analysing and monitoring social media networks (Twitter in particular).
- Second we compare the results we concluded through the first method, to results gathered from three customer satisfaction surveys performed through call interviews, by three different independent research laboratories, in order to examine if there is correlation between the results.

In this section we continue with an overview of the market, take a look at the problem of saturation in the market and present the survey based customer results. We examine the literature to find previous applications measuring customer sentiment through online media as well as state of the art applications, used by leading corporations today. In turn, we describe the methodology we followed to perform the analysis and focus on the results that were generated. The paper concludes with our conclusions and future considerations in the field.

1.2 AT&T vs. Verizon: Financial Figures in a saturated environment

According to AT&T's 2012 fourth quarter results, published on January 24, 2013, AT&T posted a net increase in total wireless subscribers of 1.1 million in the fourth quarter to reach 107.0 million subscribers in service with an annual operating revenue of $127 billion.

Verizon is number two in retail connections, with 98.2 million subscribers in service and $75.9 billion annual revenue in 2012.

According to a Columbia Graduate Consulting Club study conducted on February 13, 2012 Total U.S. Telecommunications Industry Revenues reached $985 billion during 2010 with Annualized Total Wireless Service Revenues matching $159.9 billion.

Arguably, the telecommunications sector, and the wireless services group in particular, is one of the leading drivers in the U.S economy, rendering competition between carriers very intense.

Based on a qualitative survey, of director-level and above marketing and business executives responsible for retention strategies at 40 service providers across North America, Europe, Asia Pacific and Central and Latin America, conducted from June to July 2011 by Amdocs, 66 percent of operators believed that customers are less loyal today than they were two years ago, 70 percent of service providers cited customer retention and loyalty as the critical factor for driving growth, with a strategic marketing prioritization shift from customer acquisition and market share to long-term customer engagement. Due to market saturation and increasing competition, 82 percent of service providers said that customer loyalty programs would be "very important" or "important" over the next five years to their company's strategy.

1.3 Survey Based Customer Satisfaction Results

On January 2013, Vocalabs published the National Customer Service Survey (NCSS) on Mobile Phones based on data collected from 2009 through 2012, through independent research, tracking results for AT&T, Sprint, Verizon and T-Mobile. The study draws on some pretty insightful results which are presented in brief below and along with results from the Forrester study, acted as a benchmark of comparison to the sentiment analysis performed on the collection of tweets.

The National Customer Service Survey for Mobile Phone Customer Service is a continuous survey beginning July 2009, and data from all four years are presented. Customers were interviewed immediately after a customer service call to one of the companies in the report. The survey measured customer perceptions of the quality of the customer support they received from AT&T, Sprint, T-Mobile, and Verizon.

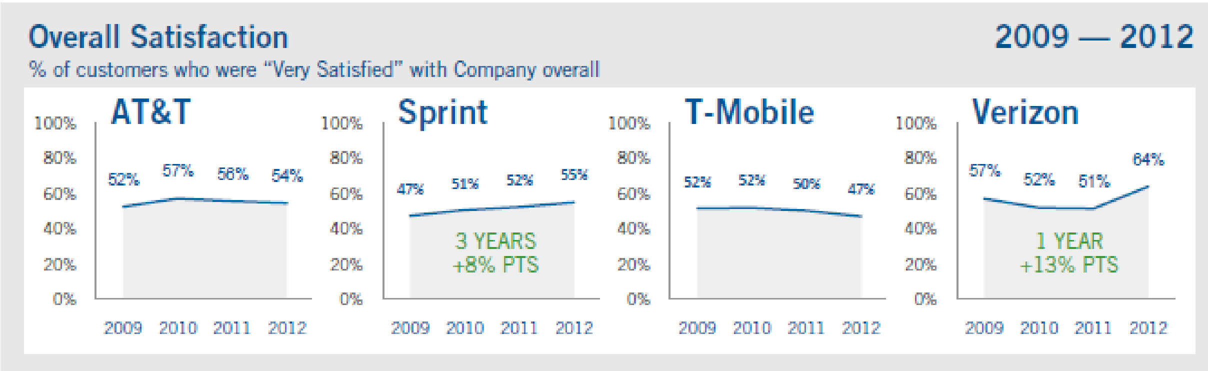


Figure 1. Overall customer satisfaction scores from NCSS as provided by Vocalabs on January 2013

The most dramatic trend over the past year is the significant improvement in Verizon's customer satisfaction and loyalty. The company posted a very significant 13-point gain in overall customer satisfaction and a six-point increase in the percentage of customers who would buy another Verizon phone if given the chance.

