Collaborative Analysis of Customer Feedbacks using Rapid Miner

Snigdha Dixit
M Tech
Computer Science & Engineering
GGITS
Jabalpur, India

Santosh Kr
Vishwakarma/Associate Professor
Computer Science & Engineering
GGITS
Jabalpur, India

ABSTRACT
Today, In the period of focused advertising environment, promoting organizations attempt their best to pick up the client consideration for buying of item and attempt to support the great position among their rival. They are giving distinctive plan and offers to pull in the client and after effect of their endeavours can be measured by interest of items among client, however the most critical part of each promoting organization is to know the client input about their item since client are fulfil with the marking of item as well as they have confidence in client audit or criticism of the individuals who have been utilizing a specific product. Now hence every e-business site requesting that customer give review about the item they purchased, so they might have many reviews or a large number of feedbacks which are not in basic shape so it is troublesome for any other new customer to get the last decision about any item weather it is good or bad based on these feedback. So this paper shown the collaborative analysis of customer feedback on certain item .To acquire the criticism, gather reviews of customers from the e-trade site. These reviews is in common dialect i.e. English dialect, so with a specific end goal to get the some valuable data from these reviews there is need to apply data mining system, in this strategy information is given as these content reviews which changed over as helpful data which is then use to develop the classifier that can anticipate whether good reviews, bad reviews or mixed reviews has been given by customers. Rapid Miner is tools which assembled the classifier and in addition ready to apply on testing dataset.

General Terms
Data-mining, Text-mining, K-NN algorithm ,Naïve Bayes algorithm.

Keywords
Collaborative customer feedback analysis, Review, natural language, e-commerce website. Nokia Lumia, Sony Xperia, Samsung.

1. INTRODUCTION
Collaborative Customer Feedback analysis is use to decide the overall feedback of customers on certain item. With the assistance of reviews. what organization conveyed can precisely be look at by foresee the attitude, satisfaction, dissatisfaction, happiness, angriness of customers on specific item .And through these response new customer can take the last decision to purchase it or not .Every e-business site ask to customers to give reviews yet they are in expansive in number ,customers can read few reviews and might take one-sided choice so to finish up the feedback of customers. This paper suggest to performing community oriented analysis the find out which classifier gives the best result in terms of better precision and recall ratios and in the given conditions. This customers feedback which require to work with information mining to constructed the classifier and afterward text mining system to change over the dataset from regular dialect to some significance data so it can be utilized further.

Data mining-Data mining is the process of analysis of large database in order to find uncover fact or hidden pattern from it(also called at knowledge discovery).It involves several fact to find hidden pattern from large database. Data pre-processing, data analysis, data interpretation process and data classification are main process of data mining. Data classification is the method which classify the data into various form .The algorithm which is used for classification called as classifier, classifier is a mathematical function which is implemented by algorithm that map input to data category.

Text mining is a burgeoning new field that attempts to glean meaningful information from natural language text. It may be loosely characterized as the process of analyzing text to extract information that is useful for particular purposes. Compared with the kind of data stored in databases, text is unstructured, amorphous, and difficult to deal with algorithmically. Nevertheless, in modern culture, text is the most common vehicle for the formal exchange of information

In this work, analysing feedback of customer on three different mobile brand nokia lumia ,sony xperia and Samsung for this two different classifiers has build to extract the feedback of customers, those are shared in e-commerce website and classify them broadly into 3 categories – good, bad and mixed. Here two classifier algorithm Naive Bayes and K-NN are used . Compare the precision and recall ratios of both classifiers and then also find out which classifier gives the best result in terms of better precision and recall ratios and in the given conditions.

2. RELATED WORK
The work is closely related to Michael Gamon,Anthony Aue,Simon Corston and Eric Ringger[1] have worked on the “Pulse:Mining Customer Opinions from free Text”. They have perform the text mining on review of customer on car .They have collected about 4 lack dataset and their research is focus on analysis of opinion(‘sentimental classification’)typically using supervised machine learning ,combine two dimension of topic and sentiment and present the result in an intuitive visualization.

