Analytics Insight into Customer Reviews Using Text Mining Techniques and Machine Learning Algorithms: A Case Study of SAMSUNG Customer Reviews.

Janvier Nkundukozera

Abstract— With the today's digital world and the extensive use of microblogging systems like companies' websites consumer reviews space, Twitter, Facebook and others, customers express their perception towards any products or brands. Product manufacturers can employ those reviews to analyze how customers are satisfied with their products. Though customer reviews are paramount in the manufacturer's perception analysis, they are bulky in size and often unstructured. Therefore, it is hard and time-consuming to analyze all customer reviews. In this paper, we focus on analyzing Samsung customer reviews and building predictive models, which can be used to predict the future perception of such customers about a product based on their current reviews. Our modeling employed five machine-learning algorithms namely Classification and Regression Trees (CART) for their high interpretability, Random Forests (RFs), Naïve Bayes, Support Vector Machines (SVMs) and Maximum Entropy for their improved accuracy and robustness against overfitting. In our study, maximum entropy and random forests classifiers outperformed other classifiers in F-measure and recall respectively. Evaluation of models is done based on four metrics, namely accuracy, precision, recall, and F-measure. Furthermore, our study finds that Samsung customers like Samsung products and are willing to recommend them to new customers only that some product defects and services they are offered seem to hinder their trust.

Index Terms—CART, MaxEnt, Naïve Bayes, Random Forests, Sentiment Analysis, SVMs, Text mining.

- *Janvier Nkundukozera is currently pursuing master's degree program in computer science and technology engineering in Xidian University, China, PH-15929935176. E-mail: nkundajanv02@gmail.com*

1. Introduction

Text mining or knowledge discovery from text data refers to the process of extracting interesting and non-trivial patterns or knowledge from text documents. These text documents are not-structured or semi-structured. Text mining includes branches such as information extraction and sentiment analysis among others. Information extraction task includes tokenization, identification of named entities, sentence segmentation and part-of-speech tagging [1]. Sentiment analysis is a study that focuses on mining of positive or negative attitudes, opinions, views and emotions from the text, speech, tweets, and database sources through Natural Language Processing (NLP) and machine learning techniques. Sentiment analysis searches to answer questions like what do customers think about company's products and services. Do people like or dislike the products of our company? Moreover, what would people prefer company's product to be like? [2]. In 2009, the International Data Corporation (IDC) reported that, by approximation, 80% of the data in any organization is text-based [3]. And with the current booming of blogs and microblogs on the web, people's opinions on a wide variety of topics like products, politics, events, healthcare, fraud detection … generate bulk amount of textual data [3][4]. However, it would not be possible for a human to go through all those opinions comparing their contents and classify them into their proper classes without biases and large error. Therefore, an automated way of classifying those reviews which should outperform human classification in accuracy and speed should be utilized [5]. In this paper, we employed information extraction and sentiment analysis text mining techniques to analyze Samsung customer reviews related to Samsung products prior galaxy Note7 fiasco.

We used simple information extraction techniques like term co-occurrences, the corpus count-based evaluation, and word-cloud representation to analyze customer reviews [6][7]. Sentiment analysis is a wide field with various applications and in this paper, we focus on the use of sentiment analysis to classify reviews/documents into positive or negative classes using machine learning algorithms based on the sentiment, or the overall opinion towards the subject matter as expressed in the review. Labelled reviews from Samsung customers were collected and used to train the classifiers which were later used to classify unlabelled reviews into respective categories [8][9].

SAMSUNG the world's leading manufacturer of Android smartphones and other electronic devices like computers, TV screens, air conditioners, microwave ovens, washing machines and refrigerators, et cetera is a South Korean company well known and respected worldwide for their well-reputed products [10], [11]. Samsung started in Taegu, Korea 1938 as a small export business and from there on it has become one of the world's leading electronics company, manufacturing digital appliances and media, semiconductors, memory, and system integration [12], [13][14]. Samsung expanded its product lines and reach to achieve remarkable revenue and market share. Today Samsung is not only known for its innovative and top quality products and processes which are worldwide recognized but also its mission of making life better for consumers around the world. Through their flagship company, Samsung Electronics, Samsung boasts of leadership in the global market in high-tech electronics manufacturing and digital media, innovation, reliable products and services, talented people, a responsible approach to business and global citizenship, collaboration with partners and customers and fulfilling corporate social obligations such as social welfare, environmental conservation, cultural events, or sports [15]. Nevertheless, recently Samsung got a thorn in their side by abandoning their Galaxy Note 7 flagship phone [16]. Since its acclaimed launch, the Note 7's early days were marked by glowing reviews because of its amazing features such as a larger, sharper, and richer display than the top phones of that time, less weight , easier to hold, a big phone that didn't feel big, built-in retina scanner, water resistance, rear-facing dual cameras making its specs impressive, the simplicity of its design, the striking beauty of its curved screen and a 3,500 mAH battery that was able enough to allow it to go without a charge even while being used constantly for 36 hours. Nevertheless, it is thought that it is this powerful lightweight battery which includes lithium-ions that would have been its downfall. The early produced Note7 and its replacements did not only got fire in the homes of some customers, shops, and airplanes but also the company themselves decided to halt its production, call customers to exchange Note7 for other Samsung smartphone or receive a refund and finally killed the brand completely after only 53 days of existence from August 19, 2016, to October 10, 2016 [17], [18].

Bearing in our minds that no product can satisfy all customers a hundred percent, this paper aims to analyze the reviews of Samsung customers in regard to products prior Note7 to find out how these customers perceived them using text mining techniques and evaluate the classification performance of some machine learning algorithms on such kind of reviews. We also build models which can be used to predict the polarity of any such kind of future reviews using supervised machine learning algorithms including Classification and Regression Trees (CART) for their high interpretability, Random Forests (RFs), Naïve Bayes, Support Vector Machines (SVMs) and Maximum Entropy for their improved accuracy and robustness against overfitting.

The rest of this paper is organized as follows: In section 2, we review the previous related works, section 3, describes materials and methods utilized. In Section 4, we present experimental results. Finally, in Section 5, we give the conclusions.

2. Previous works

This research involves two technology premises of text mining namely information extraction and sentiment analysis. Text mining is described as a process of extracting hidden information or knowledge or pattern from unstructured text data gathered from different sources. The text mining process for information extraction and categorization is shown [19]. Text analysis concepts links were used in [20] to analyze terms co-occurrences and associations to analyze and classify American Airline reviews. Text mining was used to identify terms that occur so often in the game reviews and how they affect the game reputation. Sentiment analysis was used to build predictive models on the existing reviews which can be used to predict whether a new review is good or bad using SAS® [21]. [22], [23] used support vector machines for text categorization to reduce the overhead required for fast retrieval of documents and easy exploration of similar documents. Unsupervised classification of reviews from Epinion by using the average semantic orientation was done in [8], this is the lexicon-based approach. Supervised classification of movie reviews and other text documents following the overall positive or negative sentiment using machine learning algorithms was studied in [9] [24]–[27][28]–[30] [5], [31], [32]. The use of lexicon to label documents and use them to classify subsequent documents was studied in [33], [34] but it resulted in raw recall compared to the hybrid method combining lexicon-based method and machine learning studied in [34] which proved to have both high precision and recall.

