Evaluation of Sentiment Data using Classifier Model in Rapid Miner Tool

Devipriya B, Kalpana Y

Abstract: Evaluation of internet and the usage of internet as websites which is for penetrating to gain a specific requirements, like group communication as social networks (such as face book, twitter,etc.,) ,blogs for opinions, online portals (such as iGoogle, MSN) for communication, experience as reviews, suggestions as opinions, combination of reviews and opinions as recommendations, ratings and feedbacks which is identified and elevating in almost all the field now-a-days. The writers of online portal, review, opinion and recommendation in any social media take measures as beneficial factor for the improvement of businesses, organization, governments and mostly individuals. When this content boost up the study of content and the need of data mining, text mining techniques and sentiment analysis is inescapable. Natural language processing and text analysis techniques are used in sentiment analysis to recognize and extract information from the text [1]. This paper provides a result of sentiment analysis with the intellectual tool named Rapid Miner to show the sentiment comments about the contents in the online traders.

Keywords : Decision tree, K-NN, Naive Bayes, Rapid Miner.

I. INTRODUCTION

Sentiment Analysis is combined with Natural Language Processing (NLP), text analysis and computational analysis techniques are used to extract information and to analyze the sentiment in the reviews. Most of the customer's decision depends on the online reviews to make a decision in purchase the products, buy any software for the business, selecting courses in e-learning portals, booking hotels, events for the organization. Analyze the reviews and examine the score is the primary goal of sentiment analysis. The structure of review is roughly mentioned as structured, unstructured and semi-structuredadverbs, adjectives, phrases, sentences, conjunction with emotional words which are the main challenges in text classification.

NLP techniques include Tokenizingwords (Splitting words and sentences into tokens), StopWords in a language (a stop word is a commonly used word such as "the", "a", "an", "in"), Stemming of words(sort of normalizing method), Part-of-Speech Tagging(labeling words in a sentence as nouns, adjectives, verbs…etc.), Chunking (Group words into hopefully meaningful chunks), Chinking (it is basically a way for you to remove a chunk from a chunk), Named Entities in Natural Languages (chunking is called Named Entity Recognition), Lemmatizing (an advanced concept of Stemming of words).The rest of the paper explains 2.literature review 3.introduction to Sentiment analysis, 4. About Rapid miner, 5.Methodology, 6. Experiment and Performance Analysis and 7. Conclusion.

Sentiment Analysis is generally explained as determining the meaning behind the statement and deciphering the layer behind the sentence. Sentiment Analysis is contextual mining of text. It identifies the text and extricate personalized information from the source content. The information serving in the business or organization to understand significance of the social sentiment of brand, product or service. Monitoring online conversations requires minimal knowledge but producing useful information from the conservation in need of lot of analyzing techniques and algorithms to be applied.

II. LITERATURE REVIEW

Devika M D et al[2]detailed the comparison between different approaches of machine learning approach, rule based approach and lexicon based approach includes both supervised and unsupervised learning in sentiment analysis.The most popular sentiment analysis methods and its various levels of sentiment analyzes have been mentionedby the author. The primarygoal is to come up with Sentiment Analysis which is categorizingwith various reviews. Machine learning famous algorithms like SVM, NB, Maximum Entropy methods were detailed with some other methods. This can reform the analysis process in any way. This paper explains the research work of better analysis methods in semantics by considering n-gram evaluation rather than word by word analysis. The author has also added up some other methods called rule based and lexicon-based methods. In the Internet world, most people confide on social networking sites to collectreasonable information, analyzing the product reviews for understanding and decision-making.

Walaamedhat et al[3] explained sentiment analysis techniques and related developing fields. From 41 articles in the year of 2010 to 2013, the author classified the journal under the major 6 categories named Sentiment Classification (SC), Emotion Detection (ED), Feature Selection (FS), Transfer Learning (TL), Sentiment Analysis (SA), Building Resource (BR). This also includes

Abhishek Kori[4] experimented with Rapid miner 7.4 version with two famous classifiers named K-NN classifier and Naïve Bayes classifier. The author focused on the text classification using two classifiers.

