DATASPARQ

G-CLOUD 14

Data Science Services

Service Definition

Contents

1 DATA SCIENCE

The application of data science in the form of machine learning (supervised learning, unsupervised learning, reinforcement learning) and artificial intelligence in the creation of new solutions and products to bring additional value in the form of increased revenue or reduced cost has the potential to yield significant value. Leaders from both technical and operational competencies are increasingly turning to data science in search of improved performance in their areas of responsibility, whether that is in sales, marketing, finance, manufacturing, logistics, human resource management, customer management or any other.

1.1 What is Data Science?

Simply put, data science is at the nexus of science, business and data. Having evolved from data mining, data science is the process by which discoveries are made from data through hypothesis-driven research methodologies. What this means is that given a set of data and possessed with business acumen or more generally subject matter expertise, we can explore data and pose specific questions. These questions will either be supported by the data or will have to be re-addressed through new more creative analyses just as empirical science works through an iterative process.

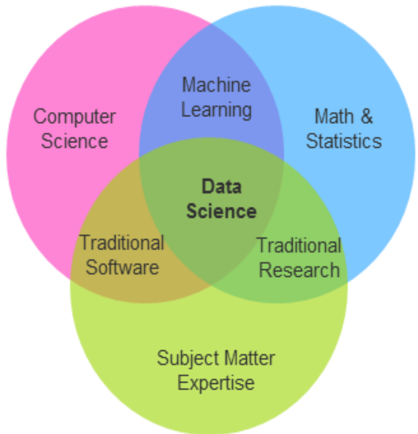


Figure 1: Emergence of Data Science

The techniques we use for addressing such questions sit at the interface of computer science, mathematics, and statistics. In the past, we have employed such machine learning techniques as artificial neural networks, classification, regression, and clustering to provide predictive power and to uncover insights into data from a set of diverse verticals.

While traditional business intelligence can supply granular insight and enable companies to develop and enhance their products and services, reinforce existing relationships with customers, and ultimately boost revenue, it does not allow them to anticipate the future. In this sense, predictive analytics is one-step ahead of traditional business intelligence and is generally based on statistical probabilities of certain events or outcomes.

1.2 Where Data Science & Artificial Intelligence sits in relation to Business Intelligence

Business Intelligence can reveal areas of under- and over-performance by slicing and dicing historical data in creative manners. Business Intelligence is the first required step in understanding the past; however, there is no predictive power in Business Intelligence and Predictive Analytics are required to optimize resources through making informed decisions and taking actions for the future.

Predictive Analytics enables organisations to evolve from "what happened" and "what is currently happening" to "what is likely to happen next." Predictive analytics provide more insightful and actionable answers to the organization's common questions than those generated by Business Intelligence alone, and more recommendations to questions that cannot be addressed at all by Business Intelligence. Shifting from historical tracking to predictive analytics enables a different outlook on today and tomorrow.

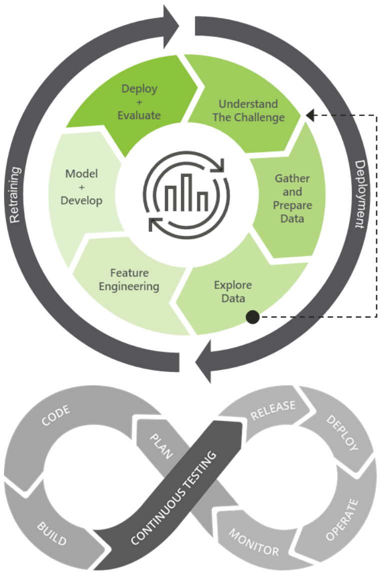
Figure 2: Business Intelligence versus Predictive Analytics

1.3 The Product approach, Agile and Dev Ops

While in traditional business intelligence and analytics we generally develop data models, reports and visualisations, in the field of data science and AI we are more frequently developing an integrated solution or product. At times we may carry out investigative analytics and calculate probabilities as part of an experiment to find value or test a hypothesis. Typically, however, the aim is to build something that can be consumed; a product. As such the structure of data science projects is different, as are the skills required and the way in which the output is consumed. While reports may be used to analyse the performance of the model in the process in which it operates, it is generally more likely that the output will be consumed by another application to assist in making an automated or manual decision.

