International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Challenges arise of Privacy Preserving Big Data Mining Techniques

M.Chalapathi Rao[1], A.Kiran Kumar[2]

[1] *Assistant Professor, Dept of CSE, CMR Technical Campus, Telangana, India.*

[2] *Assistant Professor, Dept of CSE, CMR Technical Campus, Telangana, India.*

Abstract - *Big data is being generated from various sources-transactions, social media, sensors, digital images, videos, audios and for domains including healthcare etc this data is known as big data. The characteristics of Big Data include 3 Vs they are Volume, Velocity and Variety. Useful data can be extracted from this big data with the help of data mining techniques The massive volume of big data sets are too complicated to be managed and processed by conventional relational databases the term "Big Data" was coined to address this massive volume of data storage and processing. The quality of captured data can vary greatly, affecting accurate analysis. Protecting privacy is mechanism for data processing and producing right information to favor corporate sectors, business managers, stake holders and other users make highly informed business decisions. In this paper we are proposing a big data on privacy preserving Big Data.*

Key Words: **BI, Velocity, Volume, Variety.**

1. INTRODUCTION

A combination of policy decisions, technical and legal mechanisms are use to address privacy concerns. A brief description of some of the major principles for protecting the privacy of data in its lifecycle is given below Data collection and limitations: This principle limits the unnecessary excessive collection of personal data. Once the purpose for which data is collected is known, collected data should be just sufficient enough for that purpose. This principle is clearly a policy decision on the part of collector. Usage limitations: While collecting sensitive or personal data, collector needs to specify what for and how the data is used and limit the usage of collected data for other purposes than the original one. Security of data It is an obligation of data collector to keep the data safe once collected. Adequate security mechanism should be in place to protect it from breaches. Transfer Policy Often the usage of data is governed by laws which are prevalent at the place of collection and usage. If the data is moved outside the jurisdiction where the law enforcement is not prevalent in the new place it carries the danger of misuse. Accountability when dealing with third party data, the party may ask to designate a person who is the point of contact and take processing and usage of data. Collection limitation is the policy decision on part of data collector usage limitation, securing data and transfer policy can be addressed by technical means and the last one, accountability is addresses by having a legal team sign a declaration.

2. BIG DATA

There are various Challenges that companies are facing in identifying and processing value from the "Big Data" .It is a big data challenge to preserve the privacy of end users at various stages of data life cycle. Today we live in the computerized world with drastic digitization the volume of structured and unstructured data being generated and stored in exploding. The data which is proposed is being taken from various origins In addition to business and companies, individuals spend to the data volume. For particular 40billion content are being distributed on face book every month .Big data is massive volume of both structured and unstructured data from various origins such as social data, machine generated data, traditional enterprise which is so large that it is complex to process with traditional database and software techniques. Big Data is data who's metric, diversity, and complexity require new framework, techniques, algorithms, and analytics to manage it and fetch value and hidden knowledge from it. Characteristics of Big Data include 3 Vs. They are Volume, Velocity, Variety and Veracity.

3. BIG DATA CHALLENGES

Here we have several challenges that companies are facing in identifying, processing and fetching value from the "Big Data". As in the following some of the key challenges include

1. Preserving Privacy
2. Integrating the Big Data Technologies
3. Proposing real time needs with higher data volumes and varieties
4. Data maintenance and management
5. Data distribution and Analytical systems.

4. PRIVACY PRESERVING POLICIES

4.1 Anonymization Techniques

Substitute delicate attribute values with some other values. This decreases disclosure of private data. In some situations, this clear substitution alone wills not sophisticated anonymization techniques need to be worked.

4.2 Generalization

Occasional attributes in data concepts are substituted with universal terms. For example take into consider some persons who have a PG degree are very less in employee database. A query which gathers person education qualifications and their address can be correlated and the result set of a query which lists salary and age of the person can show the uniqueness of a candidate. This can be decreased by substituting qualification with another universal procedure called graduation which makes it complex to correlate.

