Big Data Analytics on Social Media Data: A Literature Review

Prashant Sahatiya

Student-I SEM, M.Tech-Computer Engineering, Parul Institute of Engineering & Technology, Gujarat, India

--***---

Abstract – *Throughout the last years, the net has however seen a wider scope through the event of social media. Supported communication techniques and accessible to any or all, the media promote social interaction through the net. Several social networks exist and there are a unit over 900 social media sites on the market on the net. Variant of fifty eight million tweets per day. Massive information is that the border of the flexibility of an enterprise in term of storing, process and accessing all the information it desires for the effective functioning, and to create choices cut back risks, and conjointly to serve the various customers among on additionally cheap time. The main objective of this study is to spot completely different techniques of analyzing the large social media information. Distinctive these techniques can facilitate in revealing the competition's promoting strategy together with their content, audience, and messages.*

Key Words: Big data analytics, social media data, data mining, social media analysis, Hadoop, predictive analytics, text mining

1. INTRODUCTION

The one factor that social media corporations mastery lies in its information. And this they need lots of it, because of their latency to induce users to share data regarding every waking minute. The massive body of knowledge at the disposal of social media corporations mirrors however individuals act with one another, and at the center of those interactions lies priceless data regarding what people and societies hold vital. This volume of knowledge, alongside the quick rate of knowledge flow that social media is renowned for, represent the essence of big data.

By applying analytics to social media information, massive information applications in numerous industries go on the far side the mechanics of interaction to seeing however the content contained within the interactions can have an effect on business performance and people's read of a complete. Content analytics allows corporations to zero in on unjust data from the messages that users post. As an example analytics tools will be programmed to trace negative or positive sentiment a couple of complete as this might threaten name and revenue.

Just like the other trade, social media corporations find big data useful for analyzing markets and predicting client behavior. In 2012, Jay Parikh, engineering VP at Facebook, disclosed that Facebook handles over five hundred terabytes of knowledge a day, three hundred million photos daily, 2.6 billion 'Likes' and a couple of 5 billion content uploads. All this information is processed in exactly minutes giving

Facebook insight into user reactions and therefore the ability to roll out or modify its giving. [7]

What's additional, correlating content to demographics of user's age, gender, legal status, geographic location, financial gain levels, instructional accomplishment, and inclination to buy bound product permits an organization to grasp additional regarding the individuals it's managing. Such analysis conjointly reveals however adverts do among completely different client segments. This {can be} terribly helpful as advertisers can react in close to real time and modify campaigns to create additional revenue. Analysis of social media information collected by a distributor may as an example reveal that widowed females between twenty five and thirty five area unit appropriate candidates for a reduction supply on gymnasium instrumentality. Supported this data, the distributor may arrange to target these candidates with discount offers through Twitter, Facebook and different media. If analytics show the uptake and comments area unit dangerous, the supply will be refined to enhance performance.

A lot of corporations appreciate the powerful nature of social media for personal-level interaction with their customers. The product customization existed since before social media, the extent and roughness to that it's done by businesses that collect social media information is astounding. Through social media analytics tools, these corporations will build data-driven choices by the minute.

Furthermore, social media analytics tools mean that companies will look on far side the chatter contained in unstructured information to seek out purposeful data which will guide choices and action. Through analysis of applied mathematics information like impressions per post, audience distribution, interactions on mobile versus desktop, responses, click-through rates for URLs embeds, and transactional history, an organization will live the effectiveness of its social media strategy for promoting complete recognition and loyalty.

Big information conjointly makes it potential to achieve insight into the roles individuals play among social media teams. Users with an oversized range of followers as an example, will be thought of to be influencers. By singling out such individuals, an organization will monitor trends in discussion threads and even participate in such discussions.

2. IMPORTANCE OF SOCIAL MEDIA ANALYTICS

While massive information is bothered with all types of sources, it's calculable that the bulk of it comes from unstructured sources. Collectively may think, social media

constitutes maybe the most important supply of unstructured sources for giant information. Likes, tweets, views, comments, favorites, and everything else that users will do an act with in any social media platform will be collected and analyzed by interested parties. Within the digital age, social media is very important for any business. Maintaining a presence on platforms like Facebook and Twitter is important as a result of it permits individuals to act with the corporate on apparently personal level that aids businesses across multiple fronts. It's conjointly vital for the typical client. Facebook alone boasts two billion monthly users, regarding twenty sixth of the world's entire population. It's important then to think about that massive information from social media will arrive in an unbelievable range of forms. [8]

Big information analytics helps organizations harness their information and use it to spot new opportunities. That, in turn, ends up in smarter business moves, additional economical operations, higher profits and happier customers. In his report massive information in big corporations, IIA Director of analysis Tom Davenport interviewed over fifty businesses to grasp however they used massive information. He found they got price within the following ways: [9]

1. Cost reduction: Massive information technologies like Hadoop and cloud-based analytics bring important price benefits once it involves storing massive amount of knowledge and they will determine additional economical ways of doing business.

