Data Science Foundation

Accredited by IABAC™

Datamites

Who can pursue this course?

- Working Professionals, who want to switch to Data Science
- Data Professionals, Business Analyst, Software Engineers
- Senior Professionals, Managers.
- Students, Beginners from any background.
- Anyone, who wants to learn Data Science.

Datamites

Data Science Introduction

What is Data Science?

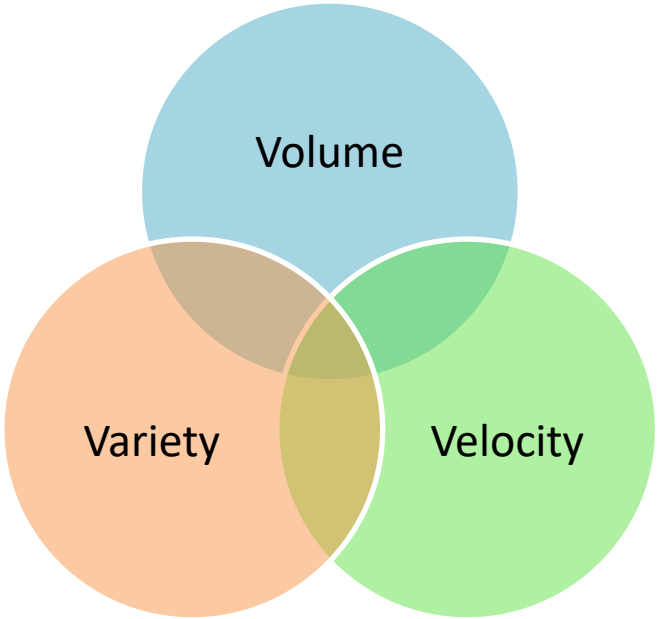
- Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structured and unstructured data. (Ref: Wikipedia)
- Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract actionable insights.
- Data Science can be simply stated as "Insights from Data"

Datamites

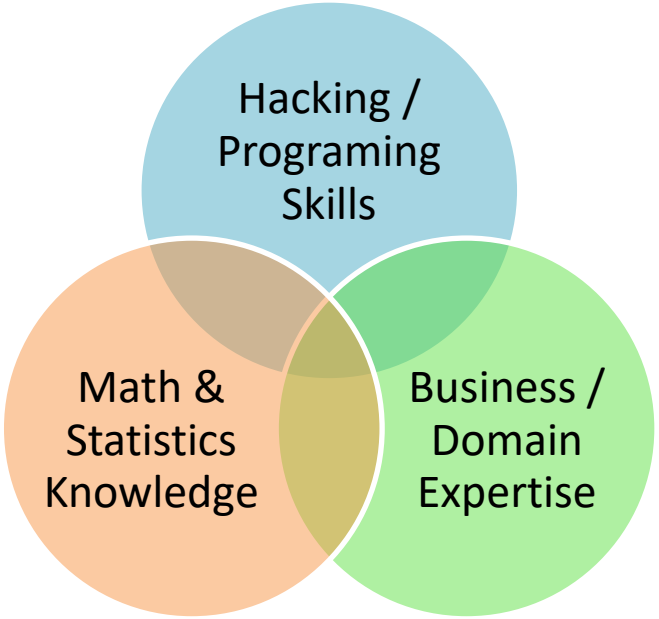
Evolution of Data Science.

Big Data vs Data Science

Big Data



Data Science



Datamites

Data Science vs Business Analytics

Data Science	Business Analytics
Involves programming / coding	Minimal/No coding required.
Data Science uses both structured and unstructured data	Business Analytics uses mostly structured data
Data Science extensively involves Machine Learning and Artificial Intelligence as tools.	Business Analytics use traditional statistical and forecasting techniques.

Datamites

Business Analytics Classification

Datamites

Machine Learning Introduction

What is Machine Learning?

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed.

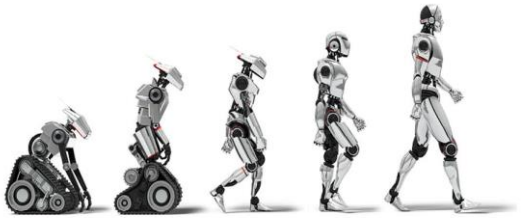
Areas of Artificial Intelligence

Computer Vision

Natural Language Processing

Machine Learning & Deep Learning

Decision Making



Robotics

Datamites

Machine Learning Use Case Demo

Amazon Return Policy Exploitation

A Machine Learning Case Study - **Demo**

Datamites

Types of Machine Learning

- Supervised machine learning
- Unsupervised machine learning
- Semi-supervised machine learning
- Reinforcement machine learning

Datamites

Supervised machine learning

- The most popular machine learning method.
- Uses labelled (past) data to learn and predict future events.
- The model can be evaluated and fine-tuned to attain accurate predictions.