  * Correspondence Author
  Devipriya B, Department of Information Technology, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India. Email: devipriya.scs@velsuniv.ac.in
  Dr Kalpana Y, Department of Information Technology, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India. Email: ykalpanaravi@gmail.com

Retrieval Number: A1323109119/2019©BEIESP
DOI: 10.35940/ijeat.A1323.109119
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

2966

Evaluation of Sentiment Data using Classifier Model in Rapid Miner Tool

The classifiers are implemented in the rapid miner tool. The rapid miner is a software platform by the company rapid miner that furnish an integrated development environment with most popular conception like machine learning, predictive analytic, text mining and deep learning. In this paper, the author experiments the data set Iris, which was given as sample data. The data set resides inside the repository panel of the tool. Validation applied by validation tool on the data set which includes training and testing operations. In the training column, the author used NaiveBayes operator and in testing column apply model and performance tools are applied.After connecting all operators, the accuracy is scored as a result. Finally, the author concludesthe experiment score explains that the accuracy of KNN is greater than Naive Bayes, but this situation is not akin, it may vary based on the data set and performance methods.

Thirunavukkarasu K et al[5] explains that there are handful number of learning algorithms, classifications algorithms, learning patterns and categorization of data are present now-a-days to analyze the data. The main purpose of the algorithm is to find the best algorithm according to the problem to be solved with suitable accepted output. This author evaluates the expected results which are experimented with number of algorithm and also taken for the comparisons study with Neural net, Support Vector Machine, Naive Bayes, Byzantine Fault Tolerant (BFT) and Decision stump algorithms. These algorithms are mostly presented in knowledge discovery and data mining mainly. The accuracy is a predicting values or events from the given data-set using handful number data mining and knowledge discovery algorithms.

III. SENTIMENT ANALYSIS

Sentiment Analysis is the most commonly used as classification tool for the context which analyses an incoming message and conclude whether the underlying sentiment is positive, negative or sometimes neutral.Sentiment Analysis is contingent mining of context that identifies and extricates personalized instruction from the source such as reviews and helping in all kind of business. To understand, develop the business and the social sentiment of their brand, product or service is full filled by monitoring online conversations. These conversations conclude into different states mentioned as positive, negative and neutral. But in this research the positive and negative states are highlighted with the help of tool.

3.1 Popular Sentiment Analysis tools
Some of the best sentiment analysis tools which are useful in real-time world or environments are listed below:

3.1.1 Brand24
Brand24 is a online responding tool. Itpresents and basic quality grip the sentiment analysis as one of the features that can be used in any business to recognize the advantage and disadvantages of the user's review which held hand in development of organization[6].
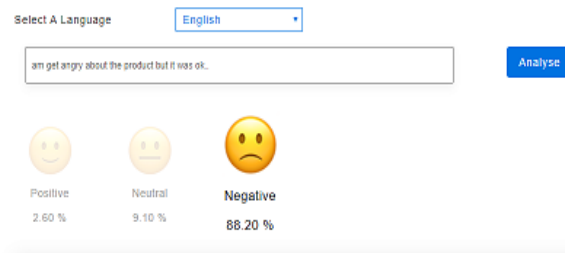

Figure 1: Sentiment analysis tool brand24

3.1.2 Repustate
Repustate furnish the solution for text analytic in different businesses. It has providedin 17 different languages. It starts with part-of-speech tagging, before the actual analysis takes placewhich is stem from the decomposition of context into meaningful grammatical parts. The other important factors like lemmatization, prior polarity, etc., also part of this Repustate tool.

3.1.3 OpenText
OpenText Sentiment Analysis Tool is belonging to OpenText Content Analytics solution. It is comfortable with identify and evaluate subjective patterns and the expressions of sentiment in the content. The analysis happensin different levels such as topic, sentence, and document. It can recognize the portions of text are factual or subjective. It also evaluates the opinion of reviews is expressed as positive, negative, mixed, or neutral.This tool support of five languages just as English, German French, Spanish, and Portuguese.

