

A Summarize of Text Data Mining in Sentimental Analysis Applications and Algorithms

J. Uma

Research scholar

Department of computer science
Periyar University PG Extension centre,
Dharmapuri.

Dr. K. Prabha

Assistant Professor

Department of computer science
Periyar University PG Extension centre,
Dharmapuri.

Abstract— Content Mining is a substantial accumulations of organized an unstructured information to inspect create new data is a content mining. Content mining as a content to information mining it comparable to content investigation is performing brilliant data from content. Content mining including the procedure of structures the whole content to information mining. Content mining includes different applications utilizing Text Analytics, Opinion Mining and so forth. Text mining difficulties to comprehend synopses feeling about for web based life. Feeling Mining is troublesome content and order errand. Content mining utilization of Text Analytics to get past new data and data extricated a database.

Keywords— *Big data, Data mining, Text mining, Text Mining Process, Summarization, and Text categorization, Information Extraction, Information Retrieval, sentiment analysis .)*

I. INTRODUCTION

Enormous information mining is most imperative job to investigation design relationship inside our information, Data concentrated processing is exceptionally practical to fulfill for hunt, break down, mine, and representation extraordinary measure of information. Enormous information approach up information step by step extensive amount of all together and we are living in the public arena that utilization dissects huge scope of huge information(1). Enormous Data is portrayed by three viewpoints: (I) the information is visit, (II) the information can't be ordered into acknowledged social databases, and (III) information are produce, caught, and prepared rapidly. Huge Data is application for trade enhancement and is rapidly wide scope of information regularly expanding as an area of the IT business. It give simultaneously vast information fields, by including control of medicinal, interpersonal organization, advertise, saving money exchanges, satellite imaging. It every now and again, data is put away in a substantial information organized configuration to expand its instructive substance. The present data information volumes are driven by both unstructured and semi organized information. In this manner, premise to depiction preparing can be produced by the transformation between organized information in social frameworks of database the board and unstructured data for examination.

II. RELATED WORK

The micro blog of-words demonstrate is a standout amongst the most generally utilized element display for practically all content arrangement errands because of its

effortlessness combined with great execution. The model speaks to the content to be named a pack or accumulation of individual words with no connection or reliance of single word with the other, for example it totally ignores sentence structure and request of words inside the content. This model is likewise exceptionally main stream in assumption examination and has been utilized by different analysts. The least difficult approach to fuse this model in our classifier is by utilizing unigrams as highlights.

As a rule n-grams is an adjoining grouping of "n" words in our content, which is totally free of some other words or grams in the content. So unigrams is only a gathering of individual words in the content to be grouped, and we expect that the likelihood of event of single word won't be influenced by the nearness or nonappearance of some other word in the content. This is a very improving suspicion yet it has been appeared to give rather great execution (for instance in [7] and [2]). One straightforward approach to utilize unigrams as highlights is to dole out them with a sure earlier extremity, and take the normal of the general extremity of the content, where the general extremity of the content could just be determined by summing the earlier polarities of individual unigrams.

Earlier extremity of the word would be certain if the word is commonly utilized as a sign of energy, for instance "sweet"; while it would be negative if the word is for the most part connected with negative meanings, for instance "fiendish". There can likewise be degrees of extremity in the model, which implies how much characteristic is that word for that specific class. A word like "amazing" would most likely have solid abstract extremity alongside energy, while "not too bad" would despite the fact that have positive earlier extremity yet presumably with feeble subjectivity.

The assignment of twitter assessment investigation can be most firmly identified with expression level slant examination. An original paper on expression level assessment investigation was displayed in 2005 by Wilson et al. [16] which distinguished another way to deal with the issue by first ordering expressions as per subjectivity (polar) and objectivity (unbiased) and after that further grouping the emotional arranged expressions as either positive or negative. The paper saw that a significant number of the target phrases utilized earlier assessment bearing words in them, which prompted poor grouping of particularly target phrases. It

guarantees that on the off chance that we utilize straight forward classifiers which accept that the logical extremity of the word is just equivalent to its earlier extremity gives a consequence of about 48%. The tale grouping process proposed by this paper alongside the rundown of brilliant highlights which incorporate data about relevant extremity brought about huge enhancement in execution (regarding exactness) of the arrangement procedure.

