Analyzing Social Media Data in Sectors like Education Using Data Mining Technique

Miss. Shelke Nikita M.[1] Miss. Bhavsar Vinita P. [2] Miss. Shinde Vaishnavi P. [3]
Miss. Somvanshi Snehal D. [4] Prof. Shaikh I. R.[5]
[1,2,3,4,5]SND College of Engineering & Research Center, Yeola, India

Abstract— The upcoming method in studying analytic and institutional data mining is on analyzing structured data taken from program management systems, or controlled online studying environments to inform educational decision-making. However, there is no proper method found to mine and analyze student posted messages from uncontrolled spaces on the social web sites with the clear goal of understanding students learning experiences. Data Mining is defined as extracting the knowledge from the huge data. This information can be used for any of the applications such as Market Analysis, Fraud Detection, Customer Retention, etc.The research goals of this study are to demonstrate a workflow of social media data for decision making in educational sectors, integrating both qualitative analysis and large scale data mining techniques and to explorestudents' informal conversations on platform like Twitter, in order to understand issues and problems encountered.

Key words: Twitter, Pattern, Structured Data, Integration, Media

I. INTRODUCTION

Social media sites such as Twitter provide a great venues for students to share their emotions, stress, and seek social support. On various social media sites, students discuss and share their everyday encounters in an informal and unstructured way. There is huge amount of data available on social media. This data is of no use until converted into useful statistical information. The extraction of information is not the only process we need to perform; it also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data representation, We chose to focus on engineering students posts on Twitter about problems in their educational experiences mainly because, Engineering schools and departments have long been struggling with student recruitment and retention issues Based on understanding of issues and problems in students' life, policymakers and educators can make more informed decisions on proper interventions and services that can help students overcome barriers in learning.

II. PROJECT SCOPE

The project will be helpful for organizations, education as well as the teachers. It will be helpful for educational organization for reviewing about the program that they are going to do or they have performed. Similarly the teachers also get review about their teaching skills, and get improved.

III. LITERATURE SURVEY

A. Analysis Models of Technical and Economic Data of Mining Enterprises Based On Big Data Analysis [1]

Published year : 2018
Author : Jian Ming

Characteristics of the technical and economic data of mining enterprises are multi-dimensionality and nonlinearity. The sales price data of mineral products is an important economic indicator of mining enterprises, and the geological data is an important technical data. The analysis method of the technical and economic data is researched using technologies of big data analysis and data mining. The fluctuation pattern and influencing factors of the mineral products price are analyzed.

B. DD-Rtree: A dynamic distributed data structure for efficient data distribution among cluster nodes for spatial data mining algorithms [2]

Published year : 2016
Author :Poonam Goyal

Parallelizing data mining algorithms has become a necessity as we try to mine ever increasing volumes of data. Spatial data mining algorithms like Dbscan, Optics, Slink, etc. have been parallelized to exploit a cluster infrastructure. The efficiency achieved by existing algorithms can be attributed to spatial locality preservation using spatial indexing structures like k-d-tree, quad-tree, grid files, etc. for distributing data among cluster nodes. However, these indexing structures are static in nature, i.e., they need to scan the entire dataset to determine the partitioning coordinates. This results in high data distribution cost when the data size is large. In this paper, we propose a dynamic distributed data structure, DD-Rtree, which preserves spatial locality while distributing data across compute nodes in a shared nothing environment.

C. Study on land use of changing district with spatial data mining method [3]

Published year : 2011
Author :Yong Wang

In this paper, the author attempted to study on the land use of Changing District (2005 and 2008) with spatial data mining through geo-statistics method, and explored new method for land use spatial data mining. The land use data (2005 and 2008) were divided into 1km1km, 2km2km, 4km4km grid cells respectively. And use spatial auto correlation to analysis the land use structure and its changes in Changing District of Beijing. The study result proved that, as time goes on, most land type enhance the performance of the spatial auto correlation, land use Structure more and more reasonable.

D. Apply word vectors for sentiment analysis of APP reviews [4]

Published year : 2016
Author :Xian Fan

Vector representations for language have been shown to be useful in a number of Natural Language Processing tasks. In this paper, we aim to investigate the

All rights reserved by www.ijsrd.com

effectiveness of word vector representations for the problem of Sentiment Analysis. In particular, we target three sub-tasks namely sentiment words extraction, polarity of sentiment words detection, and text sentiment prediction. We investigate the effectiveness of vector representations over different text data and evaluate the quality of domain-dependent vectors.

E. Sentiment analysis in a cross-media analysis framework [5]

Published year : 2016
Author :Yonas Woldemariam

This paper introduces the implementation and integration of a sentiment analysis pipeline into the ongoing open source cross-media analysis frame-
work. The pipeline includes the following components; chat room cleaner, NLP and sentiment analyzer. Before the integration, we also compare two broad categories of sentiment analysis methods, namely lexicon-based and machine learning approaches. We mainly focus on finding out which method is appropriate to detect sentiments from forum discussion posts. In order to conduct our experiments, we use the apache-hadoop framework with its lexicon-based sentiment prediction algorithm and Stanford coreNLP library with the Recursive Neural Tensor Network (RNTN) model.

