Introduction to Data Science

Outline

Introduction to Data Science

Data Science Applications

Data Science Project Life Cycle

Data Science Framework or Tools

Data Science Trends

Learning Outcomes

At the end of this topic, You should be able to;

Explain key concepts in data science, including tools, approaches, and application scenarios.

Data Science

Data Science refers to an emerging area of work that deals with the **collection, preparation, analysis, visualization, management, and preservation of large collections of information**.

Data Science Applications

- Internet Search
- Digital Advertisements (Targeted Advertising and re-targeting)
- Recommender Systems
- Image Recognition
- Speech Recognition
- Gaming
- Price Comparison Websites
- Airline Route Planning
- Fraud and Risk Detection
- Delivery logistics

What do data scientists do ?

"They need to identify insights in the data and then communicate them effectively to the business leaders."

- Define the question
- Define the ideal data set
- Determine what data you can access
- Obtain the data
- Clean the data
- Exploratory data analysis
- Statistical prediction/modeling
- Challenge results
- Distribute results to other people

Modern Data Scientist

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing package e.g. R
- Databases SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Custom reducers
- Experience with xaaS like AWS

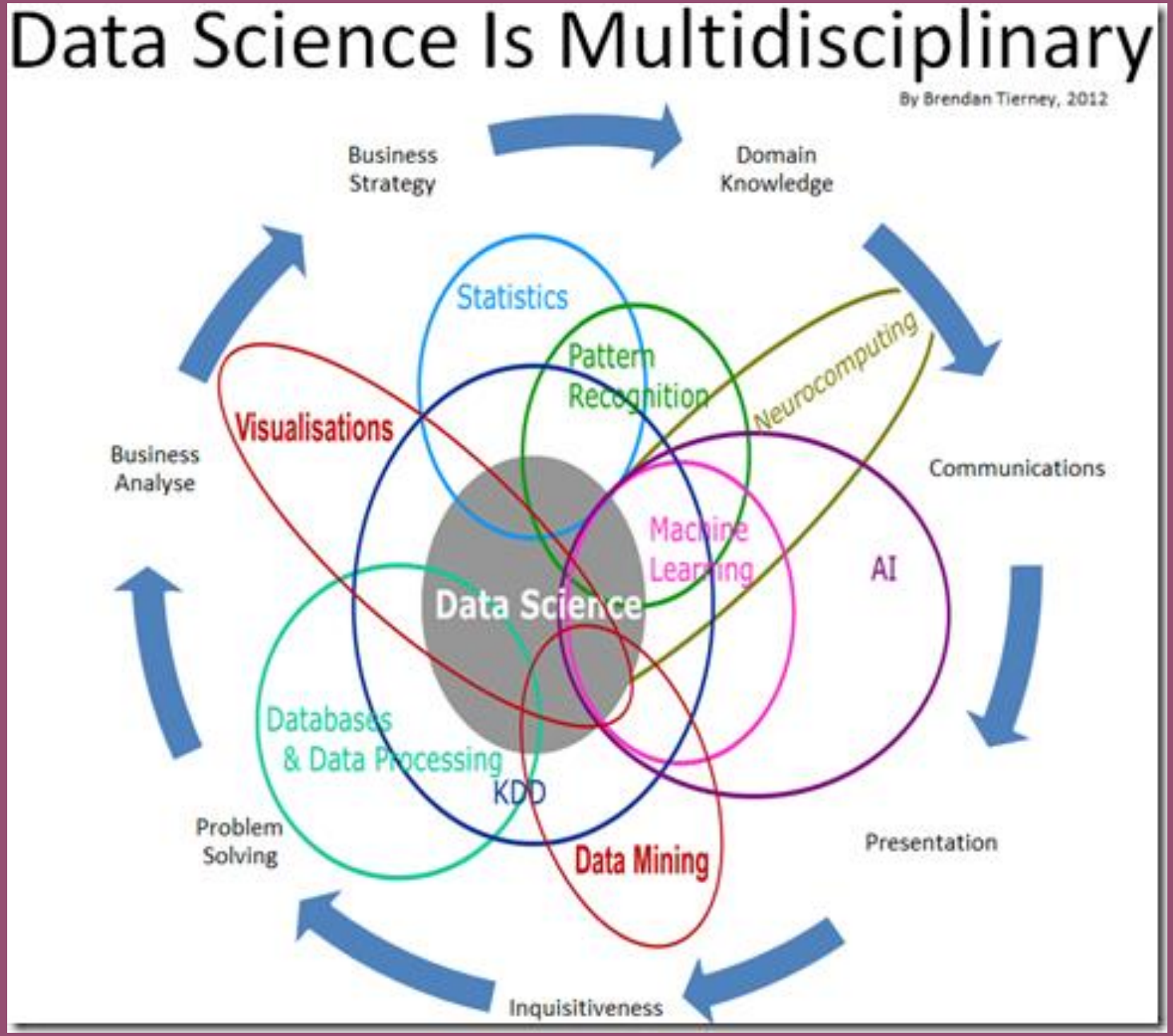
DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