Feedback of customer on qualify of camera is also analysed by Minging Hu and Bing Lu[2].They also collect the data from e-commerce website and performed text mining and classify as positive and negative feedback and compare the precision and recall ratios of both classifiers and then also research work is different from above in two aspects.1)Here performing text mining by using tool called Rapid Miner.2)For classification purpose we are using Naive Bayes.

3. METHODOLOGY

In this work, the device that is utilized is Rapid Miner[7]. Rapid Miner is a product stage created by the organization of the same name that gives a coordinated situation to machine learning, information mining, content mining, prescient examination and business investigation. This gives more than 1500 drag and drop operation with the help of which large number of data mining can be performed easily and quickly . For our work, we will utilize the text mining, classification, validation, perusing and so on.

For changing the feedback of customer which is in natural language into valuable structure for information mining utilize content preparing procedures Tokenization that parts the content of a record into a grouping of tokens to be utilized later independently. Next is Transform Cases that changes all characters in an archive to either bring down case or capitalized. Stop words is utilized to prevent words those doesn't add addition meaning. Stem (Porter) stems English words utilizing the Porter stemming calculation applying an iterative, guideline based substitution of word additions aiming to lessen the length of the words until a base length is come to

For classification purpose, two classifiers – Naive Bayes classifier[10] are in use. A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes theorem (from Bayesian statistics) with strong (naive) independence assumptions. A naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Naive Bayes classifiers can handle an arbitrary number of independent variables, whether continuous or categorical. Given a set of variables, X = {x1, x2, x3..., xd}, we want to construct the posterior probability for the event Cj among a set of possible outcomes C = {c1, c2, c3..., cd}. In a more familiar language, X is the predictors and C is the set of categorical levels present in the dependent variable. Using Bayes' rule

$$p\left(C_j \,\middle|\, x_1, x_2, \dots, x_d\right) \propto p\left(x_1, x_2, \dots, x_d\right) \,\middle|\, C_j \Big) \, p(C_j) \qquad \text{- Eq (1)}$$

Where p(Cj | x1, x2, x3..., xd) is the posterior probability of class membership, i.e., the probability that X belongs to Cj. Since Naive Bayes assumes that the conditional probabilities of the independent variables are statistically independent we can decompose the likelihood of a product of terms:

$$p(X \mid C_j) \propto \prod_{k=1}^{d} p(x_k \mid C_j) \qquad \text{- Eq (2)}$$

And rewrite the posterior as:

$$p\left(C_j \,\middle|\, X\right) \propto p(C_j) \prod_{k=1}^{d} p(x_k \mid C_j) \qquad \text{- Eq (3)}$$

Using Bayes' rule above, we label a new case X with a class level Cj that achieves the highest posterior probability[10].

Many classifier available in text mining ,here naive bayes is used along with one more classifer called K-NN algorithm to compare the result.

K-Nearest Neighbor makes predictions based on the outcome of the *K* neighbors closest to that point. Therefore, to make predictions with *KNN*, we need to define a metric for measuring the distance between the query point and cases from the examples sample. One of the most popular choices to measure this distance is known as Euclidean.

$$D(x, p) = \sqrt{(x - p)^2} \qquad \text{- Eq (1)}$$

Where *x* and *p* are the query point and a case of the examples sample, respectively.

Since *KNN* predictions are based on the intuitive assumption that objects close in distance are potentially similar, it makes good sense to discriminate between the *K* nearest neighbors when making predictions. Let the closest points among the *K* nearest neighbors have more say in affecting the outcome of the query point. This can be achieved by introducing a set of weights *W*, one for each nearest neighbor, defined by the relative closeness of each neighbor with respect to the query point.

$$W(x, p_i) = \frac{\exp(-D(x, p_i))}{\sum_{i=1}^{k} \exp(-D(x, p_i))} \qquad \text{- Eq (2)}$$

Where $D(x, p_i)$ is the distance between the query point *x* and the *i*th case p_i of the example sample. The weights defined in this manner above will satisfy:

$$\sum_{i=1}^{k} W(x_0, x_i) = 1 \qquad \text{- Eq (3)}$$

Thus, for classification problems, the maximum of y is taken for each class variables.

$$\max(y) = \sum_{i=1}^{k} W(x_0, x_i) y_i \qquad \text{- Eq (4)}$$

Figure 1 shows the flow of main process. Process documents from files operator is used for reading text data available in any document file. Validation operator is used for providing training and applying different data mining algorithms in any process. For three different brand we have follow same procedure for three time

