IJCSC 0973-7391 Volume 8 • Issue 2 March 2017 - Sept 2017 pp. 89-96 available online at www.csjournals.com

A Predictive Model for Classification of Opinion Mining

Jayanti, Dr. Anupam Bhatia
M.Phil Scholar , Assistant Professor
Department of Computer science and Applications, Chaudhary Ranbir Singh University, Jind (Haryana)
Pisces.jayanti@gmail.com, anupam.bhatia@crsujind.org,

Abstract: Opinion Mining is the task of extracting from a set of documents opinions, expressed by a source on a specified target. Opinion Mining is extracting people's opinion from the web. It analyses people's opinions, appraisals, attitudes, and emotions toward organizations, entities, persons, issues, actions, topics, and their attributes. The presented Opinion Mining approach belongs to the category of Feature-Based Opinion Mining and aims at extracting and analysing customer opinions on products in forum postings. It comprises four succeeding steps: selection, extraction, aggregation, and analysis. Classification is a Data Mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks. For data analysis, among various classification techniques one is used i.e. Naïve Classification using Cross Validation.

For data analysis, among various classification techniques one is used i.e. Naïve Classification using Cross Validation. Main aim of paper is to detect spam mails and classify whether it will be spam or not using Naïve based Classification technique, using this model the accuracy to detect spam mails is 82%. The Accuracy can be improved by using different Classification Techniques.

Keywords: JAS – Joint Aspect /Sentiment, **KDD** – Knowledge Discovery from Database, **KNN** – K Nearest Neighbor, **MRA –** Mutual Reinforcement Approach, **MRA –** Mutual Reinforcement Approach.

1. INTRODUCTION

Data Mining refers to extracting or "mining" knowledge from large amounts of data. Data Mining can be viewed as a result of the natural evolution of Information Technology. Data Mining tools predict behaviors and future trends, allowing businesses to make proactive, knowledge-driven decisions [1]. Data Mining tools can solve business questions that traditionally were too time-consuming to resolve. In addition, these tools scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data Mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data. While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data Mining software analyses relationships and patterns in stored transaction data based on open-ended user queries.Opinion Mining is a type of natural language processing for tracking the mood of the public about a particular product. Opinion Mining, which is also called sentiment analysis, involves building a system to collect and categorize opinions about a product. Opinion Mining can be useful in several ways [7] - It can help marketers evaluate the success of an ad campaign or new product launch, determine which version of a product or service are popular and identify which demographics like or dislike particular product features. Classification is used to find out in which group each data instance is related within a given dataset. It is used for classifying data into different classes according to some constrains. Several major kinds of classification algorithms including C4.5, ID3, k-nearest neighbor classifier, Naive Bayes, SVM, and ANN are used for classification. Generally a classification technique follows three approaches Statistical, Machine Learning and Neural Network for classification.

Machine Learning techniques like Naive Bayes (NB), K-Nearest Neighbor (KNN) have obtained great success in text categorization.

Naive Bayes (NB) is a simple but effective Learning & Classification algorithm. It is mostly used in Text Classification. The Classification method is based on theory of probability. It plays a vital role in probabilistic classification. It is also used in statistical method for classification and Supervised Learning method. When Bayesian classifiers apply to large databases, it exhibited high accuracy. Naive Bayes Classification method is easy to implement. It requires only a small set of practical training data to judge a standard quantity which satisfies a particular set of equations. In most of the cases, good results are acquired through this classification method [9].

K-Nearest Neighbor (KNN) is a simplest algorithm of all machine learning algorithms. It is also referred as Lazy Learning, Case-based Reasoning or Memory-based Reasoning. KNN is simple, it yet able to solve most complicated problems. It is a non-parametric method used for classification [9].

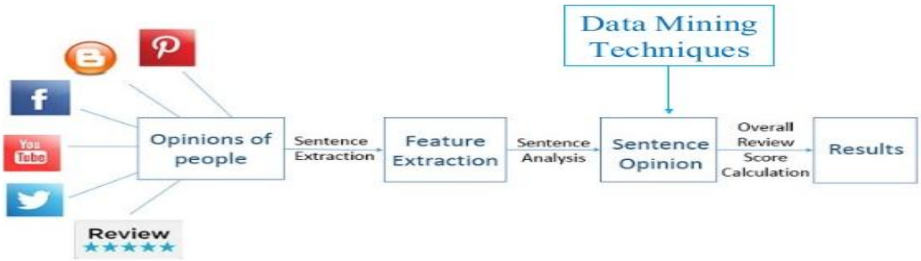


Figure 1: Process of Opinion Mining

II. CHALLENGES IN OPINION MINING

Domain-Independence: The most important challenge faced by opinion mining and sentiment analysis is that the domain dependent nature of sentiment words. One options set could provide excellent performance in one domain, at a similar time it performs terribly poor in another domain [5].

Asymmetry in Availability of Opinion Mining Software: The opinion mining software is extremely overpriced so it's on the far side from the common citizen's expectation.

Detection of Spam and Fake Reviews: The net contains each authentic and spam contents. For effective Sentiment classification, this spam content ought to be eliminated before process. This could be done by distinguishing duplicates, by detecting outliers and by considering reputation of reviewer [5].