IV. IMPLEMENTATION

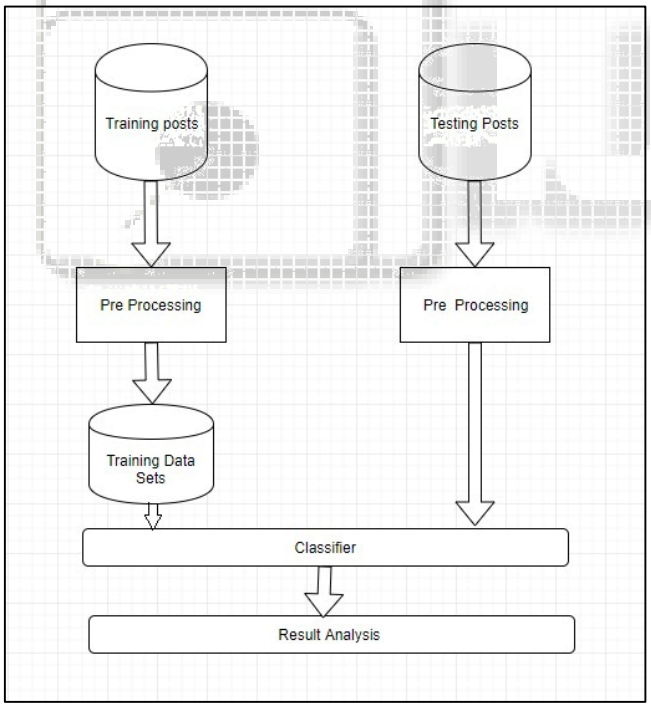

Fig. Data Flow Diagram

Software is considered to be a deliverable once its design is transformed into executable codes on selected platform subjected to specified requirements adhering to best available coding standards and practices, The platform used to develop the student learning experience systems is Windows as this Operating System supports Eclipse Luna, Net Beans and also provides server configurations, The names of the variables, methods, class have been chosen carefully so as to reflect the purpose of the respective token's use. The system uses separate file for each of the class developed with header file for every class file. The header file consists of the all the major variables to be used in the class and defines the prototype of the methods that are implemented by the class. The header comments in the files enable any third party user to understand the contents of the files and also seek help from the developer of the source file. The entire work makes use of class for each and every operation. All the codes written reside in method definition of the one or the other class. Placing all the methods in a base file and making them to accessible to all derived classes to be an efficient solution.

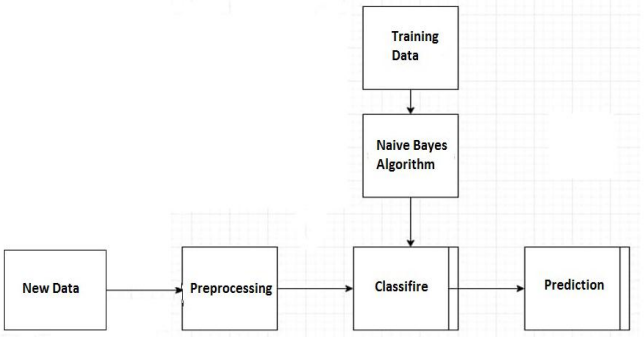

Fig. Block Diagram

V. MATHEMATICAL MODEL

S = (I,O,F)
Where,
S: System.
I= {AI, UI, PI, AC} are set of Input
Where,
• AI : Admin Information.
• UI : User Information.
• PI : Post Information.
• AC : Admin credentials.
F = {F1, F2, F3, } are set of Function
Where,
• F1 : Analyzing.
• F2 : Storing in Database.
• F3 : Retrieving data for organization.
O = {RG, NM, } are set of Output
where,
• W : Report generation.
• NM : Notification Message.
Success Condition : Should analyze the data properly and generated the report.
Failure Condition : No Database, No Internet Connection.

VI. END USERS

1) Company
2) Collages
3) Schools

VII. CONCLUSION

Hence researchers in learning analytic, educational data mining and learning technologies. It provides a workflow for analyzing social media data for educational purposes that overcomes the major limitations of both manual qualitative analysis and large scale computational analysis of user generated textual content. And the study can inform educational administrators, practitioners and other relevant

All rights reserved by *www.ijsrd.com*

decision makers to gain further understanding of engineering student's college experiences. This also provides a hope to see a proliferation of work in this area in the near future. It also advocates that great attention needs to be paid to protect students privacy when trying to provide good education and services to them

REFERENCES

[1] K. Lwin, M. Hashimoto, Y. Murayama, "Real-Time Geospatial Data Collection and Visualization with Smartphone", Journal of Geographic Information System, vol. 6, pp. 99-108, 2014.
[2] J. A. Evans, J. G. Foster, Metaknowledge, New York, N.Y.:Science, vol. 331, no. 6018, pp. 721-725, 2011.
[3] Bhavika Bathiya, Sanjay Srivastava, Biswajit Mishra, Air pollution monitoring using wireless sensor network, 2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECONECE), pp. 112-117G. King, Ensuring the data-rich future of the social sciences, New York, N.Y.:Science, vol. 331, no. 6018, pp. 719-721, 2011.
[4] J. T. Overpeck, G. A. Meehl, S. Bony, D. R. Easterling, Climate data challenges in the 21st century, New York, N.Y.:Science, vol. 331, no. 6018, pp. 700-702, 2011.
[5] O. J. Reichman, M. B. Jones, M. P. Schildhauer, Challenges and opportunities of open data in ecology, New York, N.Y.:Science, vol. 331, no. 6018, pp. 703-705, 2011.
[6] K. Lwin, Y. Murayama, "Person Trip Data Browser Analyzer and Space-Time Visu alizer", 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services ICSDM 2015, July 8–10, 2015.
[7] M. F. Goodchild, J. Glennon, "Crowdsourcing geographic information for disaster response: a research frontier", International Journal of Digital Earth, vol. 3, no. 3, pp. 231-241, 2010.
[8] M. K. McCall, "Seeking good governance in participatory-GIS: a review of processes and governance dimensions in applying GIS to participatory spatial planning", Habitat International, vol. 27, pp. 549-573, 2003

All rights reserved by *www.ijsrd.com*