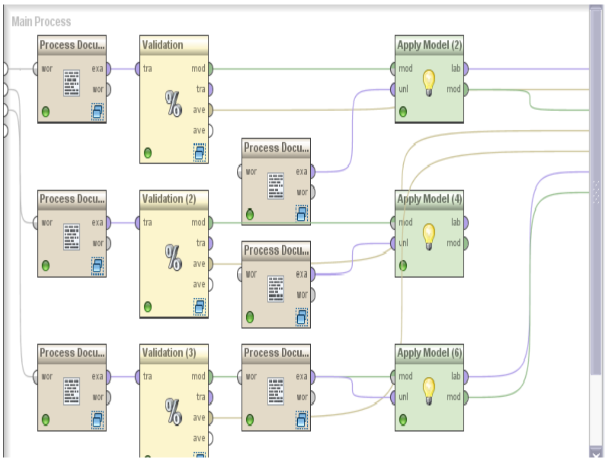

Figure 1: Main Process

Figure 2 shows the text mining operators used for pre-processing the text files before applying for training and testing. Tokenize, filter tokens, transform case filter stop words and stemming operators are used to perform text mining related operations on the training and testing dataset.

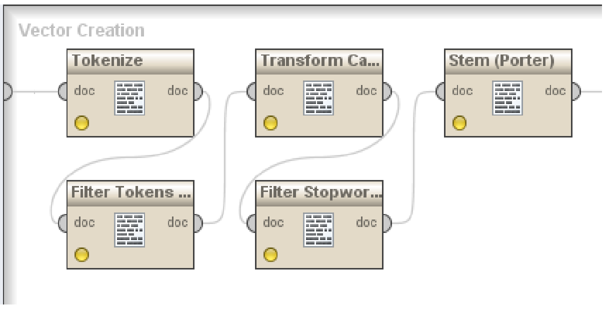

Figure 2: Text mining operators

Figure 3.1 the sub processes within the X Validation operator Naive Bayes classifier operator . and Figure 3.2 the sub processes within the X Validation operator K-NN classifier operator

Apply Model is first trained on an Example Set; information related to the Example Set is learnt by the model. Then that model can be applied on another Example Set usually for prediction. It is compulsory that both Example Sets should have exactly the same number, order, type and role of attributes. Performance operator is used for performance evaluation of only classification tasks. For evaluating the statistical performance of a classification model the data set should be labelled i.e. it should have an attribute with *label* role and an attribute with *prediction* role. The *label* attribute stores the actual observed values, whereas the *prediction* attribute stores the values of *label* predicted by the classification model under discussion

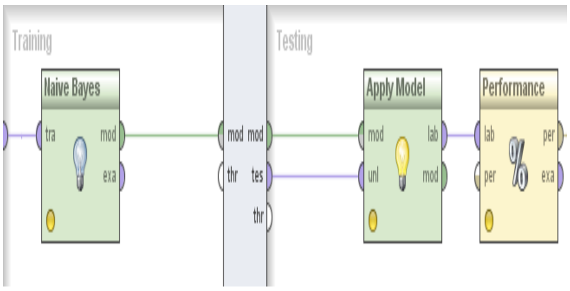

Figure 3.1: Naive Bayes Classifier

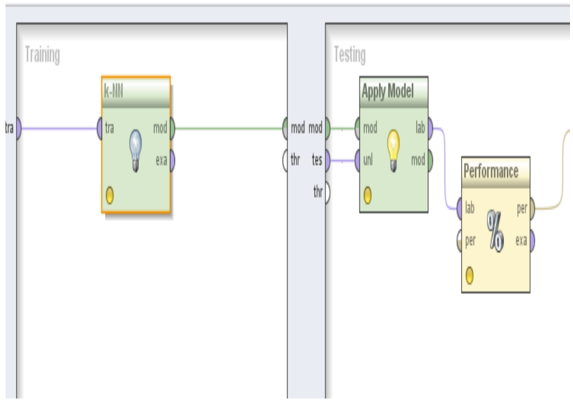

Figure 3.2: K-NN Classifier

4. EXPRIMENT AND PERFORMANCE ANALYSIS

Dataset that is used in this work, are review of customers collected from amazon.com[8]. They generally are in natural language. Hence we processed them with the text mining operators available in Rapid Miner before applying to the classifiers for training as well as testing. For providing training, need to collect review and classified them manually into 3 types of class labels– good, bad and mixed. These class labels will be used to train the classifier and then based on this learning the classifier predict the label of the testing dataset. Table 1 shows the examples. Data mining algorithms in any process. For three different brand we have follow same procedure for three time.

