Data Science vs Business Intelligence: same but completely different

What to expect from your first Data Science project? A guide for businesses. Part 2

Maxim Scherbak [Follow]
Mar 25, 2019 · 8 min read ★

Image by analogicus from Pixabay

This is part 2 of this series. See part 1 here

Unlike big tech companies, businesses, in general, are only dipping their toes into Data Science and AI. As always with early adoption, it doesn't go easy —

most of the projects do not advance beyond Proof of Concept phase, which is considered a fail from a business perspective. There is no published figure on how many Data Science projects fail, but I'm sure it is even higher than the disappointing 85% figure for Big Data projects.

Why is the failure rate so high? Mainly, because Data Science is new. It brings new types of collaboration, requirements, and culture into established corporate environments. Complications caused by these novelties are often hard to see in the beginning. But as they start unravelling, they often become too difficult at some point. Unfortunately, these complications are not usually listed in all sorts of marketing pitches selling AI and Data Science to businesses.

. . .

Business Intelligence is an umbrella term that describes concepts and methods to improve business decision making by using fact-based support systems [1]. Modern Business intelligence is not just business reporting. It is a mature system that provides interactive dashboards, what-if planning, mobile analytics, etc. It also includes large back-end parts for maintaining control and governance around reporting.

Since BI is an umbrella term, it can be different from company to company. For some, it can be just a basic Key Performance Indicator (KPI) reporting with all supporting infrastructure, other companies may use advanced predicting methods based on statistical models and advanced tools. But regardless of methods or tools used — they provide facts for decision making to business stakeholders, according to their requirements.

But hey! Isn't Data Science doing the same? Yes it is.

From a Business Process standpoint, there is not much difference between Data Science and Business Intelligence — they both support business decision making based on data facts. Probably this is why it is often assumed in businesses starting their first Data Science or AI projects, that

Data Science is the same old Business Intelligence that works much more cleverly. From this assumption, it follows that a Data Science project can be done on the top of existing BI infrastructure and processes.

This is when problems begin: it turns out that a Data Scientist wants data out of the system as a CSV file, but company security policy will not allow that; Data Scientists are building their models using weird libraries and software which infrastructure teams do not and will not support in production, and so on.

Any first Data Science project in any company will start generating new challenging requirements to existing teams right from the start. How come? Why this is happening if Data Science does the same as BI does? Why those weird requirements?

> *In reality, the difference between BI and Data Science is so fundamental, that it makes everything different: expectations, project methodologies, people involved, etc. But one has to take a different perspective to see it.*

The difference is in the type of questions that they address: BI provides new values of previously known things, using some formula that is available. Data Science works with the unknown (see the first part of this series), answering data questions that nobody have answered before and, therefore, without formula in hand.

Top highlight

In BI, business comes to a BI developer with a formula or a method of calculating a report or a KPI, that business owns. This means that business has designed the BI method and that they understand and are comfortable using it.

In Data Science it is quite different: business comes with their actual data and some question that has never been answered before. It is now up to a Data Scientist to test multiple approaches and select the best one, balancing between accuracy, simplicity, usability, and capabilities of a production

platform. Once the model is selected and agreed with the business, it becomes a known method for answering the question, it becomes a subject of Data Analytics rather than Data Science.

You may notice that above statement about BI is debatable — it does not deal with completely known things — it may have a formula or a method, but it calculates unknown KPI values or even makes predictions using approved methodology. To explain this duality, I'm using a nice concept of known unknowns and unknown unknowns, that was popularised by US Secretary of Defence Donald Rumsfeld back in 2002 in his famous answer about lack of evidence linking the government of Iraq with the supply of WMD to terrorists.

Using this concept, I could now formulate the difference much shorter:

> *BI deals with known unknowns, while*
> *Data Science deals with unknown unknowns.*

This is the biggest and fundamental difference between them. At first, it may seem a pure formalism, focusing on a difference that is not that significant, but it will change once you start thinking about the consequences.

Firstly, dealing with unknown unknowns, Data Science cannot guarantee success in the very beginning of a project, predict what the solution would look like and how difficult it will be to implement. This means it is much

more difficult to build a business case and plan a project. If you continue to unravel these implications, you may come up with something like this:

All of these differences actually arise from that distinction, which seemed insignificant at first. Indeed, when the solution is unknown from the start, a

Data Scientist will be using trial and error approach. In these circumstances, it would be wise to use tools and methods that allow quick turnaround of ideas, so that each new trial would not require too much time to prepare: new data should be readily available if needed, new software and libraries implementing the next method to try should be easily available to install and download, infrastructure must be ready to support new software or frameworks and so on.

> *That is why a typical Data Scientist's toolbox and practices are built for flexibility and agility. Programming languages, open source libraries, microservices, containers, APIs, and extreme agile are all helping Data Scientist to skim through ideas and find solutions quickly.*

All these tools, however, work best in an open, agile and fluid environment, where the use of these tools is not limited by some external factors. This type of environment can usually be found in tech companies and start-ups.