Incorporation of Opinion with Implicit and Behavior Data: For successful analysis of sentiment, the opinion words should integrate with implicit data. The implicit data determine the actual behavior of sentiment words [6].

Mixed Sentences: Suppose the word is positive in one situation may be negative in another situation. For e.g. Word LONG, suppose if customer says "The battery life of Samsung mobile is too long "so that would be a positive opinion. But suppose if customer says" That Samsung mobile take too long time to start or to charge" so it would be a negative opinion.

Way of Expressing the Opinion: The people don't always express opinions in the same way. The opinion of every individual is different because the way of thinking, the way of expressing is vary from person to person.

Use of Abbreviations and Short Forms: People using social media more and that to for chatting, expressing their views using shortcuts or abbreviations so the use of colloquial words is increased. Uses of abbreviation, synonyms, special symbols is also increased day by day so finding opinion from that is too difficult. For e.g. F9 for fine, thnx for thanks, u for you, b4 for before, b'coz for because, h r u for how are you etc.

Typographical Errors: Sometimes typographical errors cause problems while extracting opinions.

Orthographics Words: People use orthographic words for expressing their excitement, happiness for e.g. Word Sooo….. Sweeetttt….., I am toooo Haappy or if they in hurry they stress the words for e.g. comeeeee fasssssst I am waitttttnggg.

Natural Language Processing Overheads: The natural language overhead like ambiguity, co-reference, Implicitness, inference etc. created hindrance in sentiment analysis tool [6].

III. DATA EXTRACTION AND METHODOLOGY

Primary Data Collection- The data is collected from a freely available source
https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html#w1a and
http://www.cc.gatech.edu/projects/doi/WebbSpamCorpus.html

Secondary data collection: i) Journals
ii) Newspapers
iii) Books
iv) Internet etc.

Sample Size 10000 sample sizes are used out of 50000 available samples.

Software Used
Rapid Miner Studio 7.3

Initially there were about 50000 rows and 321 columns out of which 10000 samples are extracted as the software used rapid miner studio takes only 10000 values at one time shown in figure 10.Data is then extracted by applying average of the values to fill missing values.
After extraction 10000 samples are taken.

# of classes: 2
# of data: 350,000
# of features: 16,609,143

Methodology There are various methods used for opinion mining and sentiment analysis among which following are the important ones:

1) Naïve Bays Classifier.
2) Support Vector Machine (SVM).
3) Multilayer Perceptron.
4) Clustering.

Classification techniques are used. Among various classification techniques one which is used is naïve based.

NAÏVE BASED CLASSIFACTION

It's a probabilistic and supervised classifier given by Thomas Bayes. According to this theorem, if there are two events say, e1 and e2 then the conditional probability of occurrence of event e1 when e2 has already occurred is given by the following mathematical formula:

$$P\left(\frac{e1}{e2}\right) = \frac{P\left(\frac{e2}{e1}\right) * P(e1)}{e2}$$

This formulae is implemented to calculate the probability of a data to be positive or negative. So, conditional probability of a sentiment is given as:

$$\boldsymbol{P(Word|Sentiment) = \frac{P(Sentiment)P(Sentence|Sentiment)}{P(Sentence)}}$$

And Conditional Probability of a word is given as:

$$\boldsymbol{P(Word|Sentiment) = \frac{Number\ of\ word\ occurence\ in\ class + 1}{Number\ of\ words\ belonging\ to\ a\ class + Total\ no\ of\ word}}$$

IV TOOLS AND TECHNIQUES

1) RAPID MINER - Rapid Miner Studio is a code-free environment for designing advanced analytic processes with machine learning, data mining, text mining, predictive analytics and business analytics. Rapid Miner Studio is a powerful visual design environment for rapidly building complete predictive analytic workflows. This all-in-one tool features hundreds of pre-defined data preparation and machine learning algorithms to support data science projects. It is a data science software platform developed by the company of the same name that provides an integrated environment for machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, validation and optimization. Rapid Miner is developed on an open core model. The Rapid Miner (free) Basic Edition, which is limited to 1 logical processor and 10,000 data rows, is available under the AGPL license.

Rapid Miner Studio is an All-in-One Data Prep and Analytics Designer for Data Mining

- **Rapid Miner Studio is a visual workflow designer that makes it easy to build of complete analytic workflows.** It's code-optional with guided analytics, predefined connections, built-in templates, and repeatable workflows.
- **It contains a huge library of machine learning algorithms and functions to build the best possible model for any use case.** Over 1500+ built-in predefined functions.
- **Rapid Miner Studio is open and extensible with a massive user community and marketplace of add-ons.** Leverage expertise & best practices of 200,000+ users, and easily integrate existing R and Python code into your processes.

2) NAÏVE BAYES - Naive mathematical method is usually used for document level classification. A naive classifier may be a straightforward probabilistic classifier supported applying theorem with robust independence assumptions. The supervised learning technique Naive Bayes uses the Bayes theorem wherever P (A/B) is that the contingent probability of A given B, A is that the category worth and B is that the text we would like to classify.

