2204 1st Avenue South, Suite 101
Birmingham, AL 35233

6 TIPS FOR BUILDING YOUR FIRST DATA SCIENCE STRATEGY

Exponential advances in computing power and internet bandwidth in recent years have enabled the use of AI and advanced machine learning algorithms to solve problems previously considered unachievable – both in the scientific and business worlds. The data science revolution is well underway, and companies that fail to develop a strategy to maximize the value of their data risk being left behind by data-savvy competitors. So, how do you build your first data science strategy?

At StrategyWise, we've spent years helping some of America's largest companies build their own data strategies, and we've compiled our "Greatest Hits" observations below. The tips we provide here can serve as a framework to keep in mind as you work on building out your initial data science strategy, and they cover what we regard as the major criteria companies should keep in mind as they utilize modern data science techniques to maximize the value of their data. While every company's strategy should, of course, be tailored to match its specific corporate objectives, the tips outlined below will offer general guidance that we've seen prove useful time and again.

TIP 1 - RESEARCH YOUR INDUSTRY TO SEE WHAT YOUR COMPETITORS AND PEERS IN SIMILAR INDUSTRIES ARE DOING.

When it comes to data science, there's no need to reinvent the wheel. While new uses for the massive data unlocked by modern computing trends are still being discovered, many reports of successful data science tactics across a wide spectrum of industries and use cases are available. By scouring the internet and using your company's industry contacts, you should be able to gain an understanding of the approaches your competitors are taking to use data science and machine learning.

While your company's strategy for maximizing the value of its data may ultimately turn out to differ markedly from the strategies pursued by your competitors, you can still gain valuable insights from examining the methods they are using. This effort will also help you ensure thoroughness in "mapping the universe" of data's potential to change your company, leaving no stone unturned.

TIP 2 - BUILD A COMPREHENSIVE STRATEGY THAT STARTS BY TAKING INTO ACCOUNT YOUR COMPANY'S OPERATIONS FROM TOP TO BOTTOM

When helping companies optimize their use of data, we first look at understanding a company's strategy – not from a technology-focused view, but from an enterprise-wide standpoint. To accomplish this, we ask the following questions:

- What is the company's go-to-market strategy?
- How are they competing in the marketplace?
- What is their value proposition?

2204 1st Avenue South, Suite 101
Birmingham, AL 35233

- What are their core competencies and strategic competitive advantages?

Once you've answered these questions, the next step is to consider how you can leverage technology such as AI or advanced analytics to help you do a better job of positioning your company or division to achieve these strategic aims.

This approach creates a better lens for focusing on which projects are likely to deliver real results and which are just costly distractions.

"While AI applications have unlocked significant value for organizations around the world, we are currently in the midst of a machine learning hype bubble. This makes it more critical than ever to focus on projects that drive core value and measurable return on investment, not just excitement," says Joshua Jones, CEO of StrategyWise. *"In other words, don't let the technology tail wag the strategy dog."*

TIP 3 - BUILD A FUNNEL BY ROI

Once you've decided on a strategy and identified the tools or technology you want to use to execute that strategy, the next step is to filter projects by expected ROI. By ROI, we don't necessarily mean just from a financial standpoint (although that's the best place to start). In the broad sense, the return you realize on your data science investment could encompass returns in areas such as improved safety, development of key capabilities, and so forth. At StrategyWise, we often build hybrid ROI models that quantify traditionally qualitative objectives; this allows us to better evaluate which projects receive priority.

As part of this process, list all of the possible data science projects that support your company strategy that could be undertaken and estimate their results. Wherever possible, force qualitative concepts into quantitative buckets. Increased safety, while important in its own right and as a starting point, has measurable financial impact in insurance rate multipliers, time off work, etc.

Building such a funnel enables you to create a map to help you evaluate the costs and benefits of particular projects and what they are intended to help the company achieve. By comparing each project outcome to its cost, some projects will quickly rise to the top (and others will drop).

TIP 4 - CONSIDER DATA AVAILABILITY AND ACCESSIBILITY

After projects have been filtered through the ROI lens, we recommend looking at the data you have available and gauging its accessibility. In many cases, companies think they have data available but in fact there are issues with data corruption or data hygiene that make those data difficult or impossible to use.

2204 1st Avenue South, Suite 101
Birmingham, AL 35233

Thus, it is crucial to evaluate your data prior to embarking on a data science project to get an effective understanding of where it is and if it is usable. In our experience, around 70-80% of a data science project involves data hygiene or data engineering to get a company's data clean and ready to use. In many cases, the project ends early with advice on data hygiene and a year-long pause: "Start capturing variable X and call us in a year."