Datamites

Unsupervised machine learning

- Used when the past data used to train is neither classified nor labeled.
- The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Semi-supervised machine Learning

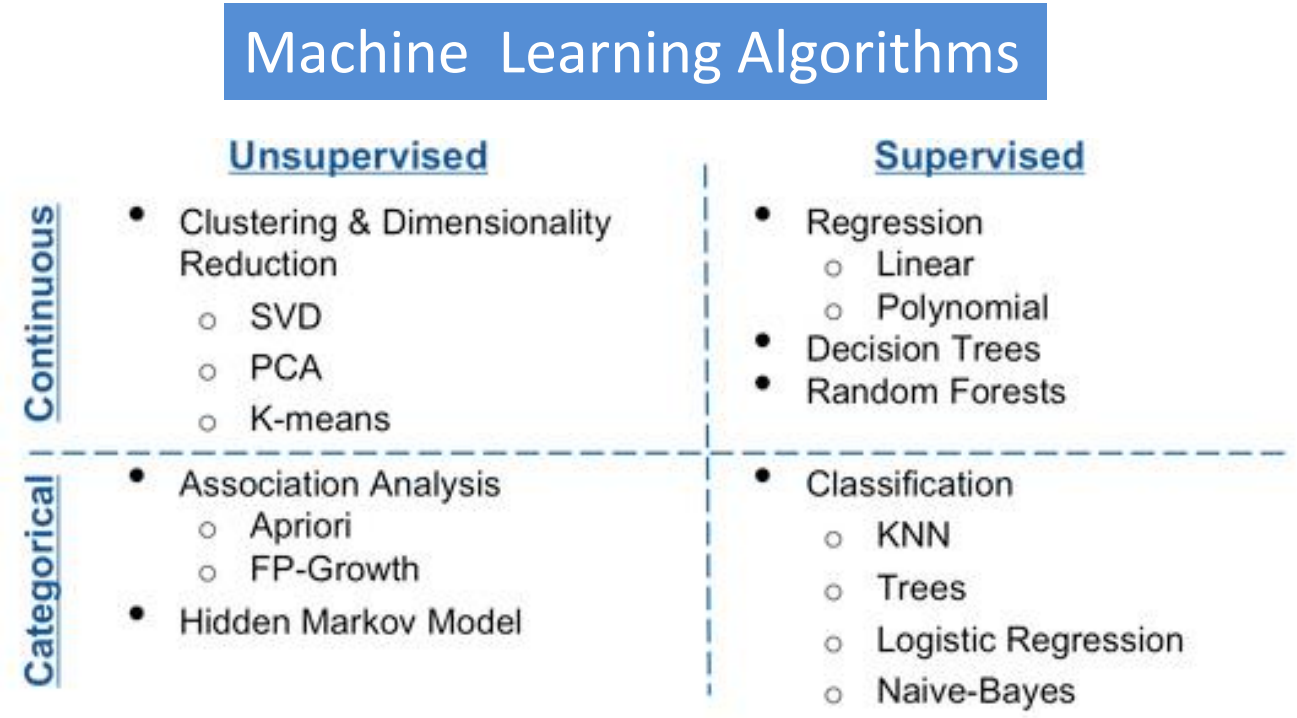
- Fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data.
- The systems that use this method can considerably improve learning accuracy.

