Role of Data mining in big Data and the Data Governance issues related to Data Ownership

Om Sadigale, w1943544

Introduction:

With the continuous growth of Big Data, Data mining has become essential for effective analysis. This data is collected from various sources such as individuals, organisations, and sensors, which have the possibility of bias. Data Mining provides valuable insights that supports informed decision-making, as well as applications like credit risk management and fraud detection. Data governance is essential in Big Data as it ensure vast amounts of data collected are managed securely and ethically to maintain data integrity. Data governance is important for organizations that use data to drive business growth and make decisions. It helps ensure that the right data is accessible to the right people, and that data is used in a way that is compliant with regulations.

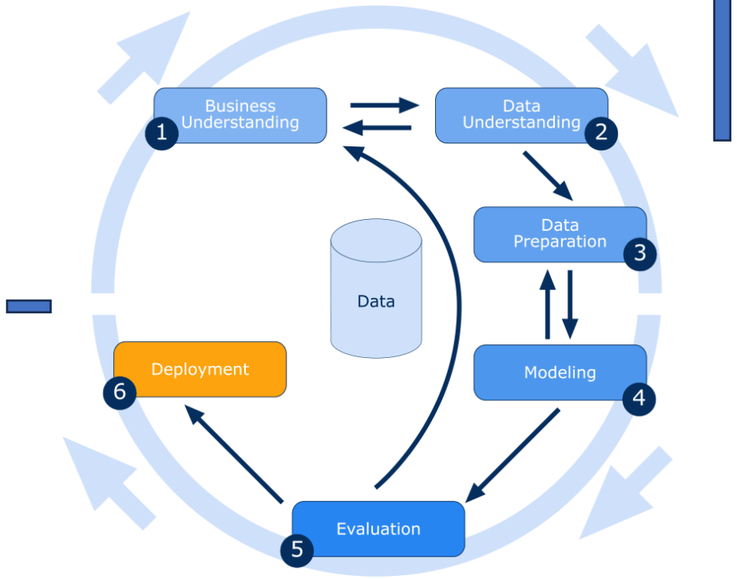
Aim:

This research aims to explain the role of Data Mining in Big Data and the Data Governance issues related to Data Ownership.

Analysis:

Considering the sheer volume of unstructured data combined with velocity, variety and poor quality of data creates significant problems which can be addressed through Data mining. The process is to collect, preprocess and analyse this data to make it useable. Though Data mining is complex and expensive, it can be applicable to wide selection of data and improving effectiveness of data. Without these processes, the data remains unreliable due to various factors, including individual bias such as race, gender, social status, and organisational bias like work culture. The two biases are connected to each other considering some organisational biases can be decided upon individual bias. Furthermore, errors may arise not only from human biases but also machine -generated data, specifically from sensors. To combat such issues the role of Data Mining is required to filter out such issues and gain insight and knowledge for firm decision making, which is done by uncovering any trend or patters.

There are various techniques used for Data Mining such as Association analysis, Classification, Prediction, Clustering and many more. In the article" Mining big Data in biomedicine and health care" such techniques were mentioned and are useful for healthcare system to make decisions faster.



Data governance is a set of processes, policies, roles, metrics, and standards that ensure data is: secure, private, accurate, available, and usable.

Here are some things to consider when implementing data governance for big data:

- Regulatory compliance
- Data classification
- Classify data based on sensitivity and regulatory requirements.
- Data lifecycle policies
- Establish, execute, and audit data lifecycle policies, including when to remediate expired data or apply policies to retain data for legal holds

Without clear data ownership, effective data governance becomes difficult to implement. A lack of accountability and responsibility for the data leads to data quality issues, security breaches, and compliance violations.

Ambiguity around data ownership could lead to different teams within an organisation to claim ownership of the same dataset, causing misinterpretation and irregularities in usage of data. Unclear ownership and responsibility could lead to inconsistent and unreliable data.

The data owners must have thorough understanding of protection laws such as GDPR and CCPA … to ensure data collection, storage and processing the data based on the legal standards. Failure to comply can result in significant fines and damage to the organisation's reputation.

From an ethical and social standpoint, organisations are obliged to respect user rights and uphold user control over their data privacy, ensuring transparency and integrity in their data practices.

Conclusion:

Data mining is crucial component in effective use of Big Data, uncovering insights within Big Data resulting in better decision-making and predictive analysis. Data governance issues related to data ownership often arise from unclear assignment of ownership roles and responsibilities. Data owners must be well-informed about legal requirements and ethical considerations to ensure responsible management.

References:

Wu, X., Zhu, X., Wu, G.-Q. and Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), pp.97–107. doi:https://doi.org/10.1109/tkde.2013.109.

Tse, D., Chow, C., Ly, T., Tong, C. and Tam, K. (2018). The Challenges of Big Data Governance in Healthcare. *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. doi:https://doi.org/10.1109/trustcom/bigdatase.2018.00240.

Herland, M., Khoshgoftaar, T.M. and Wald, R. (2014). A review of data mining using big data in health informatics. *Journal Of Big Data*, 1(1), p.2. doi:https://doi.org/10.1186/2196-1115-1-2.

Ekwonwune, E.N., Ubochi, C.I. and Duroha, A.E. (2022). Data Mining as a Technique for Healthcare Approach. *International Journal of Communications, Network and System Sciences*, 15(09), pp.149–165. doi:https://doi.org/10.4236/ijcns.2022.159011.

K.U, J. and M.David, J. (2014). Issues, Challenges and Solutions : Big Data Mining. *Computer Science & Information Technology (CS & IT)*. doi:https://doi.org/10.5121/csit.2014.41311.

Bruck, C. (2017). *Challenges and opportunities of Data Governance in private and public organizations*. [online] Available at: https://www.seres-unit.com/wp-content/uploads/DPA/DTA18_Arbeit_Christian_Bruck.pdf.

BarnRaisers, LLC. (2018). *6 essential steps to the data mining process - BarnRaisers, LLC*. [online] Available at: https://barnraisersllc.com/2018/10/01/data-mining-process-essential-steps/.

Kumar, A. (2023). *Data Governance Goals Explained with Examples*. [online] Data Analytics. Available at: https://vitalflux.com/data-governance-goals-explained-with-examples/.