3. Materials and Methods

In this paper, we leveraged text mining techniques to analyze Samsung customer reviews prior the release of Galaxy Note7. We also investigated the classification performance of different machine learning algorithms such CART, RFs, naïve Bayes, SVMs and maximum entropy on this kind of data and built different model classifiers to predict the polarity of such future reviews. The figure1 illustrates the materials and methods used in this paper [1]

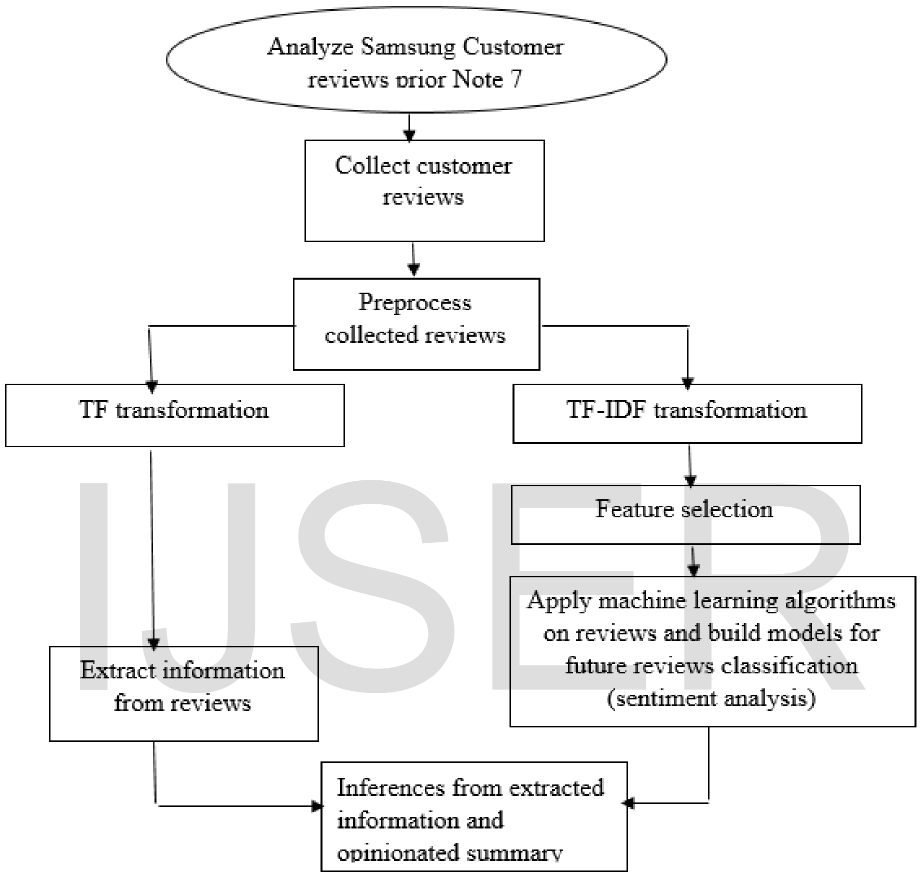

FIGURE 1: MATERIALS AND METHODS USED IN INFORMATION EXTRACTION AND MODEL CLASSIER BUILDING

Figure1 presents the method followed and techniques used in this paper. We started by collecting Samsung customer reviews and preprocessed them. Figure1 shows two main paths namely information extraction and sentiment analysis followed in this paper. We used term frequency for information extraction and term frequency-inverse document frequency for sentiment classification. Feature selection was only performed for sentiment classification. The two main sub paths of figure1 meet at the end of the figure for inferences extracted from uncovered information and reviews classification. All techniques and machine learning algorithms described by figure1 are explained as follows.

3.1 Corpus pre-processing, term weighting, extraction of information and feature selection from reviews

Our data set consists of 468 Samsung customer reviews harvested between 15th of February 2017 and 13th of October 2017 from https://www.consumeraffairs.com/cell_phones/samsung_cell_phones.html [35], a trustworthy and neutral platform for consumers to share and respond to reviews. Those reviews are almost about any Samsung phone model both the old GT-xxxx and the current SM-xxxx models [36] and they were posted on this website between 01st of August 2011 and 13th of October 2017. They are trustworthy and rightly rated by the owners of this website who claim to verify customer reviews, require contact information of customers to ensure the reviewers are real, use intelligent software that helps them maintain the integrity of reviews and employ moderators who read all reviews to verify their quality and helpfulness. Those reviews are rated on a scale of 1 to 5 [5]. Only the rated reviews by the time of data collection were harvested to make the data set, which was later divided into the training set to train the model and the testing set to evaluate the performance of the built model.

The text data set analyzed during sentiment analysis is huge, unclassified and unstructured or semi-structured. It contains different sorts of useful and useless data about science, business, health, etc. [1]. For their effective mining, a well-defined text pre-processing should be employed to extract interesting and previously hidden data patterns. The raw text data is highly susceptible to inconsistency and redundancy, pre-processing involves applying methods for cleaning up and structuring the input text for further analysis. Text pre-processing involves operations such as stop words, punctuations and numbers removal, transform text to lower case, whitespace stripping, tokenization and stemming [37] [38],[39].

The simplest approach to analyze a text corpus is to assign each term t with its weight which is equal to the number of occurrences of that term in a given document d. This weight is known as term frequency (TF) and is often used in information extraction [40], [41]. The sole use of term frequency can contribute to the less consideration of some less frequent terms but important to the study. Term frequency Inverse document frequency (TF-IDF) is used to easily determine which terms in the corpus of documents might be more favorable to the user in further processing like sentiment classification using machine learning algorithms as it avoids assigning the high score to the most frequent terms. TF-IDF is calculated as follows [42].

$$TF\_IDF_{t,d} = TF_{t,d} * IDF_t, \tag{1}$$

where TF represents the occurrence number of term t in document d, IDF=log (N/DF_t), N represents the number of documents in the corpus, and DF_t represents the number of documents containing the term t [43], [44]

To extract the hidden information from the corpus of text we used term co-occurrences, the corpus count-based evaluation techniques, and word-cloud representation. Two or more words co-occur if they appear at once in a given unit of text. A unit of text can be documents or sub-documents, paragraphs or sentences, or a window of a predefined number of words [45]. Only textual co-occurrence and a window of size four (left plus right window type) on both sides of the word of interest have been used in this paper. The corpus count-based evaluation is the simplest analysis method in text mining. It is simple and widely used since it can be interpreted nicely and is computationally inexpensive. In this method, the terms with the highest occurrence of frequency in the corpus of text are rated important. To explore the corpus we used a document-term matrix which is simply a matrix with documents as the rows and terms as the columns and a count of the frequency of words or weight of words as the cells of the matrix [46]. A word cloud or tag cloud is a pictorial representation of text data, which used particularly to show keyword metadata (tags) on websites or to visualize free form text. Tags are usually single words, and the importance of each tag is shown with font size or color. This format is useful for quickly perceiving the most important terms in the text[47]. Feature selection is the process of reducing the amount of data to be analyzed and improve the performance of machine learning algorithm by identifying those features relevant to text sentiment classification [48]. In this paper, we reduced the sparse document term matrix using a threshold of 0.99 to keep only those terms that appear in at least 2 percent of the documents.

3.2 Machine learning algorithms and model classifier construction

The aim of using machine learning algorithms in this paper was to try to understand opinions of Samsung customers towards Samsung products by leveraging the interpretability power of decision tree, evaluate the classification performance of different machine learning algorithms such CART, random forests, naïve Bayes, SVMs and maximum entropy on this special kind of reviews and build predictive models which can be used to predict the category of such future reviews.

3.2.1 Classification and regression trees (CART)

The CART decision tree is a binary recursive partitioning procedure capable of processing continuous and nominal/categorical attributes as targets and predictors. Classification and regression trees are used when the target variables are categorical or continuous respectively. The CART mechanism includes (optional) automatic class balancing by the use of priors mechanism and automatic missing value handling by the use of surrogates or substitute splitters for every node of the tree, whether missing values occur in the training data or not, and allows for cost-sensitive learning, dynamic feature construction through the automatic construction of linear combinations that include feature selection, and probability tree estimation [49]. The structure of CART is like a multilevel inverted tree with the root node known as the parent node at the top because it contains the entire set of observations to be analysed, the child nodes derived from parent node and are as pure as possible to the dependent variable and the terminal nodes known as leaf nodes holding the predicted class or numerical outcome for classification and regression problem respectively [27].