Datamites

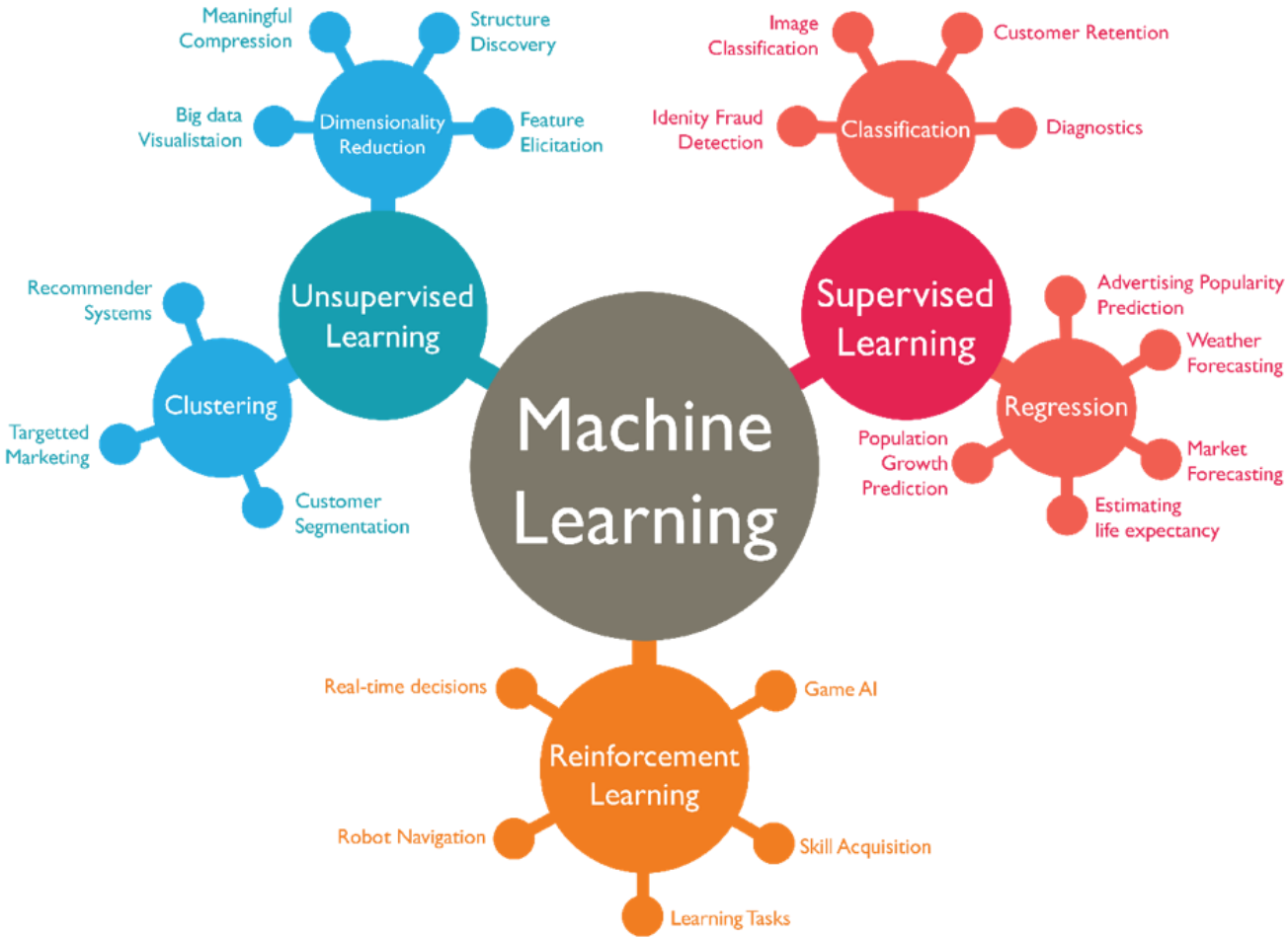
Reinforcement machine learning

- Learning by interacting with its environment by producing actions and discovers errors or rewards.
- This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance.
- Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Machine Learning



Datamites



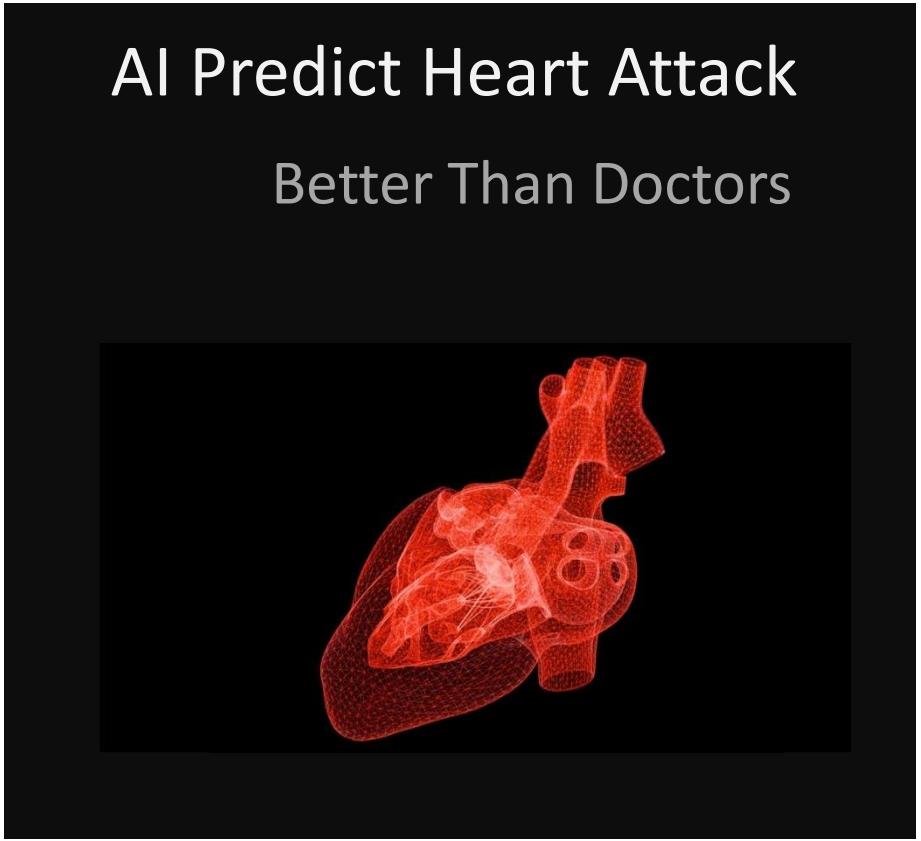
Datamites

Data Science Industry Applications, Workflow and Job Roles.