It considers every feature is conditional freelance to different options given the category. The major advantage of this technique is that a tiny low coaching dataset is enough to coach the classifier. A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be 'independent feature model'. In simple terms, a Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class (i.e. attribute) is unrelated to the presence (or absence) of any other feature.

For example, a fruit may be considered to be an apple if it is red, round, and about 4 inches in diameter. Even if these features depend on each other or upon the existence of the other features, a Naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

The advantage of the Naive Bayes classifier is that it only requires a small amount of training data to estimate the means and variances of the variables necessary for classification. Because independent variables are assumed, only the variances of the variables for each *label* need to be determined and not the entire covariance matrix

.

Advantages*:*

- This technique work well on numeric as well as textual data.
- This classifier is easy to implement and computation are simple comparing with other algorithms.
- As it can be applied to large data set, no complicated iterative parameter estimation schemes are needed.
- Easy interpretation of knowledge representation.
- Performs well and it is robust.

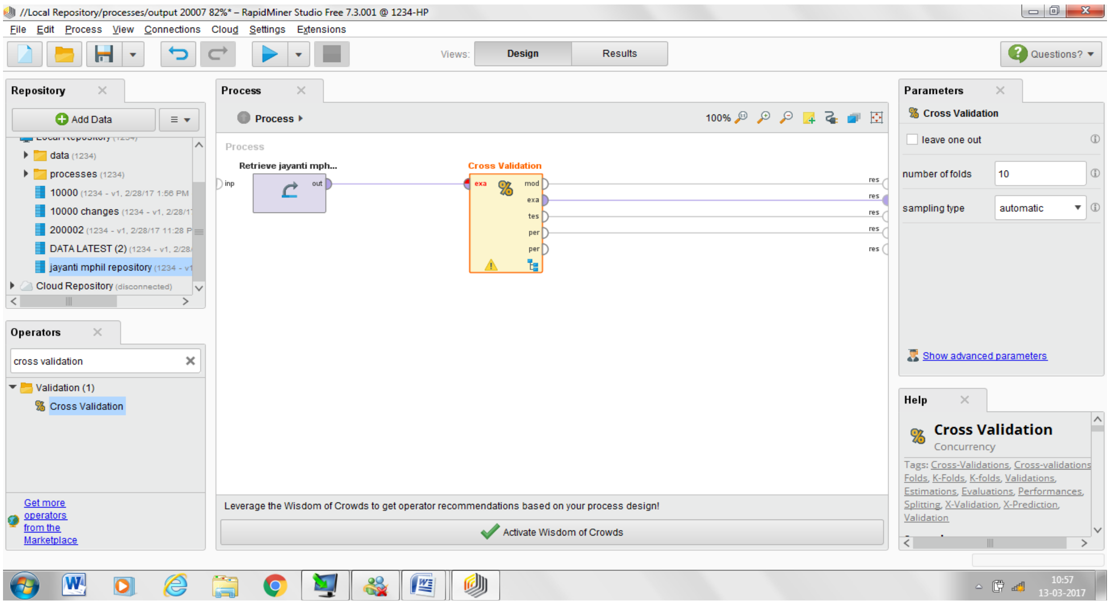
Disadvantages*:*

- Does not consider frequency of word occurrences.
- Theoretically, Naive Bayes classifier have minimum error rate when compared with other classifier, but practically it is not always true, because of the assumption of class conditional independence

V. EVALUATION AND ANALYSIS

In this paper Naïve Bayes classification technique is used using Rapid Miner Studio. Initially data contains more than 50000 samples out of which 10000 samples are extracted and missing values are filled. Extracted data is then imported into rapid miner tool. As freeware rapid miner tool only takes 10000 values therefore only 10000 values are imported. To initiate a process that runs on Rapid Miner Server from either Rapid Miner Studio or Server itself. In either case, the process must be stored in the Server (remote) repository and all process inputs (for example, data or connections) must be accessible to Server.

APPLYING CROSS VALIDATION OPERATOR - This operator performs a cross-validation in order to estimate the statistical performance of a learning operator (usually on unseen data sets). It is mainly used to estimate how accurately a model (learnt by a particular learning operator) will perform in practice. The X-Validation operator is a nested operator. It has two subprocesses: a training subprocess and a testing subprocess. The training subprocess is used for training a model. The trained model is then applied in the testing subprocess. The performance of the model is also measured during the testing phase.



