

To Analyze Student's Learning Experience in social media

Swathi Yadav

PG Student, Thakur college of engineering and technology, Mumbai

ABSTRACT:- Student's posts and remarks via social media (informal organization destinations, for example, twitter) centers into their educational experiences, for example, their issues, comments which focus on students learning process. Some important data about student taking in encounters can be construed from the information assembled from such situations. It is difficult to dissect that data. Since the online networking information continues expanding in measure it demands automation in data analysis. At that point likewise the data induced from those information needs human interpretation since it is the impression of students' emergency. A work process that joins quantitative analysis and large scale data mining systems is created. We concentrated on student's posts on comprehend issues and issues in their instructive encounters. Substantial work stack, absence of awareness of social exercises, and restlessness are a few issues that student's look as they experience circular activities. In light of these outcomes, we began to actualize a classification algorithm to group posts mirroring student's issues.

Keywords: social mining, text classification, naïve Bayes classifier.

1. INTRODUCTION

Online networking destinations for example, twitter, Facebook gatherings give great venues for students to impart their experiences, vent feeling and stress, and look for social support. Around Different Online networking sites, scholars talk about activities in casual manner . The plenitude of Online networking information gives chances understand all the students' experiences, as well as raises methodological challenges in making sense of Online networking information to instructive purposes. Just envision those sheer information volumes, the assorted qualities of web slang, the flightiness of area Furthermore timing from claiming student's posts on the web, and additionally those complexities about students' experiences. Pure manual examination can't manage those ever-growing scale of data, at the same time pure automatic algorithms typically can't catch in-depth importance inside the information.

Traditionally, educational scientists have been utilizing techniques like surveys, interviews, centering groups, Furthermore classroom exercises to gather information identified by students'. These method consume lot of time,

Subsequently can't be duplicated alternately repeated with high frequency. Those scale about such investigations may be also limited. To addition, when provoked regarding their experiences, scholars necessity to reflect around the thing that they were thinking in the past, which might have turned obscured overtime.

2. TEXT MINING

Text mining can avail an organization derive potentially valuable business insights from text-predicated content such as word documents, email and postings on gregarious media streams like Facebook, Twitter and LinkedIn. Mining unstructured data with natural language processing (NLP), statistical modeling and machine learning techniques can be arduous, however, because natural language text is often inconsistently erratic. It contains ambiguities caused by inconsistently erratic syntax and semantics, including slang, language categorical to vertical industries and age groups, double entendres and mordancy.

Text analytics software can avail by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques. With an iterative approach, an organization can prosperously use text analytics to gain insight into content-categorical values such as sentiment, emotion, intensity and pertinence. Because text analytics technology is still considered to be an emerging technology, however, results and depth of analysis can vary wildly from vendor to vendor.

3. NAÏVE BAYES

Naive bayes expects a specific probabilistic generative model for text. Those model is a specialization of the mixture model, and hence also makes the two presumptions examined there. Additionally, naive bayes makes expression independence presumptions that permits the generative model should be portrayed with a extraordinarily diminished amount of parameters. Whatever remains of this subsection portrays the generative model All the more formally, giving an exact determination of the model parameters, Furthermore inferring that likelihood that an specific record may be produced provided for its class name.

First let us introduce some notations to describe text. Consider A document, d_i , is an ordered list of word events, $w_{d_i,1}, w_{d_i,2}, \dots$. then We can write $w_{d_i,k}$ for the word w_t in position k of document d_i , where w_t is a word in the vocabulary $V = \{w_1, w_2, \dots, w_{|V|}\}$.

When a document is to be created by a specific mixture component, c_j , a document length, $|d_i|$, is selected independently of the component. Then, the chosen mixture component creates a word sequence of the specified length. We furthermore assume it creates each word independently of the length.

Hence, the likelihood of a archive provided for a mixture part As far as its constituent Characteristics camwood be communicated similarly as those record length and the expressions in the archive. Those likelihood of a expressions occasion must make molded once every last one of expressions that precede it.

$$P(d_i | c_j; \theta) = P((w_{d_i,1}, \dots, w_{d_i,|d_i|}) | c_j; \theta) = P(|d_i|) \prod_{k=1}^{|d_i|} P(w_{d_i,k} | c_j; \theta; w_{d_i,q}, q < k)$$

Following we make those standard credulous bayes assumption: that those expressions of a record are produced freely for context, that is, freely of the other expressions in the same report provided for those population mark. We further expect that those likelihood of an expression is autonomous from claiming its position inside the document; thus, to example, those likelihood about seeing the statement "homework" in the initially position of an archive is the same concerning illustration seeing it to whatever viable position. We could express these presumptions as:

$$P(w_{d_i,k} | c_j; \theta; w_{d_i,q}, q < k) = P(w_{d_i,k} | c_j; \theta)$$

Joining together these a two equations provides for those naive bayes statement for the likelihood of a report provided for its class.

$$P(w_{d_i,k} | c_j; \theta) = P(|d_i|) \prod_{k=1}^{|d_i|} P(w_{d_i,k} | c_j; \theta)$$

Thus the parameters of an particular mixture component are a multinomial distribution over words, i.e. the gathered of word probabilities, each written $\theta_{wt | c_j}$, such that $\theta_{wt | c_j} = P(w_t | c_j; \theta)$, where $t = \{1, \dots, |V|\}$ and $\sum_t P(w_t | c_j; \theta) = 1$. because we assume that for all classes, document length is distributed same, there is no need to parameterize for classification.