Apart from the overall customer satisfaction survey in regards to company services, we looked at results examining customer satisfaction for AT&T and Verizon's Wireless services (4G & LTE) as presented by Forrester on the fourth quarter of 2012, as these services are being used by the carriers to attract new customers, focusing on high broadband speed rates as a form of competitive advantage.

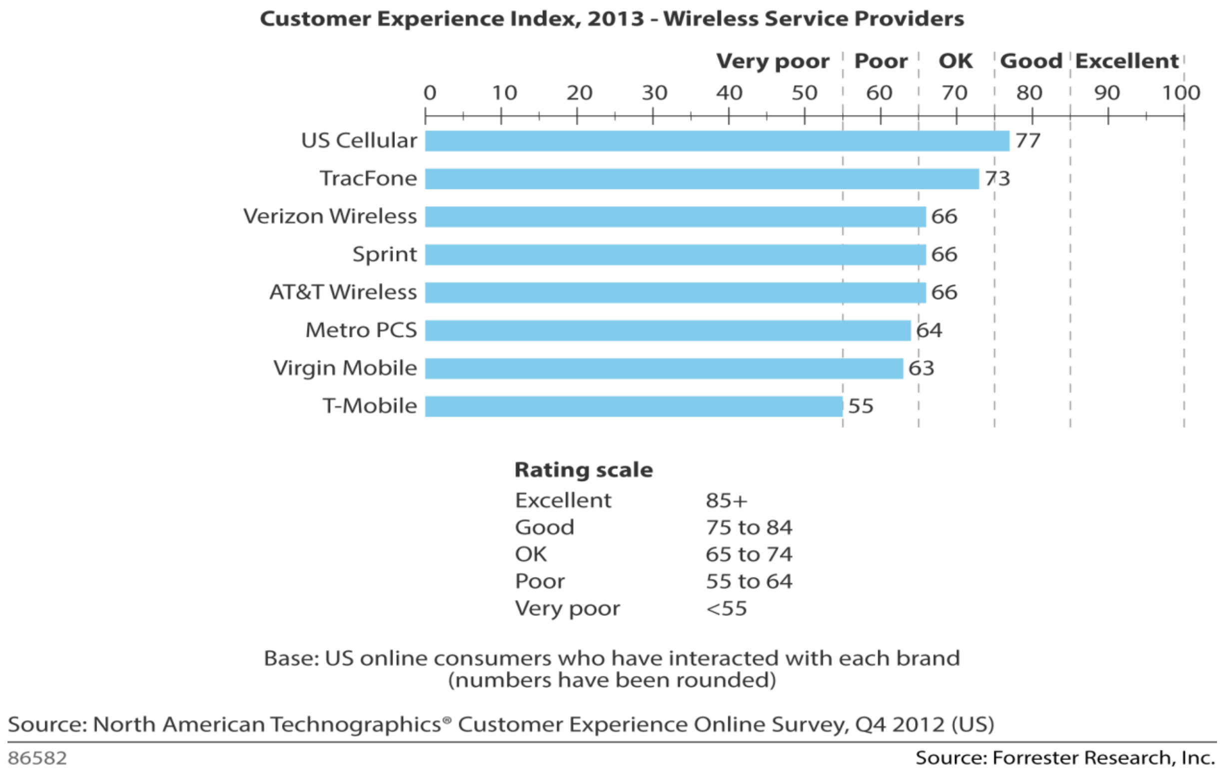


Figure 2. Customer satisfaction scores from Customer Experience Index 2013 on Wireless Service Providers.

According to Forrester, both AT&T and Verizon score 66 (ranked as OK on the rating scale).