3.1.4 ParallelDots
ParallelDots is an applied on Artificial Intelligence research group. The sentiment analysis tool is one of their services which offer the service based on Products, APIs and Plug-ins. The sentiment analysis Application Programming Interface (API) uses Long Short-Term Memory (LSTM) algorithms which explains about artificial recurrent neural network (RNN) architecture. This algorithm used to classify a text blob's sentiment into positive and negative. This tool is trained with largeamount of social media data and news data.

3.1.5 Lexalytics
This will explain how to do text analytics workin three steps:
1. The first step is done by Text deconstruction and natural language processing which break the phrases and sentences intoextricated semantics, syntax, etc.
2. The second step is Sentiment analysis which combines with categorization, name entity recognition, intention detection, etc.,
3. And the last step is Business intelligence tools, where the structured data and conclusions are transferred into visualization dashboards or business intelligence tools for predictive analytics.

3.1.6 Hi-Tech BPO
Hitech is a potent solution for data analytics, market intelligence, data processing and a sentiment analysis tool.
This tool has 4 segments in sentiment analysis:
- Opinion mining - focuses on assessing an opinion and its polarity
- Text mining –concentrate on evaluate the content
- Social sentiment- emotion behind media mentions
- Social listening- computing popularity of the brand or organization by information which is extracted from the social media

3.1.7 Social Mention

Social Mention is search engine.It is a free tool. This free social media tool aggregates user-generated content from the web.

3.1.8 Social Searcher

Social Searcher more or similar to Social Mention.This product combines with a variety of tools just as Social Buzz, Media Monitoring, Google Social Search engine, etc.

3.1.9 Sentiment Analyzer

Sentiment Analyzer is a free sentiment analysis tool.It supports content written in English and scales between negative and positive values. This tool is built with the use of more than 8000 writing samples of spoken conversations.

3.1.10 Sentigem

The tool is in the state of development version and encountering some errors is possible.It is in phase of public beta

3.1.11 SentiStrength

Sentistrength provides automatic sentiment analysis with human level English upto 16,000 social web test per second. It's showed by two sentiment strengths:
1. not negativeto extremely negative
2. not positive to extremely positive

3.1.12 Meaning Cloud

The Sentiment Analysis tool is one of the products of meaning cloud. Sentiment Analysis Application Programming Interface(API) analyzes the context by identifying individual phrases and evaluating the relationship between them. Some of the features are global sentiment, detection of irony, difference in sentiment, agreement and disagreement between analyzed textcontents.

3.1.13 Tweet Sentiment Visualization

This tool allows choosing a keyword want to analyzemust be placed into the tool. Then, keywords are visualized on several different graphs, starting with the sentiment.

3.1.14 Rapid Miner

Rapid Miner is data science platforms that combine with analytics and artificial intelligence. This tool includes text mining also

3.1.15 Hoot-suite Insights

This tool used in business decisions and have many products. This is one of the products and biggest proposal is the social media management legend. Sentiment analysis tool is furnished by Hoot-suite to understand other people's feelings about the brand or products.

3.1.16 Talkwalker's Quick Search

Talk walker's product is a social media search engine which is used in brand analysis. Quick Search offers image recognition. This tool compares sentiment between two different brands andin products

3.1.17 Clarabridge

Clarabridge's sentiment analysis tool is effective to Customer Experience Management solution, which consists of CX(Customer eXperience) Analytics and CX(Customer eXperience) Social [2].Customer eXperience-Analytics used by analytic Value Management tools.This tool is for the card and lending industry. Our proprietary depend framework brings unprecedented insight into your lending portfolio and significantly increases the effectiveness of your analytic resources.

If you are using *Word,* use either the Microsoft Equation Editor or the *MathType* add-on (http://www.mathtype.com)

for equations in your paper (Insert | Object | Create New | Microsoft Equation *or* MathType Equation). "Float over text" should *not* be selected.

