International Journal of Scientific Research in Computer Science and Engineering **Research Paper**

Vol.8, Issue.1, pp.98-104, February (2020) **E-ISSN:** 2320-7639

Resolution of E-Commerce Market Trend Using Text Mining

Rajeev Tripathi[1*], Santosh Dwievdi[2]

[1,2]Dept. of Computer Science, SRMGPC, Lucknow, India

[*]*Corresponding Author: rajeev2363@gmail.com, Tel.: +00-983849-1149*

Available online at: www.isroset.org

Received: 22/Jan/2020, Accepted: 10/Feb/2020, Online: 28/Feb/2020

Abstract— In the present mechanical complex world a few web based business sites like Amazon, Jabong.com, Myntra.com and Flipkart, and other web based shopping locales shelter gather item audits from clients to determine the fulfillment level on exact items. Information examination are down to earth on item surveys so as to realize useful investigative data as measurements that can bolster individuals working in an association for business examination in settling on very good quality choices so as to look for out the interest of client against their current business rivals. Enormous organizations around the globe understand that online business isn't simply purchasing and selling over Internet, rather it improves the fitness to contend with different monsters in the market. For this aim information mining in some cases called as information disclosure is utilized. We accomplish this by managing various patterns in the content information like content representation, content mining methods in this way examining the topic on which a book has been created by perusing a couple of html records from a nearby organizer. In Product Ranking System, audits assumes basic job in deciding client satisfaction just as market tendency for that fastidious item, state on the off chance that regarding electronic items to get pertinent information in less time.

Keywords— Electronic commerce, Data Mining, Text Mining, Business intelligence and analytics, Big data analytics

I. INTRODUCTION

Data analytic are used to analyze huge volumes of stored data generated by an organization to extract useful information which helps business people to make better decisions and thereby improve the profit from products they manufactured. Now with exhaustive growth among competent organizations, new technologies and tools should be used to analyze huge volumes of information stored in various heterogeneous sources which contains both structure and unstructured data to predict future business trends.

There are lots of users who purchase products through e-Commerce websites. Through online shopping several e-commerce enterprises were incapable to know whether the customers are satisfied by the services provided by the firm. This boosts to develop a system where various customers give reviews about the product and online shopping services, which in turn help the e-Commerce enterprises and manufacturers to get customer estimation to improve service and commodities through mining customer reviews.

In today's world social media has become a valuable source of information, where people express their opinions. Analysis of such opinion-related data can provide industrious insights. When these opinions are relevant to a company, accurate analysis can provide them with information like product quality, influences affecting other customer decisions, early feedback on newly launched products, trends and also knowledge about their competitors.

e-Commerce is the trading of goods and services, or the transmitting of funds or data, over an Internet. There are lots of resources of news available through different sources. It can either be available on the e-news paper that can be read on the internet or hard copies of the same newspaper, or articles available online. All these resources come under text. If the talk is about analyzing numerical or categorical data, we are at a very safe side and we can take a chance of working on the data. But the condition here is different all together.

We are dealing with loads of data that is presented to us in the form of text, which definitely has numbers but not the traditional rows and columns that make things easy for us. It has got paragraphs of text that just brings in the most complicated issues. Luckily for us, statistics has been very lenient in allowing us to use the conventional methodologies of machine learning and data mining that we generally use for standard data, that can be applied for text data as well(6).

The immergence of the digital age has resulted in a host of new information that online retailers can use to improve their marketing efforts and differentiate themselves from competitors to gain customers' business. Several others are turning to technology such as text mining software to tailor the shopping experience to consumers' personal preferences.

Text mining solutions are used to analyze digitized text from different written sources (e.g., search engines, blogs and forums) and social media platforms (e.g., Twitter and Facebook) to identify patterns and trends on brand affinity, product preferences, consumption patterns and more.

Sentiment analysis is used to conclude if a conversation about a brand is positive or negative. To help retailers make better use of text mining solutions, we surveyed consumers to find out how they experience about retailers' use of this technology to improve their e-commerce experience. This report highlights our most important findings.

One major benefit of text mining software is that it can assist retailers understanding which products customers are most possible to purchase.

