LEARNING TREE ™
INTERNATIONAL

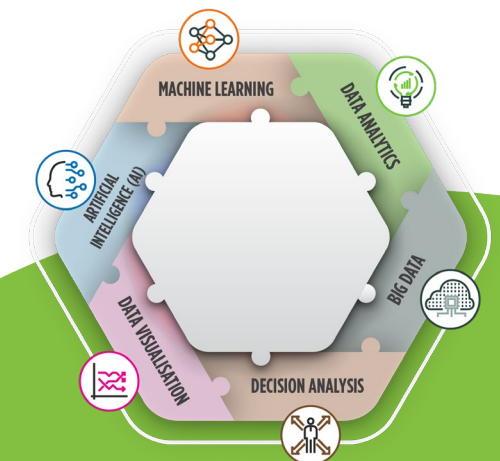# Building Your Data Science Dream Team

eBook

Contents

Contents

Contributors

Chris Mawata, Ph.D. *Learning Tree Data Science Curriculum Dean*

Chris has over 30 years of IT experience, including 17 years of teaching at the university level, and 15 years of training Java and Big Data programmers. As a Learning Tree instructor, Chris has authored four courses. As a consultant, he runs a 20-node cluster on which he has several Big Data frameworks installed. He has published peer-reviewed papers in image processing, artificial intelligence, and pure mathematics.

Andrew Tait *Learning Tree Data Science Instructor*

Andrew Tait is Chief Technology Officer of Decision Mechanics. Decision Mechanics specialies in decision science and the design/development of related technology. With a background in Artificial Intelligence and Operations Research he is currently involved in helping organisations make use of machine learning to tease insights from their data lakes.

LearningTree.co.uk/DSDreamTeam
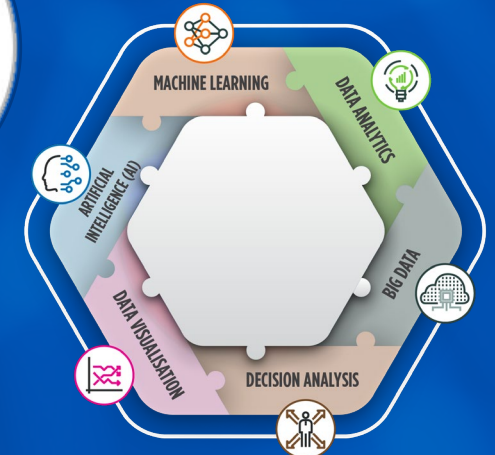
Introduction

The drive to increase confidence in business decisions has never been more important than it is now in today's uncertain digital economy. Most organisations these days aren't struggling with a lack of data, but instead the question of how to transform data into clarity, peace of mind, and valuable business insight. But where to start?

Organisations that are looking to improve their data science strategy are frequently encountering two obstacles:

1. **Hiring/Skills:** Identifying the skills and job roles needed to create a data science dream team can be a moving target, but it's necessary to get right. It's not good enough to just know how to process and work with data — the business must be confident in the conclusions its drawing or else the data cause more harm than good.

2. **Stakeholder Collaboration & C-Suite Buy-In:** Data science success requires a culture shift – not just a technology shift. Stakeholder and C-Suite buy-in from the top down, as well as a clear understanding of business goals and needs, is critical to success.

This eBook takes a deeper look into the skills and job roles that contribute to data science success, plus recommends ways in which a CEO can provide leadership in the realm of data science by championing data hygiene best practices and decision analysis. Finally, we'll see how one weather forecasting organisation was able to revolutionise their predictive capabilities through a targeted data science skills enablement programme, so you can see how all the pieces of the data science puzzle fit together to create a picture of success.

