Emotions: Extracting Sentiments from Customers Reviews

Justin R. Namba
Rochester Institute of Technology
Henrietta, New York
jrn1325@rit.edu

Anmol K. Jaising
Rochester Institute of Technology
Henrietta, New York
akj9051@rit.edu

Vivek M. Gupta
Rochester Institute of Technology
Henrietta, New York
vg4838@rit.edu

ABSTRACT

Sentiment analysis has been widely studied in the field of data mining. It refers to the systematic extraction of subjective information such as emotions. There are various types of sentiment analysis, but in this research, we focus on the polarity of the customers' reviews. We ran three algorithms to detect if a review is positive, negative, or neutral. Those algorithms were VADER, Naive Bayes, and K-Nearest Neighbors (K-NN). After implementing them, we discuss our findings and enumerate future steps to improve the results we obtained.

CCS CONCEPTS

• **Computing methodologies** → **Topic modeling**; • **Applied computing** → **Consumer products**.

KEYWORDS

datasets, lexicon-based approach, text tagging

ACM Reference Format:
Justin R. Namba, Anmol K. Jaising, and Vivek M. Gupta. 2018. Emotions: Extracting Sentiments from Customers Reviews. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Sentiment analysis, also known as opinion mining, refers to the systematic extraction of subjective information such as emotions [2]. There are four main types of sentiment analysis: (1) fine-grained, (2) emotion detection, (3) aspect-based, and (4) intent analysis. The first one involves determining the polarity (positive, negative, neutral) of an opinion. Emotion detection is used to identify specific emotional states (happy, angry, sad). Aspect-based identifies an opinion regarding a specific element of a product. The intent analysis seeks to determine the intention that is expressed in a message. Moreover, there are two major sentiment analysis algorithms: (1) rule-based approach and (2) automatic approach. Rule-based identifies the subjectivity and polarity of an opinion using several operations including stemming, tokenization, Part of Speech (POS) tagging, parsing, and lexicon analysis. Automatic sentiment analysis uses machine learning to extract the gist of a text through classification algorithms such as k-nearest neighbors (K-NN), linear regression, Naive Bayes, and support vector machines (SVMs). Many companies use these types of sentiment analysis approaches and algorithms to obtain the customers' opinions through surveys responses, reviews, online and social media for several purposes such as marketing, customer satisfaction, or political campaigns.

In this research, we decided to apply fine-grained sentiment analysis to hotel review datasets. For this purpose, we collected three datasets on hotels and mobile phones from the data.world and Kaggle websites. The three datasets are medium-size and consist of a dozen features. The first dataset is about reviews of hotels. It consists of 26 columns and nearly 10,007 rows, which represent the number of attributes and the number of records, respectively. The attributes are *id, dateAdded, dateUpdated, address, categories, primaryCategories, city, country, keys, latitude, longitude, name, postalCode, province, reviews.date, reviews.dateAdded, reviews.dateSeen, reviews.rating, reviews.sourceURLs, reviews.text, reviews.title, reviews.userCity, reviews.userProvince, reviews.username, sourceURLs, websites.* From a quick observation of the dataset, we presume that *reviews.rating, reviews.text*, and *reviews.title* will be the most pertinent features for extracting customers' opinions. The second dataset contains 400,000 reviews of unlocked mobile phones sold on Amazon. It has six attributes: *Product Name, Brand Name, Price, Rating, Reviews*, and *Review Votes.* The third dataset is a collection of reviews of hotels again. It consists of 10K entries and six attributes. The attributes are *Id, reviews.date, reviews.title, reviews.text, reviews.rating* and *reviews.username.*

After extracting these datasets from their source, we apply suitable data cleansing and data pre-processing steps to prepare the data for fine-grained sentiment analysis. In terms of implementation, we use both a rule-based approach and a machine learning approach to determine the polarity of the customers' reviews. We want to know whether the opinion customers' opinions are positive, negative, or neutral. For the rule-based approach, we use VADER [1], a lexicon sentiment analysis tool designed to analyze social media posts. For the machine learning approach, we use classification techniques like k-nearest neighbors (K-NN) and Naive Bayes for comparison.

