DATA SCIENCE ISSUES

Major Challenges faced by Data Scientists

According to acuvate.com organizations across the globe are looking to organize, process and unlock the value of the torrential amounts of data they generate and transform them into actionable and high value business insights. Hence, hiring data scientists – highly skilled professional data science experts, has become super critical. Today, there is virtually no business function that cannot benefit from them. In fact, the Harvard Business Review has labeled data science as the "sexiest" career of the 21st century.

However, no career is without its own challenges, and being a data scientist, despite its "sexiness" is no exception. According to the Financial Times, many organizations are failing to make the best use of their data scientists by being unable to provide them with the necessary raw materials to drive results. In fact, according to a Stack Overflow survey, 13.2% of the data scientists are looking to jump ship in search of greener pastures – second only to machine learning specialists. Having helped several data scientists solve their data problems, we share some of their common challenges and how they can overcome them.

1. Data Preparation

Data scientists spend nearly 80% of their time cleaning and preparing data to improve its quality – i.e., make it accurate and consistent, before utilizing it for analysis. However, 57% of them consider it as the worst part of their jobs, labeling it as time-consuming and highly mundane. They are required to go through terabytes of data, across multiple formats, sources, functions, and platforms, on a day-to-day basis, whilst keeping a log of their activities to prevent duplication.

One way to solve this challenge is by adopting emerging AI-enabled data science technologies like Augmented Analytics and Auto feature engineering. Augmented Analytics automates manual data cleansing and preparation tasks and enables data scientists to be more productive.
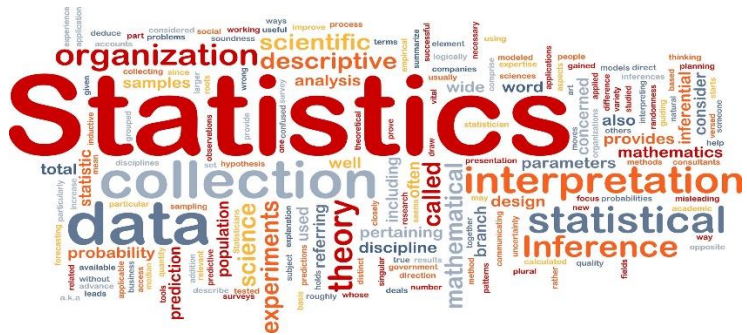
2) Multiple Data Sources

As organizations continue to utilize different types of apps and tools and generate different formats of data, there will be more data sources that the data scientists need to access to produce meaningful decisions. This process requires manual entry of data and time-consuming data searching, which leads to errors and repetitions, and eventually, poor decisions.

Organizations need a centralized platform integrated with multiple data sources to instantly access information from multiple sources. Data in this centralized platform can be aggregated and controlled effectively and in real-time, improving its utilization and saving huge amounts of time and efforts of the data scientists.

3) Data Security

As organizations transition into cloud data management, cyber-attacks have become increasingly common. This has caused two major problems –

a. Confidential data becoming vulnerable
b. As a response to repeated cyber-attacks, regulatory standards have evolved which have extended the data consent and utilization processes adding to the frustration of the data scientists.

Organizations should utilize advanced machine learning enabled security platforms and instill additional security checks to safeguard their data. At the same time, they must ensure strict adherence to the data protection norms to avoid time-consuming audits and expensive fines.

4) Understanding the Business Problem

Before performing data analysis and building solutions, data scientists must first thoroughly understand the business problem. Most data scientists follow a mechanical approach to do this and get started with analyzing data sets without clearly defining the business problem and objective.

Therefore, data scientists must follow a proper workflow before starting any analysis. The workflow must be built after collaborating with the business

stakeholders and consist of well-defined checklists to improve understanding and problem identification.

5) Effective Communication with Non-Technical Stakeholders

It is imperative for the data scientists to communicate effectively with business executives who may not understand the complexities and the technical jargon of their work. If the executive, stakeholder, or the client cannot understand their models, then their solutions will, most likely, not be executed.

This is something that data scientists can practice. They can adopt concepts like "data storytelling" to give a structured approach to their communication and a powerful narrative to their analysis and visualizations.

6) Collaboration with Data Engineers

Organizations usually have data scientists and data engineers working on the same projects. This means there must be effective communication across them to ensure the best output. However, the two usually have different priorities and workflows, which causes misunderstanding and stifles knowledge sharing.

Management should take active steps to enhance collaboration between data scientists and data engineers. It can foster open communication by setting up a common coding language and a real-time collaboration tool. Moreover, appointing a Chief Data Officer to oversee both the departments has also proven to have improved collaboration between the two teams.

7) Misconceptions about the Role

In big organizations, a data scientist is expected to be a jack of all trades – they are required to clean data, retrieve data, build models, and conduct analysis. However, this is a big ask for any data scientist. For a data science team to function effectively, tasks need to be distributed among individuals pertaining to data visualization, data preparation, model building and so on.