4. RELATED WORKS

The authors, Kamal Nigam, Andrew Kachites Mccallum, Sebastian Thrun, Tom Mitchell, Says that those exactness about figured out how quick classifiers might be enhanced Toward augmenting An little number from claiming marked preparing documents with an extensive pool of unlabeled documents. This may be paramount Since in large portions quick arrangement issues acquiring preparing labels may be expensive, same time expansive amounts about unlabeled documents would promptly accessible. He introduces an calculation for Taking in starting with marked Furthermore unlabeled documents In light of the blending for Expectation-Maximization (EM) Furthermore an naive bayes classifier.

The algorithm to start with trains a classifier utilizing the accessible marked documents, and probabilistically labels those unlabeled documents. It At that point trains another classifier utilizing those labels for every last one of documents, Furthermore iterates should joining. This fundamental em system meets expectations great at the information fit in with those generative presumptions of the model. Nonetheless morals these presumptions are frequently all the disregarded done act Also poor execution camwood result. These writers introduce two extensions of the calculation that enhance order correctness under these conditions: An weighting element will regulate the commitment of the unlabeled data, Furthermore the utilization about numerous mixture parts for every population. Test results, acquired utilizing quick starting with three separate real-world tasks, demonstrate that the utilization of unlabeled information lessens arrangement lapse toward up to 30%.

Bo Pang and Lillian Lee Concerning illustration the creators inspected those connection the middle of subjectivity identification Also polarity classification, demonstrating that subjectivity identification could layer reviews under substantially shorter extracts that at present hold polarity majority of the data toward a level tantamount to that of the full survey.

On fact, to those naïve bayes polarity classifier, those subjectivity extracts would demonstrated to make more viable input over those starting document, which prescribes that they are not main shorter, as well as “cleaner” representations of the planned polarity. They need indicated that utilizing those minimum-cut skeleton brings about the improvement from claiming productive calculations for assumption Investigation. Using relevant majority of the data through this framework might prompt statistically huge change over polarity-classification precision. Currently these creators finish up expressing the naïve bayes may be relatively best.

Tina R. Patil, Mrs. S. S. Sherekar Sets out to make similar assessment of classifiers naïve bayes and J48 in the connection of bank dataset will expand valid certain rate Furthermore minimize false sure rate from claiming defaulters instead of accomplishing main higher arrangement correctness utilizing WEKA apparatus. Those investigations outcomes demonstrated in this paper need aid around order accuracy, affectability Furthermore specificity. Those brings about those paper with respect to this dataset likewise demonstrate that those effectiveness and correctness from claiming j48 will be superior to that of Naïve bayes. Yet our issue needs the multi-level classifier. The framework needs various information qualities. Thereabouts we utilize naïve bayes with handle the ordinary developing dataset with those numerous quality values.

5. PROPOSED METHOD

Traditionally there is no such researches and implementation to analyze the posted data for the educational purpose in the social media by the students and to merge both analysis and mining techniques. The survey that has been proposed is to monitor the student activity by permitting the necessary privacy. The issues and problems of the students can be analyzed out through their posts and conversations in the social media. If the issues are negative and serious then the notice has to be sent to the concerned staff and their parents. The student’s problem and the feedback can be considered in the higher level decision making the proposed method has seven steps to get the appropriate result are listed below,

1. Collecting Data
2. Sampling
3. Analyzing Data Qualitatively
4. Result
5. Training And Evaluation of model
6. Model Adaption
7. Large Scale Data Analysis Result

In the Fig.1. Data is gathered as a data set (i.e.,) the posts from the social media is collected as a sample and analyzed. The issues and problems of the students can be analyzed out through their post. If the issues are negative and serious then the message has to be sent to the concerned staff and their parents. The student’s issues, problems and the feedback can be considered in the higher level decision making. The proposed method has following modules,

Computational Attribute Management

Those quality era module completely works on the director end Also it gathers the quality information starting with admin for administering the dataset to future motivation. Once those attributes are produced it will make simple to dissect those considerations of the student dependent upon their provided for input starting with their portal. To every last one of quality management is a dataset oversaw economy procedure which holds those data in regards the keywords and the particular qualities for mining the students performance and learning experience.

Qualitative Attribute Analysis

Once the attributes are produced the student effective mining portal is in other hand to gather the information from the student’s end. This port can be analyzed only by the administrator to know the thoughts of the students. Every time when the student tweet with some data that all data posted by them will be compared with the created attribute set and then the mining process takes place for analyzing the qualitative summary of the student. Specifically the qualitative analysis module helps the administrator to analyze the qualitative summary of the student performance.

