

Data Mining Actual Problems

Ihor Volokhovych, MMAI-1

In the last two years, over 90% of the world's data was created, and with 2.5 quintillion bytes of data generated daily, it is clear that the future is filled with more data, which can also mean more data problems.

Whilst it is clear that companies can benefit from this growth in data, executives must be cautious and aware of the challenges they will need to overcome, particularly around:

- Collecting, storing, sharing and securing data.
- Creating and utilizing meaningful insights from their data.

Luckily, there are pragmatic solutions that companies can take to overcome their data problems and thrive in the data-driven economy.

Lack of Understanding

Companies can leverage data to boost performance in many areas. Some of the best use cases for data are to: decrease expenses, create innovation, launch new products, grow the bottom line, and increase efficiency, to name a few. Despite the benefits, companies have been slow to adopt data technology or put a plan in place for how to create a data-centric culture. In fact, according to a Gartner study, out of 196 companies surveyed, 91% say they have yet to reach a “transformational” level of maturity in their data and analytics.

Solution: One way to combat slow adoption is to take a top-down approach for introducing and training your organization on data usage and procedures. If your in-house team doesn't have the resources to take this on, consider bringing in IT specialists or consultants and holding workshops to educate your organization.

High Cost of Data Solutions

After understanding how your business will benefit most from implementing data solutions, you're likely to find that buying and maintaining the necessary

components can be expensive. Along with hardware like servers and storage to software, there also comes the cost of human resources and time.

Solution: To make the most informed decision for what kind of data solution will provide the most ROI, first consider how and why you want to use data. Then, align your reasoning with your business goals, conduct research for available solutions, and implement a strategic plan to incorporate it into your organization.

Too Many Choices

According to psychologist Barry Schwartz, less really can be more. Coined as the “paradox of choice,” Schwartz explains how option overload can cause inaction on behalf of a buyer. Instead, by limiting a consumer’s choices, anxiety and stress can be lessened. In the world of data and data tools, the options are almost as widespread as the data itself, so it is understandably overwhelming when deciding the solution that’s right for your business, especially when it will likely affect all departments and hopefully be a long-term strategy.

Solution: Like understanding data, a good solution is to leverage the experience of your in-house expert, perhaps a CTO. If that’s not an option, hire a consultancy firm to assist in the decision-making process. Use the internet and forums to source valuable information and ask questions.

Complex Systems for Managing Data

Moving from a legacy data management system and integrating a new solution comes as a challenge in itself. Furthermore, with data coming from multiple sources, and IT teams creating their own data while managing data, systems can become complex quickly.

Solution: Find a solution with a single command center, implement automation whenever possible, and ensure that it can be remotely accessed 24/7.

Security Gaps

The importance of data security cannot go unnoticed. However, as solutions are being implemented, it's not always easy to focus on data security with many moving pieces. Data also needs to be stored properly, which starts with encryption and constant backups.

Solution: You can take a few low effort steps to dramatically increase the security of your data, like: automate security updates, automate backups, install operating system updates (which often include better security), use firewalls, etc.

Low Quality and Inaccurate Data

Having data is only useful when it's accurate. Low quality data not only serves no purpose, but it also uses unnecessary storage and can harm the ability to gather insights from clean data.

A few ways that data can be considered low quality is:

- Inconsistent formatting (which will take time to correct and can happen when the same elements are spelled differently like "US" versus "U.S."),
- Missing data (i.e. a first name or email address is missing from a database of contacts),
- Inaccurate data (i.e. it's just not the right information or the data has not been updated).
- Duplicate data (i.e. the data is being double counted)

If data is not maintained or recorded properly, it's just like not having the data in the first place.

Solution: Begin by defining the necessary data you want to collect (again, align the information needed to the business goal). Cleanse data regularly and when it is collected from different sources, organize and normalize it before uploading it into any tool for analysis. Once you have your data uniform and cleansed, you can segment it for better analysis.