2 Literature Review

Performing sentiment analysis and opinion mining through Twitter is a research subject that has drawn the interest of many research teams throughout the world during the past few years. The challenge to accurately predict social mood based on text mined from Twitter, still remains a big challenge and is currently being explored in various market and academic segments. O'Connor et al. (2010) connected measures of public opinion measured from polls with sentiment measured from text and found that opinions measured from polls correlate to sentiment word frequencies in contemporaneous Twitter messages. The study concludes with the potential of the use of text streams as a substitute and supplement for traditional polling. Jansen et al. (2009) investigated the overall structure of micro-blog postings, types of expressions, and sentiment fluctuations discussing the implications for organizations in using micro-blogging as part of their overall marketing strategy and branding campaigns. Mishne et al. (2006) in their study, show that, in the domain of movies, there is good correlation between references to movies in weblog posts— both before and after their release—and the movies' financial success. Furthermore, they demonstrate that shallow usage of sentiment analysis in weblogs can improve this correlation. Tumasjan et al. (2010) used the context of the German federal election to investigate whether Twitter is used as a forum for political deliberation and whether online messages on Twitter validly mirror offline political sentiment. In more detail the study found that the mere number of messages reflects the election result and even comes close to traditional election polls.

Bollen et al. (2011) argue that Twitter mood predicts the stock market. In their study they conclude that changes in the public mood state can indeed be tracked from the content of large-scale Twitter feeds by means of rather simple text processing techniques and that such changes respond to a variety of socio-cultural drivers in a highly differentiated manner which in turn are correlated or even predictive of DJIA values.

Sentiment Analysis of online text content is now in a mature state and a big part of market business analytics software such as Radian6 or IBM Cognos Consumer Insight. L.A. Times, IBM and the University of Southern California Annenberg Innovation Lab have used sentiment analysis in twitter feeds to predict the Oscars, in the 2012 ceremony. IBM along with USC Annenberg Innovation Lab (2012) performed sentiment analysis on Super Bowl XLVI analysing fan sentiment across 600,000 tweets to determine which players and teams have the most support.

3 Methodology

3.1 Data Description

We collected and analysed a set of over 135,000 tweets during the time period between February 2nd, 2013 and 26 February 2013, by utilizing the Streaming API of Twitter. The data collection process was focused on gathering tweets that were explicitly referring to the two leading mobile broadband carriers located in the broader North American area, namely AT&T and Verizon. This task was accomplished by parsing the official streaming API of Twitter through keyword filtering on the terms "AT&T" and "Verizon". The resulting dataset contained a total number of 66,000 and 70,000 tweets for AT&T and Verizon respectively, which was subsequently submitted to a series of data clearing and pre-processing operations. The data preparation process, in particular, involved text tokenization into words, elimination of English stop-words and words with less than three characters, and stem extraction from each word. Therefore, the final version of our corpus was formed by a collection of purified documents where each document contained the text from a single tweet.

3.2 Corpus Vectorization

A fundamental prerequisite in order to perform sentiment analysis through the exploitation of any machine learning algorithm is to obtain a mathematical representation of our corpus, so that each document can be treated as a point in a multi-dimensional vector space. A natural approach towards this end was the employment of the standard Vector Space Model (VSM) for our corpus, which was originally introduced by Salton, G et al (1975). The main idea behind VSM is to transform each document d into a vector containing only the words that belong to the document and their frequency by utilizing the so called "bag of words" representation. According to VSM, each document is represented exclusively by the words it contains by tokenizing sentences into elementary term (word) elements losing the associated punctuation, order and grammar information. The underlying mathematical abstraction imposed by VSM entails a mapping which transforms the original purified document to its corresponding bag of terms representation. This transformation can be formulated by the following equation:

$$\varphi: \mathrm{d} \rightarrow \varphi(\mathrm{d}) = [\mathrm{tf}(\mathrm{t}_1, \mathrm{d}), \mathrm{tf}(\mathrm{t}_2, \mathrm{d}), \ldots, \mathrm{tf}(\mathrm{t}_\mathrm{n}, \mathrm{d})] \in \mathbb{R}^\mathrm{n} \quad (1),$$

where $tf(t_i, d_j)$ is the normalized frequency of term t_i in document d_j given by the following equation:

$$tf(t_i, d_j) = \frac{f(t_i, d_j)}{\max\{f(t, d_j) : t \in d_j\}} \ (2),$$

given that $f(t_i, d_j)$ is the absolute frequency term t_i in document d_j.