Given the feature vector X that maps an instance to class Y, CART uses the binary recursive partitioning procedure known as the greedy algorithm to minimise the following cost function.

$$C(y) = \sum_{i=1}^{n} (y_i^{true} - y_i^{predict})^2 \ , \qquad (2)$$

The tree start from the node known as the root node, then data are split into two children, and each of the children is in turn split into grandchildren using the Gini splitting rule especially for classification problem. The feature and the split point with the lowest Gini value is selected. The objective function for Gini splitting rule for any leaf with class-k is defined as,

$$G_k(y) = \sum_{i=1}^{n} y_k(1 - y_k) \ , \qquad (3)$$

3.2.2 Random forests (RFs)

As their name shows, random forests is one of ensemble machine learning classifiers. RFs use multiple CART decision trees to obtain improved predictive performance that could not be obtained by any single decision tree. Random forests are known for significant improvements in classification accuracy over the CART and avoiding overfitting which results from growing an ensemble of trees and letting them vote for the most popular class. This improved accuracy depends on the strength of the individual tree classifiers and a measure of the dependence between them [50]. The two most popular strategies for introducing diversity into RFs are randomly sub-sampling the examples (bagging) and sub-sampling the features (feature selection) [51]. To improve accuracy, this randomness injected into RFs has to minimize the correlation between individual trees while maintaining strength. The right kind of randomness injected in RFs makes them accurate classifiers and regressors. It is worth highlighting that building ensemble decision trees in RFs have added benefits of providing an analysis of data such estimation of generalization error, case proximities and variable importance [51].

3.2.3 Naïve Bayes

Naïve Bayes classifiers is a family of classifiers based on the well-known Bayes probability theorem that was formulated by Thomas Bayes. These classifiers are simple to create and well computationally performing supervised models [52]. A naïve Bayes classifier can be stated in as follows [39]:

$$P(c \mid d) = \frac{P(d \mid c)P(c)}{P(d)} \ , \qquad (4)$$

In the equation4 the conditional probability (posterior probability) $p(c|d)$ is the probability of document **d** being in class **c**. This probability is equal to the probability of a document given the class (likelihood probability) $p(d|c)$**,** times the probability of occurrence of the class $p(c)$ over the probability of the document $p(d)$ **.** Here $p(d)$ and $p(c)$ are the priors of document and class respectively.

Following the naïve Bayes assumption, using add-one Laplace smoothing and using the sum of logs of probabilities, the best class by the Naïve Bayes assumption (C_{NB}) as our base equation for the multinomial Naïve Bayes Classifier is stated as follows [53] [30].

$$C_{NB} = \arg\max_{c_j} (\log P(c_j) + \sum_{i \in positions} \log P(x_i \mid c_j) \ , \qquad (5)$$

In this paper, we used the binary multinomial or binary NB Naïve Bayes classifier which is a variant of the standard Naïve Bayes classifier. The term binary means the classifier works with only Boolean features**,** it will not take into consideration any duplicate word within a document and only counts distinct words for a document since in sentiment classification and other text classification tasks the presence or absence of a word seems to matter more than its frequency. Note that the same word or feature may occur in different documents [54].

ISSN 2229-5518

The major advantage of using the Binary multinomial naïve Bayes classifier is that it improves the overall performance of classification by reducing the time of calculation. It also avoids the illusion that counting multiple occurrences of the same word should lead to the conclusion that the document belongs to the class supported by that word.

3.2.4 Support vector machines (SVMs)

SVMs aka the widest street approach uses the largest margin/decision boundary/widest street to separate the documents into positive and negative accordingly. Where the largest margin γ is regarded as the confidence of prediction which corresponds to the distance of a document from the separating hyperplane defined by the support vectors. The goal of SVMs is to find those support vectors which are the points laying on the positive and negative planes as shown on figure2. Normally, for any d dimension data, there are $d+1$ support vectors [55].

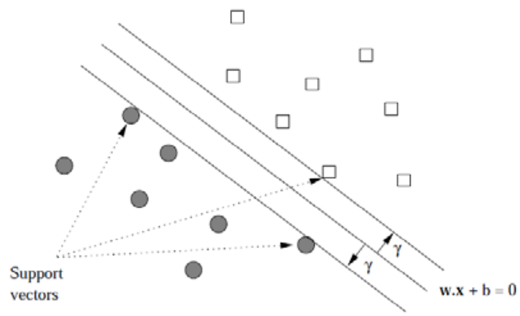

Figure 2: An SVM selects the hyperplane with the largest possible margin γ between the hyperplane and the training points. Courtesy [55]

Like in our case of two-category case, the aim of SVMs is to find a hyperplane represented by vector $\vec{w}$ which separates the document vectors in one class from the other with the largest possible margin. For example, for a training set$(x_{1,}y_1), (x_2, y_2), \dots, (x_n, y_n)$, the goal of SVMs is to find a solution of the following optimization problem:

$$\min_w \frac{1}{2}\|w\|^2 , \qquad (6)$$
$$S.t. \forall i, y_i(w \cdot x_i + b) \geq 1$$

The maximization problem in equation6 turns to be a quadratic programming problem and it is much difficult and time-consuming to solve it by hand [56]. Therefore every user of SVMs can solve it using his/her favorite quadratic programming solver. In this paper, we used e1701 package which is an R implementation of a library for support vector machine (LIBSVM) package [57], [58].

After the quadrating programming solver has found α_1^* , α_2^*,…, α_n^* for instance, as the solutions of the SVMs optimization problem, the value of weight vector w can be computed as

$$w = \sum_{i=1}^{N} \alpha_i^* y_i x_i \ , \qquad (7)$$

and the value of the bias b can be computed referring to

$$b = -\frac{\max_{i:y_i=-1} w^T x_i + \max_{i:y_i=+1} w^T x_i}{2}, \qquad (8)$$

SVMs are known to work well for text categorization task. SVMs can handle high dimensional input data by the use of overfitting protection. There are very less irrelevant features in text categorization, therefore, SVMs with their potential to handle large feature space have good performance in text classification. SVMs handles well the sparsity of document vectors problem [22].

Every user of SVMs faces the problem of the choice of the kernel function to train the data. This choice depends on the number of instances compared to the number of the features. If the number of features is large like in text classification problem, one may not need to map data to a higher dimensional space. In that case, the nonlinear mapping does not improve the performance. Therefore, the linear kernel is a good choice [59].

3.2.5 Maximum entropy

Multinomial logistic regression commonly known as maximum entropy or maxent classifier in language processing is a probabilistic classifier which belongs to the class of exponential or log-linear classifiers which is based on the principle of maximum entropy. This classifier works in a way that from all the models that fit the training data, it selects the one which is uniform as possible, that is, the one that maximizes entropy. The maximum entropy classifier has been often used to solve a large variety of text classification problems such as language detection, topic classification, sentiment analysis etc. [60].

Maximum entropy and naïve Bayes are both linear probabilistic classifiers but are not the same at all. First, naïve Bayes is a generative model and the maximum entropy is a discriminative model. Secondly, naïve Bayes is based on the assumption that all features are independent but maximum entropy does not [60].