Data Science Skill Areas

"Data science" is an umbrella term which covers a large amount of conceptual ground and can encompass a wide array of related business and technical skills. Here we take a deeper look at the six major skill areas of data science to see how these skills contribute to the overall success of an organisation.
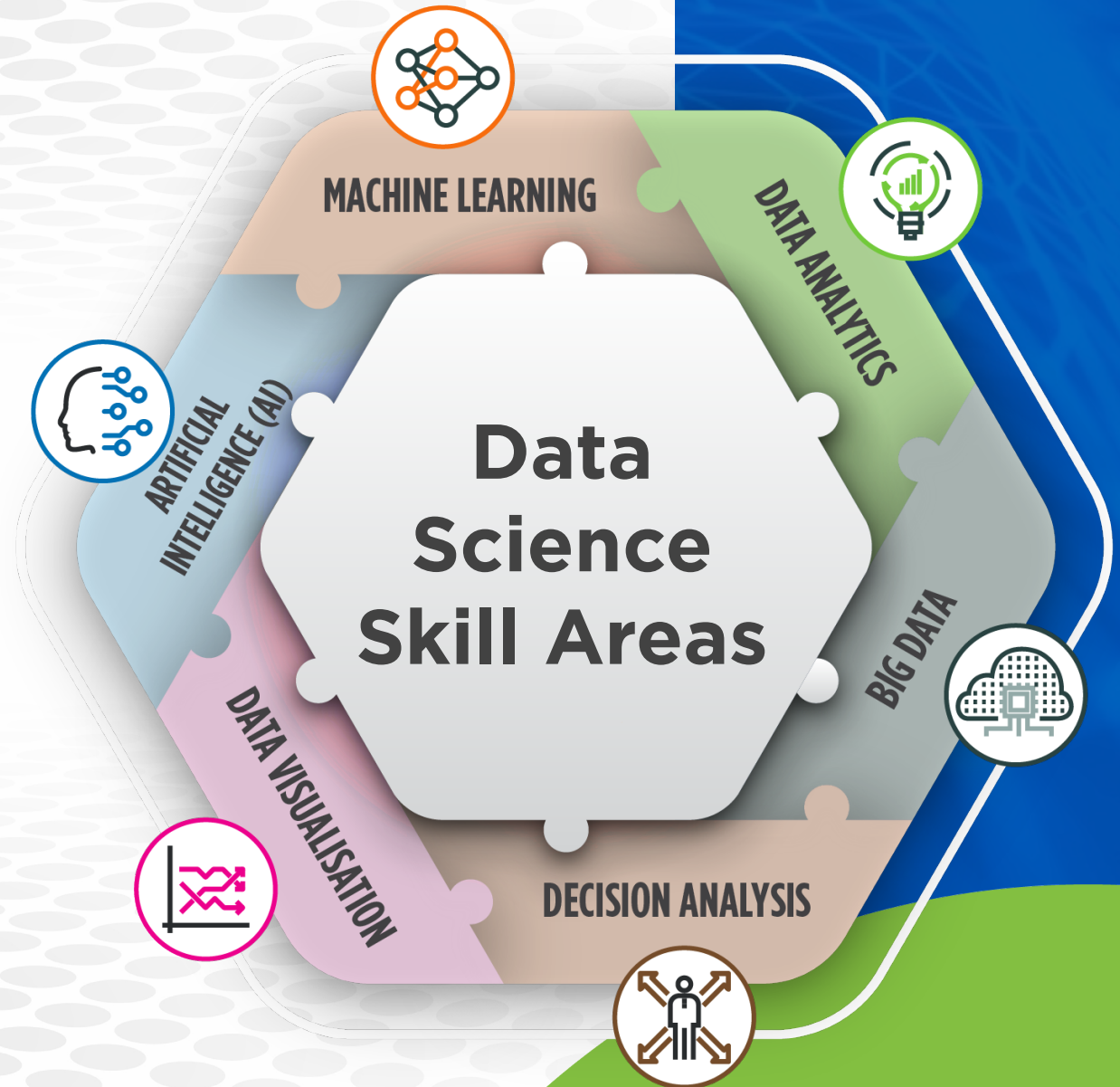
Data Analytics

Data analytics is the most common use of data, involving extracting data from relational databases (e.g., SQL Server, Oracle) and presenting findings as reports and corporate dashboards. Data Analytics professionals should have the following skills:

- Manipulate databases using SQL

- Use dashboard tools and design effective dashboards

- Utilise statistical tools to maintain data integrity

- Produce effective, clear charts that inform, rather than confuse, decision makers

RELATED JOB TITLES

- Business Analyst
- Data Analyst
- Data Scientist
- Data Engineer
- Quantitative Analyst
- Project Manager