Based on the adopted mathematical formulation for the fundamental notions of corpus and dictionary, such that a corpus **D** of l documents and a dictionary **T** of N terms may be represented as $\mathrm{D} = \{\mathrm{d}_1, \mathrm{d}_2, \dots, \mathrm{d}_l\}$ and $\mathrm{T} = \{\mathrm{t}_1, \mathrm{t}_2, \dots, \mathrm{t}_N\}$. Having in mind Eq. 1 and the formal definitions for the notions of corpus and dictionary, the mathematical representation for corpus in the context of VSM can be done through the utilization of the document-term matrix given by the following equation:

$$D = \begin{bmatrix} tf(t_1, d_1) & \cdots & tf(t_N, d_1) \\ \vdots & \ddots & \vdots \\ tf(t_1, d_l) & \cdots & tf(t_N, d_l) \end{bmatrix} (3),$$

where N, is typically, quite large resulting in a sparse VSM representation such that a few matrix entries are non-zero. In our approach, in order to mitigate the effect relating to the compete loss of context information around a term, we incorporated the term-frequency inverse document frequency (tf–idf) weighting scheme according to which each term t_i is assigned a weight of the form:

$$w_i = idf(t_i, D) = log \frac{|D|}{|\{d \in D : t_i \in d\}|} \ (4),$$

so that the relative importance of each term for the given corpus is taken into consideration.

3.3 Support Vector Machines

Sentiment analysis was conducted through the utilization of a state-of-the-art classifier, namely Support Vector Machines (SVMs). SVMs are non-linear classifiers that were initially formulated by Vapnik (1995), operating in higher-dimensional vector spaces than the original feature space of the given dataset. Letting $S = \{(\vec{x_i}, y_i) \in \mathbb{R}^n \times \{-1, +1\}, \forall i \in [m]\}$ be the set of m training patterns with associated binary labels, such that -1 denotes the class of negative sentiment and +1 the class of positive sentiment, the learning phase of the SVMs involved solving the following quadratic optimization problem:

$$\min_{\vec{w}, \xi, b} \frac{1}{2} ||\vec{w}||^2 + C \sum_{i=1}^{m} \xi_i \ (5.1)$$

$$s.t.\, y_i(\langle \vec{w}, \vec{x_i} \rangle + b) \geq 1 - \xi_i, \forall i \in [m] \ (5.2)$$

$$and\ \xi_i \geq 0, \forall i \in [m] (5.3).$$

Eqs. 5 correspond to the Primal optimization problem whose solution may be obtained by considering the Dual optimization problem of the following form:

$$\max_{\vec{\alpha}, \vec{\beta}} \min_{\vec{w}, \xi, b} L(\vec{w}, \vec{\xi}, b, \vec{\alpha}, \vec{\beta}) \ (6.1)$$

$$s.t.\, \alpha_i \geq 0, \forall i \in [m] \ (6.2)$$

$$and\ \beta_i \geq 0, \forall i \in [m] (6.3),$$

where

$$L(\vec{w},\vec{\xi},b,\vec{\alpha},\vec{\beta}) = \frac{1}{2}\langle\vec{w},\vec{w}\rangle - \sum_{i=1}^{m} \alpha_i y_i \langle\vec{w},\vec{x_\iota}\rangle - b\sum_{i=1}^{m} a_i y_i + \sum_{i=1}^{m} (C - a_i - b_i)\xi_i \ (7),$$

such that, $\vec{\alpha} = [\alpha_1, \dots, \alpha_m]^T$and $\vec{\beta} = [\beta_1, \dots, \beta_m]^T$ are the matrices of the non-negative Lagrange multipliers. The optimal solution to the dual optimization problem defined in Eq.6 gives rise to a discrimination function of the form:

$$g(\vec{x}) = \sum_{i=1}^{m} \alpha_i^* y_i \langle\vec{x},\vec{x_\iota}\rangle + b^* \ (8),$$

where $\{\alpha_i^*, i \in [m]\}$ and b^* denote the optimal solutions for the corresponding optimization variables. Eq.8 can be reduced to:

$$g(\vec{x}) = \sum_{i \in SV}^{m} \alpha_i^* y_i \langle\vec{x},\vec{x_\iota}\rangle + b^* \ (9),$$

where SV is the subset of training patterns associated with positive Lagrange multipliers. Given that the training patterns appear only in dot product terms of the form $\langle\vec{x_\iota},\vec{x}\rangle$, a positive definite kernel function such as $K(\vec{u},\vec{v}) = \Phi(\vec{u})\Phi(\vec{v})$ can be employed in order to implicitly map the input feature space into a higher-dimensional vector space and compute the dot product. In this paper, we utilized the Gaussian kernel function defined by the following equation:

$$K(\vec{x},\vec{y}) = exp\left(-\frac{||\vec{x}-\vec{y}||^2}{2\sigma^2}\right) \ (10).$$