II. RELATED WORK

As of late information based subject models were utilized which had the essential of clients giving area information preceding mining. Different inquire about works are done to discover an answer for deep rooted or steady learning. The outcomes were seen as drastically superior to anything best in class mining calculations. Content examination additionally incorporates notion investigation. It is the way toward deciding assessment in a book. A few classifiers are accessible for conclusion investigation of client assessment. It is anything but difficult to group content into classes of enthusiasm utilizing such classifiers. A specific use of item suggestion are utilized in enterprises these days that the possibility of information examination used to lead clients of different electronic item that would ready to impact clients bound to purchase that specific item. Previous explore in the zone of item proposal incorporates attempting to get statistic data from notices of item adopters from online audits. Item adopters are the individuals who have purchased the item for another person. These adopters are then classified into client bunches dependent on their physical areas. The outcomes are then joined by prescribing items to those in comparable client bunches utilizing regularized lattice factorization.

AI approaches mimic the manner in which people gain from their past encounters to get information and apply it in settling on future choices. These learning methods are generally utilized in man-made brainpower and record characterization. The order utilizing AI can be summarized in two stages: 1) Learning the model utilizing the preparation dataset 2) Applying the prepared model to the test dataset Sentiment examination is a book arrangement issue and hence any current managed grouping strategy can be applied. Our work utilizes the Naive Bayes classifier and Support Vector Machines for grouping the film surveys and thinks about the outcomes got utilizing the two methodologies. Credulous Bayes classifier is a basic probabilistic classifier that depends on the Bayes hypothesis. This arrangement method accept that the nearness or nonattendance of any component in the archive is autonomous of the nearness or nonappearance of some other element. Gullible Bayes classifier thinks about a report as a pack of words and expect that the likelihood of a word in the record is free of its situation in the archive and the nearness of other word.For a document d and class c:

$$p(c|d)=p(d|c)P(c)/p(d)$$

Support vector machines have been the most important way for document classification. These are large margin classifiers and perform better than Naïve Bayes and Maximum Entropy in almost all cases. The fundamental idea behind SVM classification is to find a maximum margin hyper plane that separates the document vector in one class from the other with maximum margin.

Text Mining

Content Mining is a type of information mining strategy that can empower associations to find any sort of conceivable significant business related contributions from information that is gotten from content which can be of the type of articles, messages, tweets or Facebook posts or an occupation notice on LinkedIn. It is a huge errand to mine content information, for example unstructured information with the assistance of Natural Language Processing (NLP), Machine Learning or any factual demonstrating method, in any case, there are a few inconsistencies with NLP. So frequently, the NLP is the last stop we might want to visit for content mining. It will have anomalies which are ascribed its not all that ideal sentence structure and it contains ambiguities regarding language related issues and numerous dishes of perceptions .

The essential favorable position of embracing a content investigation delicate product can help with the change of words and expressions in the content information into numerical information that can be mapped to organized information in a social database and afterward it is com-pared with the ordinary information mining techniques. By fol-lowing a tedious methodology, any organization can achieve the usage of content examination so as to assemble an information on content explicit perceptions like conclusion, force, humor and so forth to give some examples. Be that as it may, content investigation isn't generally executed crosswise over various ventures. So the final products and the exhaustiveness of the investigation might be not quite the same as one to the next.

Normally, to play out the activity of grouping effectively, for example to check if a classifier component is ordered relevantly, we have to required train it with some pre-grouped archives from each classification, just to ensure that it is in a situation to sum up the model it has gain from the past reports and apply a similar model to precisely arrange the records that are not seen.(2)

The essential goal of the NLP is to comprehend the how the machines are assessing the data from the normal language of the people and perform models with air conditioning curacy of careful accuracy. It is an exceptionally sound

propensity for passing on a message which is very much organized and has some reasonable data, with the execution of not all that efficient and unstructured information. Another examination expresses that content mining thinks about the data/information from where the information should be removed understanding that it very well may be utilized for explicit undertakings. It is very conceivable that the content mining is thinking about including NLP process into the model to ensure that the human language is effectively assessed and the unstructured information designs are efficient. With the headways in the innovation with each other minute, the idea of content mining will discover its application to effective execution and this legitimizes why individuals are settling on content mining.

