# Using Sentiment Analysis to Detect Customer Attitudes in Social Media Comments

Ahmet Süerdem[1,],Eda Kaya[1]

[1]İstanbul Bilgi University, İstanbul, Turkey
asuerdem@bilgi.edu.tr, edakaya@gmail.com

**Abstract .**ig data analytics is increasingly replacing or complementing classical data collection methods like surveys. Business intelligence tools and systems now play a key role in decision-making and find applications in the areas such as customer profiling, customer support, market research, segmentation, brand monitoring. Transforming unstructured text data for structured quantitative analysis is an important challenge for business analytics. Sentiment analysis is a promising method in terms of transforming words into numbers to detect the tone of subjective expressions. The aim of this paper is scrutinize the role of sentiment analysis for emulating classical attitude and market research tools. We used sentiment analysis on product reviews data occurring in the e-market place in Turkey. We have applied a variety of machine learning algorithms and some term selection algorithms. Unigram approach for the term selection and Naive Bayes approach for machine learning have performed better than others. Our results suggest that sentiment analysis can be applicable for Turkish language after a rigorous text preprocessing and term selection process.

**Keywords:** Business Intelligence • Sentiment Analysis • Text Classification • Machine Learning

## 1 Introduction

Social media sites provide opportunities for gathering vast amounts of marketing intelligence information from naturally occurring data such as comments and evaluation about brands, individuals and products. Contrary to the surveys where attitude data can be collected as a one-off occasion, social media users share their opinions on a continuous basis in a naturally occurring environment. This provides management decision makers with instantly updated information and opportunity to collect data from many people which would physically impossible to do so. Although decision makers increasingly gather information from social media about the issues such as what consumers think about a new advertisement campaign, customer complaints and how consumers compare their brand to their competitors, systematic analysis of this information is still an emerging issue. Systematic analysis can establish the links between this information and corporate knowledge management systems.

The high cost of media monitoring of vast amounts of texts through human processing brings forth a need for data mining tools that can automatically analyze texts.

Accordingly, business intelligence and text analytics tools increasingly complement classical data collection instruments such as surveys and polls. These tools find growing application in the fields such as customer profiling, customer support, marketing research, market segmentation, and brand monitoring. However, although social media applications present a rich information source for data mining, data in this form is usually available as unstructured text. In this respect, Natural Language Processing and text mining techniques offer opportunities for structuring those data. The aim of this paper is to emulate conventional attitude and market research tools by using these techniques in the Turkish language context. For this purpose, we made machine learning based sentiment analysis of Facebook comments about two competing retail brands a priori categorized as positive or negative by the customers.

## 2 Theoretical Background

Automatic text analysis finds its grounds in classical content analysis [1]. However, this method usually focuses on determining factual attributes of the content by ignoring the evaluative elements in the language. Subjective aspects such as opinions, attitudes, evaluations, styles, etc. are essential for obtaining the gist of a text [2]. Accordingly, sentiment analysis appeared as a promising alternative in recent years to complement content analysis for detecting the evaluative tone of subjective expressions [3, 4]. Using text mining methods such as statistical natural language processors, part-of-speech tags, parsers and advanced lexicons, sentiment analysis claims to automatically detect whether a text contains a positive or negative tone [5]. Although sentiment analysis has started with the analysis of movie reviews [5, 6, 7, 8, 9], it now extends its domain to a variety of applications. Reviews and discussions on cars, banks, travel resorts [10], music, books, mobile phones, kitchenware, hotels [11], cameras, printers and baby strollers [12] are amongst the various topics that sentiment analysts are interested in.

Habitual method in sentiment analysis concerns developing automatic classifiers with supervised machine learning procedures applied to human annotated texts. This method is preferred because of its compelling performance for determining positive or negative tones in texts without encountering the constraints of linguistic analysis [5]. A notable alternative to this method is building semantic thesauri, which are based on processes like the counting of keywords expressing emotional polarity in a text [10]. Between these two methods, machine learning is preferred to the latter because it also covers the relevant context around the words [3, 13]. While keyword method can detect more precisely whether the chosen aspects belong to the predicted class (Precision: true positive rate), machine learning is strong in terms of retrieving the relevant instances (Recall: false negative rate) [14].

