Comparitive study on Efficient Analyses and Inference of Geo-social media to make Real-time decisions in Big-Data

Ms.S.Sindhuja[#1], Ms.M.A.Thansira banu[#2], Ms.M.Sowndharya[#3], Ms.C.Gokulapriya[#4]

[#1]*Assistant Professor, Department of Information Technology, A.V.C College of Engineering, TamilNadu, India.*

[#2-#4]*UG Students, Department of Information Technology, A.V.C College of Engineering, TamilNadu, India.*

Email: sindhusm77@gmail.com, thansi611997@gmail.com, kavimariyappansv96@gmail.com, c.gokulapriya20@gmail.com

*Abstract-*Geo-social networking is a type of social networking in which geographic services and capabilities such as geo-coding and geo-tagging are used to enable additional social dynamics. User-submitted location data or geo-location techniques can allow social networks to connect and coordinates users with local people or events that match their interests. Geo-location on web-based social network services can be IP-based or use hotspot. Our theme is to identify and collect information about current events such as disasters, fatal diseases, and so on. From the gathered information, optimal decision will be taken for particular scenario. In part of our proposed work, various related papers are reviewed and discussed. In this paper, we equipped enormous details such as technology used, algorithm proposed, merits and demerits of existing papers.

Keywords - Geo-social Network, Big Data, Hadoop, NLP (Natural Language processing)

I.INTRODUCTION

Geo-social networking allows users to interact relative to their current locations. Web mapping services with geo-coding data for places can be used with geo-tagged information to match users with a place, event or local group. Popular Geo-social media allows users to share their locations as well as recommendations for locations[6]. The technology has obvious implications for event planning and coordination. People at the event can communicate with each other and the larger world using a mobile device connected to the Internet[3]. When a user posts any activity that is related to some events, the user acts as a sensor that sends data to its station i.e., the Geo-social Network that performs analysis on the data to find information about what is occurring in the world. Geo-social Networks can be an asset for governments in terms of providing facilities and safety from disasters through proper management and reduction of the fear of the spread of an infections[4].

II. ESSENTIAL CHARACTERISTICS OF GEO-SOCIAL MEDIA

(a) **Mobile web:**
Understanding and exploiting user behaviors from their historical interactions in Geo-social media.

(b) **Data mining:**
Mining the complex data containing social graphs of users, spatial deployment of locations and time evolving information flows.

(c) **Information retrieval:**
Integrating Geo-social media data with location information sensored by mobile devices to provide Geo-social search engine .

(d) **Social network analysis:**
Revisiting the conventional SNA tasks considering both social connections and moving behaviors of users.

(e) **Recommender system:**
Item recommendation with geographical and social factors.

(f) **Marketting:**
Devising the marketing strategies based on how users move and interact with each other in the city.

(g) **Advertisement:**
Placing the ads on mobile devices according to the visiting preferences of users.

III. ISSUES ON EXISTING WORK

a) Most of the existing system, have used **NLP (Natural Language Processing)** on social media data. It is applied for text mining particularly for Sentiment analysis.
b) Consumes large time to process when data size is billions.
c) Time series data model not supported.
d) Doesn't support huge inflow of data.
e) Consumes a large amount of time to perform operations on larger set of objects.

IV. LITERATURE REVIEW

Eleanna Kafeza et al. [1] proposed, Marketing Campaigns in Twitter using a Pattern Based Diffusion policy. In this paper they introduce a novel methodology to achieve information diffusion within a social graph that activates a realistic number of users. In a social media marketing campaign they were interested to identify the initial users who are willing to pass the message to their friends and achieve a maximum diffusion, this is called the seeding process. After seeding information within the network, customers communication result in information spread using history information of their previous patterns of diffusion. The main advantage is effective marketing. In this work, they looked into the problem of finding the pattern behavior of the users that can achieve an information diffusion result similar to the expected actual diffusion.

Nadiya Straton et al. [2] proposed, Big Social Data Analytics for Public Health: Predicting Facebook Post Performance using Artificial Neural Networks and Deep Learning. In this paper they used clustering algorithm. This research study aims at health and care organizations to improve information dissemination on social media platforms by reducing clutter and noise. At the same time, it will help users navigate through vast amount of information in direction of the relevant health and care content. They have tested using different algorithms, models and statistical approaches to find the most effective method to evaluate post performance and then make predictions based on the number of relevant attributes with Artificial Neural Networks (ANN) and Deep Neural Networks (DNN)..