Industry Applications

Health Care

University of Nottingham researchers created an AI system that scanned routine medical data to predict which patients would have strokes or heart attacks within 10 years.



https://futurism.com/confirmed-ai-can-predict-heart-attacks-and-strokes-more-accurately-than-doctors/

https://hbr.org/2017/05/how-machine-learning-is-helping-us-predict-heart-disease-and-diabetes

Retail

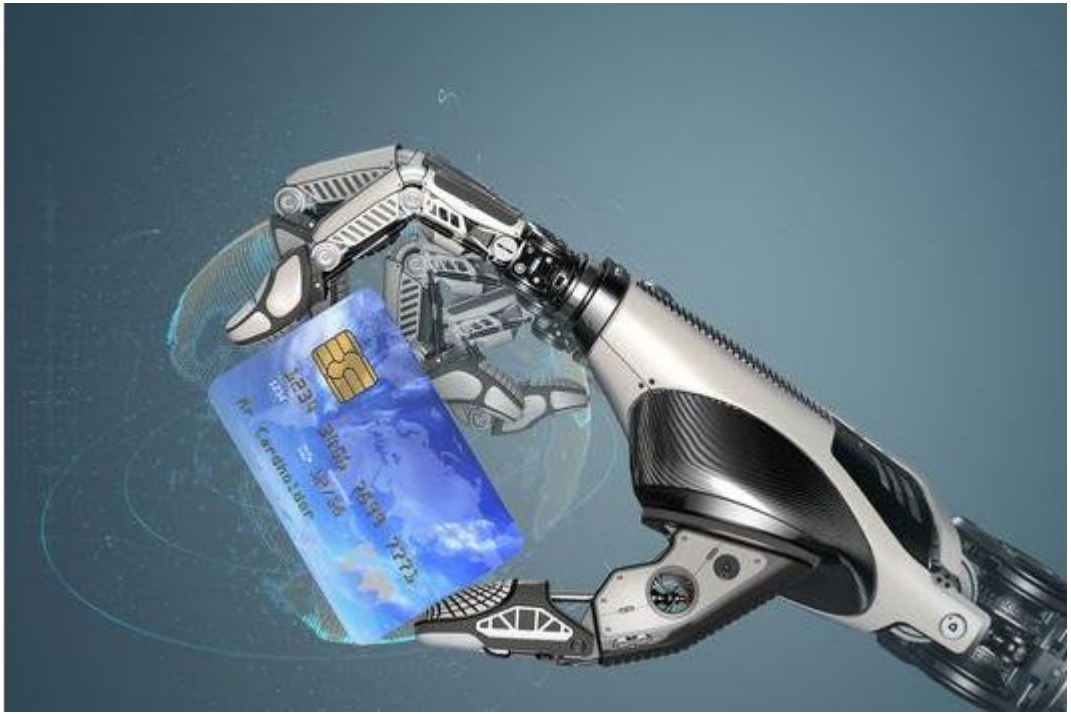
Machine Learning in Retail

what to stock, how much to buy, what products to suggest to repeat customers. But doing more with that data using machine learning is just what retailers need to really succeed in the current market.

http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world

https://www.forbes.com/sites/bobevans1/2017/06/20/how-google-and-amazon-are-torpedoing-the-retail-industry-with-data-ai-and-advertising/#766d80cb5c66

AI in Banking



- Fraud Transactions Detection
- Predicting risk of asset class based on future context models.
- More personalized and faster customer experiences

Datamites

AI will fight challenging diseases

We'll Soon Trust AI More Than Doctors

From Diagnosis to Treatment

Datamites

Financial Services - Experian

- With approximately 3.6 petabytes of data (and growing) about individuals around the world, credit reference agency Experian gets its extraordinary amount of data from marketing databases, transactional records and public information records.
- They are actively embedding machine learning into their products to allow for quicker and more effective decision-making. Over time, the machines can learn to distinguish what data points are important from those that aren't. Insight extracted from the machines will allow Experian to optimize its processes.

