The Ethics Of Data Analytics

BUSINESSES MUST FOLLOW THEIR OWN POLICIES & CODES IN ADDITION TO THE LAW

KEY POINTS

• Ethical data usage concerns range from privacy issues and making poor business decisions to potentially endangering lives.

• Following the law is a good start, but maintaining legality won’t necessarily prevent you from being unethical.

• It’s important to not only create a strong internal ethics policy, but to also build a culture of honesty and loyalty.

• Data scientists should consider following a professional code of conduct to help them avoid unethical practices and properly guide the business.

THE GROWTH OF the big data analytics market is exciting for businesses because it promises to bring new insights and potential competitive advantages to the forefront, but it can also introduce certain ethical challenges. In fact, a recent Gartner study showed that by 2018, half of all business ethics violations will happen as a result of improper big data analytics usage. For that reason, companies need to be very careful that as they embrace the idea of using analytics to make business decisions, they don’t put themselves or their customers in jeopardy.

Potentially Harmful Analytics

Using data, especially the personal data of consumers, for analytics and making business decisions can quickly become unethical or immoral if proper care isn’t taken regarding the types of information gathered and used. If a business gets too personal with customer information, it may be viewed as an invasion of privacy and could result in the loss of a customer. And even though the U.S. doesn’t have the strongest laws around privacy, the European Union and other countries “have more stringent privacy laws,” says Michael Walker, co-founder and president of the Data Science Association. This means, depending on where you’re practicing analytics, that it can be easy to shift from unethical to illegal.

Although privacy is a major concern for some consumers, it isn’t necessarily the most significant risk associated with big data analytics. In fact, according to Walker, perhaps the biggest issue has to do with improper use of data, and making faulty assumptions that lead to business leaders making ill-informed decisions. However, this works both ways. On one hand, a data scientist may make the mistake and give the executive team bad information, which will lead to a less than sound business
strategy. On the other hand, there may be a strategy that the executive wants to implement, but they need justification, so they go to the data scientist and tell them to find information that supports the initiative and essentially ignore any detracting data.

The problem with poor decision-making based on bad data is that it sometimes not only affects the business, but it can also be passed along to the public. One example that Walker uses is that of dietary guidelines established in the 1950s. The U.S. government warned the public about the dangers of high cholesterol, high-fat diets and instead recommended a high carbohydrate diet. However, Walker says these reports were based on “weak correlations” and we’ve only recently started to fix those mistakes.

“Now they’re finding out that you can eat a high-fat diet . . . ,” says Walker. “It’s the high-carbohydrate diet that really creates a lot of ill effects, and you’re better off eating more of a Mediterranean diet or other types of diets that are lower in carbohydrates. I think a side effect from that is that people became obese. . . . and I think that creates a lot of the environment for heart attacks and other types of bad health effects. That all stems from negligent science, and they’re just now coming around to correct it.”

There are countless other examples for data, in general, being used in unethical ways from “judges allowing junk science into the courtroom that can skew a lot of legal cases” to quantitative analysts (aka quants) on Wall Street building flawed predictive analytic models that led to the housing market crash and economic crisis in 2008, Walker says. However, some of these examples cross the line from unethical into illegal territory, which is a distinction businesses need to learn to identify and take into consideration.

Illegal, Unethical & Everything In Between

The problem with thinking in legal terms with data analytics is that because the U.S. doesn’t have stringent privacy laws, simply following the law isn’t enough to prevent the unethical use of information. In fact, there was a recent case about the National Security Agency collecting data from phones that was first deemed illegal by one court and then deemed legal by another court later on, which shows how the law can be a major gray area. There are obvious situations where gathering data is illegal, such as hacking into devices and stealing information directly from them, but when it comes to business analytics and big data, it’s better to not only rely on the law, but also come up with your own reasons for why you should or shouldn’t use data in a certain way.

Frank Buytendijk, research vice president and distinguished analyst at Gartner, points out four ways to look at digital ethics, both generally and specifically as it relates to data usage. According to Buytendijk, the lowest level, interestingly enough, is the legal one. The reason for that is that there are strategies that may keep you out of legal trouble but that still aren’t necessarily ethical. For that reason, you have to consider the other three levels of data ethics in order to make a more well-informed decision. That starts with the second layer after legal, which is deciding whether or not to do something because of risk. “You’re afraid of offending someone or you’re afraid that there are negative consequences for you if you do,” says Buytendijk. “It’s acting out of some kind of fear. That’s also in the lower stages of moral development.”