The main objective of the NLP is to recognize the how the machines are evaluating the facts from the natural language of the human beings and perform fashions with ac- curacy of surgical precision. It is a very wholesome dependancy of conveying a message which is very well structured and has some smart information, with the implementation of now not so well geared up and unstructured data. Another lookup states that textual content mining studies the information/ statistics from where the data needs to be extracted understanding that it can be used for unique tasks. It is pretty possible that the textual content mining is thinking about involving NLP method into the mannequin to make positive that the human language is efficiently evaluated and the unstructured records patterns are well organized. With the developments in the technological know-how with each different moment, the notion of textual content mining will discover its software to profitable implementation and this justifies why people are opting for text mining. (17).

III. METHODOLOGY

TEXT MINING ALGORITHMS LIST

1. K-MEANS CLUSTERING

The essential objective of the NLP is to understand the how the machines are evaluating the statistics from the herbal language of the humans and perform fashions with ac- curacy of surgical precision. It is a very healthful dependancy of conveying a message which is very nicely structured and has some good information, with the implementation of now not so nicely equipped and unstructured data. Another research states that textual content mining studies the information/ data from where the statistics needs to be extracted understanding that it can be used for specific tasks. It is pretty feasible that the textual content mining is thinking about involving NLP manner into the mannequin to make positive that the human language is efficiently evaluated and the unstructured data patterns are nicely organized. With the advancements in the technology with every different moment, the notion of textual content mining will find its utility to successful implementation and this justifies why people are opting for textual content mining.

2. Naive Bayes Classifier

Naive Bayes is regarded one of the most fantastic data mining algorithms. It is a easy probabilistic algorithm for the classification tasks. The Naive Bayes Classifier is primarily based on the so-called Bayesian theorem and offers excellent and reliable effects when it is used for text data analytics. Naive Bayes classifier is now not a single algorithm however a household of algorithms which expect that values of the elements used in the classification are independent. It is very easy to code with the standard programming languages such as PHP, JAVA, C#, etc. As one of the pleasant textual content classification techniques, Naive Bayes has a range of applications in e-mail unsolicited mail detection, report categorization, email sorting, age/gender identification, language detection and sentiment analysis.

3. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is also one of the most used textual content mining algorithms due to the fact of its simplicity and efficiency.KNN is a non-parametric approach that we use for classification. In a few words, KNN is a easy algorithm that shops all current facts objects and classifies the new information objects based on a similarity measure. In the text evaluation domain, it is used to check the similarity between archives and ok education data. The aim is to determine the category of the take a look at documents. One of the largest text mining functions of KNN is in “Concept Search” (i.e. searching for semantically similar documents) – a function in software program tools, which is used for helping companies find their emails, enterprise correspondence, reports, contacts, etc.

4. Support Vector Machines (SVM)

This method is one of the most accurate classification text mining algorithms. Practically, SVM is a supervised desktop learning algorithm in most cases used for classification issues and outliers detections. It can be additionally used for regression challenges.SVM is used to type two records units by using comparable classification. This statistics analysis algorithm draw traces (known as hyperplanes) that separate the businesses according to some patterns.The purpose of SVM is to create this hyperplane. The hyperplane with the maximum margin from each corporations is best. In the real world, SVM can mannequin complex problems such as text and photograph classification, hand-writing recognition, face detection, and biosequence analysis.

When it comes to textual content mining, SVM is broadly used for textual content classification things to do such as detecting spam, sentiment analysis, report classification into classes as news, emails, articles, web pages, etc.

Tune Parameters of SVM?

Tuning parameters value for machine learning algorithms effectively improves the model performance. Let’s look at the list of parameters available with SVM.

sklearn.svm.SVC(C=1.0, kernel='rbf', degree=3, gamma=0.0, coef0=0.0, shrinking=True, probability=False,tol=0.001, cache_size=200,


```
class_weight=None, verbose=False, max_iter=-1, random_state=None)
```

I am going to discuss about some important parameters having higher impact on model performance, "kernel", "gamma" and "C".

kernel: We have already discussed about it. Here, we have various options available with kernel like, "linear", "rbf","poly" and others (default value is "rbf"). Here "rbf" and "poly" are useful for non-linear hyper-plane. Let's look at the example, where we've used linear kernel on two feature of iris data set to classify their class.