//Local Repository/processes/output 20007 82%* – RapidMiner Studio Free 7.3.001 @ 1234-HP

Parameters: Cross Validation

☐ leave one out

number of folds: 10

sampling type: automatic

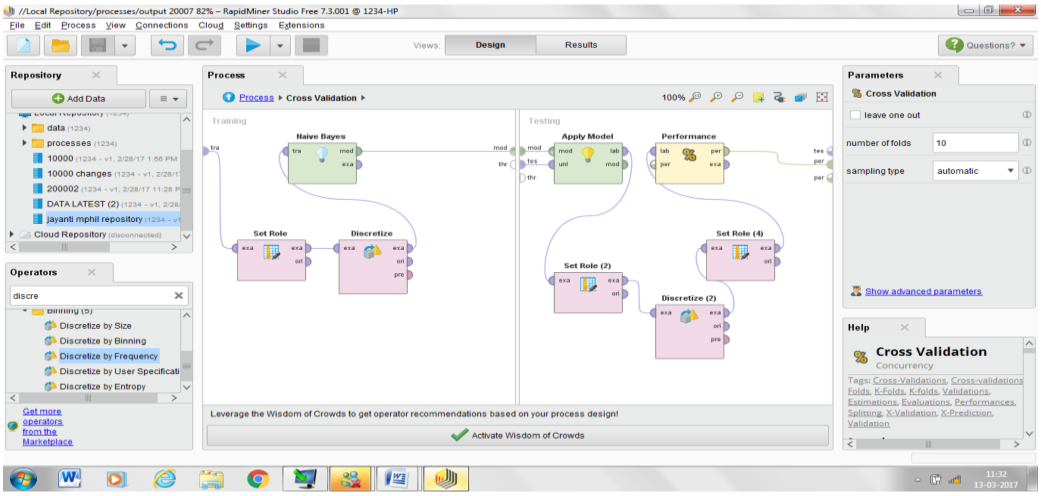
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Cross Validation
Concurrency
Tags: Cross-Validations, Cross-validations, Folds, K-Folds, K-folds, Validations, Estimations, Evaluations, Performances, Splitting, X-Validation, X-Prediction, Validation

Figure 2. Applying Cross Validation Operator

APPLYING TRAINING AND TESTING OPERATOR -

This operator is used for statistical performance evaluation of classification tasks. This operator delivers a list of performance criteria values of the classification task. This operator should be used for performance evaluation of only classification tasks.


//Local Repository/processes/output 20007 82% – RapidMiner Studio Free 7.3.001 @ 1234-HP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Questions?
Repository
Add Data
data (1234)
processes (1234)
10000
10000 changes
200002
DATA LATEST (2)
jayanti mphil repository
Cloud Repository (disconnected)
Operators
discre
Binning (5)
Discretize by Size
Discretize by Binning
Discretize by Frequency
Discretize by User Specificati
Discretize by Entropy
Get more operators from the Marketplace
Process
Process ▸ Cross Validation ▸
100%
Training
Testing
Naive Bayes
Set Role
Discretize
Apply Model
Performance
Set Role (4)
Set Role (2)
Discretize (2)
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Parameters
Cross Validation
Show advanced parameters
Help
Cross Validation
Concurrency
Tags: Cross-Validations, Cross-validations, Folds, K-Folds, K-folds, Validations, Estimations, Evaluations, Performances, Splitting, X-Validation, X-Prediction, Validation
11:32
13-03-2017

□ leave one out

number of folds	10
sampling type	automatic

Figure 3: Applying Training and Testing Operator

ANALYSIS:

To evaluate the result following measures are used:

- □ Accuracy
- □ Precision
- □ Recall
- □ Relevance

Following contingency table is used to calculate the various measures.

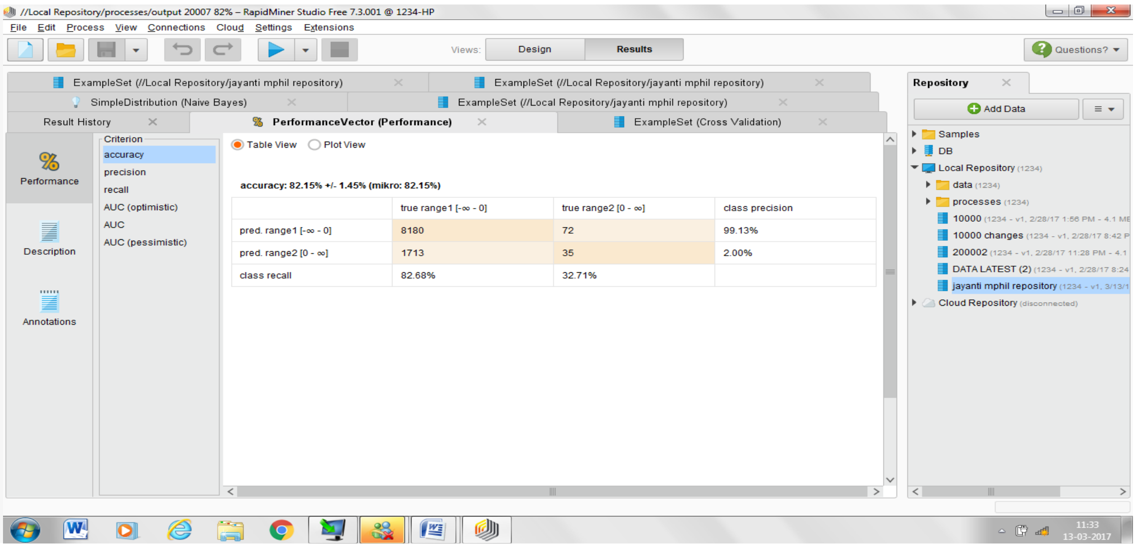
	Relevant	Irrelevant
Detected Opinions	True Positive(tp)	False Positive(fp)
Undetected Opinions	False Negative(fn)	True Negative(tn)

The accuracy comes **82.68%**

Relative number of correctly classified examples or in other words percentage of correct predictions. Accuracy is calculated by taking the percentage of correct predictions over the total number of examples. Correct prediction means examples where the value of the prediction attribute is equal to the value of the label attribute.