66,000+

Active Job Postings for Data Analytics Skills[1]

[1] *Job posting numbers come from a LinkedIn search for relevant job posting in EMEA as of February 2021.*

DATA ANALYTICS

Machine Learning

Machine Learning (ML) involves training computers, through repeated presentation of observations and outcomes, to make predictions that are not obvious to a person. Machine learning experts on your team should have the following skills:

- Expertise in modeling techniques and tools

- Understanding of the statistical basis of algorithms

- Programming skills (Python or other machine learning language)

- Ability to work closely with data scientists to ensure machine learning technology delivers results for the organisation

RELATED JOB TITLES
- Data Scientist
- Machine Learning Engineer
- AI Engineer
- Marching Learning Researcher
- Machine Learning Developer
- Data Engineer

35,000+
Active Job Postings for Machine Learning Skills
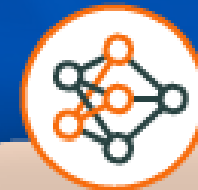
MACHINE LEARNING

Artificial Intelligence (AI)

Artificial Intelligence (AI) has been around since the 1950s, but today's AI researchers use cutting-edge technologies such as deep learning (previously known as neural networks), Natural Language Process, or NLP (used in conversational user interfaces), and image processing (as used in products such as self-driving cars). Important AI skills include:

- Programming and writing code for machines

- Designing and developing machines and systems that can learn and apply knowledge without specific direction.

- Research, analytical, organisational, and critical thinking skills

- AI professionals are typically highly specialised by area of research and niche expertise

RELATED JOB TITLES
- Machine Learning Engineer
- Data Scientist
- Business Intelligence Developer
- Research Scientist
- Big Data Engineer/Architect
- Software Engineer

31,000+
Active Job Postings for AI Skills

Big Data

Big data describes working with data that is too large to be processed using standard (e.g., workstation, single server) tools. Popular platforms include Hadoop, Spark, and Tableau. Big data professionals must know how to:

- Operate and manage clusters of networked computers

- Maintain high availability of the cluster

- Understand how cyber security issues can affect big data

- programme using enterprise languages, such as Java and Scala

RELATED JOB TITLES
- Data Scientist
- Data Engineer
- Data Analyst
- Security Engineer
- Database Manager
- Data Architect

35,000+

Active Job Postings for Big Data Skills

BIG DATA

DATA SCIENCE SKILL AREAS

Data Visualisation

Data visualisation involves representing information visually (e.g., charts, tables, diagrams, illustrations) in order to bring data to life in a way that's easy to understand to drive better business decisions. Necessary data visualisation skills include:

- Design visualisation, create and implement complex visualisations from scratch

- Show clear, impactful data by bringing the big picture to life

- Understand stakeholder needs and communicate effectively with non-technical audiences

- Integrate statistical and verbal description of the data

RELATED JOB TITLES
- Data Analyst
- Database Architect
- Data Visualisation Engineer
- Data Visualisation Scientist
- Data Engineer

10,000+
Active Job Postings for Data Visualisation Skills

DATA VISUALISATION

Decision Analysis

Decision analysis is the process of applying a standardised and tested methodology for decision-making which helps organisation make decisions more easily and transparently. Necessary decision analysis skills include:

- Formalise the decision-making process

- Frame decisions, think creatively, analyse, and implement informed and justifiable decisions

- Understand model uncertainty with data

- Evaluate alternatives against multiple criteria

- Leverage skills/tools like emotional intelligence, management techniques, decision trees, diagrams, and more

RELATED JOB TITLES
Decision analysis is a beneficial skill for anyone in an organisation who is responsible for making decisions.

5,000+
Active Job Postings for Decision Analysis Skills

DECISION ANALYSIS

Your Data Science Dream Team

Now that we've established which skills are critical to data science strategy success, it's time to explore the job roles which make up the successful data science team. Not every data science team will have all roles represented on their team, and the exact skillset needed can vary based on business initiatives and goals. All the same, the roles outlined in the following section can provide a general framework for the players who contribute to an all-star data science dream team.

- **Business Analyst** The main goal of the business analyst is to understand organisational, business, and stakeholder needs and be able to translate and communicate these requirements for the data analyst. The challenge for the business analyst lies in the difficulty in asking the right questions to gain a clear understanding of the business requirements — a task simpler in theory than in practice.