III. TEXT ANALYTICS PROCESS

3.1 Text Analytics

Content examination portray a lot of semantic measurable and machine learning procedure that display and organized the data substance of literary hotspots for correspondence different methodology in Business insight, exploratory information examination, research and examination. Science considers dialect for examination dialect shape in dialect importance in setting.

3.2 Text Analytics Process

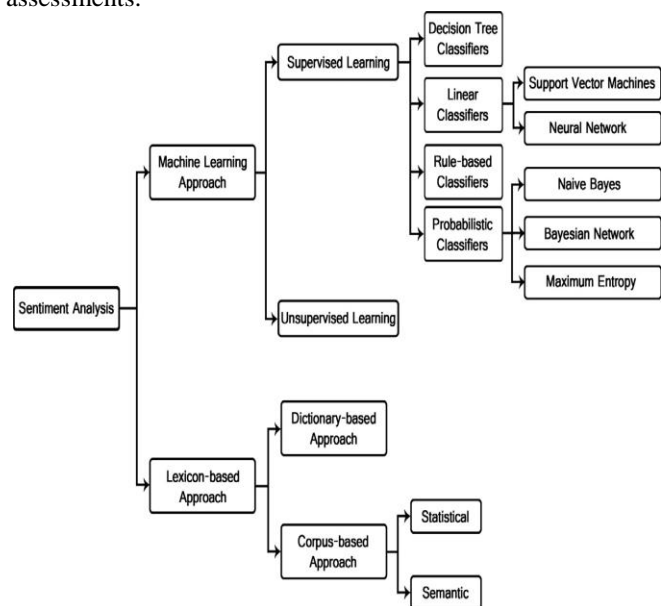
Text Analytics Information recovery or order of a corpus is a preliminary advance: gathering or recognizing a lot of printed materials, on the Web or held in a record framework, database, or substance corpus supervisor, for examination. Although some content investigation frameworks apply solely progressed measurable techniques, numerous others apply progressively broad characteristic dialect preparing, for example, grammatical feature labeling, syntactic parsing, and different sorts of semantic analysis.[citation needed]

Named element acknowledgment is the utilization of gazetteers or measurable procedures to distinguish named content highlights: individuals, associations, put names, stock ticker images, certain condensing, etc. Disambiguation the utilization of logical pieces of information—might be required to choose where, for example, "Portage" can allude to a previous U.S. president, a vehicle maker, a motion picture star, a stream intersection, or some other substance. Recognition of Pattern Identified Entities: Features, for example, phone numbers, email addresses, amounts (with units) can be recognized by means of standard articulation or other example matches. Co reference: ID of thing phrases and different terms that allude to a similar item. Relationship, reality, and occasion Extraction: recognizable proof of relationship among substances and other data in content

Sentiment examination includes recognizing abstract (rather than real) material and separating different types of attitudinal data: supposition, assessment, mind-set, and feeling. Content investigation procedures are useful in breaking down, assumption at the substance, idea, or subject dimension and in recognizing sentiment holder and feeling object.[8].Quantitative content examination is a lot of strategies originating from the sociologies where either ad human judge or a PC extricates semantic or syntactic connections between words so as to discover the importance or complex examples of, normally, an easygoing individual content with the end goal of mental profiling etc. [9]