The equation for calculating accuracy is:

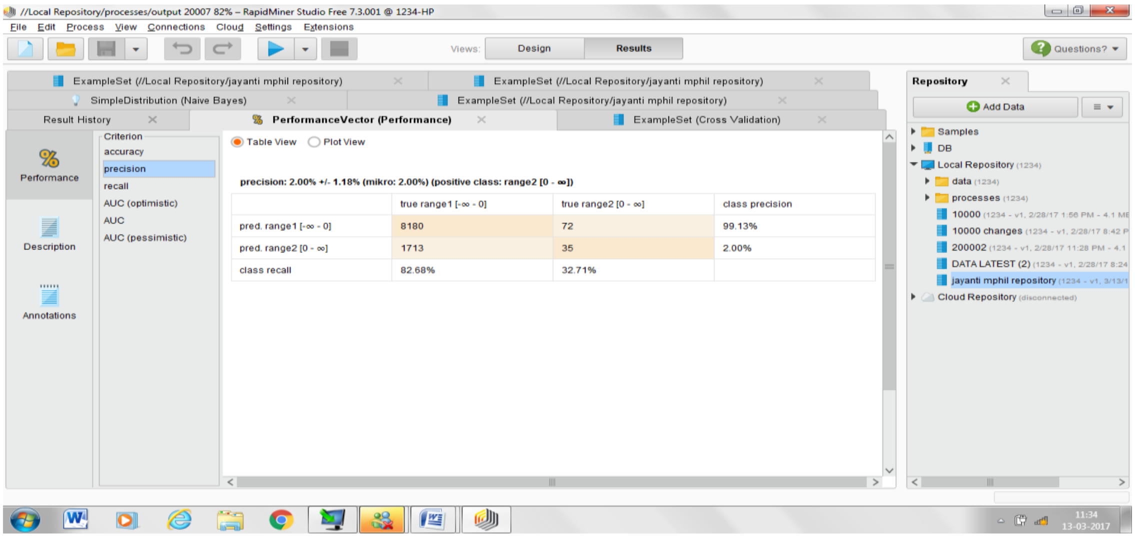
$$\text{ACCURACY}=\frac{(TP+TN)}{TP+TN+FP+FN}$$


//Local Repository/processes/output 20007 82% – RapidMiner Studio Free 7.3.001 @ 1234-HP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Questions?
ExampleSet (//Local Repository/jayanti mphil repository)
ExampleSet (//Local Repository/jayanti mphil repository)
SimpleDistribution (Naive Bayes)
ExampleSet (//Local Repository/jayanti mphil repository)
Result History
PerformanceVector (Performance)
ExampleSet (Cross Validation)
Performance
Description
Annotations
Criterion
accuracy
precision
recall
AUC (optimistic)
AUC
AUC (pessimistic)
● Table View ○ Plot View
Repository
Add Data
Samples
DB
Local Repository (1234)
data (1234)
processes (1234)
10000
10000 changes
200002
DATA LATEST (2)
jayanti mphil repository
Cloud Repository (disconnected)
11:33
13-03-2017

accuracy: 82.15% +/- 1.45% (mikro: 82.15%)

	true range1 [-∞ - 0]	true range2 [0 - ∞]	class precision
pred. range1 [-∞ - 0]	8180	72	99.13%
pred. range2 [0 - ∞]	1713	35	2.00%
class recall	82.68%	32.71%	

Fig 4: Accuracy table view


//Local Repository/processes/output 20007 82% – RapidMiner Studio Free 7.3.001 @ 1234-HP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
ExampleSet (//Local Repository/jayanti mphil repository)
SimpleDistribution (Naive Bayes)
Result History
PerformanceVector (Performance)
ExampleSet (Cross Validation)
Performance
Description
Annotations
Criterion
accuracy
precision
recall
AUC (optimistic)
AUC
AUC (pessimistic)
○ Table View ○ Plot View
Repository
Add Data
Samples
DB
Local Repository (1234)
data (1234)
processes (1234)
10000
10000 changes
200002
DATA LATEST (2)
jayanti mphil repository
Cloud Repository (disconnected)
11:34
13-03-2017