In supervised learning, the learning algorithm is trained by texts whose classes are annotated by humans. The algorithm derives a model from the texts selected as a training set through the relationship between input features (terms) and their class assignments. Features are weighted to the extent they demarcate between classes. The model is cross-validated through a test set not covering the text units in the training set. The overlap between actual and model predicted classes is accepted as the accuracy criteria [15].

While machine learning is commonly used in English, in other languages keyword method is the dominant approach. For example, Mihalcea [16] conducted sentiment analysis by translating mood state words from English to Romanian; Mathieu [17] to French; Cho and Lee [18] to Korean, and Abbasi [19] to Arabic. The limited success of the keyword approach turned the researchers of non-English languages to machine learning. For example, Kanamaru [20] has developed a machine learning algorithm for determining emotional states for Japanese language, while Xu [21] did the same for the Chinese. Pang [5] remarked that success of automatic classification for the languages other than English depends on the feature type selection such as unigrams or n-grams as well as the selection of best classifier algorithms. In this vein, this paper aims to contribute to the literature by experimenting with different machine learning algorithms on Turkish texts and testing the effect of different feature types (unigrams and n-gram) on the classification accuracy.

Sentiment analysis in Turkish is rare except a few cases (i.e. Boynukalın [22] classified Turkish texts according to emotional state expressions through machine learning; Eryiğit [23] compared machine learning with dictionary based methods). They found that feature selection process is the most critical decision for machine learning in Turkish language. Performance of feature selection depends on text pre-processing procedures such as stemming and lemmatizing, detection of synonyms, exclusion of function words. However this is a complex procedure for morphologically rich languages such as Arabic and Turkish, which requires special processes [19]. Moreover, unconventional language use in the social media context makes this issue more complex. Hence, in this paper we propose different algorithms to solve some of these problems. Our method consists of three steps: text pre-processing, feature selection and modeling.

## 3 Method

### 3.1 Text preprocessing

On preprocessing phase, we have particularly focused on the symbol use in the social media. For example, emoticons like smileys express information on user's emotions. To include them as features, we converted facial expressions (emoticons) into meaningful words such as:

:-) positive; :=) positive; :D very positive ; <3 love

Besides emoticons, unconventional abbreviations are frequently used in social media. We converted them into their synonyms for the algorithm not to detect them as different features:

ltf: please (lütfen); pls: please (lütfen); tşk: thanks (teşekkürler); inş: hopefully (inşallah); abv: damn (allah belanı versin).

Another unique social-media-only expression that defines the strength of sentiment is repeating characters. For example, nooooo is stronger than no.

The next pre-processing step involves spell checks. For this purpose we have used "Zemberek", an open source Turkish NLP library (https://github.com/ahmetaa/zemberek-nlp ) which has a spelling dictionary with more than one million words. The example output presents original word and several suggestions (Table 1).

**Table 1.** The representative output for Zemberek's results

| originalWord | suggested1 | suggested2 | suggested3 |
|---|---|---|---|
| Tartisilir | tartışılır | Tartışılır | |
| Ihtiyacimiz | ihtiyacımız | | |
| Paylasınca | paylaşınca | Paylaşınca | |
| Seninle | Seninle | Şeninle | şeninle |
| Arkadaslarn | | | |

In the "originalWord" area the actual form of the word is seen. The fields that begin with "suggested.." label, show alternative suggestions for the word.

Hence, during the preprocessing phase we have converted frequently used emoticons and abbreviations into words, corrected misspellings and edited the repeating characters.

## 3.2 Feature selection

This step involves the generation and selection of features to be used in the classification algorithms. We compared two different approaches: i) "Bag of words": features are generated disregarding syntax and word sequences (unigrams) for keeping the variability in the text. ii) N-gram: takes into account word sequence patterns (syntax) and phrases. For each syntactic pattern and phrase, it generates a separate feature. Social media texts are usually short and do not include many N-grams. While N-grams have been found to perform better for longer texts, they can discount some semantic information in shorter texts in accordance with the Zipf's law [5]. In table 2. we compare the "information gain" scores obtained from both approaches. Information Gain is an established and empirically tested method for high-dimensional feature selection. It tells us the information contribution of a feature within a vector [24]. In our analysis, Information Gain score obtained from the "bag-of-words" approach for the unigram word "beautiful" is almost equal to the sum of the word "beautiful" and the n-gram "very beautiful". After repeating the same operation for other features as well, we have found that the "n-gram" approach does not provide efficient results for social media data. Hence, we have decided to proceed to the modeling phase with the "bag-of-words" approach.