The most important steps in the use of maximum entropy are to identify a set of feature functions that will be useful for classification and then for each feature measure its expected value (weights) over the training data and take this to be a constraint for the model distribution. Those weights are learned by choosing the parameters that make the classes of the training examples more likely. Therefore, maximum entropy is trained with conditional maximum likelihood estimation. In short, maximum entropy searches for those parameters λ_i that maximize the probability of the c class labels in the training data given the documents$d$. Maximum entropy estimates the probability $p(c|d)$ using the following exponential form (to avoid negative numbers) [61] [9].

$$P(c \mid d) = \frac{1}{Z(d)} \exp(\sum_i \lambda_i f_i(d,c)), \tag{9}$$

where $f_{i,c}(d,c)$ is a feature/class function, λ_i is the parameter to be estimated and $Z(d)$ is the normalizing factor to ensure a legal probability form

$$Z(d) = \sum_c \exp(\sum_i \lambda_{i,c} f, c_i(d,c)), \tag{10}$$

and the feature/class function for feature $f_{i,c}$ and class c is defined as

$$F_{i,c}(d,c') = \begin{cases} 1 & n_i(d) > 0 \ \ and\ c' = c \\ 0 & otherwise \end{cases}, \tag{11}$$

Finding the solution to this objective in equation9 turns out to be a convex optimization problem. In this paper, we used improved iterative scaling (IIS) hillclimbing algorithm to find the maximum entropy distribution that is consistent with the given constraint [62].

4. Experimental results

In this paper, we leveraged information extraction techniques to reveal hidden information from large text data and employed machine learning algorithms to build predictive models which can be used to classify future unlabelled reviews based on current labelled reviews. The CART was used to get some understanding of information hidden in the reviews due to its power of representation. Finally, we evaluated the performance of those machine learning algorithms using different measures namely accuracy, precision, recall and F1 as they can be found in the confusion matrix. The receiver operator characteristics (ROC) was used to generate the area under the curve (AUC) to show the probability by which each classifier

ISSN 2229-5518

decides whether previously unseen review is positive. The time performance was used as measure to assess the time taken for each algorithm to train the model. All the found results are shown as follows.

4.1 Information extraction results

4.1.1 Terms co-occurrences results

The following are some important co-occurrences for each of the selected words such as **phone, battery, screen, service, warranty,** and **support**. Those co-occurrences among others were generated using NGramTokenizer from Rweka package to split strings into n-grams with 4 and 9 as given minimal and maximal numbers of grams respectively [63].

TABLE 1: SAMPLE TARGET WORDS CO-OCCURRENCES

Target word	Co-occurrences
Phone	"after repair, we noticed phone was still overheating to"
	"want to replace its phone that's burning its customers"
	"two weeks with no phone for them to fix"
Battery	"it started hanging and battery getting drain very fast"
	"Samsung lists the replacement battery at two different prices"
	"about am the phone battery exploded landing on our"
Screen	"extreme heat and the screen having a cloudy smoke haze"
	"was ridiculous my s screen start peeling off after"
	"My Samsung galaxy phone screen was frozen and would"
Service	"poor quality and customer service is unacceptable I will"
	"the person in the service station is reluctant to"
	"with Samsungs level of service, I will never buy"
Support	"heard of such poor support than this now looking"
	"item also their customer support is the worst part"
	"not offer live tech support I have also asked"
Warranty	"Samsung they refused to warranty the repair they said"
	"does not offer any warranty they just charge you"
	"have to honor their warranty and fix the phone"

These co-occurrences in the table1 convey meaningful information hidden in a long text of customer reviews. They show that Samsung customer complain about their phone having overheating and burning, and charging problems. Samsung also takes a long time to repair customer phones. The batteries explode and drain the power faster than expected. Samsung is also accused of not replacing the defaulted batteries for free. The screens overheat, freeze and some peel off after short time. The customer service is poor and the persons to service Samsung clients are reluctant, some customers also claim about the negligence of warranty. All these scupper the trust that customers have for Samsung and cause reluctance in their way of buying and recommend Samsung products to others.

4.1.2 The corpus count-based evaluation results

The corpus and the term document matrix are the main tools in this step.

TABLE 2: DOCUMENT TERM MATRIX

documents	terms	Non-/sparse entries	Sparsity	Maximal term length	Weighting
360	3320	22448/1172752	98%	27	Term frequency (tf)

Table2 shows a matrix containing 360 rows each corresponding to a specific document/review and 3320 terms. The maximal term length is 27. It is so sparse with the degree of sparsity of 98%.

We could also identify the least frequent terms and the most frequent terms. Normally, a term is more or less important according to a simple counting of frequencies. The term with high frequency tends to be more important than less frequent terms. The distribution of term frequencies helped us to understand how terms were distributed across documents; how frequent was the term. Below are the commonly talked about frequencies. We showed the first 20 high frequent terms and the last 20 low frequent terms distribution:

TABLE 3: FIRST 20 HIGH FREQUENT TERMS DISTRIBUTIONS

Terms	phone	call	will	back	problem	get	time	service	repair	day
Frequency	1647	353	308	304	284	282	281	273	265	247
Terms	told	customer	replace	purchase	work	galaxy	charge	new	use	warranty
Frequency	247	228	213	205	197	196	188	187	187	187

Table3 contains the first 20terms which appear most frequently. These are the terms that were much more likely to be of interest to us. Terms like problem, back, call and phone were really significant to our study.

TABLE 4: LAST 20 LOW FREQUENT TERMS DISTRIBUTIONS

Terms	‘till	‘ll	‘re	‘ve	‘m	Didn’t	abe	fortify	abide	acclaim
Frequency	1	1	1	1	1	1	1	1	1	1
Terms	absorbe	absurd	occur	accomplish	abil	abras	accoustiment	acceptmak	aberdeen	adapt
Frequency	1	1	1	1	1	1	1	1	1	1

Table4 contains the last 20 terms which just appear once and are probably not the terms that are of great significance to us. Most of them are the short forms of verbs.

All the above results we retrieved from a high sparse therefore it was convenient to reduce its sparsity. Here we used the threshold of 0.99 that means keeping the terms that appear in at least 2 percent of the documents

TABLE 5: LESS SPARSE DOCUMENT TERM MATRIX BY THE FACTOR OF 0.99

documents	terms	Non-/sparse entries	Sparsity	Maximal term length	Weighting
360	1031	19266/351894	95%	12	Term frequency (tf)

After reducing the sparsity of the document term matrix as shown in table5, we could list some of the most frequent terms by the threshold of 100 as follows.

TABLE 6: SOME OF THE MOST FREQUENT TERMS BY THE THRESHOLD OF 100

fix	**galaxy**	**ask**	**back**	**battery**	**bought**	**buy**
call	phone	center	charge	custom	damage	repair
device	don't	problem	fix	product	purchase	replace
warranty	water	month	Day	week	new	never

The following is the histogram that shows the distribution of the terms that appear at least 150 times

