**Trends in the Data Analytics Industry**

Michael Adams

University of Maryland Global Campus

UCSP 615: Orientation to Graduate Studies

Dr. Pamela Czapla

June 6, 2022

**Three Trends: Data Analytics Industry**

Trends in the Data Analytics Industry change throughout the years. Data Analytics use cases in businesses, environment, social, and government fields remain relevant. Data Analytics and not instincts, drives most business decisions (Srinivasan 2021). Top three trends in the Data Analytics Industry are Machine learning (ML), data democratization, and Data Fabric. Although advancements in data technology also gave rise to threats in the industry, data security is a serious emerging issue in the data analytics industry.

**Discussion of Three Trends**

**Machine Learning**

Machine learning is a new trend to the field of data analytics. Machine learning is a new trend in the data analytics industry that can automate the data analytic process. Machine learning is a new methodology in computing where we do not write programs but collect data, the concept is to learn the algorithms for the task automatically from data and to process that data into useful knowledge (Alpaydin, 2016). In other words, Machine learning is a branch of Artificial Intelligence (AI) and Computer Science that uses data by learning from past data to create insights and predictions from data automatically. Traditional data analytics method can be very inaccurate and time consuming especially on projects with very large datasets. An example where machine learning is relevant in the data analytics industry is in combat operations, with vast amounts of data becoming available to intelligence analyst, new tools will help then sift and interpret it all (Knight, 2022). One of the most promising uses of machine learning applications for intelligence analyst is the use in unmanned aerial vehicles. In addition, machine learning is extremely relevant in data analysis because machine learning algorithms are going to continue evolving in the coming decades as its use cases for intelligence analyst continue to grow.

**Data Fabric**

According to IBM a data fabric is an architectural approach to simplify data access in an organization to facilitate self-service data consumption, there are four main components of a data fabric; insights and semantics, unified governance and compliance, intelligent integration, orchestration, and lifecycle – automates data lifecycle by infusing Artificial Intelligence (AI). Data fabric also described as data mesh can have many use cases, one of which can be found in the field of financial engineering. A data fabric is relevant in financial engineering because it can be used to find insights in data for institutional investors. Investors can use a data fabric to guide their decision and mitigate risk factors.

**Data Democratization**

Data democratization is a framework for assisting end users understand in exploring data and creating insights and data storytelling that are impactful (Leonardo et. al, 2021). In other words, data democratization in the data analytics industry is not only about making data more accessible, but to have access to tools to access data that can make gaining insights for data more accessible. For instance, leaders armed with an improved awareness of the strengths and weaknesses of their information can ask better informed decisions at the strategic level (Baker, 2021). Data democratization has been brewing as a trend and is becoming more prominent as institutions become more globalized. Lack of data constrains people; first, they are unable to use the readily available data themselves, and second, they are unable to tell when others are using data poorly or detect bias in data analysis (Ransbotham, 2019). Data democratization promotes collaboration across different disciplines. Conversely, as data becomes more widely available across different disciplines, an issue of data security begins to emerge. Data security becomes an issue when people exploit the data.

**The Case for Information Security, Data Analytics**

A prominent issue surrounding the field of data analysis is the issue of data security. Improper use of can lead to data exploitation. The execution of best practices is te key to a systems success in ensuring that key data sources are protected; this includes privacy. For instance, banking information should be confidential (Ludin, 2022). Data security is an emerging issue for analysts which in turn extends into other industries. One might argue that strict data policies prevent exploitation of the data while others might argue that limiting the use of data, restrictive policies will hinder progress. One might argue that enacting strict data sharing policies limits collaboration between researchers and scientists. This in turn limits the rate of discoveries and innovations. This issue of data theft and exploitation is prominent across disciplines. In the field of scientific research, many countries share genetic data, but some countries like China and other low-income countries are becoming more uncertain about sharing data sharing and therefore enacting strict policies on how the usage of their genetic data is carried out due to data exploitation. China began increasing control on their country’s genetic data from people in the country, including for scientific research. Government officials say limitation on the use of genetic data are intended to improve protection of this resource. They were established in reaction to companies exporting genetic data without permission, or an extreme situation of an explosive revelation by Chinese researcher He Jiankui that he had created the first babies with edited genomes in 2018 (Mallapaty, 2022). Strict data policies prevent incidents like the one where a Chinese researcher created genetic engineered babies with edited genomes.

Artificial Intelligence can have a positive impact on our daily lives. One might argue that the positives far outweigh the negative. Serval threats in machine learning have been addressed by experts. Training data poisoning, a well-designed backdoor, model stealing attack, and a carefully crafted disturbance. Training data poisoning can result in a decrease in the accuracy of a machine learning model or lead to other attack purposes. A well-designed backdoor in the training data used in a machine learning model can trigger dangerous consequences of a system. A carefully crafted disturbance in the test input of a machine learning train data set can make the model go wrong. A model stealing attack or model inversion attack can steal the model constraints to recover the sensitive training data which can lead to serious consequences to machine learning systems, especially in security and safety critical applications, such as self-driving vehicles, natural language processing, and banking systems (Xue et al, 2020). Furthermore, another approach to the issue of machine learning security is to filter data before feeding it to training models or to perform data encryption.

Encryption is the process of encoding information. End users can take measures to secure their personal data. Consumer data is often to referred to as the new currency because of how the insights one can gather about a person. Data security is a serious emerging issue in the data analytics industry that should be shed more light on.

**Conclusion**

Three main trends prominent in the field of data analytics are the use of machine learning, data democratization, and introduction of new data analytics methodology- data fabric. These trends will continue to have a lasting impact on the industry for years to come. Advancements in data technology also gave rise to threats in the industry. Data theft and exploitation are two of the most common emerging issue in the data analytics industry.

References

Baker, Posted byJeffrey E. “A Few Things Leaders Need to Know about Data and Analysis – Part 1.” *War Room - U.S. Army War College*, Army War College, 10 June 2021, <https://warroom.armywarcollege.edu/articles/data-analysis-pt1/>.

Christino, L., Ferreira, M. D., Jalilvand, A., & Paulovich, F. V. (2021). *Explainable Patterns: Going from Findings to Insights to Support Data Analytics Democratization*.

Ethem Alpaydin. (2016). Machine Learning : The New AI. The MIT Press.

Knight, W. (2022, April 4). As Russia plots its next move, an AI listens to the chatter. Wired. Retrieved May 31, 2022, from https://www.wired.com/story/russia-ukraine-war-ai-surveillance/

Lundin, L. L. (2022). Information security. *Salem Press Encyclopedia*.

Mallapaty, S. (2022, May 6). China expands control over genetic data used in scientific research. Nature News. Retrieved June 3, 2022, from https://www.nature.com/articles/d41586-022-01230-z

Ransbotham, Sam. (2021) “The New Data Republic: Not Quite a Democracy.” *MIT Sloan Management Review*, MIT Sloan Management Review, 30 June 2015, https://sloanreview.mit.edu/article/the-new-data-republic-not-quite-a-democracy/.