IV. SENTIMENT ANALYSIS

Sentiment analysis , likewise called feeling mining, is the field of concentrate that breaks down individuals' assessments, notions, assessments, evaluations, frames of mind, and feelings towards elements, for example, items, administrations, associations, people, issues, occasions, points, and their traits. It speaks to an extensive issue space. There are additionally numerous names and marginally unique errands, e.g., assumption examination, conclusion mining, feeling extraction, assessment mining, subjectivity investigation, influence examination, feeling examination, survey mining, and so on. Be that as it may, they are currently all under the umbrella of conclusion examination or supposition mining. While in industry, the term notion examination is all the more regularly utilized, however in the scholarly world both supposition investigation and conclusion mining are much of the time utilized. They fundamentally speak to a similar field of study. The term feeling investigation maybe first showed up in (Nasukawa and Yi,2003), and the term sentiment mining previously showed up in (Dave, Lawrence and Pennock, 2003). Be that as it may, the examination on conclusions and sentiments seemed before (Das and Chen, 2001; Morinaga et al., 2002; Pang, Lee and Vaithyanathan, 2002; Tong, 2001; Turney, 2002; Wiebe, 2000). In this book, we utilize the terms slant investigation and conclusion mining conversely. To disentangle the introduction, all through this book we will utilize the term conclusion to mean assessment, slant, assessment, examination, frame of mind, and feeling. Be that as it may, these ideas are not proportionate. We will recognize them when required. The importance of feeling itself is still extremely expansive. Assumption investigation and conclusion mining predominantly centers around assessments.



V. OPINION MINING AND SENTENCE ANALYSIS

Assessment Mining or Sentiment Analysis is the assessment model to learning of popular conclusions, demeanors and emotions toward anything, item or dealer. The

article can describe people, items or themes [1]. Conclusion Mining is one of the best powerful research zones in Natural Language Processing. Notion Analysis characterizes a technique of mining, ordering, examining and depicting the emotions or assessments as word-based information utilizing Machine Learning, Natural Language Processing or Statistics. The two wordings assessment investigation or sentiment mining are increasingly substitutable. Assessment Mining mines the printed information and assesses open's frame of mind around an item while estimation examination arranges the supposition enunciated in content at that point inspects it. Opinion examination can be estimated a scientific classification process.

There are three fundamental classifications in supposition examination:

- Document-Level
- Sentence-Level
- Aspect-Level

Document-level SA: its primary target is, to sort a frame of mind message as articulating a positive or negative mentality or slant. It ponders the total content a fundamental information unit.

Sentence-level SA: its principle objective is, to sort assessment verbalized in individual sentence. The underlying stage is to characterize either the sentence is neither emotional nor objective. In the event that the sentence is abstract, Sentence-level slant examination will choose whether the sentence expresses a positive or negative sentiment.

Aspect-level SA: its principle objective is, to classify the slant through component to the correct highlights of articles. The essential stage is to arrange the items and their highlights. The assessment holders can give divergent feelings for unique highlights of a similar article like this sentence "The camera of this telephone isn't great, yet the voice clearness is fantastic". The substance sets are most imperative issue in the estimation examination field. The primary wellsprings of information are the item surveys, news articles or political discussions and so forth. These are assumes the significant job in the determination of business basic leadership in with the end goal that they dissect the aftereffects of clients assessments about their things. The audit sources are fundamental survey sites. The conclusion examination in utilized in things or item audits as well as can be utilized in securities exchanges, news articles or political discussions.

5.1 Opinion/Sentiment Components:

There are three fundamental segments in the assessment/notion.

i) Opinion holder: Individual who gives a remark. Ex. The camera nature of this telephone is magnificent.

ii) Opinion object: Article on which remark communicated. Ex. The supposition object is "the camera nature of this telephone is great".

iii) Opinion orientation: Discover the remark either positive or negative or impartial ex. The camera nature of this telephone is fantastic

5.2 Opinion/Sentiment Types:

Regular type: An ordinary conclusion is frequently all added just as a feeling in the writing and it has two subtypes.

Direct Opinion: An immediate supposition indicates to a mentality explained straight on an article or an item viewpoint. For instance, "The battery life of this cell phone is great".

Indirect Opinion: It denotes to an opinion that is articulated indirectly on an object or an object aspect. For example, "After taking this syrup, my body pains relieved".

Comparative type: It means to a supposition that is verbalized by implication on an article or an item viewpoint. For instance, "In the wake of taking this syrup, my body torments alleviate.