  The business analyst can also be known as business architect.

  Relevant Skills: Decision Analysis, Data Visualisation, Data Analytics, Communication Skills
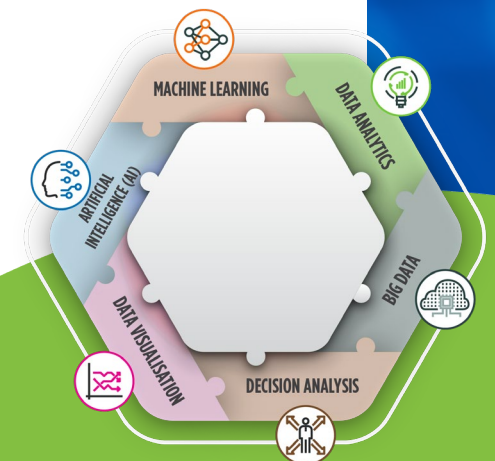
Your Data Science Dream Team

- **Data Analyst** The mission of the data analyst is to understand the business requirements obtained by the business analyst and decide what data is needed to address the requirements. Much like the business analyst, the data analyst must ask right questions, so he truly understands the business requirements to collect the appropriate data. Data analysts rely on specialised skills related to visualisation tools like Tableau and Power BI. Knowing the best tool for the job based on type and volume of data is under the purview of the data analyst.

Relevant Skills: Decision Analysis, Data Visualisation, Data Analytics, Programming (Python or other data science language), Communication Skills

- **Data Engineer** The data engineer is responsible for collecting data through data pipelines, cleaning and transforming the data, and aggregating the data into a database or data source. They must also deliver the data back to the stakeholders. A successful data engineer must determine which tools and programmes are best for their data environment. The data engineer is also responsible for ensuring there is a good match between how data is stored and the way it is queried.

Alternative titles for the data engineer include data architect, big data developer, IoT engineer developer, and data designer.

Relevant Skills: Big Data, Programming (SQL, Python, or other big data language), Data Wrangling

Your Data Science Dream Team

- **Data Scientist** Your team's data scientist is responsible for creating models of the system, entering all parameters of the data received, and predicting what will happen given the analysis of the incoming data. It is imperative that the data scientist be able to select the right model for their analysis that will give the most accurate results. The data scientist is highly focused on accuracy, so knowing the risk of any prediction is important, and it is their job to clearly present those risks to stakeholders when necessary.

Relevant Skills: Machine Learning, Artificial Intelligence, Data Visualisation, Programming (Python or other data science language), Deep Learning, Statistics, Risk Management, Communication Skills

- **Data Wrangler** The data wrangler's goal is to create inputs for data science models based on provided raw data. For those with programming experience like Python and can massage data into the right shape, the data wrangle job role can be a good entry point into a data science career. The data wrangler must know that the data they're using is real, and to know how to extract the relevant information from raw data sources like log files, sensor data, text, and more. Another name for the data wrangler is data science programmer.

Relevant Skills: Machine Learning, Artificial Intelligence, Big Data, Programming (Python, SQL or other data science language)

Your Data Science Dream Team

- **Framework Administrator** Your framework administrator is responsible for managing and backing up the data, as well as managing security and access. Understanding data models to be able to perform data mapping is important, as well as the technical capabilities to perform database queries and data analysis. The framework administrator can also be known as database administrator.

Relevant Skills: Big Data, Database Management, Programming Skills (Python, SQL, or other database management language)

- **Quantitative Researcher** The quantitative research is a math and stats person of the data science team with a primary goal of bringing numbers to life through compelling, accurate visualisations. They also have knowledge about the ethics of data science. "What is the cost of an error," is a question the quantitative research will focus on. The quality of information and predictions that are being passed back to stakeholders is their number one priority. The quantitative research can also be known as data visualisation design or quantitative analyst.