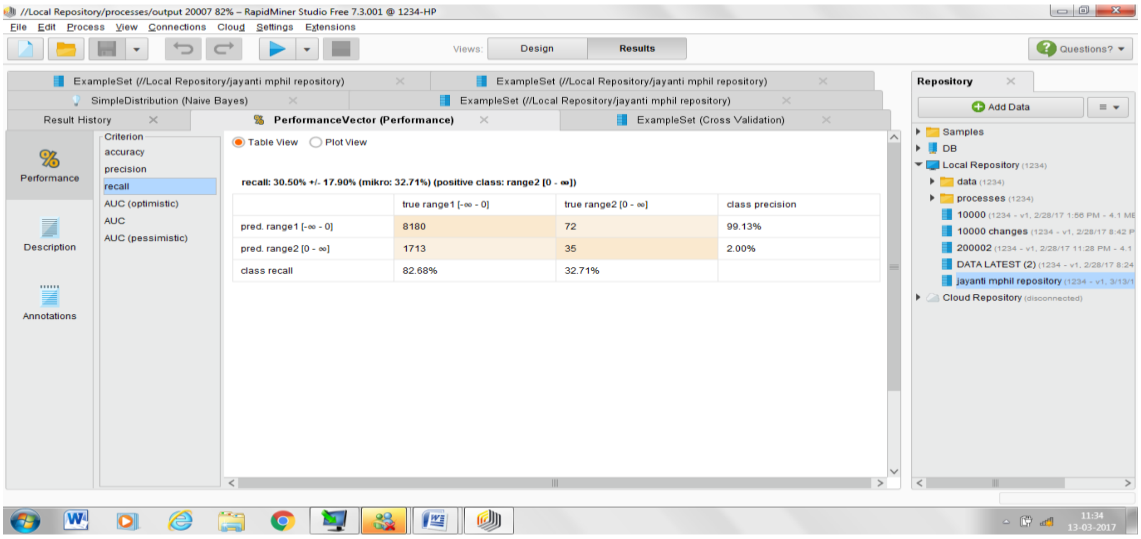
precision: 2.00% +/- 1.18% (mikro: 2.00%) (positive class: range2 [0 - ∞])

	true range1 [-∞ - 0]	true range2 [0 - ∞]	class precision
pred. range1 [-∞ - 0]	8180	72	99.13%
pred. range2 [0 - ∞]	1713	35	2.00%
class recall	82.68%	32.71%	

Fig 5: Precision Table View

The precision can be calculated as :

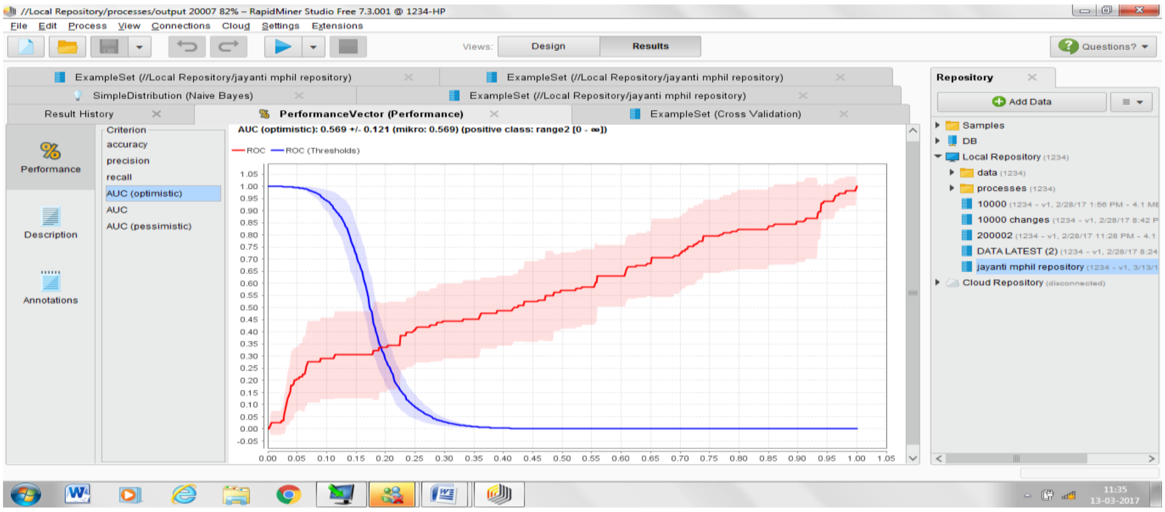
$$\text{PRECISION}=\frac{TP}{TP+FN}$$


//Local Repository/processes/output 20007 82% – RapidMiner Studio Free 7.3.001 @ 1234-HP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
ExampleSet (//Local Repository/jayanti mphil repository)
SimpleDistribution (Naive Bayes)
Result History
PerformanceVector (Performance)
ExampleSet (Cross Validation)
Performance
Description
Annotations
Criterion
accuracy
precision
recall
AUC (optimistic)
AUC
AUC (pessimistic)
○ Table View ○ Plot View
Repository
Add Data
Samples
DB
Local Repository (1234)
data (1234)
processes (1234)
10000
10000 changes
200002
DATA LATEST (2)
jayanti mphil repository
Cloud Repository (disconnected)
11:34
13-03-2017

recall: 30.50% +/- 17.90% (mikro: 32.71%) (positive class: range2 [0 - ∞])

	true range1 [-∞ - 0]	true range2 [0 - ∞]	class precision
pred. range1 [-∞ - 0]	8180	72	99.13%
pred. range2 [0 - ∞]	1713	35	2.00%
class recall	82.68%	32.71%	

Figure 6: Recall Table View



AUC (optimistic): 0.569 +/- 0.121 (mikro: 0.569) (positive class: range2 [0 - ∞])

Figure 7: AUC (Optimistic)

AUC (optimistic) is the extreme case when all the positives end up at the beginning of the sequence. *Range: Boolean.* AUC is used to assess the performance of a binary classifier. AUC is literally the area under the ROC curve, meaning the percentage of the ROC "box" that is below the ROC curve.