Compliance Hurdles

When collecting information, security and government regulations come into play. With the somewhat recent introduction of the General Data Protection Regulation (GDPR), it's even more important to understand the necessary requirements for data collection and protection, as well as the implications of failing to adhere. Companies have to be compliant and careful in how they use data to segment customers for example deciding which customer to prioritize or focus on. This means that the data must: be a representative sample of consumers, algorithms must prioritize fairness, there is an understanding of inherent bias in data, and Big Data outcomes have to be checked against traditionally applied statistical practices.

Solution: The only solution to adhere to compliance and regulation is to be informed and well-educated on the topic. There's no way around it other than learning because in this case, ignorance is most certainly not bliss as it carries both financial and reputational risk to your business. If you are unsure of any regulations or compliance you should consult expert legal and accounting firms specializing in those rules.

Using Data for Meaning

You may have the data. It's clean, accurate and organized. But, how do you use it to provide valuable insights to improve your business? Many organizations are turning to robust data analysis tools that can help assess the big picture, as well as break down the data into meaningful bits of information that can then be transformed into actionable outcomes.

Solution: Whether this means having a consistent reporting structure or a dedicated analytics team, be sure to turn your data into measurable outcomes. This means taking data and transforming into actions for the business to take in an effort to produce wins for the company.

Keeping Up with Growth in Data

Like scaling a company, growing with data is a challenge. You want to make sure that you can scale your solution with the company's growth so that the costs and quality don't decrease as it expands.

Solution: This is achievable by creating projections from the get go of introducing data and data management tools. Make sure that you select a robust data solution and know in advance that it can handle the capabilities you may need down the line. Another option is to rely on support systems and internal teams to manage aspects of growth. For example, you can define milestones for your team to be aware of so that only when you reach them will you consider moving to a more sophisticated system.

Accessibility

Sometimes, companies silo data to one person or one department. Not only does this put immense responsibility on a select few, but it also creates a lack of accessibility throughout the organization in departments where the data can be of use to provide a positive impact. Data silos directly inhibit the benefits of collecting data in the first place.

Solution: It sounds simple, but it's not done enough - integrate your data. Set clear expectations and create a unified system that can handle each department's needs. If it's not through finding a single integrated system, consider using APIs so that data is accessible in one, centralized location.

Pace of Technology

Inventor, author and futurist Ray Kurzweil put it best when he defined the accelerating rate of change of technology. Each subsequent technological advancement builds more quickly upon the last because they evolve at each step to become more efficient and therefore can better inform what comes next. For example, just consider how rapidly cloud computing and artificial intelligence are improving.

With the rapid advancement of technology and systems, you don't want your data tools to become outdated, especially when you're investing time, energy and human resources into them.

Solution: While you can't stop progression, you can prepare for it. This begins with staying informed of information technology and its new features, products and threats.

Lack of Skilled Workers

While the technological demand is high and artificial intelligence and data analysis tools are innovating swiftly, the lack of skilled workers is causing a bottleneck for many companies. The number of new, skilled graduates isn't keeping pace with technology, and in turn, companies are asking staff to supplement this shortfall by working multiple roles.

Solution: If the solution doesn't exist naturally, try to create it. While you can't control how many data scientists and data analysts graduate each year, you can leverage your current workforce and provide training to instill and teach the skills you need them to have. You can also look for more powerful data tools that make the analysis work less complex, which open up recruitment to a broader pool of less specialized analysts.

Data Integration

Data integration consists of taking data from various sources and combining it to create valuable and usable information.

Solution: There are a few ways to go about integrating data, including the following approaches:

- Consolidation: Combining the data from various sources in one consolidated data store
- Propagation: Leveraging applications to copy data from one location to another
- Federation: Using a virtual database to create a model to match data from different systems

- Virtualisation: Viewing data in one location, but where the data is still stored separately

In today's data-driven world, the management of your data is essential and must not be ignored. You need to be proactive in understanding and implementing data solutions that align with your business goals. By doing so, you can effectively mitigate any big data problems.

Some organizations will need to assemble a dedicated team of experts to manage their data. That being said, modern data tools offer a simple way to augment and leverage existing staff to be able to turn data into insights for the business.