It is best to illustrate this using an example, let's say Fraud Detection in an Insurance Claims department. Data science experiments will let you test the hypothesis of whether you can use statistics and data to find fraud. The product, however, is some combination of components including at a minimum the engine (executable data science model that has evolved from your experiment, and the execution framework and resources) which processes a new claim to determine whether it is likely to be fraudulent and warrants investigation or some other action. To support the training of the model, data pipelines including transformations and the generation of complex features is required. In order to feed a new case to the engine for scoring there is a requirement for an ability to send a message with the appropriate case data to the engine and to receive a response with the appropriate score (typically using an API framework). Then depending on the score, they may be a workflow to decide if a specific action is required and to automatically take that action. Where that action requires a person to interact with the case, perhaps for investigation of whether it is truly fraud, there may be a new interface that is required as part of the product. All of these components together give you a solution that can be used and that will deliver value. As such, the data science model is just a piece of your final product. Alone it does not necessarily give you all of the desired value.

We typically use the Agile methodology to get us to a minimum viable product, and then Dev Ops to continuously evolve that product to realise optimal value.



MVP Development

- Follow Agile method to support rapid, cyclical development of the initial product.
- The "product" approach ensures alignment to developing something that can be integrated with business processes even as an MVP (supporting AB testing).
- Understanding of how the product will be used in autonomous, semi-autonomous or manual business processes, and designing the MVP accordingly.
- Technical and product teams work closely together to ensure alignment and retain the ability to adjust the product as needed.
- The product owner prioritises development activities of the joint team.
- Evaluate appropriate data sets, features, algorithms and models to obtain the appropriate performance.
- Test plan evolves with the MVP to ensure full testing is complete before release 1.

Dev Ops

- Continuous integration of the solution from the initial release.
- Release management in an agile and flexible manner.
- Monitoring and operation of the product feeding back to the development cycle.

Figure 3: Agile Development and Dev Ops

1.4 Why Data Science?

Advanced analytics permit inferences to be made such that data is transformed into knowledge. Data science goes a step further by looking at real empirical data and trying to gain a deeper understanding of the whole picture in an attempt to deliver higher value to companies. Data Science teams at Datasparq work with complex data to solve specific, high value problems using statistical analytic algorithms and models.

In many organisations, the people responsible for handling data seldom interact with those running the business. This is a seriously flawed approach. Data science is the ability to tell a story and support making a decision based on data beyond just pointing to the numbers. Data scientists are a new type of analyst – part data engineer, part statistician, part business analyst. They are the people who can bridge the gap between doing advanced data analysis and using the findings to produce business results that align with an organization's goals.

Data Science is used to look for specific patterns in data which indicate a particular behaviour. Frequently the number of data elements, the volume of data or its complexity means that a human mind would not be able to compute the answer. Similarly, traditional approaches for analytics would not be able to observe the nuances in the patterns. This is where algorithmic analysis through data science comes into its own.

1.5 Examples of Applications of Data Science

Customer Segmentation

Through the use of advanced analytics, we enable businesses to segment their customer base in order to get a better grasp of distinct customer profiles that exist naturally within a population. This allows businesses to customise their products and services to better address needs of distinct customer profiles. Furthermore, it supports improvement of customer experience through tailored interaction from a digital or other perspective.

Fraud Detection

Our predictive analytics capabilities detect fraudulent and risky transactions from legitimate ones by applying machine learning methods to historical transaction data. This allows businesses to reduce risk and associated costs by adjusting their operational procedures in detecting and dealing with fraud. This model uses features from a wide selection of internal and external (publicly available) data sources. It also considers features extracted from free text using natural language processing. This broad feature space enables us to remove features which may perpetuate bias and protect us from bias that may exist in particular features of the historic true positive cases whether we are aware of it or not.

Demand and Profit Forecasting

Powered with Artificial Intelligence algorithms and historic data, we have been able to forecast customer demand and participation in specifically tailored promotional programmes empowering business to optimise their supply chain operations. In this endeavour, to really understand customer demand and shopper behaviour we have been able to draw on an array of sources and to integrate country-specific data as well as global macroeconomic metrics to elucidate cultural and regional differences in how customers approach shopping. Similarity we have used statistical modelling and predictive analytics to forecast profit and help our clients fine-tune their revenue streams and sales strategies.