It is critical for data scientists to have a clear understanding of their roles and responsibilities before they start working with any organization.

8) Undefined KPIs and Metrics

The lack of understanding of data science among management teams leads to unrealistic expectations on the data scientist, which affects their performance. Data scientists are expected to produce a silver bullet and solve all the business problems. This is very counterproductive.

Therefore, every business should have:

a. Well-defined metrics to measure the accuracy of analysis generated by the data scientists
b. Proper business KPIs to analyze the business impact generated by the analysis

9. The amount of data being collected

Clearrisk.com explained some challenges too that with today's data-driven organizations and the introduction of big data, risk managers and other employees are often overwhelmed with the amount of data that is collected. An organization may receive information on every incident and interaction that takes place on a daily basis, leaving analysts with thousands of interlocking data sets.

There is a need for a data system that automatically collects and organizes information. Manually performing this process is far too time-consuming and unnecessary in today's environment. An automated system will allow employees to use the time spent processing data to act on it instead.

10. Collecting meaningful and real-time data

With so much data available, it's difficult to dig down and access the insights that are needed most. When employees are overwhelmed, they may not fully analyze data or only focus on the measures that are easiest to collect instead of those that truly add value. In addition, if an employee has to manually sift through data, it can be impossible to gain real-time insights on what is currently happening. Outdated data can have significant negative impacts on decision-making.

A data system that collects, organizes and automatically alerts users of trends will help solve this issue. Employees can input their goals and easily create a report that provides the answers to their most important questions. With real-time reports and alerts, decision-makers can be confident they are basing any choices on complete and accurate information.

11. Visual representation of data

To be understood and impactful, data often needs to be visually presented in graphs or charts. While these tools are incredibly useful, it's difficult to build them manually. Taking the time to pull information from multiple areas and put it into a reporting tool is frustrating and time-consuming.

Strong data systems enable report building at the click of a button. Employees and decision-makers will have access to the real-time information they need in an appealing and educational format.

12. Inaccessible data
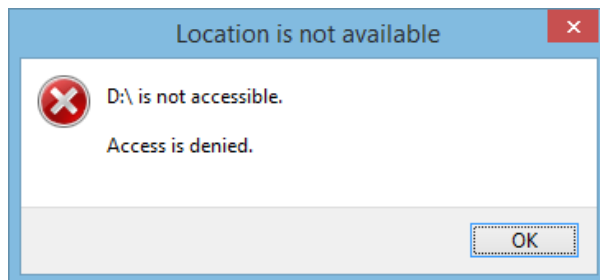
Moving data into one centralized system has little impact if it is not easily accessible to the people that need it. Decision-makers and risk managers need access to all of an organization's data for insights on what is happening at any given moment, even if they are working off-site. Accessing information should be the easiest part of data analytics.

An effective database will eliminate any accessibility issues. Authorized employees will be able to securely view or edit data from anywhere, illustrating organizational changes and enabling high-speed decision making.

13. Poor quality data

Nothing is more harmful to data analytics than inaccurate data. Without good input, output will be unreliable. A key cause of inaccurate data is manual errors made during data entry. This can lead to significant negative consequences if the analysis is used to influence decisions. Another issue is asymmetrical data: when information in one system does not reflect the changes made in another system, leaving it outdated.

A centralized system eliminates these issues. Data can be input automatically with mandatory or drop-down fields, leaving little room for human error. System integrations ensure that a change in one area is instantly reflected across the board.

14. Pressure from the top

As risk management becomes more popular in organizations, CFOs and other executives demand more results from risk managers. They expect higher returns and a large number of reports on all kinds of data.

With a comprehensive analysis system, risk managers can go above and beyond expectations and easily deliver any desired analysis. They'll also have more time to act on insights and further the value of the department to the organization.

15. Lack of support

Risk managers will be powerless in many pursuits if executives don't give them the ability to act. Other employees play a key role as well: if they do not submit data for analysis or their systems are inaccessible to the risk manager, it will be hard to create any actionable information.

Emphasize the value of risk management and analysis to all aspects of the organization to get past this challenge. Once other members of the team understand the benefits, they're more likely to cooperate. Implementing change can be difficult, but using a centralized data analysis system allows risk managers to easily communicate results and effectively achieve buy-in from multiple stakeholders.

16. Budget

Another challenge risk managers regularly face is budget. Risk is often a small department, so it can be difficult to get approval for significant purchases such as an analytics system.

Risk managers can secure budget for data analytics by measuring the return on investment of a system and making a strong business case for the benefits it will achieve. For more information on gaining support for a risk management software system, check out our blog post here.

17. Shortage of Skills

Some organizations struggle with analysis due to a lack of talent. This is especially true in those without formal risk departments. Employees may not have the knowledge or capability to run in-depth data analysis.