**Table 2.** "Information Gain" scores for bag of words and N-Gram approaches

| **Ranked attributes of N-gram:** | | | **Ranked attributes of bag of words** | | |
|---|---|---|---|---|---|
| 0.07414 | 11130 | güzel | 0.14844 | 403 | güzel |
| 0.07136 | 4191 | bir | 0.08109 | 1293 | harika |
| 0.07113 | 26802 | çok güzel | 0.07198 | 688 | pahalı |
| 0.05936 | 5013 | bu | 0.05936 | 201 | bu |
| 0.05771 | 18932 | pahalı | 0.05658 | 1466 | olmuş |
| 0.04908 | 25612 | yok | 0.0541 | 183 | bir |
| 0.04371 | 36 | 1 | 0.04908 | 940 | yok |
| 0.04358 | 27643 | ürün | 0.04614 | 869 | var |
| 0.04066 | 23712 | ve | 0.04389 | 981 | çok |
| 0.04041 | 23528 | var | 0.04371 | 4 | 1 |
| 0.03682 | 17380 | neden | 0.04358 | 1011 | ürün |
| 0.0364 | 11663 | harika | 0.0433 | 838 | teşekkürler |
| 0.03486 | 6543 | da | 0.04066 | 875 | ve |
| 0.03443 | 18367 | olmuş | 0.04033 | 783 | sorun |
| 0.03391 | 21444 | sorun | 0.03682 | 625 | neden |
| 0.03219 | 14812 | ki | 0.03352 | 1438 | mükemmel |
| 0.03076 | 246 | 2 | 0.03339 | 124 | aynı |
| 0.03076 | 13815 | kadar | 0.03219 | 533 | ki |
| 0.03076 | 1464 | alışveriş | 0.03076 | 81 | alışveriş |
| 0.02998 | 27190 | çıktı | 0.03076 | 502 | kadar |
| 0.02998 | 27102 | çünkü | 0.03076 | 14 | 2 |
| 0.02998 | 16020 | lütfen | 0.03072 | 701 | pozitif |
| 0.02998 | 4940 | bozuk | 0.03025 | 244 | da |
| 0.02998 | 2559 | aynı | 0.02998 | 574 | lütfen |
| 0.02885 | 12452 | i | 0.02998 | 986 | çünkü |
| 0.02772 | 6967 | de | 0.02998 | 197 | bozuk |
| 0.02659 | 18147 | olduğunu | 0.02998 | 987 | çıktı |
| 0.02633 | 10478 | gibi | 0.02885 | 268 | değil |

In the next step, we calculated the weight of each feature (word) according to the inverse document frequency ($idf_t$) transformation method. $Idf_t$ is a measure of the relative importance of a term calculated as the concentration of a term in particular documents compared to its appearance in all the documents:

$$idf_t = \log \frac{N}{df_t} \quad (1)$$

$Idf_t$ transformation shows us the relative frequency of a term compared to its appearance in all documents. Words with lower $idf_t$ carry lower information values. For example, while function words like "the" have lower $idf_t$ scores; opinion expression words like "beautiful" which are not distributed along all documents have higher scores.

In our study, we have tested the performance of the models including selection of features exceeding an $idf_t$ score threshold against a baseline including all features.

### 3.3 Modeling

We used Naïve Bayes and Support Vector Machines for classification of texts and compared the results with and without feature selection. Classification procedures were performed on positive and negative texts.