The rest of the paper is organized as follows. Section 2 summarizes the cleansing operations executed on the data. Section 3 describes the approach and algorithms selected to mine the data. Section 4 lists common challenges and lessons learned during the project. Section 6 concludes the paper and enumerates potential future works.

2 DATA CLEANSING & DATA INTEGRATION

We analyzed the datasets and extract pertinent attributes for sentiment analysis. For the hotel dataset, we picked five attributes: *id, name, reviews.rating, reviews.text, reviews.title.* Then we cleansed

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

the data by removing all the rows that had at least one empty cell. We took a sample of 10,000 rows.

3 DESIGN & IMPLEMENTATION

3.1 Rule-based Approach

In a rule-based approach, there are two lists of words. One of them includes only the positive one, and the other includes the negatives. The algorithm goes through the text to find the words that match the above criteria. After that, the algorithm calculates which type of words is more prevalent in text. If there are more positive words, then the text is deemed to have a positive polarity. The same goes for negative words. However, if none of the words are positive or negative, then the text is deemed neutral.

3.1.1 VADER. We chose Valence Aware Dictionary and sEntiment Reasoner (VADER) as a rule-based tool. The advantage of VADER is that it is sensitive to the polarity and intensity of the sentiments in the text. It not only tells about the positivity, negativity, and neutrality of the text, but also how positive, negative, or neutral a sentiment is. Moreover, it uses the compound score, a metric that calculates the sum of the valence scores of each word in the lexicon which have been normalized between -1 (most extreme negative) and +1 (most extreme positive). The overall sentiment of the text is positive if the compound score is greater or equal to 0.05, negative if the compound score is less than or equal to -0.05, and neutral otherwise.

3.2 Machine Learning Approach

3.2.1 K-Nearest Neighbors. K-Nearest Neighbors (K-NN) is an unsupervised machine learning technique that classifies the k-closest points to a given data point in terms of a given metric. We classified sentences nearest to the closest review/rating. At every iteration and by classifying the data into several steps, we can find an optimal new sample point.

3.2.2 Naive Bayes. Naive Bayes model applies the Bayes theorem along with an assumption that all the features are independent and there is no relationship between them. Naive Bayes works well in text classification methods. It requires fewer data to train the model and identify the features of the model. Moreover, it proves to be fast than other sophisticated models to get itself trained. In Naive Bayes, the probability of a label given a text document can be found using the joint probability of labels and words. Mathematically speaking,

P(label|text) = P(text|label)/P(text).

Furthermore, by assuming that the probability of features is independent, the probability can be calculated as:

P(label|x) = P(label).$\prod_{i=1}^{n} P(x_i|C)$

4 RESULTS

We tested three algorithms on three datasets and described the results below.

4.0.1 VADER. For the first hotel dataset, we applied the VADER algorithm on two attributes: *review.title* and *review.text*. The results that we obtained for *review.title* can be seen in Table 1. For the sake of space, we only show the results of five reviews. The table contains the title of the review, as well as its polarity and intensity. The polarity can be positive, negative, and/or neutral. The intensity represents the level of positivity or negativity of the text. For Doc 1, 2, 3, and 5, the overall polarity corresponds to the different levels of sentiment extracted from *review.title*. However, there was an anomaly for the fourth document. The *review.title* was 55% neutral, but the overall polarity ended up positive.

Table 2 shows the results from using *review.text*. For this feature, we only reported the mode (most frequent polarity) because VADER is only applicable to a sentence and not a paragraph.

Table 3 shows the results using *review.title* from the second hotel dataset. This output was like the one we saw in the first hotel dataset. Moreover, there are two instances that we see the distribution of polarity does not correspond to the overall polarity. The overall polarity should be neutral for the first two documents, but the negative and positive, respectively. The phenomenon can mean that when the percentage of neutral polarity is like either the percentage of positive or negative polarity, then the algorithm ignores the neutrality.

Table 4 has similar results as Table 2. We see that most of the reviews are positive.

The third dataset was about Amazon products and only had one string feature: *Review*. The results after applying VADER are seeing in Table 5. Most of the reviews in the sample selected were neutral.