IV. RAPID MINER TOOL

Rapid Miner studio is visual workflow designer, which is used for predictive analysis. This combines machine learning and big data science for the better analysis. Understanding of the rapid miner studio is important to design the process diagram. The IDE (Integrated Development Environment) has repository, Operators, Views, Global Search, Parameters, help panel, Functional ports and Process panel with usual tool such as file, edit. The benefits of the RapidMiner explained in detail as it offers a robust and very powerful tool and have plenty of features. In rapid miner tool each component is user-friendly environment interface which helps to realize the users about massive productivity gains. It is a visual work-flow designer tool is specially contribute the users with an environment. This environment enables the user to create, design and deploy analytics processes, visual presentations, and models. The guidance of Rapid Miner tool operators uncluttered, meaningless, disorganized, expressions and seemingly useless data also enhanced into very valuable and visualizable. The system facilitates data access, manage accessed data, load and evaluate of data such as texts, images, and audio tracks. Rapid Miner allows the user to structure data to be mending to arrange. Rapid Miner used to create models and plans with a strong set of tools and functionalities so that the user has the possibilities to extricate or extract critical statistics and information. The user is fully allowed to utilize data exploration capabilities and descriptive statistics like graphics and visualization. This software is more powerful which offers analytics related to real-world data transformation settings which gives the mastery of formatting and creating the most favorable data set for user predictive analytics[7].

Explicit Features of the Rapid Miner tool is explained in the below diagram

1. Visual Workflow Designer - Rich library with over of 1500 machine learning algorithms, drop and drop visual interface, pre-built templates, possibility of proactive recommendations.
2. Data Access and Management- Accessing files including SAS (Statistical Analysis System), ARFF(Attribute-Relation File Format),stata(software for statistics and Data science) and via URL.
3. Data Exploration- A label in column target variable.
4. Descriptive Statistics- Uni-variate statistics and plots, Distribution plots, Bivariate statistics and plots.
5. Graphs and Visualization- visualization module is created as an alternative to the well known and old module Plot View.
6. Data Prep- Turbo or data Prep is a new modulein RapidMiner that speed up time-consuming data prep tasks to speed the productivity of analytics teams.
7. Data Sampling -The number of examples in the sample can be specified on absolute, relative or probability basis depending on the setting of the sample parameter.Moreover, the samples are generated randomly.

8. Data Partitioning - In this partitioning the parameter is used to separate the number of subsets or partitions and the relative size of each partition. It is named as partitions parameter.

9. Data Replacement - The data replacement operator enables to select attributes to use replacements in and to a specify regular expression. Attribute values of selected attributes that match this regular expression are replaced by the specified replacement. The replacement can be empty and can contain capturing groups.

10. Weighting and Selection - Select by Weights operator allows selecting attributes using the weights of the attribute.

11. Similarity Calculation - A similarity calculate object measure the calculated similarity between each data to similarity and similarity to data. The Data to Similarity operator and vice verse can generate a similarity calculation or measure object.

12. Clustering - Clustering groups Examples together which are similar to each other. As no Label Attribute is necessary, Clustering can be used on unlabeled data and is an algorithm of unsupervised machine learning.

13. Market Basket Analysis - Market Basket Analysis is a association analysis method which is used to discover attractive relationship between the variables in the given set of data set. The modeling of association between variables is based on the set of items, frequently appeared together.

14. Bayesian Modeling - Naive Bayes model is a high-bias and low-variance classifier. It has the capability to build a better model even with a minimum number of data. It is very simple to use and understand. The use cases mostly involved with text categorization, spam detection, sentiment analysis, and popularly known for recommender systems.

15. Scoring - Rapid Miner Time Scoring is an additional service to Rapid Miner. It is designed for fast scoring use cases through the web services which fulfilled by the components, they are called as Scoring Agents. It access minimal memory and have fast response times.

16. Automation and Process Control - It execute multiple process in parallel. The long-term process can be run in the background, while the other process execute in the front. The main services are background process execution, automatic optimization, scripting in data preparation, logging in process, macros, process control and process based reporting.