Example: Have linear kernel

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn import svm, datasets
# import some data to play with
iris = datasets.load_iris()
X = iris.data[:, :2] # we only take the first two features. We could
# avoid this ugly slicing by using a two-dim dataset
y = iris.target
# we create an instance of SVM and fit out data. We do not scale our
# data since we want to plot the support vectors
C = 1.0 # SVM regularization parameter
svc = svm.SVC(kernel='linear', C=1,gamma=0).fit(X, y)
# create a mesh to plot in
x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1
h = (x_max / x_min)/100
xx, yy = np.meshgrid(np.arange(x_min, x_max, h),
np.arange(y_min, y_max, h))
plt.subplot(1, 1, 1)
Z = svc.predict(np.c_[xx.ravel(), yy.ravel()])
Z = Z.reshape(xx.shape)
plt.contourf(xx, yy, Z, cmap=plt.cm.Paired, alpha=0.8)
plt.scatter(X[:, 0], X[:, 1], c=y, cmap=plt.cm.Paired)
plt.xlabel('Sepal length')
plt.ylabel('Sepal width')
plt.xlim(xx.min(), xx.max())
plt.title('SVC with linear kernel')
plt.show()
```

Example: Have rbf kernel

Change the kernel type to rbf in below line and look at the impact.

```
svc = svm.SVC(kernel='rbf', C=1,gamma=0).fit(X, y)
```

I would suggest you to go for linear kernel if you have large number of features (>1000) because it is more likely that the data is linearly separable in high dimensional space. Also, you can RBF but do not forget to cross validate for its parameters as to avoid over-fitting.

gamma: Kernel coefficient for 'rbf', 'poly' and 'sigmoid'. Higher the value of gamma, will try to exact fit the as per training data set i.e. generalization error and cause over-fitting problem.

Example: Let's difference if we have gamma different gamma values like 0, 10 or 100.