Public Web Conversation Port

The hypothetical establishment for the worth of casual information on the web can be drawn from Goffman’s hypothesis of social execution. Despite formed to clarify face to face interactions, Goffman’s hypothesis of social execution may be generally used to demonstrate mediated collaborations on the web today. A standout amongst the practically essential viewpoints from claiming this hypothesis will be those thought from claiming front-stage and back-stage from claiming people’s social performance. If a social setting is front-stage or back-stage may be a relative matter. To students, compared for formal classroom settings, online networking is a relative casual and unwinding back-stage. When student’s post content with respect to Online networking sites, they normally post the things that they feel during that moment.

Mining Student Conversation

From the different fields about a large number existing meets expectations specialists bring broke down the mining content with produce particular information to their particular liable domains. To example, Gaffney analyzes tweets for hash tag iran race utilizing histograms, client networks, Furthermore frequencies from claiming highest point keywords will quantify on the web activism. Comparative investigations bring been led over other fields including healthcare, promoting Also athletics, only on sake a couple. Investigation systems utilized in these investigations typically incorporate qualitative substance analysis, semantic analysis, organize analysis, and a few oversimplified systems for example, such that saying clouds and histograms. In this module, an order model may be manufactured In light of inductive substance dissection. This model might have been that point connected Also approved for a brand new dataset. Therefore, not main the insights picked up from particular case dataset need aid emphasized, as well as those provision of the order algorithm will other datasets for identifying person issues. The human exert will be in this way increased for extensive scale information investigation.

Text Preprocessing

Many social mining users may use some special symbols to convey their messages. For example, # is used to show a hashtag, is used to show an account of the user, and RT is used to show a re-tweet. Social Mining users may also repeat letters in words so that to emphasize the words, for example, “hungryyy”, “soooo muuuchh”, and “Monndayyyy”. Common stop words such as “a, an, and, of, he, she, it, the”, non-letter symbols, and punctuation also bring noise to the text. Therefore the pre-processed texts are needed completely for analyzing the student mining data and through this module it would be simple to attain.

6. ARCHITECTURE

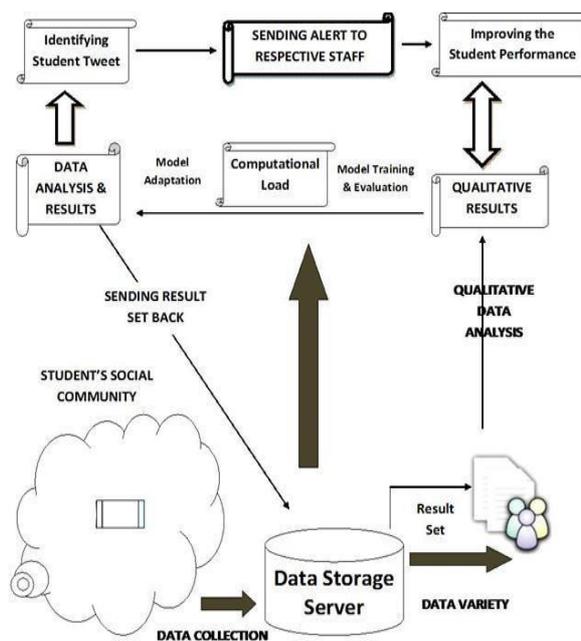


Fig 1: Architecture

The above Fig 1 shows the architecture of the proposed method. The architecture shows the overall flow of the proposed method. The flow starts with gathering of the students tweet and identifying those tweets which shows the student’s

learning experiences. After gathering those tweets the emotions of the student is analyzed by Naïve bayes algorithm. According to the analysis the result which is generated is send as an alert message to the respective teacher. Then the teacher will try to improve the students performance.

7. CONCLUSION

This paper provides a flow for analyzing the contents posted by the student in social media for educational purposes that overcomes the major limitations of both manual qualitative analysis and large scale computational analysis of user generated textual content. And also it can give essential facts to the educational administrators, practitioners and other relevant decision makers to gain further insights of engineering student's college experiences. In this project, the community like structure is planned between students and staff. Manually the admin has to gather the data from the student's post and then the alert message has to be send to the staff, if the posts are negative.

8. REFERENCES

- 1) Xin Chen, Mihaela Vorvoreanu and Krishna Madhavan, " Mining social media data for understanding students learning experience" , IEEE transaction on Learning Technologies, vol.7, no.3, Pp,16-22 July - September 2014
- 2) Kamal Nigam, Andrew Kachites Mccallum, Sebastian Thrun, Tom Mitchell, "Text Classification From Labeled And Unlabeled Documents Using EM", Machine Learning, 39, Kluwer Academic Publishers. Printed In The Netherlands, Pp. 103-134, 2000.
- 3) Bo Pang And Lillian Lee,"A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based On Minimum Cuts", Morgan & Clay Pool Publishers, Pp. 54-58, 2008.
- 4) Tina R. Patil, Mrs. S. S. Sherekar,"Performance Analysis Of Naive Bayes And J48 Classification Algorithm For Data Classification", J.sci.Education, Vol.86, No.1, Pp.7-15, 2000
- 5) Swathi Yadav, Shwetha Yadav,"Text Mining of Voot Application Reviews on Google Play Store ",IRJET, Vol.05, Issue.1,Pp.56-72,2018.