V. METHODOLOGY AND PROPOSED FRAMEWORK

Statistical text mining using rapid mining model, includes loading the data, Pre-processing the data, generating term-by-document matrix, building models and lastly applying the model on new data to predict the outcome. A process diagram also created for similarity-based methods and clustering techniques for measuring the similarity between the documents.
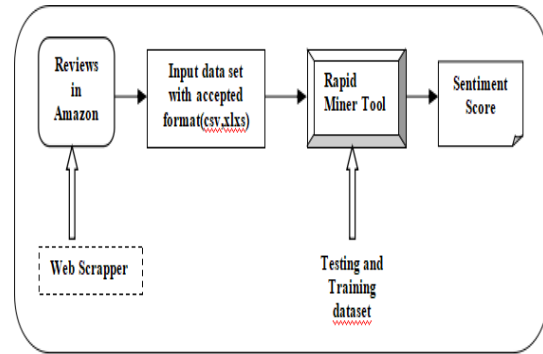

Figure 2: methodology diagram for the proposed framework

The above figure2 explains the idea of proposed framework clearly. First, the data needed to be analysis are gathered from the various online shopping portals. Here, the Amazon reviews are the main data-set to analysis. The book reviews are the main sources. The review data are gathered in the acceptable format using webscrapper. Web scraping or web harvesting is data scraping technique which is manipulated for extracting data from the various websites. Web scraping is software, itenters the World Wide Web directly using either by the Protocol, or by the web browser. The acceptable dataset is entered the tool to analysis based on the training data if the algorithms are depending on supervised learning. Otherwise the data set can be directly accessed by the built-in model depicted in the sentiment analysis tool. Before that the Pre-processing methods are used to the acceptable data-set. Pre-processing includes the transform the given case into acceptable case, Tokenizing of the data, removing the stop words, and stemming the given words. Some of the extension like text processing, web mining and rosette text analysis, etc., also helpful to build the model. The data model is a construction of the algorithm with specified data set. The data-set can be analyzes based on the model of the tool. Finally, based on the model the sentiment score can be scored, using the performance metrics such as precision, recall.

VI. EXPERIMENT AND PERFORMANCE ANALYSIS

Rapid Miner is a great tool to the non-programmers which is especially suitable for data mining and text analysis. This can be explained in the below paragraph, how to do sentiment analysis with Rapid Miner. Rapid Miner is a great tool which has packed with text processing capabilities and third party Application Programming Interface (API) also easy to connect to this tool. The below figure.3 explains the process diagram for the sentiment analysis of Amazon book reviews which is scrapped by the web tool web scrapper.
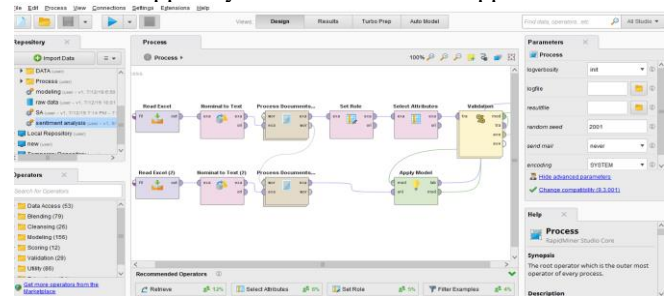