Nicolai H. Egebjerg et al. [3] proposed, Big Social Data Analytics in Football: Predicting Spectators and TV Ratings from Facebook Data. Machine learning algorithm were used and it explores the predictive power of big social data in regards to football fans off-line and on-line behaviors using Data analysis achieved by predictive and textual data analysis. Finally they concluded this approach is possible to set up a model to predict the number of spectators and TV ratings of football viewers**.**

Adel Assiri et al. [4] proposed a Real-Time Sentiment Analysis of Saudi Dialect Tweets Using SPARK. This work was based on lexicon-based algorithm that will be implemented using big data techniques and sentiment analysis for Saudi dialect in twitter. The results showed significant performance in running time compared to the implementation of the lexicon based algorithm.

Sien Chen et al. [5] proposed, Big Data Analytics on Aviation Social Media: The Case of China Southern Airlines on Sina Weibo. This paper proposed that how big social data analysis can help an airline companies to understand the passengers better and improve customer relationship management. To achieve this, three methods data collection, data analysis and findings with Machine-learning algorithm were used. As a result, with the combination of a passenger's social-media value, the airline company can improve flight arrangements and their service and also improve customer satisfaction and loyalty.

Manuel Rodriguez-Martinez et al. [6] proposed their experiences with the Twitter Health Surveillance (THS)System. Machine learning algorithm were used. THS was designed as a platform to allow end-users to monitor a stream of tweets and process the stream with a combination of built-in functionality and their own user-defined functions.

Andy Januar Wicaksono et al. [7] developed a method for Predicting US Presidential Election by Analyzing Sentiment in Social Media. This can achieved by Pre-processing, Sentiment analysis and Visualization using graph representation with advanced Naive Bayes algorithm. This method w a s created

by referring the previous researches and adding some value to make prediction more accurate.

Dandan Jiang et al. [8] proposed, Sentiment Computing for the News Event Based on the Social Media Big Data. The sentiment computing of news event was a significant component of the social media big data. It has also attracted lots of researches which could support many real-world applications, such as public opinion monitoring for governments and news recommendation for websites. Using word emotion computation through word sentiment association network and word emotion refinement through standard sentiment. This is helpful for the corporations and governments to make specific decisions or global strategies.

Yanish Pradhananga et al. [9] developed, High Performance Analytics of Big-data with Dynamic and Optimized Hadoop Cluster. Users demand a service platform that can store and handle large quantities of data with some features such as easy accessibility, fast performance, durable and secure. These features can be availed without having to spend too much on hardware, upgrading, configuring to perform analysis of big data. Using Cluster, Optimization Bootstrap Action and Map Reduce with R-Programming language and R hadoop. This system supports dynamic and optimized cluster nodes size as per the desired time, user does not need to calculate and estimate the number of nodes. It also calculates the optimal number of nodes required in the cluster for the processing of input data within a fixed time frame.

Nikos Tsirakis et al. [10] proposed, Large scale opinion mining for social, news and blog data. This system collect and analyze data from social media, news and other data streams are faced with several challenges that concern storage and processing of huge amounts of data. They first used Greek stop-words in order to collect only Greek tweets from the Twitter API and then performed text analysis in order to create any metadata for these documents. The documents were indexed in the elastic search cluster.

V.COMPARITIVE STUDY

RESEARCH PAPER	ALGORITHM	DESCRIPTION
[1]Marketing Campaigns in Twitter using a Pattern Based Diffusion policy	Pattern Based Diffusion Algorithm	Identifies the initial users who are willing to pass the message to their friends and achieve a maximum diffusion this is called the seeding process. It is useful for marketers who are intended to use social influence and run effective marketing campaigns.
[2]Big Social Data Analytics for Public Health: Predicting Facebook Post Performance using Artificial neural network and Deep Learning	1)Artificial Neural Network and Deep learning. 2)Clustering algorithm	Testing different algorithms, models and statistical approaches to find the most effective method to evaluate post performance and then make predictions based on the number of relevant attributes with Artificial Neural Networks (ANN) and Deep Neural Networks (DNN).
[3]Big Social Data Analytics in Football: Predicting Spectators and TV Ratings from Facebook Data	Machine learning algorithm	Uses Machine learning algorithm to explores the predictive power of big social data in regards to football fans off-line and on-line behaviors using Data analysis which it is achieved by predictive and textual data analysis.
[4]Real-Time Sentiment Analysis of Saudi Dialect Tweets Using SPARK	Lexicon-based algorithm	Real-time solution for Saudi dialect in twitter is developed. The developed solution supports the process of stream data in real-time.
[5]Big Data Analytics on Aviation Social Media: The Case of China Southern Airlines on Sina Weibo	Machine-learning algorithm	Uses Machine-learning algorithm to help an airline company to understand the passengers better and improve customer relationship management.