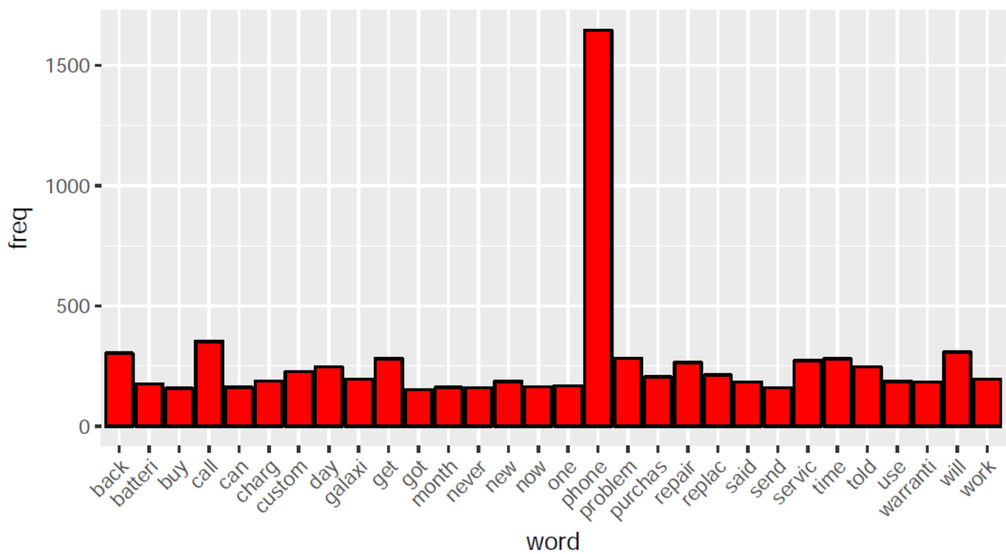

FIGURE 3: HISTOGRAM OF TERMS FREQUENCY DISTRIBUTION APPEARING AT LEAST 150 TIMES

Figure3 shows that the term phone has the highest frequency but our other selected terms including battery, service, and warrant have considerable frequencies as well.

4.1.3 Association of terms in the document term matrix

Here we illustrate correlations of our selected terms with other terms in the document term matrix. The higher the correlation the more the two terms are associated.

TABLE 7: PHONE ASSOCIATIONS WITH A THRESHOLD OF 0.28

Term	back	drop	repair	fix	damage	warranty	water	charger
Correlation	0.48	0.33	0.33	0.33	0.32	0.31	0.29	0.28

TABLE 8: SCREEN ASSOCIATIONS WITH THRESHOLD OF 0.39

Term	crack	bother	brighten	consent	hacker	article	listen
Correlation	0.46	0.4	0.4	0.4	0.4	0.4	0.4

TABLE 9: WARRANTY ASSOCIATIONS WITH A THRESHOLD OF 0.3

Term	break	honor	cover	charger	damage	port	phone
Correlation	0.47	0.43	0.41	0.36	0.36	0.36	0.31

TABLE 10: SERVICE ASSOCIATIONS WITH THRESHOLD OF 0.38

Term	center	controversy	obsolete	patient	persist	problem	redirect	reimbur se
Correlation	0.71	0.56	0.56	0.53	0.46	0.42	0.40	0.40

TABLE 11: BATTERY ASSOCIATIONS WITH A THRESHOLD OF 0.23

Term	charge	defectthi	paperweight	advantage	lunch	postal	gadget
Correlation	0.34	0.33	0.33	0.28	0.28	0.28	0.25

The following is the correlation plot that summarizes and illustrates the correction between terms

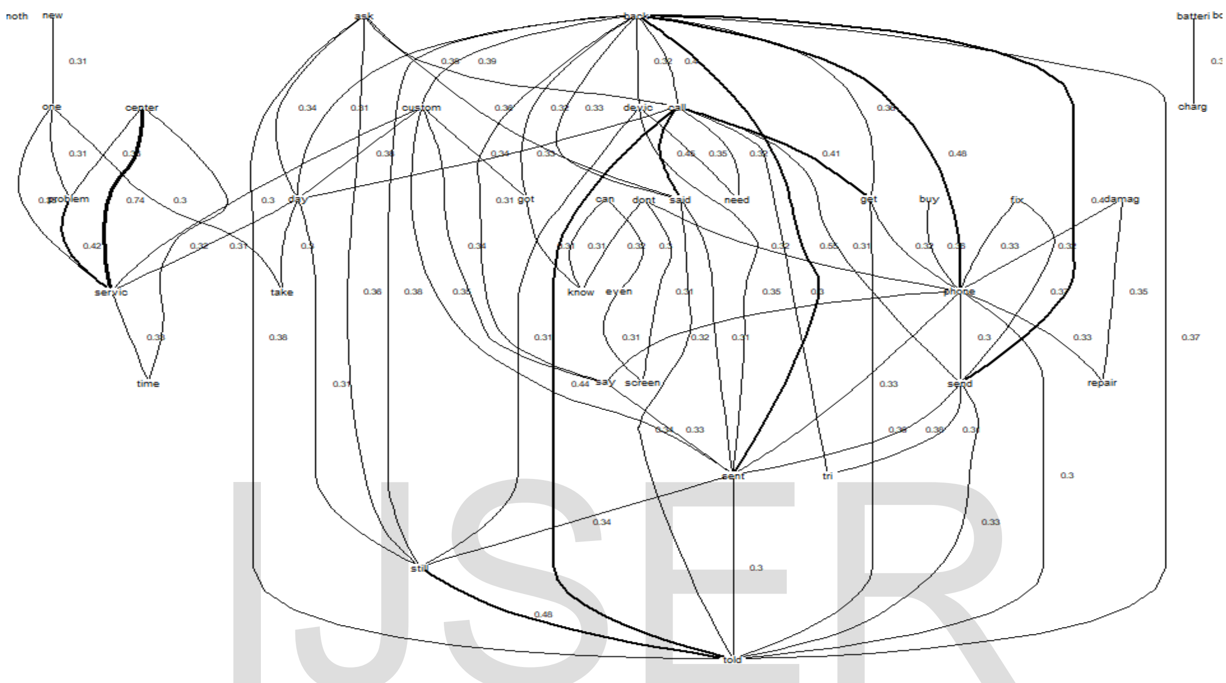

FIGURE 4: CORRELATION PLOT INDICATING CORRELATION BETWEEN TERMS

Figure4 shows correlations among 50 most frequent words. It consists of the nodes as words, links connecting correlated words and the value of correlation between every pair of words. There are strong correlations between center and service, sent and back, phone and back, repair and damage, product and purchase and more other correlations can be observed from this plot.

4.1.4 Word cloud

This format is useful for quickly perceiving the most prominent terms.

FIGURE 5: WORD CLOUD

Figure5 shows that the terms battery, problem, repair, call, charger, warranty, etc. are prominent terms in the reviews under our study.

4.2 Machine learning predictive models

The aim of this paper was to analyze Samsung customer reviews to get inferences on their opinions and how they perceive Samsung products prior the release of Galaxy Note7 and build machine learning models to predict whether customers are satisfied or not with Samsung products they have purchased. We trained the algorithms with the factor variable of 0 to indicate satisfaction (FALSE) or 1(TRUE) to indicate no satisfaction. Since the reviews on https://www.consumeraffairs.com/cell_phones/samsung_cell_phones.html [35] are rated on a scale of 1 to 5, we viewed any rate equal or less than 2 as no satisfaction and satisfaction otherwise.