4 Experimental Results

4.1 Labelled Data

Applying SVM in such a large amount of tweets requires a reasonable amount of labelled data (i.e. tweets already classified as positive, negative or neutral, based on a business perspective classification). This ensures that the SVM algorithm runs with accuracy, providing robust results that limit the amount of fault. These labelled data are in turn used by the SVM algorithm as a benchmark, in order to score the number of tweets that are in scope of the sentiment exercise. In order to create a reasonable amount of labelled data, we manually labelled a set of 7,223 collected tweets, in terms of sentiment, as positive (1), neutral (0) or negative (-1). Example tweets of each category are presented below.

Sample of Tweets marked as positive

RT @xxxxxx: AT&T is the best.	1
Just got great service from AT&T over the phone thats never happened with any other phone company #HappyHappyHappy	1
Does anyone else think those AT&T commercials with the little kids are funny and cute or is it just me?	1

Sample of Tweets marked as negative

I dont hate kids but those AT&T commercials makes me hate them	-1
Apple and AT&T is pissing me off.	-1
AT&T why u hate me	-1

Sample of Tweets marked as neutral

At the AT&T Store with my Homegirl! Next stop, StoneCrest Mall!	0
@zzzzzs: I just saw @xxxxx name on the iPhone 4S screensaver! AT&T store?	0
Do yall pay attention to these AT&T commercials?	0

The results broken down per carrier and sentiment are presented below.

Total tweets assessed (Labelled data): 7,223

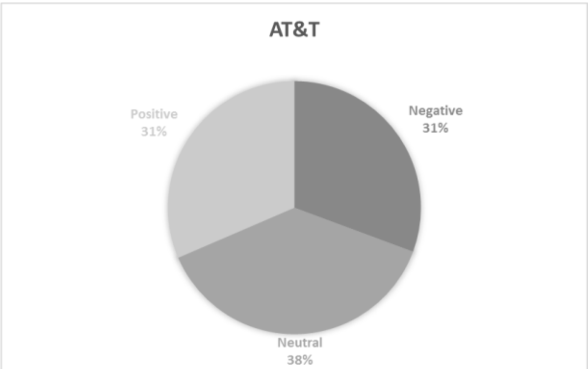
AT & T tweets assessed: 2,939

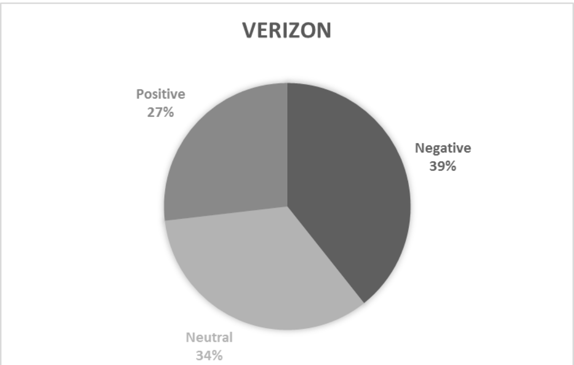
Negative (-1)	901
Neutral (0)	1113
Positive (1)	925

Verizon tweets assessed: 4,284

Negative (-1)	1684
Neutral (0)	1450
Positive (1)	1150

The diagrams below represent the above figures in percentages (rounded).





4.2 Training & Testing Data Results

In order to test the accuracy and validity of the SVM algorithm, we split the total amount of labelled tweets (7,223 Tweets) to a 95% training data - 5% testing data ratio. This resulted in the following breakdown.