Figure 3: Rapid miner process diagram

2969

Rapid Miner sentiment analysis is starts from read excel operator which is import the excel sheet need to analysis. This read excel operator is main task is to import data from Microsoft Excel spreadsheets. This operator can read data from all compatible Excel sheets. The tool is defining the spreadsheets in the workbook and the table, which have a format such that each row represents as an example and each column represents an attribute. Nominal to text operator changes the type of nominal attributes to readable text, which can be given as an input to the next operator. It mainly concentrate on mapping values of the attributes to respective string values. So the nominal to text operator is mainly used in the place of converting all nominal attributes to string attributes. The main purpose of nominal value is simply used as a string value of the new attribute even though the value is missing in the nominal attribute, the new value will also be missing. Set Role operator is used to change the role of attributes which has several role. The default role is regular and usual option to set the attributes and the other roles are classified as special. If any attributes is assigned as special role, and the other attributes are changed to regular default role. The main purpose of the Select Attributes operator is mainly used to select attributes of the data-set which is to be displayed. This operator has default parameter which is for defining filter conditions using Add Filters option in dialog window. It is also allows to customize the filters with the help of custom filter option. It has condition class used for alter the attributes and variables with different options like all, correct predictions, wrong predictions, no missing attributes, missing attributes, no missing labels, missing labels, attribute value filter, expression, custom filters.
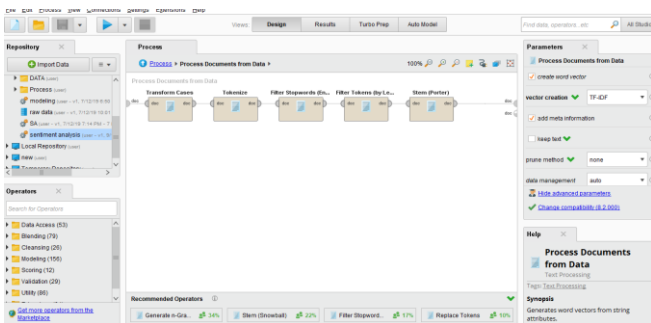

Figure 4: Rapid miner pre-processing process diagram

Pre-processing technique in rapid miner includes five steps which shown in the figure 4. Transform case is used to change one case to another, either upper to lower or vice-verse. Tokenization is an important step in natural language processing tasks. It is used in statistical analysis of text; parsing, word counting, corpus generation and spell checking. This operator dissect the text of a document into a sequence of tokens. There are several options which explains that how to specify the splitting points and the most important three options are discussed below. The first is default settings will result in tokens consisting of one single word, which is the most eligible option before finally building the word vector. And the second option is split mode option. It helped in the splitting of the complete sentences into interesting tokens. This is possible only by setting to specify character and enter all splitting characters. The third option used to define as regular expressions and is the most flexible for very special cases. The non-letter character is treated as a separator. As a result this concept, each word in the text is represented by a single token. Another fascinating step in

pre-processing is Filtering Stop words, which can be explained in two different ways based on the language and number of characters present in the words. The special characters and operators also removed by using the filter stop words operator. This operator filters tokens based on their length when the filter is set as length option. One of the famous operator, the Porter stemmer is used for English words. This operator ruled based on Porter stemming algorithm to stems the English words by applying an iterative, rule-based replacement of word suffixes which concentrates in reducing the length of the words to reach the minimum length.
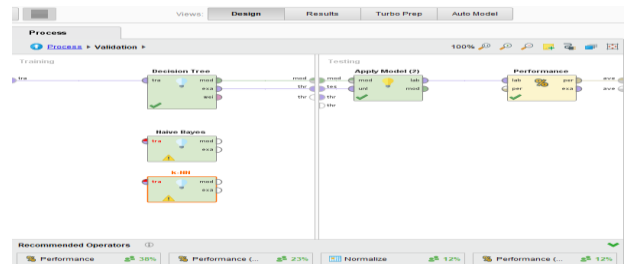

Figure 5: Rapid miner validation diagram

Figure 5 explains the machine learning algorithm used for sentiment analysis in rapid miner. Decision tree is a tree structured algorithm which is collection of node that expresses the values of decision in the shape of tree. Each node represents a specific Attribute based on splitting rule. This splitting rule intended to separates the values into different classes for classification. And for the numerical prediction regression, the splitting rule separates value in order to reduce the error in an optimal way for the selected parameter criterion. The building of new nodes is recursive process until satisfy the basic criterion of prediction. A prediction for the class label Attribute is depending on the Example data-set. The estimation for a numerical value is acquired by averaging the values in a leaf of the tree. These operators mainly operate on training sets which carry both nominal and numerical attributes. The label attribute act as nominal for classification and numerical for regression. After generation of the tree, the decision tree model can be applied by the Apply Model Operator option. Each data values follow the branches of the tree in accordance with the splitting rule which turns the attributes into leaf. The next algorithm used for finding the positivity and negativity in the context. The k-Nearest Neighbor algorithm is intended to compare testing data-set with the k-training data-set which are the nearest neighbors of the testing data-set. The first step in the k-Nearest Neighbor algorithm on a testing data is prepared to find the k- closest training sets. The closeness between the training and testing data-set is defined in terms of a distance between the dimensional space and attributes used in the training set. The Euclidean distance metric is used to calculate the distance between the testing data and the training set. The distance between the closeness is depends on absolute values. It is often recommended to normalize data before training and the normalized data is applied to the k-Nearest Neighbor algorithm. The metric used in k-NN algorithm and its configuration of the hardware is defined based on the parameters of the apply model operator.