4.2.1 Generated decision tree

The CART tree constructed from the reviews is represented as follows

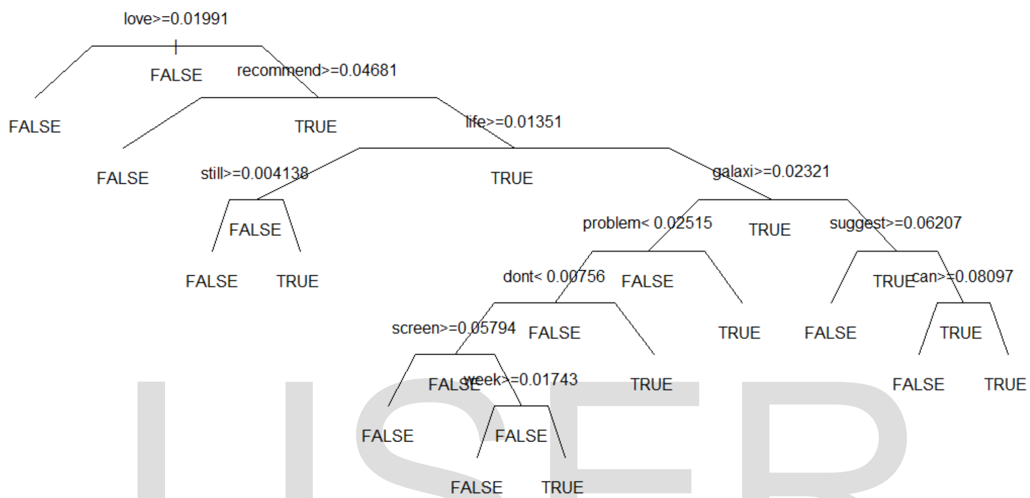

FIGURE 6: CART TREE GENERATED FROM CUSTOMER REVIEWS

Figure6 shows our constructed CART tree. The root node with love attribute splits the tree with the Gini index of roughly 0.02. If the value of Gini is greater than or equal to approximately 0.02 the review is negative else we consider other attributes. From this tree we can observe important attributes like love, recommend, life, galaxy, suggest, screen and problem.

4.2.2 The prediction performance of the built models

After building classification models, we used them to predict the categories (class labels) of new observations. The prediction power of our models is validated by generating the confusion matrix to determine the measures like accuracy, recall, precision and F1 as shown below. The performance time and the receiver operator characteristic (ROC) curves were also used as follows.

TABLE 12: THE 2X2 CONFUSION MATRIX WITH CLASSIFIER METRICS ILLUSTRATED

		Prediction condition	
	Total population	Prediction positive	Prediction negative
True condition	Condition positive	True Positive(TP)	False Negative(FN) (Type II error)
	Condition negative	False Positive(FP) (Type I error)	True Negative(TN)

Table12 shows a confusion matrix which is a table that is often used in machine learning mostly in supervised learning to evaluate the classification performance of a classifier model on a given set of test data for which the conditional true values are known. Each column and row of the matrix represents the data points in a predicted and actual class respectively (or vice versa) [64][65][66]. In the table12, True and False Positives (TP/FP) refer to the number of predicted positives that were correct/incorrect, and similarly for true and False Negatives (TN/FN), and these four cells sum to N (the total population). The commonly used metrics from the above matrix are [67]:

a. **Accuracy ((TP+TN)/total population)**: it shows in the overall correctness of the classifier.

b. Sensitivity, recall, hit rate, or true positive rate (TPR= TP/FN+TP): it shows the ability of a classifier to predict positive/true when the data point is in reality positive.

c. **Specificity or true negative rate (TNR= TN/TN+FP)**: When it's actually negative, how often does it predict negative?

d. **Precision or positive predictive value (PPV= TP/FP+TP):** When it predicts positive/true, how often is it correct?

e. **F_1 score** (also **F-score** or **F-measure=2xPrecision x Recall)/(Precision + Recall):** The F_1 score can be interpreted as a weighted average of the precision and recall, where an F_1 score reaches its best value at 1 and worst at 0.

Following the metrics found in the confusion matrix and the performance time, the machine learning algorithms are evaluated as shown in the table13.

TABLE 13: THE CONSOLIDATION TABLE COMPARING ALL FIVE MODELS

Measures	CART	RF	Naïve Bayes	SVM	MaxEnt
Accuracy	0.63	0.67	0.69	0.72	**0.73**
Recall	0.95	**1**	0.98	0.93	0.93
precision	0.60	0.62	0.64	0.68	**0.69**
F1	0.74	0.77	0.77	0.78	**0.79**
Area under the curve	0.59	**0.93**	0.92	0.79	0.82
Performance time(sec)	3.0	27.9	2.68	**2.1**	2.38

Table13 illustrates the performance of machine learning algorithms under our research. Numbers in bold show the best performing algorithm in each row per performance measure. Maximum entropy beat other algorithms in accuracy, precision, and F1. SVMs beat maximum entropy in time performance by fractions of seconds and are also very close to other measures. Random forests beat other algorithms in accuracy but their performance time is worse almost 10 times compared to SVMs. Our study showed that the performance of naïve Bayes as a baseline method in text mining is not far from one of best performers like maximum entropy and SVMs. This confirms the findings of [68] [31] that naïve Bayes works even better on small datasets or short documents and is easy and fast to train. The CART has the least but not so bad performance and its power of generating a model tree is its unmatched weapon. It is worthy to note that in case of large documents or datasets and for the tasks where combinations of features are important that the best choice of the algorithm would be SVMs, Maxent and RFs [60] as the literature confirms.

Finally, we used the AUC as shown by the ROC to evaluate the performance of machine learning algorithms. The receiver operator characteristics (ROC) graphs are two-dimensional graphs in which *tp* rate is plotted on the *y*-axis and *fp* rate is plotted on the *x*-axis, they are used for selecting classifiers based on their performance. An ROC graph depicts relative tradeoffs between benefits (trues positives) and costs (false positives). Several points in ROC space are important to note. The lower left point (0, 0) represents the strategy of never issuing a positive classification, such a classifier commits no false positive errors but also gains no true positives. The opposite strategy, of unconditionally issuing positive classification, is represented by the upper right point (1, 1). The point (0, 1) represents perfect classification. Informally, one point in ROC space is better than another if it is to the northwest (*tp rate* is higher, *FP rate* is lower, or both) of the first. For instance, the figure7 below shows that the classifier D's performance is better than all other shown on the same ROC [64].

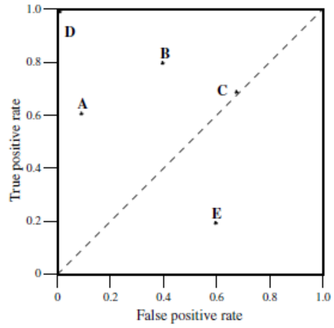

FIGURE 7: A BASIC ROC GRAPH SHOWING FIVE DISCRETE CLASSIFIERS. COURTESY[64]

In order to compare the performance of classifiers using ROC, we reduced the ROC performance to a single scalar value representing the expected performance. We calculated the area under(AUC) the ROC curve [69]. The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. It's important to note that no ideal classifier should have an AUC less than 0.5 [70]. The figure8 shows the AUC for each algorithm.