5.3 Supervised Learning

A near feeling states a connection of likenesses or contrasts between at least two elements. For instance, the sentences, "Lift tastes superior to Horlicks" and "Lift tastes the best" express two relative sentiments.

5.4 Unsupervised Learning

The directed procedures assemble utilization of an increasingly number of named prepared reports. Directed calculations require people to supply each information and wanted yield, moreover to outfitting Input concerning the precision of expectations all through preparing. When instructing is finished, the algorithmic program standard can apply the educated angles to new information. The unsupervised methodologies utilized once it's troublesome to seek out these named preparing archives. Unsupervised calculations don't be prepared with wanted result information. Rather, they utilize relate approach alluded to as profound figuring out how to audit information and get ends. Unsupervised learning calculations are utilized for a great deal of muddled process assignments than administered learning frameworks.

5.5 Sentiment Analysis Twitter Data

The aim while performing twitter sentiment analysis is classifies the tweets in different sentiment classes accurately. In this field of research, various techniques have evolved, which come up with methods to train a model and then test it to check its effectiveness. Performing sentiment analysis is challenging on twitter tweets. Here we define some reasons for this:

Use of slang: With only 280 characters close by, decreased proclamations are created, which results meager arrangement of highlights.

Twitter features: These words are not exactly equivalent to English words and it can make a methodology out dated as a result of the transformative utilization of slangs.

User variety: It allows the utilization of hash labels, client reference and URLs. These require diverse preparing in contrast with different words

5.6 Sentiment Classification

Notion investigation can be performed at four unique dimensions word level, state level, sentence level, and archive level [3]. Wiebe et al [6] proposed Subjectivity and significance are both essential properties of dialect. Word speculations are Subjectivity is a property, it very well may be related with word detects. Subjectivity comments are specifically [4] used to word sense disambiguation. Each record is displayed as an arrangement of perceptions (words) and fundamental states, prompts increment in time. Theresa et al [13] proposed Phrase-level assumption investigation. It decides if an articulation is unbiased or polar and after that disambiguated the extremity of the polar articulations. In this methodology, naturally recognize the relevant extremity for an expansive subset of accomplishing results, assessment articulations that are fundamentally superior to anything gauge yet it requires greater investment. Ache et al [8] proposed an original methodology in conclusion grouping. Basically, they presume that machine learning procedures, similar to NB and SVM, don't accomplish exactness as great on slant grouping as on conventional theme based arrangement.

The characterization precision coming about utilizations just unigrams as highlights. Yi et al [19] proposed Sentence level extremity classification endeavors to characterize positive and negative suppositions for each or whether a sentence is abstract or target. There has additionally been work on expression level classification so as to catch various opinions that might be available inside a solitary sentence. In this methodology, we can't precisely foresee the supposition, to defeat this issue we go for the following methodology. Ache and Lee [9] additionally proposed characterizing sentences as being either abstract or goal, and afterward apply supposition characterization on the emotional bit of the content. It isn't adequate for recognizing estimation of substances. Turney [14] proposed record level slant order there are two sorts of methodologies: term-tallying methodologies and machine learning approaches. Term-tallying approaches typically include inferring an opinion measure by figuring the all out number of negative and positive terms. Ache and Lee [9] proposed machine learning approaches recast the feeling arrangement issue as a measurable characterization assignment. Contrasted with term-checking approaches, machine learning approaches for the most part accomplish better execution, and have been adjusted to progressively entangled situations, for example, space adjustment, multi-area learning and semi supervised

learning for assessment characterization. Whitelaw et al [15] proposed considering descriptive articulations as an imperative sign of the assumption extremity in literary surveys. Slant order dependent on removing and breaking down examination gatherings, for example, "great" or "not terribly terribly". Wang et al [5] proposed managed learning strategies have been broadly utilized and demonstrated compelling in supposition characterization.