```
svc = svm.SVC(kernel='rbf', C=1,gamma=0).fit(X, y)
```

5. Decision Tree

Decision Tree algorithm is a normal machine mastering method for data mining that creates classification or regression fashions in the structure of a tree structure. The shape consists of a root node, branches, and leaf nodes. Each internal node indicates a check on an attribute and every department indicates the result of a test. Finally, each leaf node shows a category label. Decision Tree algorithm is nonlinear and simple. As a text mining algorithms, Decision Trees has many functions such as examining all the textual content that comes from client relationship management. It is additionally used in making medical predictions based totally on clinical records files and etc.

6. Generalized Linear Models (GLM)

Generalized Linear Models is a popular statistical method used for linear modeling. Actually, GLMs combine a large variety of fashions such as linear regression models, logistic regression, Poisson regression, ANOVA, log-linear fashions and etc. Combining the linear approach with records mining equipment has many advantages such as accelerating the modeling procedure and achieving higher accuracy. Some of the fine content material evaluation software program carriers (such as Oracle) use GLM as one of the key textual content mining algorithms.

7. Neural Networks

Neural networks are nonlinear models which represent a metaphor for the functioning of the human brain. Despite that Neural networks have a complex structure and lengthy training time, they have their location in facts evaluation and text mining algorithms.In the domain of textual content analytics, Neural network can be used for grouping comparable patterns, for classifying patterns, and etc.The utility of the neural network is essential in records mining due to the fact of some traits such as self-organizing adaptiveness, parallel performance, fault tolerance, and robustness.When it comes to text data analysis, neural networks are famous in the location of scientific lookup documents, finance, and advertising content mining.

8. Association Rules

Association regulations are simply if/then statements that purpose to discover some relationships between unrelated data in a given database. They can find relationships between the items which are often used together. Popular applications of affiliation rules are basket information analysis, cross-marketing, clustering, classification, catalog design, etc. For example, if the client buys eggs then he may additionally also purchase milk.

Using this approach in the are of text data mining, can assist customers to achieve knowledge from the collection of the extraordinary kind of content material such as internet documents (to limit the time for analyzing all those documents).Another instance is, the association guidelines used for identifying fine or negative associations between symptoms, medications, and laboratory results and clinical text statistics reports.

9. Genetic Algorithms

Genetic algorithms or evolutionary algorithms are a family of stochastic search algorithms witch mechanism is inspired by means of the manner of neo-Darwinian evolution. Naturally, GAs have applied binary strings (chromosomes) to encode the points that structure an individual. They essentially attempt to imitate the human evolution. The purpose for the usage of GAs for statistics mining is that they are adaptive and sturdy search techniques. GAs can clear up numerous textual content statistics mining problems such as clustering, the discovery of classification rules, attribute choice and construction.

10. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation is one of the techniques which currently is used in theme text modeling. In fact, latent Dirichlet Allocation (LDA) is a generative probabilistic mannequin designed for collections of discrete statistics (to recognize what is discrete information see our discrete vs non-stop facts post).To put in every other way, LDA is a method that mechanically finds subjects that given documents contain.LDA has a number of advance variations (dynamic, correlated, and etc.) which have a range of functions in an data retrieval.For example, if you have a ton of archives (such as emails) and you favor to find out what they are about barring the want to read them. In this case, LDA can give you numerous matters which are characterized with the aid of the most probably words.

ADVANCED METHODS

Now that we've touched upon the basic techniques of text analysis, we'll introduce you to the more advanced methods: text classification and text extraction.

Text Classification

Text classification is the process of assigning predefined tags or categories to unstructured text. It's considered one of the most useful Natural Language Processing (NLP) techniques because it's so versatile and can organize, structure and categorize pretty much anything to deliver meaningful data and solve problems.

Sentiment Analysis

Emotions are essential to effective communication between humans, so if we want machines to handle texts in the same way, we need teach them how to detect emotions and classify text as positive, negative or neutral. That's where sentiment analysis comes into play. It's the automated process of understanding an opinion about a given subject from written or spoken language.

Other uses of sentiment classifiers include assessing brand reputation, carrying out market research, and improving products with customer feedback.

Topic Analysis

Another common example of text classification is topic analysis or, more simply put, understanding what a given text is talking about. It's often used for structuring and organizing data. For example:

"The journal is really good and easy to use"

This journal feedback can be classified under ease of use.

IDENTIFYING E-COMMERCE IN ENTERPRISES BY MEANS OF TEXT MINING AND CLASSIFICATION ALGORITHMS