AT & T	Verizon
TOTAL LABELED NEGATIVE PATTERNS: 901 **LABELED NEGATIVE PATTERNS FOR TRAINING:** 856 **LABELED NEGATIVE PATTERNS FOR TESTING:** 45 **TOTAL LABELED POSITIVE PATTERNS:** 925 **LABELED POSITIVE PATTERNS FOR TRAINING:** 879 **LABELED POSITIVE PATTERNS FOR TESTING:** 46	**TOTAL LABELED NEGATIVE PATTERNS**: 1684 **LABELED NEGATIVE PATTERNS FOR TRAINING:** 1600 **LABELED NEGATIVE PATTERNS FOR TESTING:** 84 **TOTAL LABELED POSITIVE PATTERNS:** 1150 **LABELED POSITIVE PATTERNS FOR TRAINING:** 1093 **LABELED POSITIVE PATTERNS FOR TESTING:** 57

The percentage of tweets, which are labelled data that have already been classified as positive or negative, is named Training Data. The percentage of tweets to be scored by the SVM algorithm is named Test Data. This produced a confusion matrix letting us compare how accurately the SVM algorithm classified the testing data in accordance to our already classified labelled data. The results for AT&T and Verizon on this subset of data are presented below.

Subset of Data Results	
AT& T Results	**Verizon Results**
SVM LABELED TRAINING ACCURACY: 0.914697 **SVM LABELED TRAINING CONFUSION MATRIX:** [[800 56] [92 787]] **SVM LABELED TESTING ACCURACY:** 0.923077 **SVM LABELED TESTING CONFUSION MATRIX:** [[44 1] [6 40]]	**SVM LABELED TRAINING ACCURACY**: 0.897141 **SVM LABELED TRAINING CONFUSION MATRIX:** [[1484 116] [161 932]] **SVM LABELED TESTING ACCURACY:** 0.638298 **SVM LABELED TESTING CONFUSION MATRIX:** [[63 21] [30 27]]

This experiment proved that the SVM algorithm could achieve a very good testing accuracy percentage, given the nature of the data that it ran the classification on. With this result in hand, we ran the SVM algorithm for the full data set of collected tweets (135K tweets) in order to classify all collected tweets according to their sentiment. The results for the full data set of tweets are presented below.

Full Data Set of Tweets Data Results	
AT& T Results	**Verizon Results**
Minimum decision value: -2.078615	**Minimum decision value:** -2.320534
Maximum decision value: 2.424328	**Maximum decision value:** 2.090736
Minimum Negative Decision Value -2.078615	**Minimum Negative Decision Value** -2.320534
Maximum Negative Decision Value -0.000032	**Maximum Negative Decision Value** -0.000018
Minimum Positive Decision Value 0.000050	**Minimum Positive Decision Value** 0.000048
Maximum Positive Decision Value 2.424328	**Maximum Positive Decision Value** 2.090736
Absolute Threshold Value 1.000000	**Absolute Threshold Value** 1.000000
Decision Value Based Estimated Negative Patterns 5361	**Decision Value Based Estimated Negative Patterns** 7236
Decision Value Based Estimated Positive Patterns	**Decision Value Based Estimated Positive Patterns**

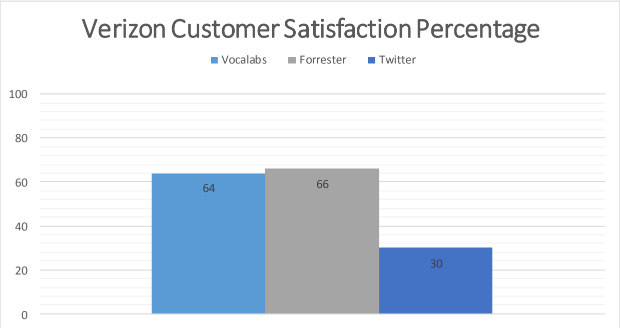
5475	3127
Decision Value Based Estimated Neutral Patterns 55135	**Decision Value Based Estimated Neutral Patterns** 60213

4.3 The Bottom Line: Comparison with Survey Results

From the results derived above, we can calculate customer satisfaction as measured from Twitter as:

[Decision Value Based Estimated Positive Patterns / (Decision Value Based Estimated Positive Patterns + Decision Value Based Estimated Negative Patterns)] * 100

Thus the figures for AT&T and Verizon are 51% and 30% respectively. The comparison with the surveys presented in the previous section, can be visualized as Figure 4 and Figure5 depict below.