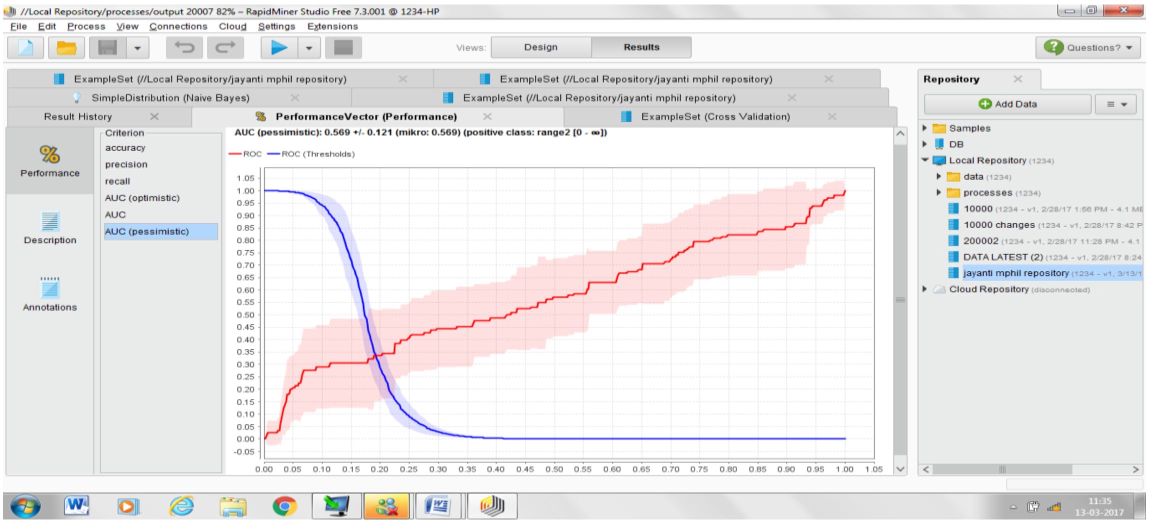
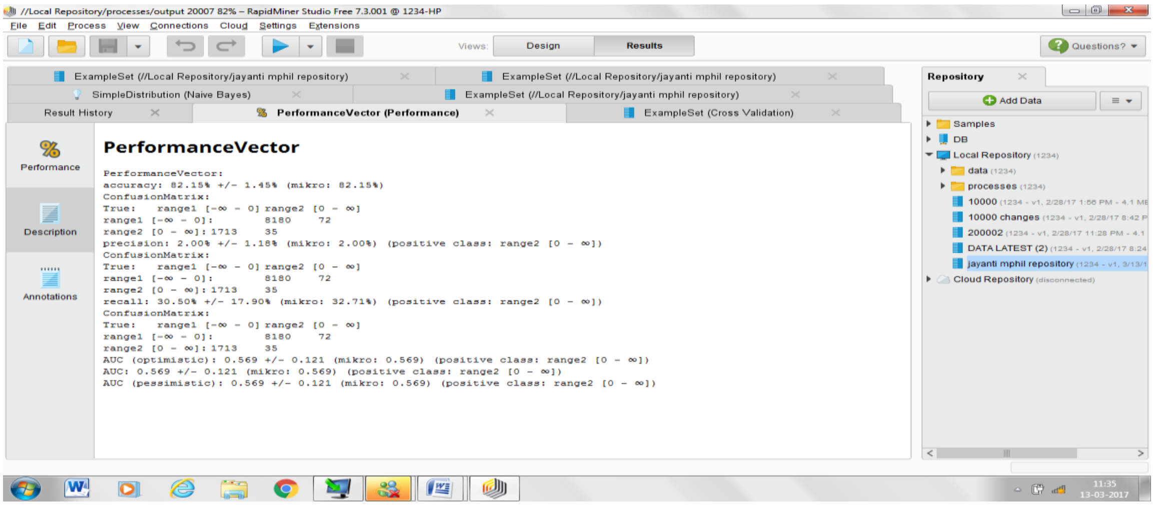


Figure 8: AUC (pessimistic)

AUC (pessimistic) is the extreme case when all the negatives end up at the beginning of the sequence.