4.3 Randomization

In this Randomization technique buzz is appended to the data fields of records. This controls fetching true personal information however the cumulate results are protected. For example name and address of employees is unplanned however queries like name and address will result in correct result set. One of the advantage of unplanned technique is it can be used with independent data records and does not require more discovered knowledge of other record values; hence this technique is most closed for data which is created as a stream. There is variety of unplanned techniques such as additive concepts, multiplicative concepts and data swapping and manipulation techniques in this domain.

5. BUSINESS INTELLIGENCE

This is the big community and encloses the possible three concepts here. BI is a uniqueness of a process. It contains aggregation, analytical and disclosure of data to express and help business logics and plans. All the other things spotted to some kind of how data is cumulated, while BI goes inside the data to have what business officials usually do. In previous years has been away from systems that rely on IT staff to facilitate reports and graphs for decision-makers BI data can hold old information, as well as new data cumulated from origin systems as it is generated, enabling BI analysis to help both strategic and decision-making processes. Primarily, BI tools were used by data analysts and other IT experts who ran analyses and produced summary with query results for business users. However, business officials and workers are using BI tools themselves. The desirable benefits of business intelligence sessions include stimulating and developing decision making; internal business processes; increasing process efficiency; driving new revenues; and gaining competitive advantages over business methods. BI systems can also help companies finding market movements and mentioned business problems that need to be addressed. Business intelligence combines some set of data analysis applications including

1. Ad hoc analysis and querying
2. Online Analytical Process
3. Real-time BI
4. Operational BI
5. Cloud
6. Location intelligence.

BI technology also contains data judgment software for designing charts and other relevant graphs as well as tools for building BI control panel and performance chart that display data on business calculations and key performance symbols in an easy-to-understand the way. BI applications can be bought separately from various dealers or as part of a consolidated BI environment from a single dealer. BI plays very important role in the big data environments with implies some suitable applications these are vary from the numerous domains that arise the challenges of big data

6. CONCLUSION

The Big Data analytics offers large amount of opportunities to improve business values. The main applications of Big Data analytics is for business intelligence to enhance decision making proficiency, speed decision making, understanding of customer demands, developing plans for inviting new products and services, exploring new markets and improving staff proficiency and efficiency. To that objective, this study examines significance of Big Data analytics of data collected from Social Media for enhanced business intelligence in various businesses. Social Media in India and online business in India have grown exponentially over the last decade and therefore present an opportunity to look into the how data collected from these channels can be usefully utilized for further business enhancements. Data mining techniques provides the backbone to harnessing information quickly and efficiently on big data. However this also means there is a potential for extracting personal information by compromising on user privacy. In this paper we initially describe principles that can be used to protect the privacy of end users at various stages. Subsequently we explore technical aspects of protecting privacy while processing Big Data.

REFERENCES

[1] D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.

[2] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[3] K. Elissa, "Title of paper if known," unpublished.

[4] D.J. Abadi, Data management in the cloud: Limitations and opportunities,

[5] IEEEDataEngineeringBulletin32(1)(2009)3–12.

[6] Amazonredshift,http://aws.amazon.com/redshift/.Amazondatapipeline,http://aws.amazon.com/datapipeline/.

[7] AmazonElasticMapReduce(EMR),

[8] http://aws.amazon.com/elasticmapreduce/.

[9] AmazonKinesis,http://aws.amazon.com/kinesis/developer-resources/.

[10] R. Ananthanarayanan, K. Gupta, P. Pandey, H. Pucha, P. Sarkar, M. Shah.

BIOGRAPHIES

M.CHALAPATHI RAO, Currently working as an Assistant Professor in CMRTC, Hyderabad. He completed his M.Tech degree in J.N.T.U. Hyderabad. He received B.Tech Degree from J.N.T.U.Hyderabad. He had 9 years of Teaching Experience. Interested domain is Big Data.

Adepu Kiran Kumar presently working as a Assistant Professor in CMR Technical campus, Hyderabad, Telangana, INDIA. He received his M.Tech (CSE) degree from Medak Engineering College in the year 2014. He received B.Tech. (CSE) degree from Jayamukhi Institute of Technological Sciences, in the year 2005. His fields of interests are Cloud Computing, Web Technologies, computer Networks, etc.