**Table 3.** Classification results for Brand-1 (355 texts)

| | Accuracy rate | ROC area | Prec. (Pos) | Prec. (Neg) | Recall (Pos) | Recall (Neg) |
|---|---|---|---|---|---|---|
| Support Vector Mach | 90.44% | 0.905 | 0.924 | 0.883 | 0.893 | 0.917 |
| Naive Bayes | 92.53% | 0.983 | 0.901 | 0.958 | 0.966 | 0.879 |
| SVM (selected features) | 85.97% | 0.856 | 0.836 | 0.893 | 0.916 | 0.796 |
| Naive Bayes (selected features) | 91.44% | 0.964 | 0.878 | 0.957 | 0.966 | 0.847 |

**Table 4.** Classification results for Brand-2 (425 texts)

| | Accuracy rate | ROC area | Prec. (Pos) | Prec. (Neg) | Recall (Pos) | Recall (Neg) |
|---|---|---|---|---|---|---|
| Support Vector Mach | 85.41% | 0.789 | 0.838 | 0.923 | 0.972 | 0.606 |
| Naïve Bayes | 86.82% | 0.912 | 0.856 | 0.909 | 0.969 | 0.657 |
| SVM (selected features) | 88.23% | 0.823 | 0.858 | 0.968 | 0.99 | 0.657 |
| Naïve Bayes (selected features) | 86.82% | 0.92 | 0.858 | 0.901 | 0.965 | 0.664 |

Models with feature selection were found to give more accurate results. We select the Naïve Bayes model with feature selection to build the sentiment classifiers. Naïve Bayes approach is more robust to the number of features, hence more efficient for detecting classes with the lower number of features (Table 3 and 4).

### 3.4 Results

The results from application of the Naïve Bayes model results on Brand-1 and Brand-2 are shown in Table 5 and 6.

**Table 5.** Application of the Naïve Bayes model results on Brand-1

| Brand-1 | Frequency | Percentage |
|---|---|---|
| Total | 2834 | |
| Negative | 846 | 30% |
| Positive | 1594 | 56% |
| Negative comments about the brands | 90 | 10,6% |
| Positive comments about the brands | 55 | 3,5% |

**Table 6.** Application of the Naïve Bayes model results on Brand-2

| Brand-1 | Frequency | Percentage |
|---|---|---|
| Total | 1906 | |
| Negative | 360 | 19% |
| Positive | 1392 | 73% |
| Negative comments about the brands | 52 | 14% |
| Positive comments about the brands | 70 | 19% |

As the tables suggest, Brand-2 products received more positive comments compared to Brand-1. Our results suggest that sentiment analysis can be an alternative or complementary method for attitude polls or surveys. Of course this approach carries some limitations in terms of sampling and instrument design. For example, the results of the analysis depend on the individuals who write comments on the Web pointing to a potential sampling bias. Moreover, measurement instrument (sentiment classifier) although highly reliable (giving the same results on the same sample) can pose some validity problems (measuring attitudes rather than other constructs such as idiosyncracies in language use). After rigorous cross-validating with other instruments like surveys and interviews, sentiment analysis has a potential for attitude measurement in the social media.

## 4 Conclusion

Application of sentiment analysis to the Turkish language and social media context is a challenging issue. Our results provide some evidence that with comprehensive preprocessing effort and machine learning algorithms, sentiment analyses can be effectively performed in Turkish Language. Future studies can improve the performance not only by analyzing the general attitudes towards brands, but also the attitudes towards specific brand attributes. Furthermore, adding a third neutral value to negative-positive poles can enrich the scale for precision purposes. The results of this study provide support to the argument that sentiment analysis can be used to complement survey based attitude research.

## References

1. Krippendorf, K., Content Analysis; An Introduction to its Methodology, 3rd Edition, Thousand Oaks, CA: pp. 441, Sage (2012)
2. Shanahan, J., Qu, Y., & Wiebe, J. (Eds.). Computing attitude and affect in text: Theory and applications. Dordrecht, the Netherlands: Springer (2006)
3. Choi, Y., Breck, E., & Cardie, C. Joint extraction of entities and relations for opinion recognition. In Proceedings of Empirical Methods in Natural Language Processing (EMNLP). Sydney, Australia (2006)