Predictive Maintenance

At Datasparq we develop predictive maintenance solutions using sensor data from IoT (Internet of Things) devices in machines. Typically, we use two approaches to do this. Firstly, by analysing the patterns in the data leading up to previous failures we can use supervised machine learning to teach the solution to identify similar failures likely to occur in the future. This approach captures the previously known failures but misses new failures not previously observed in the data used to train the model. For “new” failures we use anomaly detection. In this scenario we look for patterns of behaviour that are not deemed “normal” for the machine. Combining these two we can help businesses to decrease the down time of machines, optimise maintenance regimes, re-design products and improve customer experience.

Churn Prevention

Working with a variety of customers, we have developed algorithms that predict customer churn and employee attrition. The models combine customer data and usage data with a variety of external data sets to forecast the propensity of each individual to churn. The solution allows marketing, customer retention and HR activities to be targeted at individuals with highest return rates. The approach we follow results in an output which indicates which elements in the data are mostly influencing the score predicting the churn so that retention team can understand causality. An extension to the churn model is a best action model. This is a recommendation engine which informs the retention team what action is best used to retain a customer like the one you are dealing with who is interacting in a similar way with your service. This solution prevents churn and therefore improves revenue, customer experience and margin (through a reduction of cost of customer acquisition).

Product Recommendation

At Datasparq we have significant experience in developing personalised product recommendation solutions in the retail space. Product Recommenders are used in consumer business for a variety of reasons including to increase basket size, to increase basket revenue, to improve basket margin and to improve customer experience. In developing these solutions, we use a mix of approaches including market basket analysis, matrix factorisation and collaborative filtering. As the number of historic transactions of an individual increase, the recommendations become more personalised. In fact, the solution selects the appropriate model in the engine to execute depending on the individual shopping and their shopping history.

2 Service Definition Detail

2.1 Pricing

We provide flexible pricing models to accommodate our solution and product offerings ranging from, T&M consulting, (see SFIA rate card for prices), performance based fixed fee sprints, service subscriptions or benefit share. All details provided within the pricing document.

2.2 Service Constraints

Specific service constraints will be agreed at initial engagement, however, typically no maintenance or modification to developed solutions shall be undertaken without direct approval from the customer. Full and extensive customisation will be accommodated within the constraints of the project in terms of budget, time and resource. If operationally necessary, any scheduled modifications and enhancements to the developed solutions will be undertaken outside of typically working hours or in scheduled system downtime.

2.3 Service Levels

Specific levels of service and expected performance criteria will form part of any initial engagement. Datasparq will endeavour, where possible, to meet these requirements within the constraints of the project in terms of budget, time and resourcing.

2.4 Digital Experience

Today, all businesses are 'Digital Businesses', servicing their users across multiple channels and devices. We do not just see design as the visual component to a project, but more a way you engineer your solution to build true competitive advantage using technology, creating differentiation where it matters.

Our Digital Experience experts are passionate about making design not only look great but deliver tangible benefits through the application of User-Centered Design principals. We understand that to our clients this is an investment to ensure their solution is designed and built against a robust business strategy / vision, but more importantly the needs and wants of their users. We focus on users to design a contextual, personal and relevant digital experience.

2.5 Backup/Restore and Disaster Recovery (DR)

Any developed solutions will be regularly backed up and the number of days to retain backups for will be agreed with the customer to ensure the system can be restored to a point within that period.

Disaster recovery options will be considered on a case-by-case basis from the needs of the customer.

Factors to consider include:

- Mission critical severity of the solution
- Previously agreed uptime targets
- Size and geographical location of the user base

Solutions range from automated failover capabilities with redundant systems to take load capacity in the event of total failure, to secure off site data storage that can be utilised to recover manually.

2.6 Training

Following initial engagement, training documentation will be provided as a matter of course. If additional training is identified for system end users, this can be structured into any service support delivery offering. Typical training operations could consist of workshops and mentoring at service go-live to longer term solutions such as an extended on-site presence and support.

Any training solution will be developed in conjunction with the customer to achieve a best fit for the required skill levels.

2.7 Ordering and invoicing process

Initial ordering of services will typically, be confirmed by receipt of confirmed purchase order to Datasparq Limited.

2.8 Termination terms

Termination of services would typically be at the conclusion of the pre-arranged engagement. Terminations prior to this incur penalties for early termination to reflect the allocated resource and potential losses incurred by Datasparq Limited.

Following the conclusion of the initial engagement, continued service support will be provided to which consumers may choose to terminate consumption at the end of one contractual cycle.

2.9 Consumer responsibilities

No specific consumer responsibilities have been identified for this specialised service offering. These would typically be discussed on an individual engagement basis and be tailored to the expectations and capabilities of the client.