> *A typical corporate environment is very different — it is built for control and reliability, which are delivered through strict process rules, shared responsibilities, multilevel decision making, etc.*
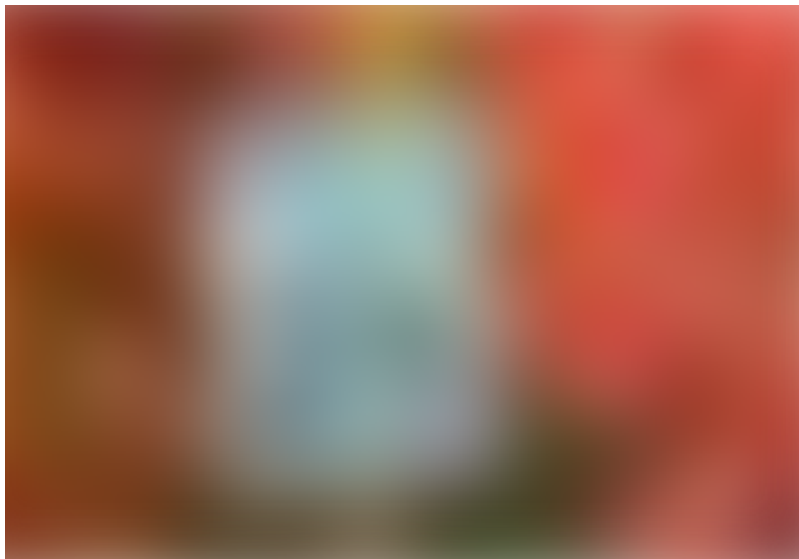

Photo by Joshua Hoehne on Unsplash

This means that IT systems of most big non-tech companies are very regulated and slow to implement changes. Any Data Science project in these companies will face multiple distracting hurdles: a lengthy process of getting data out of business systems for analysis, inability to operationalize the solution because current IT architecture cannot support containers and microservices,

etc.

People maintaining corporate systems have very different priorities and mindset too. They will not necessarily be excited about making changes to their systems or adding new, they might be worried about security compliance when signing off access to company data, and so on. Business end users might not be very excited about introducing AI or Data Science to their roles too. They might be worried about lack of knowledge in the new field or about the security of their jobs because of a false perception of the AI threat they got from media.

Nevertheless, all those people are critical for any Data Science project. Project leaders and sponsors have to consider those issues to avoid "clash of cultures", that can ruin any project.

> *Project leaders have to provide strong support to Data Scientists, who would otherwise find themselves outnumbered and disadvantaged in a hostile over regulated environment. They also have to communicate why a Data Science project is needed for the company and to reassure people that it will not impact their jobs security. Project sponsors will also need to be capable of pushing exceptional requests through all layers of corporate structure quickly if needed.*

In contrast, Business Intelligence is already a well-established part of a typical corporate landscape and BI project is mostly free from those issues by definition. BI functionality is usually provided by a single or very few platforms that are already incorporated into IT architecture and processes. Business users are largely familiar and confident with it. BI projects will deal with known unknowns, which means there is a method of finding those unknowns and therefore the project can be well planned in advance. There is not much trial and error in BI. On top of that, a company would usually have a good experience and track record of successful BI projects and would have good project expertise available.

Summary

From a business perspective, both Data Science and Business intelligence play the same role in the Business Process — they both provide fact-based insights to support business decisions. But they are fundamentally different from another perspective, which makes everything different: expectations, methods, tools used, etc.

The difference is in the type of questions they are addressing: BI works with known unknowns, when a known formula is used to calculate a new value of a known KPI, while Data Science works with unknown unknowns, answering data questions that no one has answered before. This little difference in definition means a lot. Without a formula or a method given, Data Scientists use a trial and error approach. In these circumstances, generally speaking, Data Science cannot guarantee success before a project begins, it cannot predict how many steps would be needed to find a solution and what it will look like.

In order to find solutions as quickly as possible, Data Science employs tools and methods optimized for speed: programming languages, libraries, Docker containers, microservices architecture, etc. It is very different from a typical corporate environment where IT systems are built for control and reliability. This difference alone creates many difficulties for the first Data Science projects in an established company.

But that's not all! There is another problem lurking around the corner — the use of *Machine learning*. Introducing ML into a business environment can be a big cultural shock for business analysts, who's life is designing and maintaining business rules. For solutions that use Machine Learning, these rules will no longer be required! Who would like a change like that? In the next short part, I will touch on the tectonic shift that Machine Learning brings to an established corporate culture.

. . .

References:

[1] D. J. Power, "A Brief History of Decision Support Systems, version 4.0". (10 March 2007), DSSResources.COM.

. . .

You can find me on LinkedIn, Twitter, Facebook

Data Science   Business Intelligence   Project Management   Data Analytics   Artificial Intelligence

Discover Medium

Welcome to a place where words matter. On Medium, smart voices and original ideas take center stage - with no ads in sight. Watch

Make Medium yours

Follow all the topics you care about, and we'll deliver the best stories for you to your homepage and inbox. Explore

Become a member

Get unlimited access to the best stories on Medium — and support writers while you're at it. Just $5/month. Upgrade

This is your last free story this month. Sign up and get an extra one for free.   ✕