4.0.2 K-Nearest Neighbor. For each of the three datasets, we used *ratings* and *reviews* as the train and test data for the k nearest neighbor algorithm, respectively. Each dataset was split on a 70:30 ratio. For ease of understanding, a sample set of five instances are selected and tests are performed on them.

The k-Nearest Neighbour algorithm is applied to the first dataset, Hotel 1. The parameters to be considered were the *review.rating* and *review.text*. Results from Table 6 show that most of the reviews were very positive, with a few very negative and neutral ones.

Table 7 shows resemblance in terms of results from Table 1. Although the number of more positive reviews is greater than the number of less negative reviews, the polarity of the reviews was evenly balanced.

The k-Nearest Neighbour algorithm is applied to the third dataset, the Amazon mobile reviews. Results from Table 8 show that most of the reviews were very positive, thus meaning customers were satisfied with what they had.

4.0.3 Naive Bayes. For all three datasets, we considered two attributes: *review rating* and *review text* to classify a review using Naive Bayes. While training, the labels were assigned according to the rating where reviews with rating greater than 3.8 were considered positive, those between 3.0 to 3.8 were considered neutral and the ones having a rating smaller than 3.0 were considered negative. We trained the Naive Bayes classifier on a sample data, then determined its accuracy when applying to the test data. We obtained the results on four custom reviews for each dataset and the top 5 most informative features that helped in classifying the reviews. The results of the three datasets can be seen below.

Table 9 shows results on Hotel1 dataset. The accuracy was 62% and the top features identified were *unacceptable, spray, mildew, filthy, mess* labeled as negative features. The results show that the fourth review was classified as positive instead of being neutral as

Doc	Title Review	Negative	Neutral	Positive	Compound Score	Overall Polarity
1	measure.	0.0%	100.0%	0.0%	0.0%	Neutral
2	disappointed	100.0%	0.0%	0.0%	-47.7%	Negative
3	great hotel would stay	0.0%	42.3%	57.7%	62.5%	Positive
4	fantastic hotel ..... let staff	0.0%	52.6%	47.4%	55.7%	Positive
5	seattle excellence	0.0%	19.6%	80.4%	62.5%	Positive

Table 1: VADER: Title Reviews from Hotel 1

Doc	Text Review	Mode
1	first last stay property...	Negative
2	wife stayed hotel two nights, arriving tuesday...	Positive
3	family 5 travelling australia...	Neutral
4	visted san diego hotel solamar first leg honeymoon...	Positive
5	hotel staff service great...	Positive

Table 2: VADER: Text Reviews from Hotel 1

Doc	Title Review	Negative	Neutral	Positive	Compound Score	Overall Polarity
1	ok place short visit	0.0%	57.7%	42.3%	29.6%	Positive
2	staff accommodating problems	47.4%	52.6%	0.0%	-40.2%	Negative
3	terrible customer service	60.8%	39.2%	0.0%	-47.7%	Negative
4	lake delton resort	0.0%	100.0%	0.0%	0.0%	Neutral
5	150 day ridiculous was	55.6%	44.4%	0.0%	-36.1%	Negative

Table 3: VADER: Title Reviews from Hotel 2

Doc	Text Review	Mode
1	inn lake ok tourist area standards...	Positive
2	bad: came evening long business journey...	Positive
3	stayed hotel several times every time...	Negative
4	helpful staff check...	Positive
5	book prepaired!...	Neutral

Table 4: VADER: Text Reviews from Hotel 2

Doc	Title Review	Average Rating
1	All great around	Very Positive
2	We Miss You Already	Neutral
3	Wonderful visit	Very Positive
4	The Most Disgusting Motel!!!!!	Very Negative
5	This hotel was perfect for us	Very Positive

Table 6: K-Nearest Neighbours on Hotel 1

Doc	Text Review	Mode
1	first nokia phone owned love...	Neutral
2	exactly expecting!...	Neutral
3	phone 3 weeks mostly love aside annoying little things...	Positive
4	've bought factory unlocked device, received locked version...	Negative
5	liked phone...	Neutral

Table 5: VADER: Text Reviews from Amazon

Doc	Title Review	Average Rating
1	It is also good to here	Very Positive
2	We hope to see you back..	Neutral
3	great hotel would stay	Very Positive
4	fantastic hotel .....let staff	Very Positive
5	Settle excellence	Very Positive
6	We would hesitate to ..	Very Negative

Table 7: K-Nearest Neighbours on Hotel 2

the hotel1 dataset consisted mostly of positive and negative labels providing less information for neutral reviews to get trained on.