Evaluation of Sentiment Data using Classifier Model in Rapid Miner Tool

In the next step, the k-Nearest Neighbor algorithms arrange the unknown data set according to neighbor attributes. In regression, the predicted value from the help of training set find neighbor average value. It can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones [8]. The figure 6 explains the accuracy of sentiment analysis in different algorithm like decision tree, k-NN classifier, Naive Bayes.

Decision Tree

accuracy: 93.10%

	true negative	true positive	class precision
pred. negative	13	2	86.67%
pred. positive	0	14	100.00%
class recall	100.00%	87.50%	

Figure 6a: Rapid miner –decision tree

Figure6b: Rapid miner –Decision tree performance graphical representation

B) K-NN Classification

accuracy: 72.41%

	true negative	true positive	class precision
pred. negative	10	5	66.67%
pred. positive	3	11	78.57%
class recall	76.92%	68.75%	

Figure 6c: Rapid miner –KNN Classification

Naïve Bayes

accuracy: 51.72%

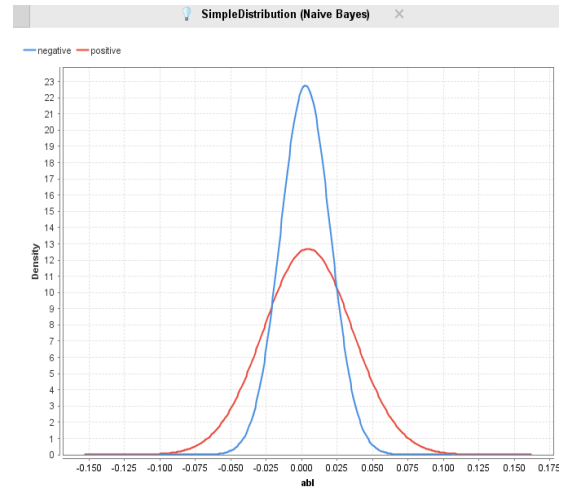|	true negative	true positive	class precision
pred. negative	6	7	46.15%
pred. positive	7	9	56.25%
class recall	46.15%	56.25%	

Figure 6d: Rapid miner –Naive Bayes

Figure 6e: Rapid miner –Naïve Bayes performance graphical representation

Naive Bayes is a high-bias and low-variance classifier. It has the capability to build a best model even with a small quantity of data set. It is very ease to work and understand. Also it is computationally inexpensive compare to other. Text categorization, spam detection, sentiment analysis, and recommender systems are the test cases provided by the bayesian model. The fundamental belief of Naive Bayes the difference between the value of the label and the value of any Attribute is independent of the value of any other Attribute. This belief is sometimes create great contradictory , but the experience of using bayes model shows that the Naive Bayes classifier works well often based on this assumption, that is, it better than any other model often. This Operator uses Gaussian or gauss or probability is a continuous probability method used in this attributes [8]. The above 6b and 6e figure explains the graphical representation of the decision tree model result node with leaf and Bayesian performance graphical representation of result shows the positive and negative words in the reviews.