They typically rely upon a lot of named information, which as a rule includes staggering expense in labor and time. To beat this issue, different semi-administered learning techniques are proposed to viably use a little size of marked information alongside a bigger measure of unlabeled information. Semi-directed strategies [17] for feeling characterization are to use earlier lexical information related to the marked and unlabeled information and to utilize some bootstrap methods, for example, self-preparing and co-preparing. Vapnik [15] proposed SVM are a gathering of directed learning strategies that performs order by building a N-dimensional hyper plane that ideally isolates the information into two classes. Suzuki et al [12] proposed SVM display utilizing a sigmoid portion work is proportionate to a two-layer, recognition neural system. SVM[7] has been appeared to perform great on a wide assortment of characterization issues that require expansive scale input space, for example, written by hand character acknowledgment, confront identification, and in particular for this situation, content classification. Rodrigo et al [11] proposed for robotizing the assignment of ordering a solitary theme literary survey, archive level assessment arrangement is utilized for communicating a positive or negative opinion. Directed learning techniques comprise of two phases, extraction/determination of educational highlights and arrangement of audits by utilizing learning models like SVM. Proposed anticipating the hotspots dependent on assumption examination in online discussions. Exactness in the order of positive or negative film audit is less in notion investigation. In view of the general population sentiment, the thought regarding the motion picture may differ that is a similar motion picture might be great or terrible rely on the conclusion of the people. SVM [18] utilizes $g(x)$ as the segregate work. It clients express their assessments in an alternate ways, some utilizing diverse dialect in the middle of, while others utilizing rehashed words or images to express their feeling. All the above issues are required to be looked in the pre-handling segment. The work stream for feeling examination .The framework comprises of the four principle modules: information accumulation module, information handling module, arrangement module and investigation of yield.

$$g(x) = w^T f(x) + b \quad (1)$$

Where w is the weights vector, b is the bias, and $f(x)$ denotes nonlinear mapping from input space to high-dimensional feature space. The parameters w and b are learned automatically on the training dataset following the principle of maximized margin .where N denotes the slack variables and C denotes the penalty coefficient. The above problem directly, it is converted to an equivalent quadratic optimization problem by Lagrange multipliers. The training

sample (xi,yi) is called a support vector. Due to the dimension of feature space is quite large in text classification tasks, the classification problem is always linearly separable [13] and therefore linear kernel is commonly used.

VI. EVALUATION OF SENTIMENT CLASSIFICATION

Conclusion investigation can be characterized as a procedure that mechanizes mining of frames of mind, feelings, perspectives and feelings from content, discourse, tweets and database sources through Natural Language Processing (NLP). Assumption investigation includes ordering conclusions in content into classifications like "positive" or "negative" or "unbiased". It's additionally alluded as subjectivity investigation, conclusion mining, and examination extraction. The words conclusion, opinion, view and conviction are utilized reciprocally however there are contrasts between them.

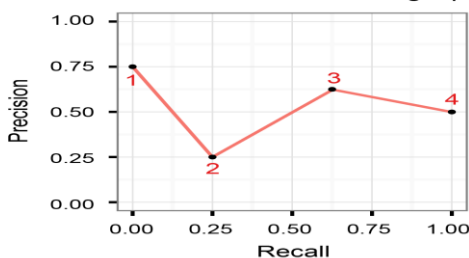
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad \text{---(1)}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{---(2)}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{---(3)}$$

$$\text{F1} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{---(4)}$$

A Precision-Recall curve connecting 4 points



For the most part, the execution of conclusion arrangement is assessed by utilizing four files: Accuracy, Precision, Recall and F1-score. The regular path for processing these lists depends on the perplexity grid.

Opinion: A conclusion opens to dispute (because different experts have different opinions)

View: subjective opinion

Belief: deliberate acceptance and intellectual assent

Sentiment: opinion representing one's feelings
 An example for terminologies for Sentiment Analysis is as given below,

<SENTENCE> = The story of the movie was weak and boring
 <OPINION HOLDER> = <author>
 <OBJECT> = <movie>
 <FEATURE> = <story>
 <OPINION >= <weak> <boring>
 <POLARITY> = <negative>

Mathematically we can represent an opinion as a quintuple (o, f, so, h, t), where
 o = object;
 f = feature of the object o;
 so= orientation or polarity of the opinion on feature f of object o;
 h = opinion holder;
 t = time when the opinion is expressed.