Datamites

Financial Services - American Express

- American Express processes $1 trillion in transaction and has 110 million AmEx cards in operation.
- They rely heavily on data analytics and machine learning algorithms to help detect fraud in near real time, therefore saving millions in losses.
- Additionally, AmEx is leveraging its data flows to develop apps that can connect a cardholder with products or services and special offers.
- They are also giving merchants online business trend analysis and industry peer benchmarking.

Datamites

Healthcare - Infervision

- AI and deep learning is being put to use to save lives by Infervision.
- In China, where there aren't enough radiologists to keep up with the demand of reviewing 1.4 billion CT scans each year to look for early signs of lung cancer.
- Radiologists need to review hundreds of scans each day which is not only tedious, but human fatigue can lead to errors.
- Infervision trained and taught algorithms to augment the work of radiologists to allow them to diagnose cancer more accurately and efficiently.

inferVISION 推想科技

Datamites

Manufacturing - Volvo

- Cars are increasingly connected and generate data that can be used in a number of ways.
- Volvo uses data to help predict when parts would fail or when vehicles need servicing, uphold its impressive safety record by monitoring vehicle performance during hazardous situations and to improve driver and passenger convenience.
- Volvo is also conducting its own research and development on autonomous vehicles.

Datamites

Manufacturing - BMW

- BMW has big data-related technology at the heart of its business model and data guides decisions throughout the business from design and engineering to sales and aftercare.
- The company is also a leader in driverless technology and plans for its cars to deliver Level 5 autonomy—the vehicle can drive itself without any human intervention—by 2021

Datamites

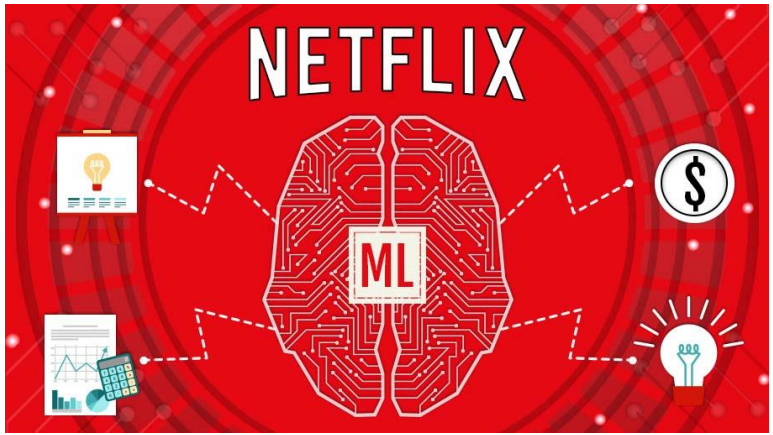
Media - BBC

- The BBC project, Talking with Machines is an audio drama that allows listeners to join in and have a two-way conversation via their smart speaker.
- Listeners get to be a part of the story as it prompts them to answer questions and insert their own lines into the story.
- Created specifically for smart speakers Amazon Echo and Google Home, the BBC expects to expand to other voice-activated devices in the future.

Datamites

Media - Netflix

- Big data analytics is helping Netflix predict what its customers will enjoy watching.
- They are also increasingly a content creator, not just a distributor, and use data to drive what content it will invest in creating.
- Due to the confidence they have in the data findings, they are willing to buck convention and commission multiple seasons of a new show rather than just a pilot episode.



Datamites

Retail - Burberry

BURBERRY®
LONDON

- When you first think of Burberry, you likely consider its luxury fashion and not first consider them a digital business.
- However, they have been busy reinventing themselves and use big data and AI to combat counterfeit products and improve sales and customer relationships.
- The company's strategy for increasing sales is to nurture deep, personal connections with its customers.
- As part of that, they have reward and loyalty programs that create data to help them personalize the shopping experience for each customer

Datamites

Retail - Walmart

- As the world's second-largest retailer, Walmart is on the cutting edge of finding ways to transform retail and provide better service to its customers.
- They use big data, machine learning, AI and the IoT to ensure a seamless experience between the online customer experience and the in-store experience (with 11,000 brick-and-mortar stores) something rival Amazon isn't able to do.
- Enhancements include using the Scan and Go feature on the app, Pick-up Towers and they are experimenting with facial recognition technology to determine if customers are happy or sad.

Datamites

Service - Microsoft

- Central to everything Microsoft does is leveraging smart machines.
- Microsoft has Cortana, a virtual assistant; chatbots that run Skype and answer customer service queries or deliver info such as weather or travel updates and the company has rolled out intelligent features within its Office enterprise.
- Other companies can use the Microsoft AI Platform to create their own intelligent tools.
- In the future, Microsoft wants to see intelligent machines with generalized AI capabilities that allow them to complete any task.