Table 1.1 Examples of reviews in the labels

Label	Review
Good	“Last week, “The phone was very good & it's fulfillment my expectation”
Bad	“camra is not good not atall satisfied”
Mixed	“Camera and sound clarity is best ,but battery back up is little down”

The text files provided for the testing are being predicted in one of the predefined labels - good, bad and mixed using naive bayes classifiers. The results are shown in the figures below.

Results from Naive Bays

Row No.	label	metadata_file	metadata_p...	metadata_d...	confidence(mixed)	confidence(good)	confidence(bad)	predicion(label)
1	testing	t1.txt	E:\thesis_wc	8 Mar, 2016	0	0	1	bad
2	testing	t10.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
3	testing	t100.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
4	testing	t101.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
5	testing	t102.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
6	testing	t103.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
7	testing	t104.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
8	testing	t105.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
9	testing	t106.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
10	testing	t107.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
11	testing	t108.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
12	testing	t109.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
13	testing	t11.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
14	testing	t110.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
15	testing	t111.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
16	testing	t112.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good

Figure 4: Precision and Recall ratio for the classification of Samsung

accuracy: 47.00% +/- 21.32% (mikro: 47.12%)

	true mixed	true good	true bad	class precision
pred. mixed	8	14	4	30.77%
pred. good	13	34	7	62.96%
pred. bad	4	13	7	29.17%
class recall	32.00%	55.74%	38.89%	

Figure 5:Accuracy of Naïve Bayes Classifier for Samsung

Row No.	label	metadata_file	metadata_p...	metadata_d...	confidence(mixed)	confidence(good)	confidence(bad)	prediction(label)
1	testing	t1.txt	E:\thesis_wc	8 Mar, 2016	1	0	0	mixed
2	testing	t10.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
3	testing	t100.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
4	testing	t101.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
5	testing	t102.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
6	testing	t103.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
7	testing	t104.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
8	testing	t105.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
9	testing	t106.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
10	testing	t107.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
11	testing	t108.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
12	testing	t109.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
13	testing	t11.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
14	testing	t110.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
15	testing	t111.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
16	testing	t112.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good

Figure 6: Precision and Recall ratio for the classification of Sony xperia

accuracy: 47.00% +/- 21.32% (mikro: 47.12%)

	true mixed	true good	true bad	class precision
pred. mixed	8	14	4	30.77%
pred. good	13	34	7	62.96%
pred. bad	4	13	7	29.17%
class recall	32.00%	55.74%	38.89%	

Figure 7:Accuracy of Naïve Bayes Classifier for Sony

Row No.	label	metadata_file	metadata_p...	metadata_d...	confidence(mixed)	confidence(good)	confidence(bad)	prediction(label)
1	testing	t1.txt	E:\thesis_wc	8 Mar, 2016	0	1	0	good
2	testing	t10.txt	E:\thesis_wc	18 Feb, 201	1	0	0	mixed
3	testing	t100.txt	E:\thesis_wc	18 Feb, 201	0	0	1	bad
4	testing	t101.txt	E:\thesis_wc	18 Feb, 201	0	0	1	bad
5	testing	t102.txt	E:\thesis_wc	18 Feb, 201	0	1	0	good
6	testing	t103.txt	E:\thesis_wc	18 Feb, 201	0	1	0	good
7	testing	t104.txt	E:\thesis_wc	18 Feb, 201	1	0	0	mixed
8	testing	t105.txt	E:\thesis_wc	18 Feb, 201	0	0	1	bad
9	testing	t106.txt	E:\thesis_wc	18 Feb, 201	0	1	0	good
10	testing	t107.txt	E:\thesis_wc	18 Feb, 201	1	0	0	mixed
11	testing	t108.txt	E:\thesis_wc	18 Feb, 201	0	0	1	bad
12	testing	t109.txt	E:\thesis_wc	18 Feb, 201	1	0	0	mixed
13	testing	t11.txt	E:\thesis_wc	18 Feb, 201	0	0	1	bad
14	testing	t110.txt	E:\thesis_wc	18 Feb, 201	1	0	0	mixed
15	testing	t111.txt	E:\thesis_wc	18 Feb, 201	1	0	0	mixed
16	testing	t112.txt	E:\thesis_wc	18 Feb, 201	0	0	1	bad