Table 10 shows results on Hotel2 dataset. The accuracy was 71% and the most informative features identified were *crack, daylight, messy, walls, poorly* out of which the first four were labeled as neutral and the last feature was considered neutral. The second review was misclassified as positive instead of negative.

Table 11 shows results on the Amazon dataset. The accuracy was 74% and the most informative features were *waste, disappointed, garbage, horrible, excellent* out of which the first four were labeled as negative and the last feature was labeled as positive. The second

Doc	Review.Text	Overall Polarity
1	Cute and very good windows phone, it's very simple	Very Positive
2	After using this phone for approximately 2 months, the screen started..	Very Positive
3	i bought this phone to use when i travel to the carribbean and it worked..	Negative
4	Everthing was ok, but dont have the instruction. Regards	Very Positive
5	seattle excellence	Positive
6	nice	Very Positive

Table 8: K-Nearest Neighbours on Amazon dataset

Doc	Text Review	Polarity
1	The food was delicious...	Positive
2	The rooms were not clean...	Negative
3	The hotel was expensive but...	Positive
4	The quantity of the food was not enough	Positive

Table 9: Naive Bayes on Hotel1 dataset

Doc	Text Review	Polarity
1	The food was delicious...	Positive
2	The rooms were not clean...	Positive
3	The hotel was expensive but...	Positive
4	The quantity of the food was not enough	Neutral

Table 10: Naive Bayes on Hotel2 dataset

Doc	Text Review	Polarity
1	Very clean set up and easy set up.	Positive
2	The camera quality was not that great...	Positive
3	There was only one little blemish on the side, but who cares	Negative
4	I'm really disappointed about my phone and service	Negative

Table 11: Naive Bayes on Amazon dataset

review was misclassified as positive instead of negative. The reason is that it is unable to handle the negations like 'not' in the reviews.

5 LESSONS LEARNED

We performed a thorough fine-grained sentimental analysis on three different datasets. We learned that understanding the data before performing any analysis was necessary because domain knowledge plays an influential part in deciding the work to be done. Also, data cleaning and pre-processing proved to be a necessity while performing analysis on all three datasets. It followed a pattern of tokenization, domain-based cleaning, lemmatization, and POS-tagging. In terms of implementation, VADER gives more precise information because it not only gives polarity but also the intensity of an opinion, in addition to the compound score for accuracy. The algorithm K-NN is time-consuming, but it gives high accuracy. Naive Bayes proved to be fast than most of the models but assumes that features are independent. This is generally not the case and as a result, gives an average accuracy.

6 CONCLUSION & FUTURE WORKS

Sentiment analysis is a field of natural language processing (NLP) that identifies the sentiment of public opinions and reviews which could help businesses make better decisions. In this project, we used two approaches to determine the polarity of text reviews using two hotel datasets and one Amazon product dataset. The first approach was rule-based using the algorithm VADER that gives us a normalized valence score between +1 and -1 defining the polarity of a sentiment. The other approach consisted of using machine learning classification algorithms like Naive Bayes and k-Nearest Neighbors. None of the accuracy values were high; they centered around 70%.

For future works, we plan to use more rule-based approaches like regular languages and expressions that can be used as a search text. We will also try other machine learning approaches such as Random Forest and Decision Trees to obtain a higher accuracy.

REFERENCES

[1] C. Hutto and E. Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *ICWSM*.
[2] Azeema Sadia, Fariha Khan, and Fatima Bashir. [n.d.]. An Overview of Lexicon-Based Approach For Sentiment Analysis. ([n. d.]).