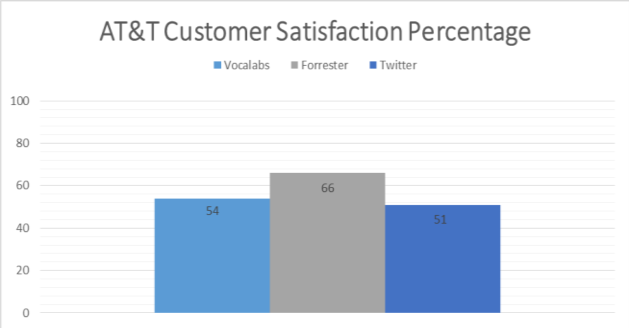


Figure 4. Customer satisfaction scores for Verizon. *Figure 5. Customer satisfaction scores for AT&T.*

The results indicate two very important findings. Running SVM algorithm for sentiment detection, in very large sets of data, can prove highly accurate if provided with a good labelled set of already scored data. The algorithm proved the ability to learn and score accurately based on the labelled data provided, given the assumption that the data set is properly cleaned, as referred in paragraph 3.1.

Second, the results showed that mining twitter for customer satisfaction can prove a very big asset for any organization if used appropriately. Given the nature of the medium, sentiment analysis in sets of data of a specific time frame, can provide useful insights about the specific period. We propose that such exercises be ran during monitored periods that the organization expects high load of conversation to arise in twitter triggered by specific events. In our case, we found that twitter users showed a positive tendency towards a specific commercial AT&T had recently launched, while showing a negative tendency towards Verizon, due to broadband problems that the service was facing for a few hours during the period monitored. This reveals that, although this approach can prove very insightful for drawing conclusions during the specific period, it shouldn't be compared with results from customer satisfaction surveys that the data collection timeframe spans during large periods of time.

5 Conclusions & Future Work

This paper aimed at examining how the two leading mobile broadband carriers located in the broader North American area, AT&T and Verizon, can benefit from monitoring and performing sentiment analysis, on tweets sent from users in Twitter, mentioning keywords in scope of, or related to the two carriers.

Our research results showed that data gathered from Twitter, if mined, cleaned and scored appropriately can prove of outmost importance, as this information depicts customer sentiment towards the respective carrier on a real-time and a more intimate or straightforward basis. Mobile broadband carriers will benefit from and improve customer satisfaction if they include such an activity in their customer satisfaction methodology.

We propose that carriers perform such an activity during periods of events that trigger twitter users to actively participate in discussions and express their opinions. These activities could be during a launch of a new commercial, launch of new services, or even disaster situations where wireless services are not responding. The amount of information that could be gathered in such situations in such a small period of time can prove salutary in situations where quick responses may need to be taken in order to maintain the churn rate low.

We conclude that comparing results from customer surveys with results gathered from twitter could prove useful as a benchmark of the validity of the results that are generated through offline telephone conversations but in no means can one method replace the other. This is due to the difference of nature of each medium used to gather customer opinion and from the authors' perspective both methods should be used complementary.

6 Acknowledgments

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Aristeia, SocioMine

References

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010, May). From tweets to polls: Linking text sentiment to public opinion time series. In Proceedings of the International AAAI Conference on Weblogs and Social Media (pp. 122-129).

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009, April). Micro-blogging as online word of mouth branding. In Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (pp. 3859-3864). ACM.

Mishne, G., & Glance, N. (2006, March). Predicting movie sales from blogger sentiment. In AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006) (Vol. 30, No. 2, pp. 301-304).

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010, May). Predicting elections with twitter: What 140 characters reveal about political sentiment. In Proceedings of the fourth international aaai conference on weblogs and social media (pp. 178-185).

Senti-meter. L.A. Times, IBM and the University of Southern California Annenberg Innovation Lab predict the Oscars. http://graphics.latimes.com/senti-meter/

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

SuperBowl analysis takes us beyond the tweets. http://asmarterplanet.com/blog/2012/02/super-bowl-analysis-takes-us-beyond-the-tweets.html

Salton, G., Wong, A., & Yang, C. S. (1975, November), A Vector space model for automatic indexing. Commun. Journal (Vol. 18, No. 11, pp. 613–620).ACM

Vapnik, V. (1995). The Nature of Statistical Learning Theory, Springer, New York.