Monitoring specific points of the enterprises, for example, the adoption of e-commerce, is an important and basic assignment for numerous monetary activities. This type of information is typically bought by using potential of surveys, which are pricey due to the amount of personnel involved in the task. An automated detection of this records would allow constant savings. This can truely be performed with the aid of relying on laptop engineering, on the grounds that in accepted this records is publicly handy on line through the company websites. We classify the obtained dataset by means of skill of 4 classification algorithms: Support Vector Machines; Random Forest; Statistical and Logical Analysis of Data; Logistic Classifier. This turns out to be a challenging case of classification problem. However, after a careful format and set-up of the entire procedure, the outcomes on a sensible case of Italian businesses are encouraging.

Why e-commerce companies using text mining

We all nicely knew about 'E-Commerce'. E-commerce businesses are going to be a deafening industry in all over the world. In recent days, people are ready to buy all kinds of matters with the aid of sitting at the alleviation region of their homes. 'Time is more treasured than money'. E-commerce web sites are saving people's time. Very expedient for every individual to save anytime at anywhere in any gadget alternatively of journeying the shop in person. This is in particular proper for human beings dwelling in rustic areas who may also come across it elaborate to travel to the metropolis on a ordinary basis.

TEXT MINING INFLUENCE IN E-COMMERCE

A. Endorsing Paired Products –

Think about the current scenario. Visitors lands on an electronic retailer's internet site to buy a unique brand of Mobile phone. In a couple of seconds, the traffic will mark it and routinely redirects to the price web page to shut the transaction.

Visitors now desire a again case for their cellular phone. At the moment to their gladness, they may additionally possibly see a higher returned case. They joyously discover the returned case in the buying basket together with the cell phone, to procure both the bits and pieces. 'Text-mining plays a critical position on this.' By the use of the text-based filtering the relevant products will display on the under net page. This will motivate shoppers, to procure greater than what they have chiefly determined. The key here is aptness and exactness.

B. Personal Discount Makes Business Big-Usually, event buyer will opt for the discount. E-commerce will experience user to search for the particular product they like. Data mining will get the data from the user on the website and they scroll through the products. With this

result, E-commerce sites will pop-up with the personal discounts.

Not only the in the sites. Text mining also uses the data from the social media sites. Ads will be posted in the social media. This will induce the buyer to select the product.

C. Augmenting Newcomers

User purchasing in new E-Commerce sites, they don't direct go and purchase the products when they are new to that website. They commonly surf a lot of factor about the delivery, rate comparison, about their service, charges, product availability.

When the users sign-in with their account and purchase a product. If the person entered a wonderful review, then they will grow to be a properly customer. Text mining will give retailer attention about the new visitors. This will be the key for the sellers to get the new clients into their circle.

D. Identifying Fault Users

There are some users involving some mischiefs. They use the product for the warranty period and enjoy the product and return the product and opt for the money. When the retailer takes a closer look there will be no such faults. With text mining, we can identify such users and block them in future.

Analysis of e-Commerce Market Trend Using Text Mining

- Collects the most frequent words/ terminology from the text and visualizes them and performs clustering.
- Understands the trends in the text, makes use of statistical, machine learning techniques for implementing an analysis of the text behaviour.
- Make use of Natural Language Processing (NLP) technique which is the ideal methodology for text data (which is unstructured).
- Converts unstructured text data into numerical data, again on which the classical data mining technique is applied.

Relevant details should be given including experimental design and the technique (s) used along with appropriate statistical methods used clearly along with the year of experimentation (field and laboratory).

IV. CONCLUSION AND FUTURE SCOPE

E-commerce on the web platform has come a lengthy way, the place we don't see the real product or a mediator to advise a product. But the user can take reviews from extraordinary users; can take their one time to pick out the products. To remedy this issues text mining will be an high-quality answer to let the user go through the relevant products. Next-generation self-service textual content analytics. All the statistics sources, algorithms and AI equipment you need in one place. Interact with records via an intuitive visual interface to prep, explore, enrich, and build fashions faster than ever before.Text mining, the manner of inspecting textual statistics in order to become aware of patterns and gain insights, is increasingly being used by way of e-commerce retailers to learn greater about consumers. By identifying consumer buy patterns and opinions on unique products, e-commerce outlets can target precise people or segments with customized affords and reductions to increase sales and make bigger purchaser loyalty.

The text analysis has made it evident that any article drives completely primarily based on the keywords. So, it is always a desirable exercise to hold a music of the most time-honored words that show up in the text articles and use them as the keywords for that reason so that the customers can have an appreciation of what to focus on and it gives the agency an idea of where the e-commerce market is headed.

There is a lot of deep level evaluation that is taking location in text analysis. We have neural networks that is being utilized to the text, we have methods referred to as as pronoun processing so that we can create an analysis of the persona nature, emotion taxonomies that assist in the detection of the emotion in which the textual content has been written and operate an vital analysis (16).