2. Faster, higher cognitive process: With the speed of Hadoop and in-memory analytics, combined with the flexibility to investigate new sources of knowledge, businesses area unit able to analyze data now and build choices supported what they've learned.

3. New product and services: With the flexibility to measure client desires and satisfaction through analytics comes the facility to offer customers what they require. Davenport points that with massive information analytics, additional corporation's area unit making new product to satisfy customer's desires.

3. HOW IT WORKS AND KEY TECHNOLOGIES

There's no single technology that encompasses massive information analytics. Of course, there's advanced analytics which will be applied to massive information, however essentially many kinds of technology work along to assist you get the foremost price from your data. Here are some of the most important players:

I.   Data management: Information has to be top quality and well-governed before it will be dependably analyzed. With information perpetually flowing in and out of a company, it is important to determine repeatable processes to create and maintain standards for information quality. Once information is reliable, organizations ought to establish a master information management program that gets the complete enterprise on an equivalent page.

II. Data mining: Data processing technology helps you examine massive amounts {of information of knowledge of data} to get patterns within the data – and this information will be used for any analysis to assist answer advanced business queries. With data processing software system, you'll sift through all the chaotic and repetitive noise in information, pinpoint what is relevant, use that data to assess doubtless outcomes, and so accelerate the pace of constructing privy choices.

III. In-memory analytics: By analyzing information from system memory (instead of from your magnetic disc drive), you'll derive immediate insights from your information and act on them quickly. This technology is ready to get rid of information schoolwork and analytical process latencies to check new eventualities {and create and build} models; it isn't solely a simple manner for organizations to remain agile and make higher business choices, it conjointly allows them to run repetitious and interactive analytics eventualities.

IV. Predictive analytics: Prophetical analytics technology uses information, applied mathematics algorithms and machine-learning techniques to spot the probability of future outcomes supported historical information. It's all regarding providing a best assessment on what is going to happen within the future, therefore organizations will feel additional assured that they are creating the most effective potential business call. A number of the foremost common applications of prophetical analytics embody fraud detection, risk, operations and promoting.

V.   Text mining: With text mining technology, you'll analyze text information from the online, comment fields, books and different text –based sources to uncover insights you hadn't noticed before. Text mining uses machine learning or tongue process technology to comb through documents- emails, blogs, twitter feeds, surveys, competitive intelligence and additional- to assist you analyze massive amounts of knowledge and see new topics and term relationships.

4. TOOLS USED FOR BIG DATA ANALYTICS

Instead of puzzling over social media analytics as a noun, think about it as a verb. Specifically, its gathering information from social platforms to assist guide your promoting strategy.

This method begins by prioritizing business goals. For example, your focus could also be to double the quantity of recent guests to your web site.

The second step is decisive key performance indicators (KPIs). During this case, your chief social media KPI would doubtless be based mostly on engagement stats. These can be broken down into:

- Likes and shares your post receives
- Replies and comments
- (most importantly) clicks your links and content earn

By collecting this data, you can figure out how social media factors into meeting your business goal. From there, you can keep going in the direction you're headed or adjust your approach.

The famous tools for big data analytics are: [6]

- Keyhole
- AgoraPulse
- Brandwatch
- Buffer
- BuzzSumo
- Crowdbooster
- Edgar
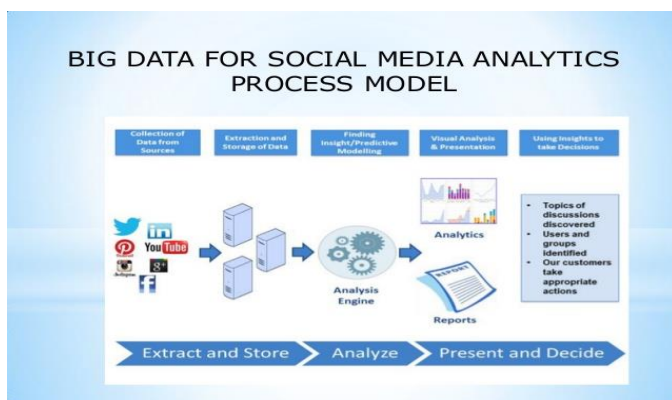- Google Analytics and many more

Fig-1: Process model for big social analytics

5. HOW BIG DATA IS ANALYZED

Big data is analyzed using different techniques which gives different outcomes. To use which technique is dependent on the user's need for analyzing the big data.
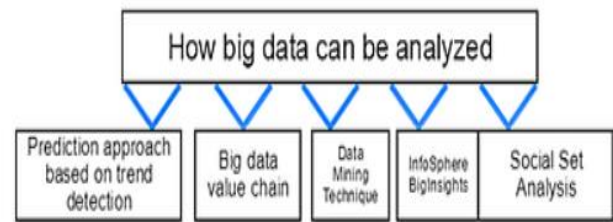

Fig-2: Ways of analyzing social bog data

Prediction approach based in trend detection: It presents the improved prediction approach supported trend detection. First, a dynamic time distortion distance-based K-medoids rule is applied to cluster programs' quality evolution into four trends. Then, four trend-specific prediction models are designed one by one using random forests regression. Consistent with the options extracted from an electronic program guide and early viewing records, freshly printed programs are classified into the four trends by a gradient boosting decision tree. Finally, by combining prediction values from the trend-specific models and the classification probability, the planned approach achieves higher predication results. [1]