Relevant Skills: Artificial Intelligence, Machine Learning, Data Analytics, Data Visualisation, Decision Analysis, Programming Skills (Python, R, or other data science language)

Your Data Science Dream Team

- **Machine Learning Engineer** The machine learning engineer knows how to move from prototype to production as efficiently as possible. They must know the right programmes and tools to use for a secure environment that can handle scale, and keep that environment monitored. The ML engineer is responsible for the accuracy of data from step to step as they progress toward the data model to satisfy business needs. Programming skills, experience with clusters, and AI knowledge are all required for success in this role. Alternative titles for the machine learning engineer are enterprise architect and machine learning researcher.

  Relevant Skills: Machine Learning Artificial Intelligence, Decision Analysis Data Visualisation, Programming Skills (Python, R, or other data science language)

- **Cloud Specialist** Your cloud specialist utilises virtual environments to create streamlined operations for the organisation and assists in the migration of information and services into the cloud. The cloud specialist must understand and be able to clearly speak on the ROI of migrating the data to the cloud, as well as keep abreast of the constantly changing cloud features within data science and machine learning. The cloud specialist may also be responsible for a portion of **managing and monitoring their cloud environment.**

  Relevant Skills: Machine Learning, Big Data, Artificial Intelligence Data Warehousing

Your Data Science Dream Team

There are many players on the idea data science dream team, and the skills necessary to success are shifting rapidly. The organisations who consistently find success are the ones who most adeptly align their data science strategy to organisational goals and initiatives, and those who invest in their workforce to stay up to speed on critical skills and the tools that up your data science game.

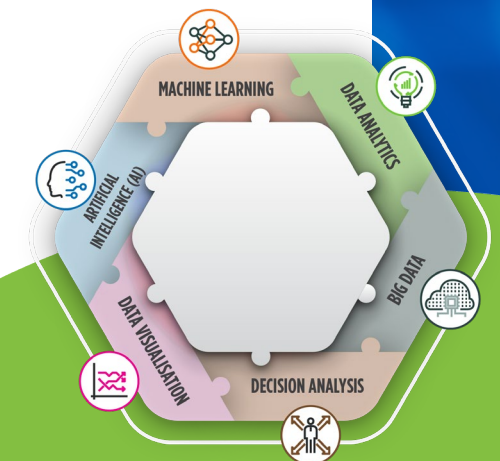# The 2 Data Science Topics Every CEO Should Care About

BY ANDREW TAIT

Many executives today know that data science is important, but they're not sure where to start, or which projects to start advocating for on the ground level of a data science strategy. There are two major topics that every CEO should read up on and care about. Early investment and understanding of these two facets will pay off in dividend of value later. They are:

1. **Data Hygiene**

2. **Decision Analysis**

Today's data protection regulations can significantly impact the type of data your organisation has and how it's managed. Regulations like COPPA, GDPR and CCPA clearly need to be adhered to. If you're in a regulated industry, like health or finance, then there are even more regulations that need to be considered. **And If you are in breach of data protection regulations, then it's not a C-Suite problem---it's a CEO problem.** You need to act.

If you are still struggling with proper handling of data, there is a good chance you're also struggling to meet the minimum requirement of these regulations. This can lead to a serious data toxicity problem, where you're essentially breaking the law through improper data storage.

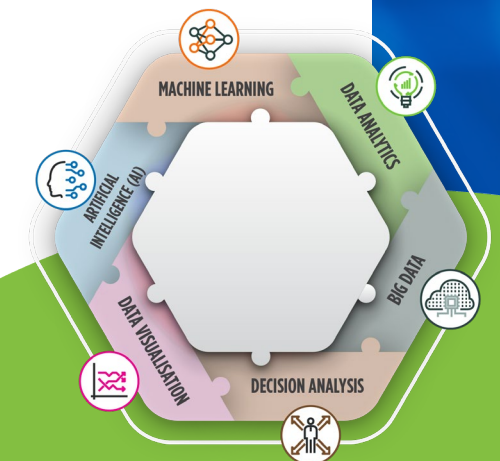# The 2 Data Science Topics Every CEO Should Care About

BY ANDREW TAIT

To prevent this nightmare, basic hygiene factors have been addressed. And it must be addressed from this perspective: **data science isn't a technology issue — it's a cultural issue**. How do you make decisions in your organisation? Do you require experiments and evidence before you sign off? If the big decisions don't demand data science, why would the small ones? **The C-Suite is responsible creating an environment in which scientific decision-making is the gold standard.**

Companies like Chevron have decision analysis as part of their DNA. It just obvious to them that you'd use that kind of process for making your decisions. It's embedded in the culture — *from the top*.