Figure 8: Precision and Recall ratio for the classification of Nokia Lumia

accuracy: 29.94% +/- 8.27% (mikro: 29.89%)

	true mixed	true good	true bad	class precision
pred. mixed	11	45	55	9.91%
pred. good	6	29	8	67.44%
pred. bad	11	4	15	50.00%
class recall	39.29%	37.18%	19.23%	

Figure 9:Accuracy of Naïve Bayes Classifier for Nokia Lumia

Results from K-NN:

Row No.	label	metadata_file	metadata_p...	metadata_d...	confidence(mixed)	confidence(good)	confidence(bad)	prediction(l...
1	testing	t1.txt	E:\thesis_wc	8 Mar, 2016	0	0	1	bad
2	testing	t10.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
3	testing	t100.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
4	testing	t101.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
5	testing	t102.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
6	testing	t103.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
7	testing	t104.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
8	testing	t105.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
9	testing	t106.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
10	testing	t107.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
11	testing	t108.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
12	testing	t109.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
13	testing	t11.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
14	testing	t110.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
15	testing	t111.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
16	testing	t112.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
17	testing	t113.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good

Figure 10: Precision and Recall ratio for the classification of Samsung

accuracy: 64.64% Changes to a plot view of the confusion matrix.

	true mixed	true good	true bad	class precision
pred. mixed	10	8	4	45.45%
pred. good	10	50	7	74.63%
pred. bad	5	3	7	46.67%
class recall	40.00%	81.97%	38.89%	

Figure 11:Accuracy of K-NN Classifier for Samsung

Row No.	label	metadata_file	metadata_p...	metadata_d...	confidence(mixed)	confidence(good)	confidence(bad)	prediction(l...
1	testing	t1.txt	E:\thesis_wc	8 Mar, 2016	1	0	0	mixed
2	testing	t10.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
3	testing	t100.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
4	testing	t101.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
5	testing	t102.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
6	testing	t103.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
7	testing	t104.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
8	testing	t105.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
9	testing	t106.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
10	testing	t107.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
11	testing	t108.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
12	testing	t109.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
13	testing	t11.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
14	testing	t110.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
15	testing	t111.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
16	testing	t112.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good

Figure 12: Precision and Recall ratio for the classification of Sony

accuracy: 52.67% +/- 11.33% (mikro: 52.67%)

	true mixed	true good	true bad	class precision
pred. mixed	49	31	35	42.61%
pred. good	1	18	2	85.71%
pred. bad	0	2	12	85.71%
class recall	98.00%	35.29%	24.49%	

Figure 13:Accuracy of K-NN Classifier for Sony

Row No.	label	metadata_file	metadata_p...	metadata_d...	confidence(mixed)	confidence(good)	confidence(bad)	prediction(l...
1	testing	t1.txt	E:\thesis_wc	8 Mar, 2016	0	0	1	bad
2	testing	t10.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
3	testing	t100.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
4	testing	t101.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
5	testing	t102.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
6	testing	t103.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
7	testing	t104.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
8	testing	t105.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
9	testing	t106.txt	E:\thesis_wc	18 Feb, 201(	1	0	0	mixed
10	testing	t107.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
11	testing	t108.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
12	testing	t109.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
13	testing	t11.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
14	testing	t110.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good
15	testing	t111.txt	E:\thesis_wc	18 Feb, 201(	0	0	1	bad
16	testing	t112.txt	E:\thesis_wc	18 Feb, 201(	0	1	0	good