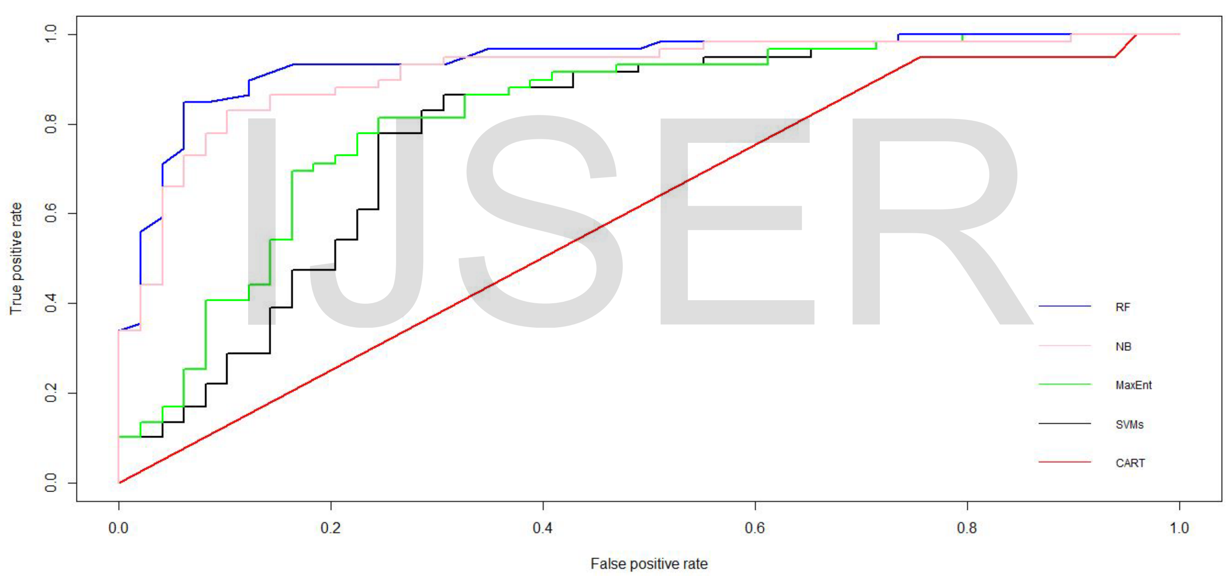

Figure 8 : The ROC FOR ALL FIVE ALGORITHMS

Figure8 shows the area under the curves of our different classifiers. Random forests have a better area under the curve of 0.93 %. Naïve Bayes though it is a baseline model, beats the remaining classifiers except for RFs with an AUC of 0.92%. The CART has the lowest AUC value of 0.59%.

5. Conclusions

It is paramount for manufacturers' marketing departments to have accurate information about their consumers 'opinions, preferences, and perception about their products. Nowadays, Companies' websites consumer reviews space, Twitter, Facebook and other microblogging systems constitute a powerful tool for communication among consumers and between them and products manufacturers. The digital world has made microblogging an online word-of-mouth branding. Though consumer reviews expressed online is bulk in size, it is also a valuable source of insight into consumers' opinions regarding available products and services. Companies should consider consumer-generated opinions a handful source of information which otherwise would tarnish the company's name and loss of power if not considered in a today's competitive world of business.

In this paper, we analyzed Samsung customers' reviews using text mining techniques such as information extraction and sentiment analysis. We used machine learning algorithms including classification and regression trees (CART), Random Forests (RFs), Naïve Bayes, Support vector machines (SVM), and Maximum Entropy to build models that can be used to automate such future unlabelled reviews classification into positive and negative classes. All these machine learning algorithms produced good accuracy but Maximum Entropy had better performance in accuracy, precision and F1 measures than other classifiers and its performance time is closer to SVMs, hence, it is our recommended classifier for such kind of reviews. The predictive models built in this paper can be used for classifying the reviews from any website or blog into positive and negative classes provided that the reviews are about cell phones and the rating is between 1 and 5. Furthermore, the CART tree generated in this paper showed that Samsung customers love Samsung products as shown by the root node and are willing to recommend the products to new customers but different problems like some devices defects including fast power drainage, excess production of heat and device freezing, lack of warranty consideration and the low quality of service and support offered to customers seem to hinder their trust as revealed by information extraction. Our interest remains in the analysis of Samsung customers' opinions and reviews to understand more about their perception of current and future Samsung products.

Acknowledgements: I would like to thank my supervisor, Prof. Jing Liu, for the patient guidance, encouragement, and advice she provided for the completion of this paper. I also thank KALINIJABO Yves for his helpful comments during this research.

Author Contributions: The author claims the full responsibility of this paper including conceptualization of the problem, data collection, data analysis, research method design, computations and paper writing.

Conflicts of Interest: The author declares no conflict of interest.

References

[1] A. Kaushik and S. Naithani, "A Comprehensive Study of Text Mining Approach," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 16, no. 2, pp. 69–76, 2016.

[2] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," *Proc. Seventh Conf. Int. Lang. Resour. Eval.*, pp. 1320–1326, 2010.

[3] G. Chakraborty, M. Pagolu, and S. Garla, "PREVIEW: Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS," *Text Min. Anal. Pract. Methods, Examples, Case Stud. Using SAS*, pp. 1–38, 2013.

[4] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for Twitter sentiment analysis," *Int. J. Electron. Commun. Soft Comput. Sci. Eng.*, vol. 89, pp. 1–8, 2015.

[5] N. Mehra, S. Khandelwal, and P. Patel, "Sentiment Identification Using Maximum Entropy Analysis of Movie Reviews," 2002.

[6] A. Caticha, "The Basics of Information Geometry," *MaxEnt 2014*, vol. 15, no. 2015, pp. 1–13, 2014.

[7] H. P. Luhn, "The Automatic Creation of Literature Abstracts," *IBM J. Res. Dev.*, vol. 2, no. 2, pp. 159–165, 1958.

[8] P. D. Turney, "Thumbs up or thumbs down? Semantic Orientation applied to Unsupervised Classification of Reviews," *Proc. 40th Annu. Meet. Assoc. Comput. Linguist.*, no. July, pp. 417–424, 2002.

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," *Proc. Conf. Empir. Methods Nat. Lang. Process.*, pp. 79–86, 2002.

[10] J. Garside, "Samsung overtakes Apple as world's most profitable mobile phone maker," 2013. [Online]. Available: https://www.theguardian.com/technology/2013/jul/26/samsung-apple-profitable-mobile-phone. [Accessed: 22-Nov-2017].

[11] Statista, "Global smartphone shipments from 2007 to 2016, by vendor (in million units)," 2017. [Online]. Available: https://www.statista.com/statistics/271539/worldwide-shipments-of-leading-smartphone-vendors-since-2007/. [Accessed: 22-Nov-2017].

[12] Samsung, "About Samsung." [Online]. Available: http://www.samsung.com/us/aboutsamsung/ir/corporategovernance/articlesofincorporation/IR_ArticlesChapter1.html.

[Accessed: 22-Nov-2017].

[13] C. Demeritt, "HTC ditches Samsung components for other suppliers, a la Apple," 2012. [Online]. Available: http://www.techradar.com/news/phone-and-communications/mobile-phones/htc-ditches-samsung-components-for-other-suppliers-a-la-apple-1103278.

[14] J. Vanian, "Samsung Dethrones Intel As World's Biggest Chip Maker." [Online]. Available: http://fortune.com/2017/07/27/samsung-intel-chip-semiconductor/. [Accessed: 22-Nov-2017].

[15] Samsung, "About samsung," *about samsung*. [Online]. Available: http://www.samsung.com/us/aboutsamsung/corporateprofile/history06.html. [Accessed: 07-Jun-2017].

[16] C. S.-H. Brian X, "Why Samsung Abandoned its Galaxy Note 7 Flagship Phone." [Online]. Available: https://www.nytimes.com/2016/10/12/business/international/samsung-galaxy-note7-terminated.html?_r=0. [Accessed: 08-Jun-2017].

[17] J. Swearingen, "The Brief, Sulfurous Life of the Samsung Galaxy Note 7." [Online]. Available: http://nymag.com/selectall/2016/10/samsung-note-7-killed-and-removed-from-market.html. [Accessed: 08-Jun-2017].

[18] S.-H. L. PAUL MOZUR, "Samsung to Recall 2.5 Million Galaxy Note 7s Over Battery Fires," 2016. [Online]. Available: https://www.nytimes.com/2016/09/03/business/samsung-galaxy-note-battery.html. [Accessed: 08-Jun-2017].

[19] R. R. Tated and M. M. Ghonge, "A Survey on Text Mining- techniques and application," vol. 3, no. 1, pp. 380–385, 2015.

[20] S. K. Choudhary, B. Analytics, and B. Analytics, "Text Analysis of American Airlines Customer Reviews," pp. 1–7, 2016.

[21] M. K. Singh, N. Batti, and M. Mcgaugh, "Text mining and sentiment analysis on video game user reviews using SAS ® Enterprise Miner ™," *SESUG 2016, Pap. EPO-280*, pp. 1–13, 2016.