VII. CONCLUSION

In the digital world the online trading is becoming one of the day-today life activities. The daily digital prints create a massive amount of data which is need some serious attention of the organization, businesses and online traders. The online traders are stand in need of the tools which segregate their customers for their improvement. That leads to sentiment analysis, text mining and data mining are essential. It opens the gateway of sentiment analysis and because of number tools present in analyze the rivalry arise. There are number tools available in market to analyze the reviews of the reviewers, in that some of them are online tools. The desktop or standalone tools also have their own strength to compete with online tools. Rapid miner is one of the best tools which have the advance features like data exploration, sampling, replacing, partitioning, Bayesian modeling, clustering and modeling evaluation. Using the rapid miner tool with three different machine learning algorithm the decision making gives accuracy than the other machine learning algorithms for the given set of books content reviews.

REFERENCES

1. Doaamohey EI-Din Mohamed Hussein, "A Survey on sentiment analysis challenges", Journal of Kings saud University-Engineering Sciences,2016, 1018-3639.
2. Devika M D, Sunitha C, Amal Ganesha," Sentiment Analysis:A Comparative Study On Different Approaches", Fourth International Conference on Recent Trends in Computer Science & Engineering,doi: 10.1016/j.procs.2016.05.124,2016, 1877-0509.
3. Walaamedhat, Ahmed Hassan, HodaKorashy, "Sentiment Analysis algorithms and Applications: A survey", Ain shams Engineering Journal, Volume 5, 2014, ISSN no : 1093-1113.
4. Abhishek Kori, "Comparative study of data classifiers using Rapid Miner", International Journal of Enginerring Development and Research, Volume 5, Issue 2, 2017, ISSN: 2321-9939.
5. Thirunavukkarasu K, Dr Manojwadhawa, "Analysis and Comparison study of Data Mining Algorithms using Rapid Miner", International Journal of Computer Science, Engineering and Applications, Volume 6, Issue no 1, 2016, DOI : 10.5121/ijcsea.2016.6102.
6. https://brand24.com/blog/best-sentiment-analysis-tools/
7. https://reviews.financesonline.com/p/Rapid Miner/
8. https://Rapid Miner.com/products/studio/feature-list/
9. file:///C:/Users/user/Desktop/desktopNEW/Journal/rapid%20miner/Rapid Miner-v6-user-manual.pdf.
10. https://community.Rapid Miner.com/discussion/48663/data-preprocessing-ideas
11. Jovica Krstevski , Dragan Mihajlov , Ivan Chorbev, "Student Data Analysis with Rapid Miner", Information and Communication Technology Innovations, 2011, Web Proceedings, ISSN 1857-7288.
12. Sheena Angra , Sachin Ahuja, "Analysis of Student's Data using Rapid Miner", Journal on Today's Ideas – Tomorrow's Technologies, Vol. 4, No. 2, December 2016 pp. 109-117, DOI: 10.15415/jotitt.2016.42007.
13. Cik Feresa Mohd Foozy , Rabiah Ahmad , M.A. Faizal Abdollah and Chuah Chai Wen," A Comparative Study with Rapid Miner and WEKA Tools over some Classification Techniques for SMS Spam", International Research and Innovation Summit (IRIS2017), doi:10.1088/1757-899X/226/1/012100.
14. D. Kerana Hanirex, K.P. Thooyamani, "An analysis on adult datset in a decision tree using rapid miner tool", Journal of Chemical and Pharmaceutical Sciences, April - June 2016 , Volume 9 Issue 2, ISSN: 0974-2115.
15. Tanu Verma, Renu, Deepti Gaur, "Tokenization and Filtering Process in Rapid Miner", International Journal of Applied Information Systems (IJAIS), Volume 7– No. 2, April 2014, – ISSN : 2249-0868.

AUTHORS PROFILE

Devipriya B is a Ph.D. candidate in Computer Science and Applications of Vels Institute of Science, Technology and Advanced studies (VISTAS). Mrs Devipriya has completed Master of Computer Applications from Anna University affiliated college and has been working as an Assistant professor in private Institution.Email id: devipriya.scs@vesluniv.ac.in

Dr Y Kalpana is working in Department of Computer Applications and Information Technology. Dr Kalpana has completed Master of Computer Applications and has 20 years of teaching experience and 7 years of research experience. Her Area of research is image processing and neural networks.Email id:ykalpanaravi@gmail.com