//Local Repository/processes/output 20007 82% – RapidMiner Studio Free 7.3.001 @ 1234-HP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Questions?
ExampleSet (//Local Repository/jayanti mphil repository)
SimpleDistribution (Naive Bayes)
Result History
PerformanceVector (Performance)
ExampleSet (Cross Validation)
Performance
Description
Annotations
Repository
Add Data
Samples
DB
Local Repository (1234)
data (1234)
processes (1234)
10000
10000 changes
200002
DATA LATEST (2)
jayanti mphil repository
Cloud Repository (disconnected)
11:35
13-03-2017

PerformanceVector

```
PerformanceVector:
accuracy: 82.15% +/- 1.45% (mikro: 82.15%)
ConfusionMatrix:
True:   range1 [-∞ - 0] range2 [0 - ∞]
range1 [-∞ - 0]:        8180    72
range2 [0 - ∞]: 1713    35
precision: 2.00% +/- 1.18% (mikro: 2.00%) (positive class: range2 [0 - ∞])
ConfusionMatrix:
True:   range1 [-∞ - 0] range2 [0 - ∞]
range1 [-∞ - 0]:        8180    72
range2 [0 - ∞]: 1713    35
recall: 30.50% +/- 17.90% (mikro: 32.71%) (positive class: range2 [0 - ∞])
ConfusionMatrix:
True:   range1 [-∞ - 0] range2 [0 - ∞]
range1 [-∞ - 0]:        8180    72
range2 [0 - ∞]: 1713    35
AUC (optimistic): 0.569 +/- 0.121 (mikro: 0.569) (positive class: range2 [0 - ∞])
AUC: 0.569 +/- 0.121 (mikro: 0.569) (positive class: range2 [0 - ∞])
AUC (pessimistic): 0.569 +/- 0.121 (mikro: 0.569) (positive class: range2 [0 - ∞])
```

Figure 9: Description of Performance Vector

VI. CONCLUSION AND FUTURE WORK

The important part of gathering information always seems as, what the people think. The rising accessibility of opinion rich resources such as online analysis websites and blogs means that, one can simply search and recognize the opinions of others. One can precise his/her ideas and opinions concerning goods and facilities. These views and thoughts are subjective figures which signify opinions, sentiments, emotional state or evaluation of someone.
Following contingency table is used to calculate the various measures.

	Relevant	Irrelevant
Detected Opinions	8180(tp)	72(fp)
Undetected Opinions	1713(fn)	35(tn)

Now , Precision = $\frac{tp}{tp+fp}$

Accuracy =$\frac{tp+tn}{tp+tn+fp+fn}$, F=$\frac{2*Precision*Recall}{Precision+Recall}$, Recall=$\frac{tp}{tp+fn}$

Using Rapid Miner tool NAÏVE BAYES classifier is used to calculate the accuracy of webspam. Cross Validation Operator is used to performing a cross-validation process. From the above analysis we have concluded that Naïve Bayes produce the accuracy of 82%. Attributes are detected to classify whether it will be a spam of not and accuracy to predict spam mails is 82%.Since opinions are always fluctuating therefore accuracy is satisfactory, however the work to improve accuracy will be done in future.

Future work:

Evolutionary techniques are Artificial intelligence techniques may be implemented to improve performance but may not deteriorate speed. By using different Operator Accuracy can be increased.

REFERENCES

[1] http://www.thearling.com/text/dmwhite/dmwhite.htm

[2]https://www.rroij.com/open-access/case-study-of-data-mining-models- andwarehousing.php?aid=46050

[3] https://books.google.co.in/books?isbn=8184501838

4] G. Sahoo, P. Pawar, K. Malvi, A. Jaladi, and K. Khithani, "Environment Monitoring System based on IoT," pp. 1083–1091, 2017.

[5] https://books.google.co.in/books?isbn=9381335052

[6] http://slidewiki.org/print/deck/11306

[7] https://www.slideshare.net/aerofoilkite/opinion-mining-using-data-mining-techniques

[8]https://www.researchgate.net/publication/274509414_Sentiment_analysis_A_review_and_comparative_analysis_of_web_services

[9] M. Sadegh, R. Ibrahim, and Z. A. Othman, "Opinion Mining and Sentiment Analysis : A Survey," *Int. J. Comput. Technol.*, vol. 2, no. 3, pp. 171–178, 2012.

[10]https://www.researchgate.net/publication/274509414_Sentiment_analysis_A_review_and_comparative_analysis_of_web_services

[11] W. Intelligence *et al.*, "Opinion Mining on Newspaper Quotations," pp. 523–526, 2009.

[12] P. Tian, Y. Liu, M. Liu, and S. Zhu, "Research of product ranking technology based on opinion mining," *2009 2nd Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2009*, vol. 4, pp. 239–243, 2009.

[13] K. Khan, B. B. Baharudin, and A. Khan, "Mining opinion targets from text documents: A review," *J. Emerg. Technol. Web Intell.*, vol. 5, no. 4, pp. 343–353, 2013

[14] X. Xu, X. Cheng, S. Tan, Y. Liu, and H. Shen, "Aspect-level opinion mining of online customer reviews," *China Commun.*, vol. 10, no. 3, pp. 25–41, 2013.

[15] W. Wang and Y. Zhou, "E-business websites evaluation based on opinion mining," *Proc. - 2009 Int. Conf. Electron. Commer. Bus. Intell. ECBI 2009*, pp. 87–90, 2009.

[16] P. Jiang, C. Zhang, H. Fu, Z. Niu, and Q. Yang, "An approach based on tree kernels for opinion mining of online product reviews," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 256–265, 2010.