The risk of holding too much data should be a C-Suite concern. There's a balance to be made between competing priorities. Data scientists want as much data as possible but holding that data results in reputation risk and increased security costs. The risk vs reward calculation must be performed and acted on.

The simplest privacy policy is "We don't collect any data. Period." So, starting from there, the next question is, "What data do we need to generate business value?" Then it becomes a question of how much value you get vs how much risk you must shoulder to generate that value.

The 2 Data Science Topics
Every CEO Should Care About

BY ANDREW TAIT

The C-Suite should be looking to promote a culture of respect for data — its value and its risks. They should be providing the resources for the business units to understand the implications. And, they should be removing any impediments the business units have in making effective use of data. But, the day-to-day work of deriving value from data is probably best conducted at the business unit level. If you have a small team, then sharing data science resources may still be necessary, but the projects should still be owned by the business units.

The Bottom Line: A successful data science programme is built on data hygiene best practices and good decision analysis processes — and it must be championed from the top down.
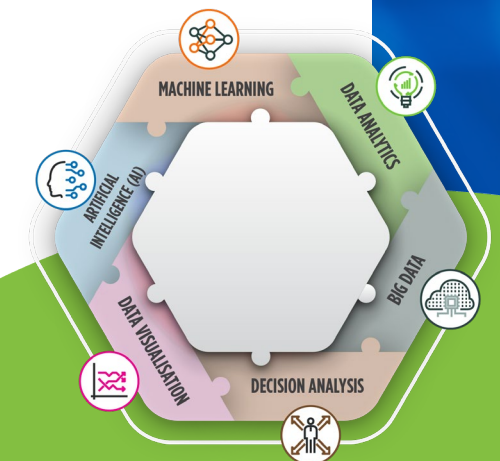
Improving Predictive Analytic Capabilities through Targeted Data Science Training Programmes

A European weather forecasting organisation operates as a research institute as well as an around-the-clock service, producing global numerical weather predictions and other data for its member states and the broader community. Their Scientists, Statisticians, and Programmers manage a high volume of rapidly changing data on a constant 24x7 basis. To achieve this, the organisation's data professionals must be at the forefront of programming skills to leverage technology for accurate forecasting. Leveraging the Learning Tree model of using practicing Consultants to deliver training, Learning Tree was invited as a trusted training partner to support this business need through developing a Data Science Academy addressing the following upskilling topics:

- Jupyter notebooks
- Supervised and unsupervised learning
- Deep learning and interpretability
- Data mining and pattern discovery
- The mathematical basis of machine learning
- Uncertainty quantification
- Gaussian processes
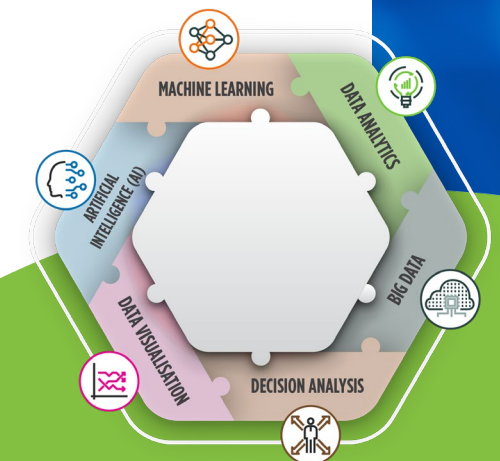
Improving Predictive Analytic Capabilities through Targeted Data Science Training Programmes

Through engagement and discussion with our Lead Data Science Consultant, a two-tiered customised solution was firstly developed as a specification and then on approval as a programme of courses.