RESEARCH PAPER	ALGORITHM	DESCRIPTION
[6]Experiences with the Twitter Health Surveillance (THS) System	Machine learning algorithm	Twitter Healthcare Surveillance is designed as a platform to allow end-users to monitor a stream of tweets and process the stream with a combination of built-in functionality and their own user-defined functions.
[7]A Proposed Method for Predicting US Presidential Election by Analyzing Sentiment in Social Media.	Advanced Naive Bayes algorithm	Research methodology is used to perform data collection and implementation. Data used in this research are collected from twitter.
[8]Sentiment Computing for the News Event Based on the Social Media Big Data	Sentiment analysis algorithm	Uses word emotion computation through word sentiment association network and word emotion refinement through standard sentiment. This is helpful for the corporations and governments to make specific decisions or global strategies.
[9]High Performance Analytics of Bigdata with Dynamic and Optimized Hadoop Cluster	R-Programming language and R hadoop	Supports dynamic and optimized cluster nodes size as per the desired time, user does not need to calculate and estimate the number of nodes. It also calculates the optimal number of nodes required in the cluster for the processing of input data within a fixed time frame.
[10]Large scale opinion mining for social, news and blog data	Fast text clustering Algorithm	Collects and analyze data from social media, news and other data streams are faced with several challenges with storage and processing of huge amounts of data. They first used Greek stop-words in order to collect only Greek tweets from the Twitter API and then performed text analysis in order to create metadata.

VI. CONCLUSION

In this paper, we reviewed some existing research related to our proposed work. In existing work above reviewed all techniques have some disadvantages. So we proposes Efficient analyses and inference of geo-social media to make real-time decisions in Big Data. In Geo-social media humans acts as sensors for monitoring activities world wide. When a user posts any activity that is related to some events, the user acts as a sensor that sends data to its station to find information about what is occurring in the world. Our decided system will collect and analyses all the tweets generated by Twitter users, in addition to Facebook statuses, filter images, and information from other Geo-social Network to make real-time decisions.

ACKNOWLEDGEMENT

REFERENCES

[1]. Eleanna Kafeza, ChristosMakris, Pantelis Vikatos, "Marketing Campaigns in Twitter using a Pattern Based Diffusion policy", Big Data (Big Data Congress), 2016 IEEE International Congress on 27 June 2016.

[2]. Nadiya Straton, Raghava Rao Mukkamala, Ravi Vatrapu , " Big Social Data Analytics for Public Health: Predicting Facebook Post Performance using Artificial neural network and Deep Learning", Big Data (Big Data Congress), 2017 IEEE International Congress on 25 June 2017.

[3]. Nicolai H. Egebjerg, Niklas Hedegaard, Gerda Kuum, Raghava Rao Mukkamala and Ravi Vatrapu, " Big Social Data Analytics in Football: Predicting Spectators and TV Ratings from Facebook Data", Big Data (Big Data Congress), 2017 IEEE International Congress on 25 june 2017.

[4]. Adel Assiri, Ahmed Emam, Hmood Al-dossari, "Real-Time Sentiment Analysis of Saudi Dialect Tweets Using SPARK", Big Data (Big Data), 2016 IEEE International Conference on 5 December 2016.

[5].   Sien Chen, Yinghua Huang, Wengqiang Huang, " Big Data Analytics on Aviation Social Media: The Case of China Southern Airlines on Sina Weibo" Big Data Computing Service and Applications (Big Data Service), 2016 IEEE Second International Conference on 29 March 2016.

[6].   Manuel Rodriguez-Martinez, "Experiences with the Twitter Health Surveillance (THS) System", Big Data (Big Data Congress), 2017 IEEE International Congress on 25 June 2017.

[7].   Andy Januar Wicaksono, Suyoto, Pranowo, "A Proposed Method for Predicting US Presidential Election by Analyzing Sentiment in Social Media", Science in Information Technology (ICSITech), 2016 2nd International Conference on 26 October 2016.

[8].   Dandan Jiang, Xiangfeng Luo, Junyu Xuan, Zheng Xu, "Sentiment Computing for the News Event Based on the Social Media Big Data", IEEE Access (Volume: 5), 15 September 2016

[9].   Yanish Pradhananga , Shridevi Karande, Chandraprakash Karande, "High Performance Analytics of Big data with Dynamic and Optimized Hadoop Cluster", Advanced Communication Control and Computing Technologies (ICACCCT), 2016 International Conference on 25 May 2016

[10]. Nikos Tsirakis, Vasilis Poulopoulos, Panagiotis Tsantilas, Iraklis Varlamis, "Large scale opinion mining for social, news and blog data", Journal of Systems and Software Volume 127, May 2017.cation With Randomized Tag", IEEE Transactions On Information Forensics And Security, Vol. 12, NO. 3, March 2017.