Big data value chain: It presents a scientific framework to decompose massive information systems into four successive models, particularly information generation, information acquisition, information storage, and information analytics. These four modules form a giant information value chain. Following that, paper presents a close survey of diverse approaches and mechanisms from analysis and trade communities. Additionally, it presents the prevalent Hadoop framework for addressing massive information challenges. Finally, paper outlines many analysis benchmarks and potential analysis directions for giant information systems. [2]

Data mining techniques: The increasing reliance on social networks needs data processing techniques that is likely to facilitate reforming the unstructured information and place them among a scientific pattern. The goal is to investigate the information mining techniques that were utilized by social media between 2003 and 2015. [3]

InfoSphere BigInsights: It presents the usage of Twitter during a range of planned subjects that is that the largest social networking web site wherever Twitter information is in increasing at higher rates a day that considers it as massive information supply. Then, describing well the manner within which massive information technology, such as, InfoSphere BigInsights allows process of this information, that are primarily collected from social networks by Apace Flume and keep in Hadoop storage. [4]

Social Set Analysis: It presents a brand new approach to massive information analytics known as social set analysis. Social set analysis consists of a generative framework for the

philosophies of procedure science, theory of social information, abstract and formal models of social information, and an analytical framework for combining massive social information sets with structure and social information sets. [5]

6. CONCLUSION

Big data is analyzed using different techniques and tools. One of the most important factor is to identify the techniques and tools for the big social data. Further handling, storage and retrieving of big data can be achieved after selecting proper tools and techniques for analyzing the big data. In this paper, different analysis techniques and tools categorized which can be helpful in building business strategies and other relevant fields.

7. CHALLENGE

Big information can still accelerate the intrusion of social media corporations into People's Privacy. As Facebook, Twitter, Instagram and Pinterest still monetize their offerings, it might seem that the advantages that massive information can have for social media within the future can become even additional personalized. A study printed by researchers from Cambridge and Stanford Universities shows that Facebook will use its information to predict people's temperament with additional accuracy than shut friends and families.
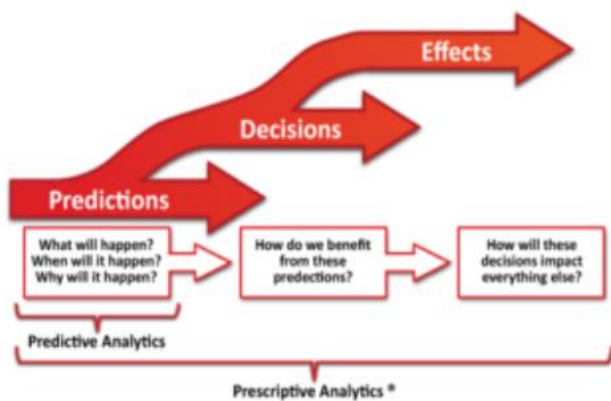

Fig-3 : Security at the risk

Every like, share, follow and comment, is information that tells social media corporations what you wish or dislike, what your actions are going to be, that cause or complete you'll support and what you're doubtless to shop for. Not to mention, any action you're taking on browsers and search engines these days can possibly link back to your social media profile, leaving behind an extended path of digital footprint which will be used for detection your next moves. This case can be solely intensify as individuals become additional dependent on social media platforms for accesing and sharing data.

REFERENCES

[1] Chengang Zhu, Guang Cheng (senior member, IEEE) and Kun Wang (senior member, IEEE), "Big data analytics for program popularity prediction in broadcast TV industries", IEEE Access, October'2017.

[2] Han Hu, Yonggang Wen (Senior member, IEEE), Tat-Seng Chua and Xuelong Li (Fellow, IEEE), "Toward scalable systems for Big Data Analytics: A technology tutorial", IEEE Access, June'2014/

[3] MohammadNoor Injadat, Fadi Salo, Ali Bou Nassif, "Data mining techniques in social media: A survey", Neurocomputing- ScienceDirect, June'2016.

[4] Marouane Birjali, Abderrahim Beni-Hssane, Mohammed Erritali, "Analyzing social media through big data InfoSphere BigInsights and Apache Flume", ScienceDirect, June'2017.

[5] Ravi Vatrapu, Raghava Rao, Mukkamala, Abid Hussain and Benjamin Flesch, "Social set analysis: A set theoretical approach to big data analytics", IEEE Access, June'2015.

[6] "A list of the top 25 social media analytic tools", https://keyhole.co/blog/list-of-the-top-25-social-media-analytics-tools/

[7] "How Social media companies use big data" https://datafloq.com/read/how-social-media-companies-use-big-data/1957

[8] "Will big data change how you use social media?" https://thenextweb.com/contributors/2017/07/06/will-big-data-change-use-social-media/

[9] "Big data analytics. What it is and why it matters" https://www.sas.com/en_us/insights/analytics/big-data-analytics.html