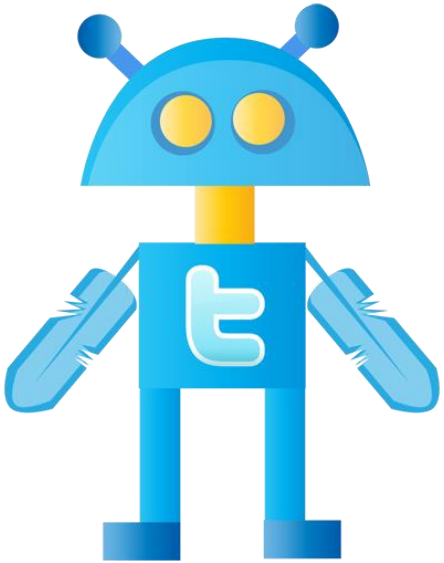
Datamites

Service - Disney

- Always at the top of delivery extraordinary service, Disney is getting even better thanks to big data.
- Every visitor gets their own Magic Band wristband that serves as ID, hotel room key, tickets, Fast Passes and payment system.
- While guest enough the convenience, Disney gets a lot of data that helps them anticipate guests' needs and deliver an amazing, personalized experience.
- They can resolve traffic jams, give extra services to guests who may have been inconvenienced by a closed attraction and data even allows the company to schedule staff more efficiently.

Datamites

Social Media - Twitter



- From what tweets to recommend to fighting inappropriate or racist content and enhancing the user experience, Twitter has begun to use artificial intelligence behind the scenes to enhance their product.
- They process lots of data through deep neural networks to learn over time what users preferences are.

Social Media - Facebook

- Deep learning is helping Facebook draw value from a larger portion of its unstructured datasets created by almost 2 billion people updating their statuses 293,000 times per minute.
- Most of its deep learning technology is built on the Torch platform that focuses on deep learning technologies and neural networks.

Datamites

Consumer goods - Barbie

- Using natural language processing, machine learning and advanced analytics, Hello Barbie listens and responds to a child.
- A microphone on Barbie's necklace records what is said and transmits it to the servers at ToyTalk.
- There, the recording is analyzed to determine the appropriate response from 8,000 lines of dialogue.
- Servers transmit the correct response back to Barbie in under a second so she can respond to the child.

Datamites

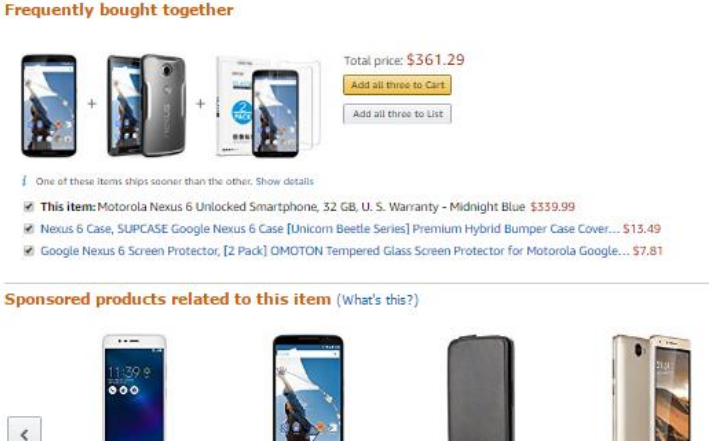
Some more..