Just like we are discussing about big data, we have huge text, which is growing even greater than the massive data. So, we can make use of the deep getting to know strategies for textual content summarization that helps the organisation come up with summarized headlines that can goal the clients going via the then market fashion and make them power towards the article or the advertisement. This will give the corporation a thorough thought of what exactly the patron is looking at.

We can also look at generating remarks systems, which locate their purposes at one-of-a-kind domains like pupil Analysis of e-Commerce Market Trend Using Text Mining feedback systems, appraisal administration systems, the place in the text that has been fed in via one of a kind sources can be analyzed according to the standards for unique individual and generate the apt remarks that can assist in the boom of the individuals.

REFERENCES

[1] S. Ananiadou, D. B. Kell, and J. I. Tsujii, "Text mining and its potential applications in systems biology. Trends in biotechnology", Vol.24, Issue.12, pp.571-579, 2006.

[2] Vishwanath Bijalwan, Vinay Kumar, Pinki Kumari, and Jordan Pascual,"Knn based machine learning approach for text and document mining. International Journal of Database Theory and Application", Vol.7, Issue.1, pp.61-70, 2014.

[3] Robert Dale, Hermann Moisl, and H. L. Somers, "editors. Handbook of natural language processing" *ISROSET Publisher*, India, Marcel Dekker, New York, NY, 2000.

[4] Jiawei Han, Chi Wang, and Ahmed El-Kishky,"Bringing structure to text: mining phrases, entities, topics, and hierarchies. In Pat Langley, editor, Proceedings of the 17th International Conference on Machine Learning",ICML,2000,pp 1968-1968, New York, NY, 2014. ACM.

[5] M. Hu and B. Liu. "Mining and summarizing customer reviews."In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168-177. ACM, August 2004.

[6] B. J. Jansen and M. Resnick,"An examination of searcher's perceptions of non sponsored and sponsored links during ecommerce web searching", Journal of the Association for Information Science and Technology, Vol.57, Issue.14, pp.1949-1961, 2006.
[7] Z. M. Jusoh and G. H. Ling.," Factors influencing consumers' attitude towards e-commerce purchases through online shopping. International Journal of Humanities and Social Science", Vol.2, Issue.4, pp.223-230, 2012.
[8] Roman Klinger, Robert Pesch, Theo Mevissen, and Juliane Fluck. Text mining in full text articlesme- thodical and represenation issues. Nature Precedings, 2009.
[9] R. LaRose,"On the negative effects of ecommerce: A sociocognitive exploration of unregulated online buying. Journal of ComputerMediated Communication", Vol.6, Issue.3, **2001**.
[10] Q. Mei and C. Zhai," Discovering evolutionary theme patterns from text: an exploration of temporal text mining", In In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge dis- covery in data mining, ACM, pp. 198-207,August 2005.
[11] W. W. Moe and P. S. Fader," Dynamic conversion behavior at e-commerce sites. Management Science", Vol.50, Issue.3, pp.326-**335**, 2004.
[12] R.S. Walse, G.D. Kurundkar, P. U. Bhalchandra, "A Review: Design and Development of Novel Techniques for Clustering and Classification of Data," *International Journal of Scientific Research in Computer Science and Engineering*, Vol.06, Issue.01, pp.19-22, 2018.
[13] M. and Heckerman, D., "An experimental comparison of several clustering and initialization method," Technical Report MSR TR-98-06, Microsoft
[14] A Novel Method for Text and Non-Text Segmentation in Document Images International conference on Communication and Signal Processing, April 3-5, 2013, India,
[15] A Novel Method for text and Non-Text segmentation in document images. International Conference on Communication and signal processing, April 3-5, 2013 978-1-4673-4866-9/13 2013 IEEE.

Authors Profile

Dr. Rajeev Tripathi received his B.Sc. degree in computer science from Lucknow University, Lucknow, India, in 2003, the M.C.A degree in computer application from AKTU University, Lucknow, India, in 2006, and the Ph.D. degree in Data mining technology and Text Mining from Sai Nath University, Ranchi, Jharkhand, India in 2019. He was a teaching assistant, lecturer and assistant professor, with Department of Computer Science Courses, SRMGPC, from 2007 till date respectively. He was an assistant professor, SRMGPC, Lucknow. He has more than twelve years of professional experience in information systems, computer Science and technologies.

His research interests include data mining, text mining, search optimization and video mining technique. His research interests include search optimization, data mining, programming, emerging trends and technologies. His current research interests in the areas of: Software quality, software measurement, software metrics, software process improvement, software project management, object oriented technologies, XML, Agile, Web Services, cognitive informatics and security. At present, He is engaged in search optimization and text mining technique in data mining application.

Dr. Santosh kumar Dwivedi received his B.Sc. degree in computer science from Lucknow University, Lucknow, India, in 1999 the M.C.A degree in computer application from AKTU University, Lucknow, India, in 2004, and the Ph.D. degree in "Software Testing process thorough Agile Process" from CMJ University, Meghalaya, India in 2013. He was a teaching assistant, lecturer and assistant professor, with Department of Computer Science Courses, SRMGPC, from August 08, 2007 till date respectively. He was an assistant professor, SRMGPC, Lucknow.

I achieved more than fifteen years of professional experience in SDLC, Data Structure, Information systems, computer Science and technologies.

His research interests include Agile Process, Text mining, Testing, search optimization and video mining technique. His research interests include search optimization, data mining, programming, emerging trends and technologies. His current research interests in the areas of: Software quality, software measurement, software metrics, software process improvement, software project management, object oriented technologies, XML, Agile, Web Services, cognitive informatics and security. At present, He is engaged in search optimization and text mining technique in data mining application through Agile.