The next highest level has to do with using or not using data out of differentiation. In this instance, a company may look at its competitors that do collect and sell information from consumers, and create a specific policy where they promise they won’t sell consumer data or even gather it at all. “With all of the personalization that’s going on in health insurance at the moment, I could totally understand if there was a large health insurance company that says, ‘with us, we don’t even do big data,’” says Buytendijk. “That could be a very competitive positioning.”

The highest level for deciding whether or not to use data in certain ways has to do with your organization’s values. Although Buytendijk admits that there is no consensus for right or wrong, there is “a hierarchy of motivation” that should aid in the decision-making process. “The hierarchy of motivations we just walked through from compliance to values is a very good measure to figure out whether you’re relatively mature or relatively immature in your discussions on ethics,” he says.

Building An Internal Data Ethics Policy & Strong Company Culture

Buytendijk says because there is no universally accepted common ground for what is right vs. wrong regarding
data analytics, and because there are so many different schools of thought on the matter, the best thing you can do when building an internal ethics policy is to listen to all opinions and try to create something that is “defendable.”

Instead of avoiding policy implementation because it’s too difficult, simply do your best to build a policy that you believe in and keep in mind that someone somewhere will undoubtedly disagree with it. And in the event of a disagreement, you can at least trace good intentions, address unintended consequences as they arise, and make changes to the policy to lessen the chance of future disagreements.

In fact, Buytendijk says, there are only three ideas that he would consider to be nearly universally accepted (meaning he couldn’t find a situation where he wouldn’t recommend them). The first is understanding that there is no universal truth; the second is building that defendable policy; and the third is performing the mirror test when making a decision and building an ethics policy. “If you look at yourself in the mirror and it doesn’t feel right, then it probably isn’t right,” says Buytendijk. “How would I feel being treated like this as a citizen, consumer, or just as a person? How would my mother, father, grandparents, or children feel? Make it personal.”

But in addition to creating policies, you also need to have a strong company culture to back them up. “If there’s a bad organizational culture, I don’t know that a code of ethics is going to help at all,” says Walker. “Some of these investment banks and hedge fund managers in the financial sector are just bad people and have bad cultures. A code of ethics isn’t going to matter because they’re going to use [data scientists] to justify whatever it is they want to do. The number one thing that an organization has to have is a good solid culture where you respect people’s work and a culture of respect and honesty.”

Choose A Code & Follow It

While it’s important for businesses to have strong policies and cultures in place to help ensure proper ethics, it’s also important for data scientists and analysts to have their own ethical codes so they can steer the business in the right direction. The Data Science Association, for example, has a Data Science Code of Professional Conduct that offers an outline for data scientist to follow in their day-to-day lives. And while there are quite a few concepts covered in the code, there are three major ideas that Walker likes to highlight: competence, client confidentiality, and loyalty to the scientific method.

**Competence.** Walker says there are quite a few “garden variety analysts” who like to call themselves data scientists in order to get a larger salary but aren’t fully trained in the scientific method. He says that his team gets calls from businesses all the time saying that their internal data science team has been leading them astray and asking for help. Unfortunately, there are no standards or best practices in place to establish a minimum level of competence required to be considered a true data scientist, but he says it’s being worked on and is maybe a couple of years away from reality.

**Client confidentiality.** Walker says that businesses are often concerned about hiring expert data scientists who are less likely to make mistakes and false assumptions because data scientists are in such high demand and could get snatched up by a rival company. The worry is that if that were to happen, the data scientist could potentially share sensitive corporate information with a competitor, which is of course highly unethical and possibly illegal in some instances. “As an external data scientist, I have a duty of loyalty to my client,” says Walker. “If my team does work for that client, they own it. I am not allowed by my ethical code, and sometimes by law, to take that and give it to one of my other clients. That’s extremely important,” he adds, “and one of the dirty little secrets is that it happens all the time.”

**Loyalty to the scientific method.** “You’re using the scientific method in you predictive or prescriptive models, and you’re using the appropriate statistical power laws,” he says. “When you’re dealing with big data, and there’s a lot of hype around big data, you’re going to get a lot of patterns and correlations. That’s where a lot of these data analytics teams go wrong. They start seeing these patterns and correlations and say, ‘Wow, look at all of this.’ The problem is that there’s very little causality. You really have to find causality to find high value. Not that patterns or correlations don’t have value—they do, and they can be very useful—but you can’t start making decisions or be certain that this result is realistic and high value until you have causality.”

---

“**Choose A Code & Follow It**

Michael Walker
Co-Founder & President
Data Science Association

“There are going to be bad actors out there that want to misuse it, but we need to try to isolate them and create a culture where people are going to use it for the betterment of all. We want data science to be used for the majority of people, not just the minority.”