Precision is the part of all obvious anticipated examples against all anticipated occasions. A precision of 100% implies that the anticipated occasions are actually equivalent to the genuine cases. Exactness is the bit of genuine positive anticipated occasions against all positive anticipated occurrences. Review is the part of genuine positive anticipated occasions against all real positive instances. F1 is a consonant normal of accuracy and review.

6.1 Naïve Bayes Classifier

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:

$$P(e_1|e_2) = \frac{p(e_2|e_1)p(e_1)}{E_2} \quad \text{---(5)}$$

This algorithm is implemented to calculate the probability of

Set	Word	Document	Class
Trainin g set	1	I Like that product iphone	Pos
	2	It's like good peace ,Nice entertainment	Pos
	3	Highly cost but fully enjoyed like kids product	Pos
	4	safety not maintained	Neg
Test set	5	Children enjoyed product but not safety and maintained	?

a data to be positive or negative. So, conditional probability of a sentiment is given as:

$$P(\text{sentiment}|\text{sentence}) = \frac{P(\text{sentiment})(\text{sentence} | \text{sentiment})}{P(\text{sentence})} \quad \text{---(6)}$$

And conditional probability of a word is given as:

$$P(\text{Word} | \text{Sentiment}) = \frac{\text{Number of word}}{\text{Total words in sentiment}}$$

occurrence in class + 1 / Number of words belonging to a class + Total no's of Word

Algorithm

Step1: Initialize P(positive) ← num – opposition (positive) / num_total_opposition

Step2: Initialize P(negative) ← num – opposition (negative) / num_total_opposition

Step3: Convert sentences into words
 for each class of {positive, negative}:
 for each word in {phrase}
 $P(\text{word} | \text{class}) < \text{num_apart}(\text{word} | \text{class}) / (\text{num_cuv}(\text{class}) + \text{num_total_cuvinte})$
 $P(\text{class}) \leftarrow P(\text{class}) * P(\text{word} | \text{class})$
 Returns max {P(pos), P(neg)}

Object: An entity which can be a, person, event, product, organization, or topic.

Feature: An attribute (or a part) of the object with respect to which evaluation is made.

Opinion orientation or polarity: The orientation of an opinion on a feature f represent whether the opinion is positive, negative or neutral.

Opinion holder: The holder of an opinion is the person or organization or an entity that expresses the opinion.

Assumption Analysis is a term that incorporates numerous assignments, for example, feeling extraction, slant grouping, and subjectivity arrangement, rundown of conclusions or assessment spam identification, among others. It plans to investigate individuals' notions, mentalities, conclusions feelings, and so on towards components, for example, items, people, points, associations, and administrations.

Assumption Analysis is a term that incorporates numerous assignments, for example, feeling extraction, slant grouping, and subjectivity arrangement, rundown of conclusions or assessment spam identification, among others. It plans to investigate individuals' notions, mentalities, conclusions feelings, and so on towards components, for example, items, people, points, associations, and administrations.

$$\gamma(\alpha|\beta) = \frac{\gamma(\alpha) * \gamma(\beta|\alpha)}{\gamma(\beta)} \dots (7)$$

Where α : specifies class, β Document wants to classify, $\gamma(\alpha|\beta)$ and $\gamma(\beta|\alpha)$:Posterior probabilities.

The value of class α might be positive or negative. Document is a review of particular theme, The multidimensional model of Naïve Bayes capture word frequency information in document .The maximum Like hood Estimate (MLE) is simply the relative frequency and corresponds to given the most likely value of each parameter given the training data. The prior probability this estimate is shown in equation 1.

$$\gamma(\alpha) = N_c / n \dots (8)$$

Where N_c : The number of document in a class α

N : Total number of documents

$$\gamma(\text{pos}) = 3/4$$

$$\gamma(\text{neg}) = 1/4$$

Calculate maximum like hood smoothing Naïve Bayes estimate by using equation 2.A unigram feature marks the presence or absence of a single word within the text .Estimate the conditional probability $\gamma(\omega|\alpha)$ as the relative frequency of term ω in document belonging to class α including multiple occurrences of a term in a document .