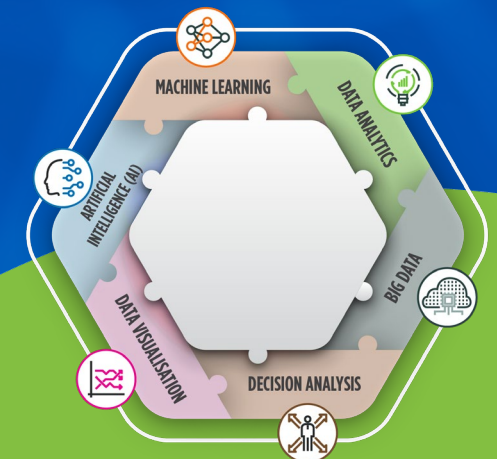| Data Science I	Data Science II	Data Science B: Random Forests & Artificial Neural Networks
Focus on Python Programming • Understand the fundamental ideas behind artificial intelligence and machine learning in particular • Choose algorithms that would be good candidates to solve a given problem • Interpret the measurements used to gauge the quality of a model • Appreciate the workflow needed to produce machine learning solutions	Understand the fundamental ideas behind deep learning • Generalise machine learning concepts to Bayesian analysis and Gaussian processes in particular • Interpret the measurements used to gauge the quality of a model	Gain the perspective of a Python programmer • Perform classification and regression using decision trees and ensembles of decision trees • Understand the fundamental ideas behind deep learning
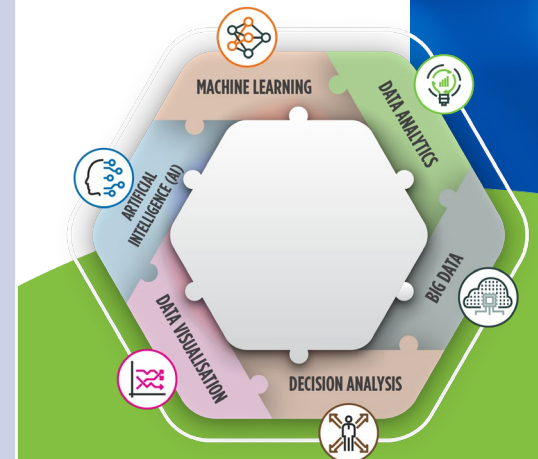
Improving Predictive Analytic Capabilities through Targeted Data Science Training Programmes

The emphasis of the programme was on a Bayesian approach to modelling, rather than the more commonly used "frequentist" approach. The Bayesian approach allows one to combine prior information you have about the phenomenon you are trying to predict with evidence from newer information to make a better prediction. This approach better fits the way a weather prediction organisation continuously gets new data. Until the recent past, the Bayesian approach has not been practical because only a few special cases can be solved by hand mathematically. With current computational resources and Monte Carlo methods implemented in Python, it is now a feasible approach to prediction.

KEY programme RESULTS:

- Learner Satisfaction in Course Quality: 94% Average
- Learning Satisfaction in Instructor Quality: 100%

Final Thoughts

As previously stated, the drive to increase confidence in business decisions has never been more important than it is now in today's uncertain digital economy. Organisations that are looking to improve their data science strategy can take the following steps:

1. Take inventory of your data team's current skillset and implement targeted, multimodal training solutions to bridge the gap between current state and where you want to go.

2. Examine the job roles listed in this eBook and reflect on where you might be missing a key member of your data science dream team.

3. Managers and members of the C-Suite should write this on a sticky note and post it in the office: "Data science success requires a culture shift – not just a technology shift." A successful data science programme requires top-down support.

Now is the time to take the next step in revolutionising your data science strategy through investment in your most value asset — your people. Explore the resources below to get started or contact Learning Tree to learn more.
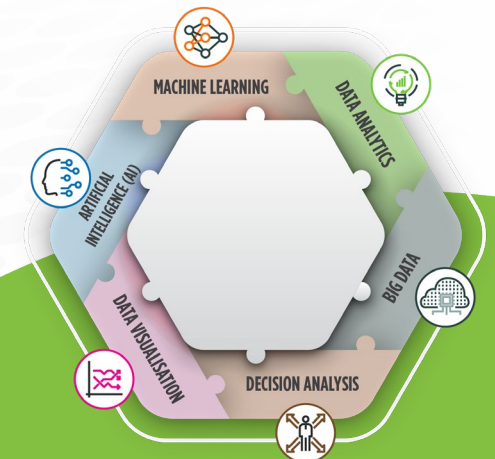
To learn more about advancing your data science strategy:
Visit LearningTree.co.uk/DSDreamTeam