Another factor to consider is that you may have more data than you think you have, albeit much of this data may not be in usable form *as-is*. Data from different sources such as sensors or social media can provide valuable raw material for a number of data science projects. The fact that companies often have more data than they know what to do with is one reason it's so critical to have a good data science strategy.

Complicating matters, though, is that many companies don't realize that the data they have is not suitable for use without at least some degree of master data management (MDM). For instance, the term "revenue" might be defined differently depending on which department or Tableau report you reference. It might mean "total sales," "same store sales," or "sales not counting gift cards," or "sales minus returns" and so on. Sadly, we've seen far too many occasions of conflicting reports all being "right". Proper Master Data Management ensures everyone is speaking the same language.

TIP 5 - BUILD YOUR DATA SCIENCE STRATEGY IN ADVANCE

As mentioned previously, companies often identify the data science projects they want to launch only to find that they need to make changes in how they capture and manage data for their strategy to be effective. In many of these cases, projects must be put on hold until sufficient data can be captured. To mitigate this "stop the presses" possibility and feed this crucial bottleneck, we recommend building your data science strategy as early as possible, even if it's before you plan to make major investments, so that you will have the right and sufficient data when you begin.

TIP 6 - TAKE PRODUCTION REQUIREMENTS INTO ACCOUNT

Taking production requirements into account is another important aspect of building a data science strategy. Machine learning algorithms today can perform complicated calculations rapidly, but as the amount of data increases, even modern servers will require greater time and resources to run models. This is particularly when acquiring data from multiple sources, where data engineering must be performed before it can be used.

***Key Insight**: Be aware of both the dollar costs of your computing requirements and the time allocated to performing a routine. The more complex the algorithm, the more resources required.*

Let's consider an example of where this commonly plays out. If you have an algorithm that predicts the price of homes, for instance, there will be typically be a variety of variables that go

2204 1st Avenue South, Suite 101
Birmingham, AL 35233

into the prediction process. Thus, your formula is built on market conditions that are changing on a regular basis. This begs the question: how often do you update your formula? Whenever a new home is sold? When a market condition, such as area crime or school rankings change?

This is where things like machine learning come in. In order to update your formula, you would need to build a data pipeline to ingest data regarding all the homes that have been sold in the area and retrain your model to take this data into account on a regular basis.

While minute-by-minute or hour-by-hour updates would likely be unnecessary for something like home prices, in other cases this type of frequency is more necessary. Fraudulent transactions in financial services, for example, need to be identified in as near real-time as possible.

Newcomers to the field often underestimate the cost or complexity of running algorithms and that can be a major impediment down the road. *"We see companies hit a wall all the time in their data science execution process when it comes to deployment,"* says Casey Phillips, Director of IT at StrategyWise. *"They've put so much energy into building the perfect model, they haven't accounted for the crucial architectural components of actually putting that model into production. It is absolutely vital for data scientists to go through the model building process with the end state in mind."*

An interesting example that demonstrates the necessity of considering the computing requirements of the algorithmic component of your data science strategy comes from the streaming video company Netflix.

In 2006, Netflix sponsored the Netflix Prize, a competition designed to generate an algorithm that would be substantially better than the algorithm the company was using to predict how much a person would enjoy a movie based on their movie preferences. The winner of the contest improved the existing algorithm by 10%. Netflix, however, decided not to implement the winning formula. The company stated: "We evaluated some of the new methods offline, but the additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment." In other words, the winning algorithm's added compute requirements would have cost the company more to implement than they would have gained in actionable intelligence from its use.

CONCLUSION

Data science projects, like their comparative efforts in IT and other similar sciences, encounter enormous odds that threaten their success. Gartner has estimated data science project failure rates as high as 85%. The good news, however, is that many of these project killers can be identified and addressed in the early days of developing your strategy. The tips we've identified

2204 1st Avenue South, Suite 101
Birmingham, AL 35233

above will go a long way in ensuring your project is successful, but it never hurts to have an industry professional help you build your strategy, and serve as an outside advisor as you put your team and plans into place. StrategyWise has helped dozens of Fortune 1000 companies build data science strategies, and we offer solutions that can generate solid strategic first steps in as little as 3 days, with one of our data science workshops. Call us today to speak with a data science expert about how we can help you build a lasting strategy for your company or division!

CALL 888-623-3282 OR E-MAIL ROADMAPS@STRATEGYWISE.COM