[22] T. Joachims, "Introduction to Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Mach. Learn.*, vol. 1398, no. LS-8 Report 23, pp. 137–142, 1998.

[23] C. W. and M. S. A Basu, "Support vector machines for text categorization," *Proc. 36th Hawaii Int. Conf. Syst. Sci.*, pp. 1–7, 2003.

[24] X. Wu, V. Kumar, Q. J. Ross, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.

[25] T. Hayes, S. Usami, R. Jacobucci, and J. J. McArdle, "Using Classification and Regression Trees (CART) and random forests to analyze attrition: Results from two simulations.," *Psychol. Aging*, vol. 30, no. 4, pp. 911–929, 2015.

[26] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision Forests for Classification , Regression , Density Estimation , Manifold Learning and Semi-Supervised Learning," *Learning*, vol. 7, pp. 81–227, 2011.

[27] L. Kuhn, K. Page, J. Ward, and L. Worrall-Carter, "The process and utility of classification and regression tree methodology in nursing research," *J. Adv. Nurs.*, vol. 70, no. 6, pp. 1276–1286, 2014.

[28] D. D. Lewis, "Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval," *Mach. Learn.*, vol. ECML-98 Te, pp. 4–15, 1998.

[29] P. Domingos and M. Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Los," *Mach. Learn.*, vol. 29, no. 1, pp. 103–130, 1997.

[30] F. Ricci, "Text Classification ; The Naïve Bayes algorithm." Free University of Bolzen-Bolzano, Bozen-Bolzano, pp. 1–48.

[31] S. Wang and C. Manning, "Baselines and Bigrams: Simple, Good Sentiment and Topic Classification," *Proc. 50th Annu. Meet. Assoc. Comput. Linguist.*, no. July, pp. 90–94, 2012.

[32] D. Patel, S. Saxena, and T. Verma, "Sentiment Analysis using Maximum Entropy," pp. 8355–8361, 2016.

[33] J. Wiebe and E. Riloff, "Creating Subjective and Objective Sentence Classifiers from Unannotated Texts," pp. 486–497, 2005.

[34] S. Tan, Y. Wang, and X. Cheng, "Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples," *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. SIGIR 08*, p. 743, 2008.

[35] Consummeraffairs, "400 Samsung Cell Phones Consumer Reviews and Complaints," 2017. [Online]. Available: https://www.consumeraffairs.com/cell_phones/samsung_cell_phones.html. [Accessed: 15-Feb-2017].

[36] Wikipedia, "Samsung Galaxy." [Online]. Available: https://en.wikipedia.org/wiki/Samsung_Galaxy. [Accessed: 18-Jan-2018].

[37] G. Williams, "Text Mining," *Hands-On Data Sci. with R*, no. January, p. 47, 2016.

[38] A. Kumar and A. Paul, "Mastering Text Mining with R. Master text-taming techniques and build effective text-processing applications with R," p. 253, 2016.

[39] S. Raschka, "Naive Bayes and Text Classification I - Introduction and Theory," pp. 1–20, 2014.

[40] J. Beel, C. Breitinger, and S. Langer, "Evaluating the CC-IDF citation-weighting scheme: How effectively can 'Inverse Document Frequency' (IDF) be applied to references?," *Proc. iConference 2017*, pp. 1–11, 2017.

[41] C. D. Manning, P. Raghavan, and H. Schütze, "Scoring, term weighting, and the vector space model," *Introd. to Inf. Retr.*, no. c, pp. 118–132, 2009.

[42] K. Ahmed, N. El Tazi, and A. H. Hossny, "Sentiment Analysis over Social Networks: An Overview," *2015 IEEE Int. Conf. Syst. Man, Cybern.*, no. September, pp. 2174–2179, 2015.

[43] B. Trstenjak, S. Mikac, and D. Donko, "KNN with TF-IDF based framework for text categorization," *Procedia Eng.*, vol. 69, pp. 1356–1364, 2014.

[44] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," *Proc. first Instr. Conf. Mach. Learn.*, pp. 1–4, 2003.

[45] R. E. Banchs, *Text mining with MATLAB®*. 2014.

[46] I. Feinerer, K. Hornik, and D. Meyer, "Text Mining Infrastructure in R," *J. Stat. Softw.*, vol. 25, no. 5, pp. 1–54, 2008.

[47] Wikipedia, "Tag Cloud." [Online]. Available: https://en.wikipedia.org/wiki/Tag_cloud#cite_note-2. [Accessed: 14-Jun-2017].

[48] N. Dwivedi and S. Yadav, "Sentiment analysis: A critical review," *Int. J. Control Theory Appl.*, vol. 9, no. 17, pp. 8607–8613, 2016.

[49] D. Steinberg, "Chapter 10 CART : Classification and Regression Trees," *Foundations*, no. December, pp. 179–201, 2009.

[50] L. Breiman, "Random Forests," pp. 1–33, 2001.

[51] M. Cord, P. Cunningham, W. F. Edv, I. Datenbanken, D. O. De, and S. Wirtschaft, "Machine Learning Techniques for Multimedia schnell und portofrei erhältlich bei," 2008.

[52] geoffrey H. Ashraf .M Kibriya, Eibe frank, Bernhard Pfahringer, "Multinomial Naive Bayes for Text Categorization Revisited." p. 14.

[53] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI/ICML-98 Work. Learn. Text Categ.*, pp. 41–48, 1998.

[54] L. Speech, Processing, Daniel, H. James, and Martin, "Naive Bayes and Sentiment Classification," 2017.

[55] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*. 2014.

[56] B. M. Jadav and M. E. Scholar, “Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis,” *Int. J. Comput. Appl.*, vol. 146, no. 13, pp. 26–30, 2016.

[57] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C.-C. Chang, and C.-C. Lin, “R Package e1071 Version 1.6-8,” *Gpl-2*. 2017.

[58] C. Chang and C. Lin, “LIBSVM : A Library for Support Vector Machines,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2. pp. 1–39, 2013.

[59] and C.-J. L. Chih-Wei Hsu, Chih-Chung Chang, “A Practical Guide to Support Vector Classification,” *BJU Int.*, vol. 101, no. 1, pp. 1396–400, 2008.

[60] J. H. M. Daniel Jurafsky, “Logistic Regression,” in *Speech and Language Processing*, Second edi., Sttanford: Prentice Hall, 2008, 2009.

[61] S. Della Pietra, V. Della Pietra, and J. Lafferty, “Inducing Features of Random Fields,” vol. 19, no. 4, pp. 1–13, 2004.

[62] adam berger, “the Improved Iterative Scaling Algorithm: a Gentle Introduction,” pp. 1–4.

[63] A. Kurt, A. Karatzoglou, and D. Meyer, “Package ‘ RWeka ,’” 2018.

[64] T. Fawcett, “An introduction to ROC analysis,” vol. 27, pp. 861–874, 2006.

[65] data school, “Simple guide to confusion matrix terminology,” 2014. [Online]. Available: http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/. [Accessed: 20-Nov-2017].

[66] D. M. W. Powers, “Evaluation : From Precision , Recall and F-Factor to ROC , Informedness , Markedness & Correlation,” no. December, 2007.

[67] Wikipedia, “Confusion matrix,” 2017. [Online]. Available: https://en.wikipedia.org/wiki/Confusion_matrix. [Accessed: 27-Oct-2017].

[68] M. I. J. Andrew Y. Ng, “On decriminative vs. Generative classifiers: A comparison of Logistic Regression and Naive Bayes,” *NIPS*, pp. 841–848, 2002.

[69] A. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.

[70] P. Flach, “Repairing Concavities in ROC Curves.”