$$\gamma(\text{like}|\text{pos}) = (3+1)/(15+21) = (4/36) = 1/9$$

$$\gamma(\text{safety}|\text{pos}) = (0+1)/(15+21) = (1/36) = 1/36$$

$$\gamma(\text{good}|\text{pos}) = (1+1)/(15+21) = (2/36) = 1/18$$

$$\gamma(\text{like}|\text{neg}) = (0+1)/(3+21) = (1/24) = 1/24$$

$$\gamma(\text{safety}|\text{neg}) = (1+1)/(3+21) = (2/24) = 1/12$$

$$\gamma(\text{good}|\text{neg}) = (0+1)/(3+21) = (1/24) = 1/24$$

Calculate posteriori probability

$$\gamma(\text{pos} | d5) = 3/4 * 1/9 * 1/36 * 1/18 * 1/28 = 0.000004592 \text{ e-9}$$

$$\gamma(\text{neg} | d5) = 1/4 * 1/24 * 1/12 * 1/24 = 0.00003616 \text{ e-s9}$$

$$\gamma(\text{pos} | d5) > \gamma(\text{neg} | d5)$$

$\gamma(\text{pos} | d5)$ is maximum means probability of positive words in document 5 is maximum so document 5 is positive. The extricating conclusions from web based life information and break down the feelings of the clients on a specific subject and furnish with the translated outcomes valuable for basic leadership. This examination built up a conclusion mining framework to remove sentiments from Twitter information on a specific space and utilizing arrangement techniques, a preparation informational index were worked to enhance the execution of the model and got upgrades in execution when contrasted and the current models. The issue looked amid the extraction of conclusions was in choosing the proportion of intriguing quality or fitting component words, which for the most part influence the understanding of the supposition communicated which impact the basic leadership.

VII. CONCLUSION

Information mining to gather expansive sum in that information step by step expanding vast information to investigate for prescient information to issuing uses of Enormous Information Examination in amusement park dataset it related with ascertain of Huge Information Arrangement of different procedures and features the greater part of the huge discoveries of positive and negative feelings, which is talked about quickly in Table I. The paper additionally features the greater part of the huge research issues related with the current informational collection utilizing calculation Credulous bayes strategies 0.000004592 e-9 (Positive) It improvement of Enormous Information Investigation in different research points of view.

REFERENCES

- [1] Baksh, K.,(2012),”Considerations for big data: Architecture and approach”
- [2] Mukherjee, A.; Datta, J.; Jorapur, R.; Singhvi, R.; Haloi, S.; Akram,W., (18-22 Dec.,2012) , “Shared disk big data analytics with Apache Hadoop”
- [3] Aditya B. Patel, Manashvi Birla, Ushma Nair, (6-8 Dec. 2012),“Addressing Big Data Problem Using Hadoop and Map Reduce”
- [4] Yu Li; Wenming Qiu; Awada, U. ; Keqiu Li.,(Dec 2012),” Big Data Processing in Cloud Computing
- [5] Garlasu, D.; Sandulescu, V. ; Halcu, I. ; Neculoiu, G. .;(17-19Jan.2013),”A Big Data implementation based on Grid Computing”, Grid Computing
- [6] Sagiroglu, S.; Sinanc, D., (20-24 May 2013),”Big Data: A Review”
- [7] Grosso, P. ; de Laat, C. ; Membrey, P.,(20-24 May 2013),” Addressing big data issues in Scientific Data Infrastructure”
- [8] Kogge, P.M.,(20-24 May,2013), “Big data, deep data, and the effect of system architectures on performance”
- [9] Szczuka, Marcin,(24-28 June,2013),” How deep data becomes big data”K. Elissa, “Title of paper if known,” unpublished.
- [10] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
- [11] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron Spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [12] M. Young, The Technical Writer’s Handbook. Mill Valley, CA:University Science, 1989