4. Kim, S.-M., & Hovy, E. Extracting opinions expressed in online news media text with opinion holders and topics. In Proceedings of the Workshop on Sentiment and Subjectivity in Text at the Joint COLINGACL 2006 Conference pp. 1–8. Sydney, Australia (2006)
5. Pang, Bo, Lee, Lillian, and Vaithyanathan, Shivakumar.. Thumbs up? Sentiment classification using Machine Learning techniques. In Proceedings of Conference on Empirical Methods in NLP, pp. 79-86 (2002)
6. Tong, Richard M. An operational system for tracking opinions in on-line discussions. In Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification, 1-6. New York, NY: ACM. (2001)
7. Dave, Kushal, Lawrence, Steve, and Pennock, David M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the Twelfth International World Wide Web Conference (WWW 2003). Budapest, Hungary (2003)
8. Mullen, Tony, and Collier, Nigel. Sentiment analysis using support vector machines with diverse information sources. In Proceedings of Conference on Empirical Methods in Natural Language Processing. Barcelona, Spain (2004)
9. Whitelaw, Casey, Garg, Navendu, and Argamon, Shlomo.. Using Appraisal groups for sentiment analysis. In Proceedings of ACM SIGIR Conference on Information and Knowledge Management (CIKM 2005), pp. 625-631. Bremen, Germany (2005)
10. Turney, Peter. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of 40th Meeting of the Association for Computational Linguistics, pp. 417-424 (2002)
11. Taboada, Maite, and Grieve, Jack. Analyzing appraisal automatically. In Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07), eds. Yan Qu, James G. Shanahan and Janyce Wiebe, 158-161. Stanford University, CA: AAAI Press. (2004)
12. Bloom, Kenneth, Garg, Navendu, and Argamon, Shlomo. Extracting Appraisal Expressions. Paper presented at NAACL HLT 2007, Rochester, NY. (2007)
13. Breck, E., Choi, Y., & Cardie, C. Identifying expressions of opinion in context. In M. M. Veloso (Ed.), Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI-07). Hyderabad, India. (2007)
14. Wilson, T., Wiebe, J., & Hoffmann, P. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing HLT-EMNLP-2005. Vancouver, Canada (2005)
15. Witten, Ian H., and Frank, Eibe. Data Mining: Practical Machine Learning Tools and Techniques. San Francisco: Morgan Kaufmann. (2005)
16. Mihalcea, R., Banea, C., & Wiebe, J. Learning multilingual subjective language via cross-lingual projections. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL'07) pp. 968–975, Prague, Czech Republic (2007)
17. Mathieu, Y. A computational semantic lexicon of French verbs of emotion. In J. Shanahan, Y. Qu, & J. Wiebe (Eds.), Computing attitude and affect in text: Theory and applications pp. 109–123, Dordrecht, the Netherlands: Springer. (2006)
18. Cho, Young Hwan, and Lee, Kong Jo. Automatic affect recognition using natural language processing techniques and manually built affect lexicon. IEICE Transactions on Information and Systems 89: 2964-2971. (2006)
19. Abbasi, Ahmed, Chen, Hsin-Hsi, and Salem, Arab. Sentiment Analysis in Multiple Language: Feature Selection for Opinion Classification in Web Forums. ACM Transactions on Information Systems 26. (2008)

20. Kanamaru, T., Murata, M., & Isahara, H. Japanese opinion extraction system for Japanese newspapers using machine-learning method. In Proceedings of NTCIR-6 Workshop Meeting. Tokyo, Japan. (2007)
21. Xu, R., Wong, K.-F., & Xia, Y. Opinmine— opinion analysis system by for NTCIR-6 pilot task. In Proceedings of NTCIR-6 Workshop Meeting. Tokyo, Japan. (2007)
22. Boynukalın, Z. Emotion analysis of Turkish texts by using machine learning methods. Middle East Technical University. (2012)
23. Eryigit, G., Cetin, F. S., Yanık, M., Bilgi, T. G., Temel, T., & Çiçekli, I. Turksent: A sentiment annotation tool for social media. In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse pp. 131-134, (2013)
24. Forman, G. An extensive empirical study of feature selection metrics for text classification JMLR 3 1289-1306 (2003)