Figure 14: Precision and Recall ratio for the classification of Nokia Lumia

accuracy: 46.23% +/- 3.80% (mikro: 46.20%)

	true mixed	true good	true bad	class precision
pred. mixed	4	2	0	66.67%
pred. good	24	76	73	43.93%
pred. bad	0	0	5	100.00%
class recall	14.29%	97.44%	6.41%	

Figure 15:Accuracy of K-NN Classifier

Figure 4 and 10 shows the classification of reviews using the Naive Bayes classifier and K-NN classifier respectively in the testing dataset into the labels, i.e. good, bad and mixed of Samsung mobile according to learning provided to the Naive Bayes classifier and K-NN classifier respectively during the time of training. Figure 5 and 11 shows the precision and recall ratios of the Naive Bayes classifier and K-NN classifier respectively for the prediction done on the testing dataset.

Figure 6 and 12 shows the classification of reviews using the Naive Bayes classifier and K-NN classifier respectively in the testing dataset into the labels, i.e. good, bad and mixed of Sony xperia mobile according to learning provided to the Naive Bayes classifier during the time of training. Figure 7 and 13 shows the precision and recall ratios of the Naive Bayes classifier and K-NN classifier respetively for the prediction done on the testing dataset.

Figure 8 and 14 shows the classification of reviews using the Naive Bayes classifier and K-NN classifier respectively in the testing dataset into the labels, i.e. good, bad and mixed of Lumia mobile according to learning provided to the Naive Bayes classifier during the time of training. Figure 9 and 15 shows the precision and recall ratios of the Naive Bayes classifier and K-NN classifier respectively for the prediction done on the testing dataset.

Table1.2- Comparing three brand according to their feedback using naive bayes

Label	Good	Bad	Mixed	Recommend
Samsung	62.96%	29.17%	30.77%	Highly
Sony	47.14%	48.78%	51.28%	Moderate
Lumia	67.44%	50%	9.91%	Low

Table1.3- Comparing three brand according to their feedback using K-NN

	Good	Bad	Mixed	Recommend
Samsung	74.63%	46.67%	45.45%	Highly
Sony	85.71%	85.71%	62.61%	Moderate
Lumia	43.93%	100%	66.67%	Low

Table1.4- Comparing accuracy between naive bayes and K-NN Classifier:

	Naive Bayes	K-NN
Samsung	47%	64.64%
Sony	48.67%	52.67%
Lumia	29.94%	46.23%

5. CONCLUSION

In this paper, the collaborative analysis on customer reviews on three brand of mobile phone has been performed. For analysis of textual data i.e. feedback of customers through the process of text mining. Text mining is performed by the tool called rapid miner and through the result we land at conclusion that Samsung is giving best organizations over nokia lumia and sony xperia or customer are more satisfied by Samsung. With this essential objective get achieved i.e. to analysis of large dataset of customer feedback and close which is best. We trust this research is progressively will be being used as more individuals give their sentiment in e-trade. What's more, looking at from reviews from site on various brand valuable for customer as well as for manufacture.

In future, research can be utilize more refine strategy to give more precision and manage the some other issue like decide the quality of opinion, additionally build the span of the testing dataset and can look at the more brand of cellular telephone as huge number of versatile brand are accessible in market. Not just with portable brand however for other thing we can perform same analysis. We can use other labels and apply classification

6. REFERENCES

[1] Michael Gamon, Anthony Aue, Simon Corston-Oliver, and Eric Ringger :Natural Language Processing, Microsoft Research, Redmond, WA 98052, USA "Pulse: Mining Customer Opinions from Free Text"

[2] Minqing Hu and Bing Liu Department of Computer Science from University of Illinois at Chicago 851 South Morgan Street Chicago, IL 60607-7053 "Mining and Summarizing Customer Reviews."

[3] Handbook of Natural Language Processing, Second Edition, edited by Nitin Indurkhya, Fred J. Damerau

[4] Data Mining Methods for the Content Analyst. By Kalev Leetaru .

[5] Data Mining Techniques, Tenth Edition by Arun K Pujari.

[6] Data Mining Concepts and techniques, Third Edition by Kamber.

[7] RapidMiner.com

[8] www.amazon.in.

[9] www.microsoft.com/enin/mobile/phone/lumia/520/ review

[10] Data Mining Concepts and techniques, Third Edition by Kamber