Overview of the Data Scientists Skills



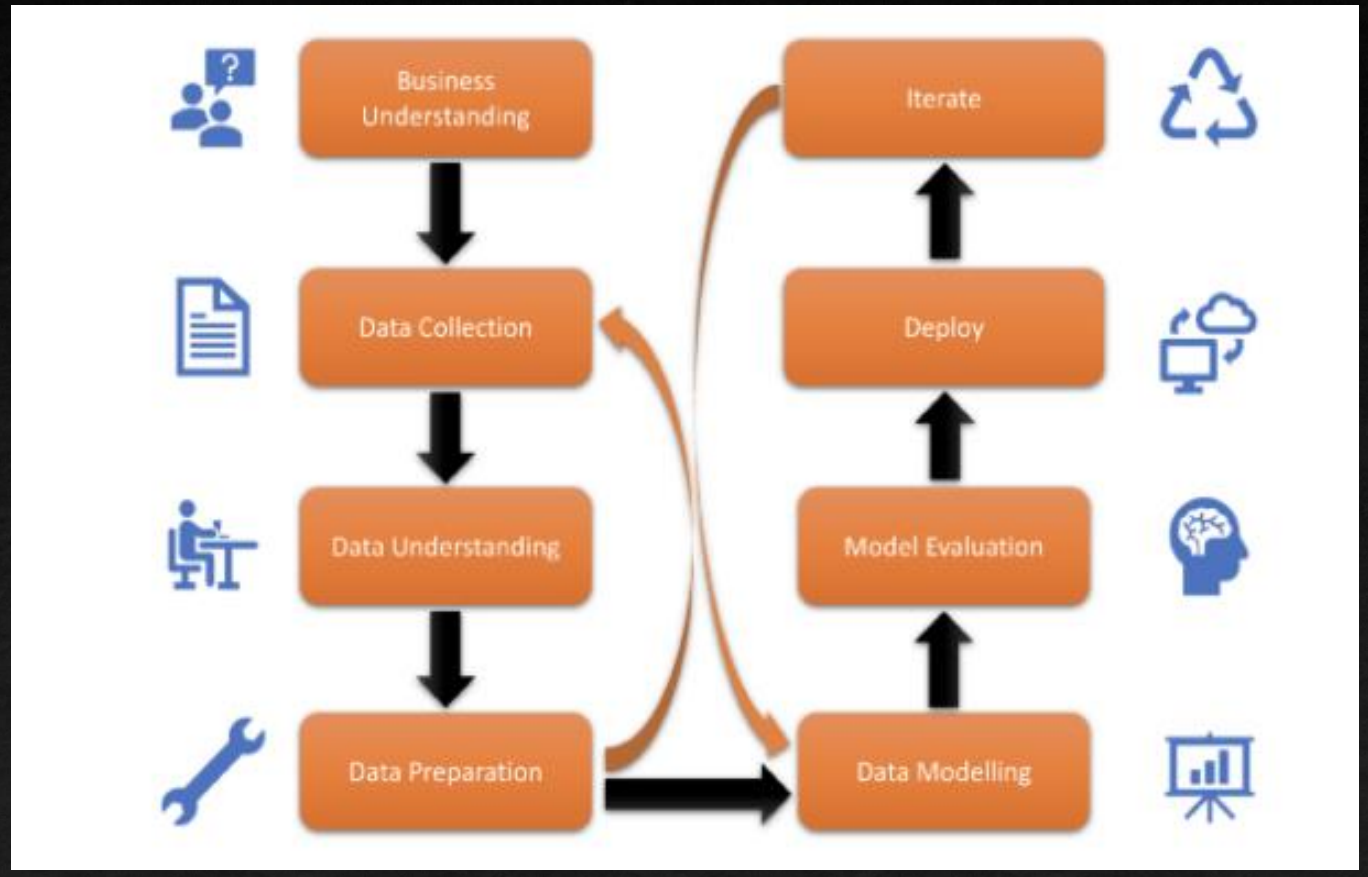
Skills to Learn in Order to Become a Data Scientist

- Domain Knowledge
- Statistics
- Data Wrangling
- SQL
- Data Visualizations & Storytelling
- Python and R Programming
- Machine Learning
- Deep Learning
- Big Data
- Unstructured Data
- Git - Version Control
- Dev-Ops
- Soft Skills

First Skill Second Skill Third Skill

Life Cycle of a Data Science Project

- Business Understanding
- Data Collection
- Data Preparation
- Data Modelling
- Model Evaluation
- Deploy and iterate



Source: https://towardsdatascience.com/life-cycle-of-a-data-science-project-3962b9670e5b

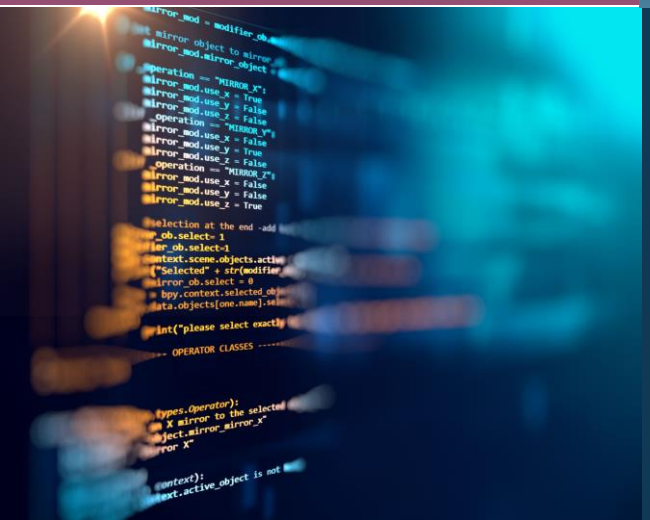
Data Science Toolkits

Machine Learning Libraries

- Sklearn
- Tensorflow
- Keras
- Pytorch

Big Data

- Hadoop
- Spark
- Cassandra
- MongoDB



Programming

- Jupyter Notebook
- Spyder
- Pycharm
- R Programming

Cloud Platforms

Iaas

Paas

Saas