HR Analytics

Facial Recognition

Skin Cancer Deduction

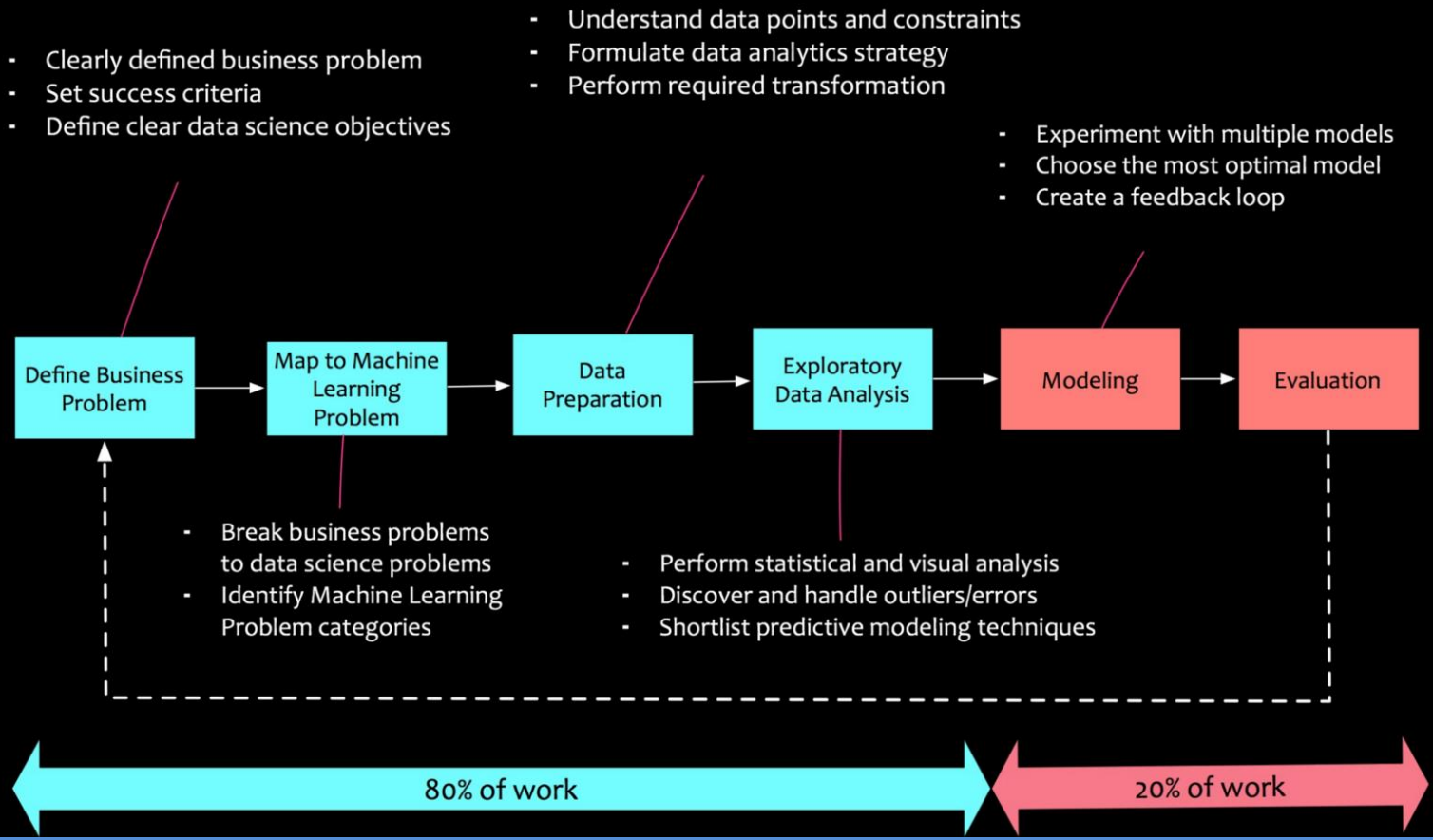
Recommendations



Datamites

Data Science Workflow

Data Science Workflow



Data Science Roles

Data Science Roles

Data Science Developers

Big Data Specialists

Infra Engineers

Data Scientist

Analysts

Researchers

Business Person

Datamites

DS Infrastructure Engineers

- Data Science Engineers are hard-core techies who deal with Data Science Infrastructure, ie., hardware, software applications and other aspects to get the back end of Data Science up and running.
- They set up the entire IT infrastructure from servers, networks to processes, also manage it such as infrastructure monitoring, application management, database administration etc.,

Datamites

Data Science Developers

- Data Science Developers are the ones, who code the models and applications through programing in R Language, Python etc., They are versatile developers, who have good knowledge on math & statistics, machine learning algorithms and related concepts.

- As this domain of data science development is evolving rapidly, these developers are expected to keep themselves updates with all latest technological advances from development perspective, so that they can use the right platform to achieve their goal in effective manner.

Datamites

Machine Learning Specialist

- These are the professional with deep knowledge in computer science and mathematics.
- They engage in machine learning and deep learning heavily.T
- hey create predictive and prescriptive models based on the machine learning algorithms, such as random forest, Artificial Neural Network, K-NN etc.,
- They are masters in all kinds of data mining techniques pruning, regularization etc., helping to create a robust data science models, which can be used in creating great business insights.

Datamites

Big Data Specialists

- Big Data Specialists focus on designing model for of Big Data Processing as one of the building block of Data Science workflow

- Their work involve architecture of Data gathering including streaming and snapshots, storing and processing in effective and efficiently manner using Big Data Technologies

Datamites

Researchers

- Researchers are domain experts. These professionals have more expertise in Data analysis, Data Science related Statistics along with significant expertise in specific domain such as HR, Marketing, Fraud Analytics, Health Care, Finance etc.,

- They have less emphasis and knowledge in backend part such as IT infrastructure, coding, computer science etc.,

Datamites

Data Science Analyst

- These professionals are focussed on day-to-day analysis of data, including website analytics, retrieving data from various data sources and creating data visualizations.
- They work closely with business person. Their role is to provide the reports from data analysis with appropriate visualizations in an easy to understand format.There by enabling the decision makers to gain valuable business insights

Datamites

Business Person

- This role is the one, who manages the entire Data Science project. We can call him as sponsor. This roles is predominantly business focussed and, formulates the problem statements for which, Data science project needs to find the answers. This role also helps in understanding, interpreting the intermediate results of the data science projects and drive the project to final solution.
- Though this role is primarily business focussed, he/she must also speak Data in order to able to perform the job well.

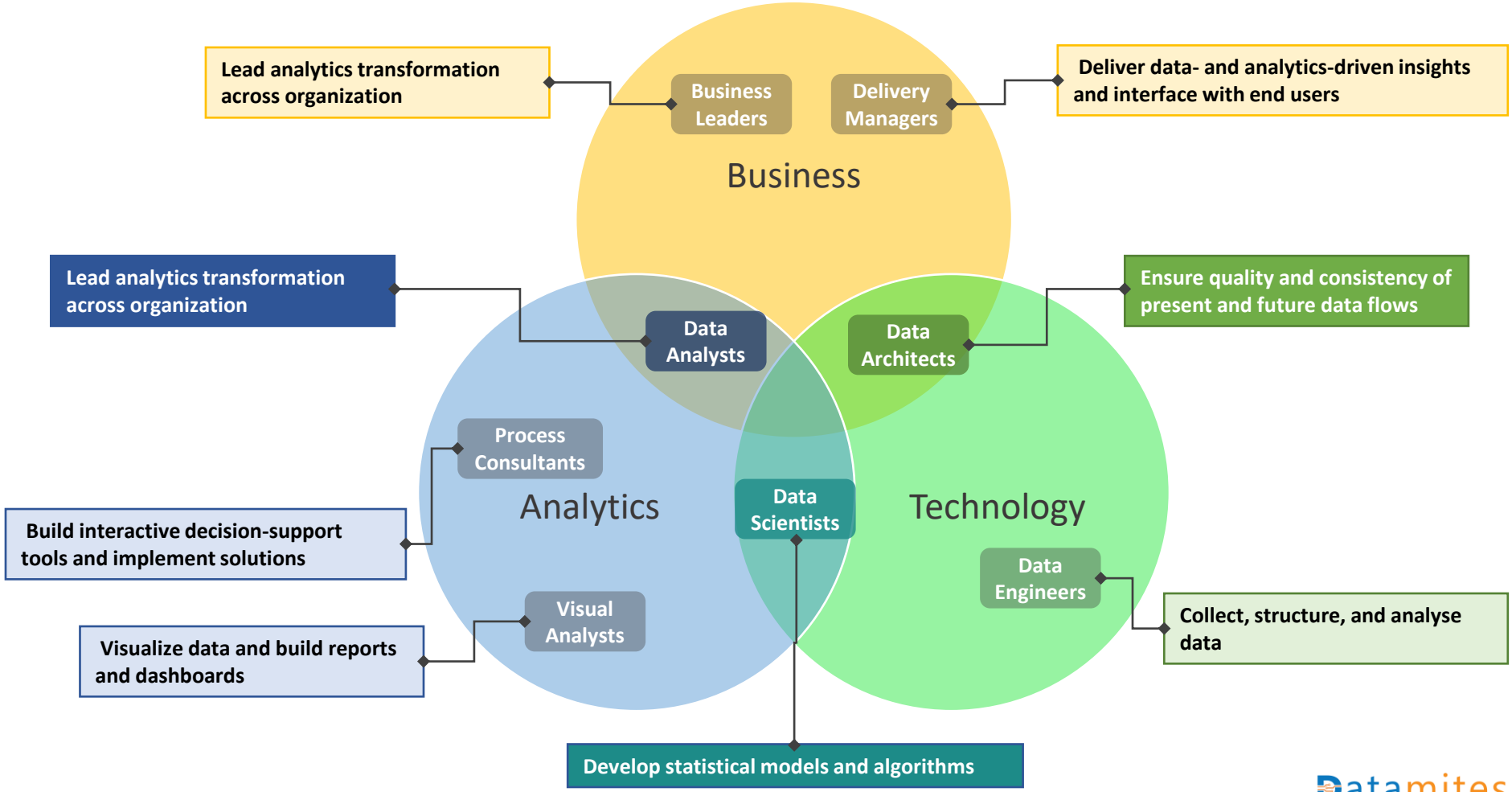
Datamites

Data Scientist

- These are the professional who is able to perform every aspect of Data Science, sometime called as full-stack Data science professionals. Well, these people are rare as mastering entire Data Science roles is difficult, if not impossible.

- If a company manages to hire full-stack Data Scientist, there would be tremendous progress in transforming the business in gain significant competitive advantage by finding solutions to important business questions.

Datamites



Datamites

Datamites