EEE356 - Data Analytics (R)
Week 2: Introduction to Data Analytics

ADANA ALPARSLAN TÜRKEŞ
SCIENCE AND TECHNOLOGY UNIVERSITY

Dr Kasım Zor

Department of Electrical and Electronic Engineering

Spring 2021

Outline

Outline

Data

Datum: A piece of information or a fixed starting point of a scale or operation. (Singular) [Ref: Google]

Data:

- Information, especially facts or numbers, collected to be examined and considered and used to help with making decisions (Plural) [Ref: Cambridge English Dictionary]
- Data, in the information age, are a large set of bits encoding numbers, texts, images, sounds, videos, and so on. Unless we add information to data, they are meaningless. When we add information, giving a meaning to them, these data become knowledge [1].

Data Analytics

Data Science: Data science is an exciting discipline that allows you to turn raw data into understanding, insight, and knowledge [2].

Data Analytics: The science that analyse crude (raw) data to extract useful knowledge (patterns or insights) from them. This process can also include data collection, organisation, preprocessing, transformation, modelling and interpretation [1].

Outline

A Short Taxonomy of Data Analytics – Part 1 [1]

Instance or Object: Examples of the concept that will be characterised.

Contact Person	Phone Number	E-Mail Address
Person 1	Number 1	Address 1
Person 2	Number 2	Address 2
Person 3	Number 3	Address 3

Table 1: Data set of a contact list

Attribute or Feature: Attributes, also called features, are characteristics of the instances.

A Short Taxonomy of Data Analytics – Part 2 [1]

Descriptive Analytics: Summarise or condense data to extract patterns. In descriptive analytics tasks, the result of a given method or technique is obtained directly by applying an algorithm to the data. The result can be a statistic, such as an average, a plot, or a set of groups with similar instances.

Predictive Analytics: Extract models from data to be used for future predictions.

A Short Taxonomy of Data Analytics – Part 3 [1]

Method or Technique: A method or technique is a systematic procedure that allows us to achieve an intended goal. A method shows how to perform a given task. But in order to use a language closer to the language computers can understand, it is necessary to describe the method/technique through an algorithm.

Algorithm: An algorithm is a self-contained, step-by-step set of instructions easily understandable by humans, allowing the implementation of a given method. They are self-contained in order to be easily translated to an arbitrary programming language.

A Short Taxonomy of Data Analytics – Part 4 [1]

Model: A model in data analytics is a generalisation obtained from data that can be used afterwords to generate predictions for new given instances. It can be seen as a prototype that can be used to make predictions. Thus, model induction is a predictive task.

A Short Taxonomy of Data Analytics – Part 5 [1]

Hyperparameters and Parameters: Assume that in the induction of a model, there are both hyperparameters and parameters whose values are set. The values of the hyperparameters are set by the user, or some external optimisation method. The parameter values, on the other hand, are model parameters whose values are set by a modelling or learning algorithm in its internal procedure. When the distinction is not clear, the term parameter is used.

Outline

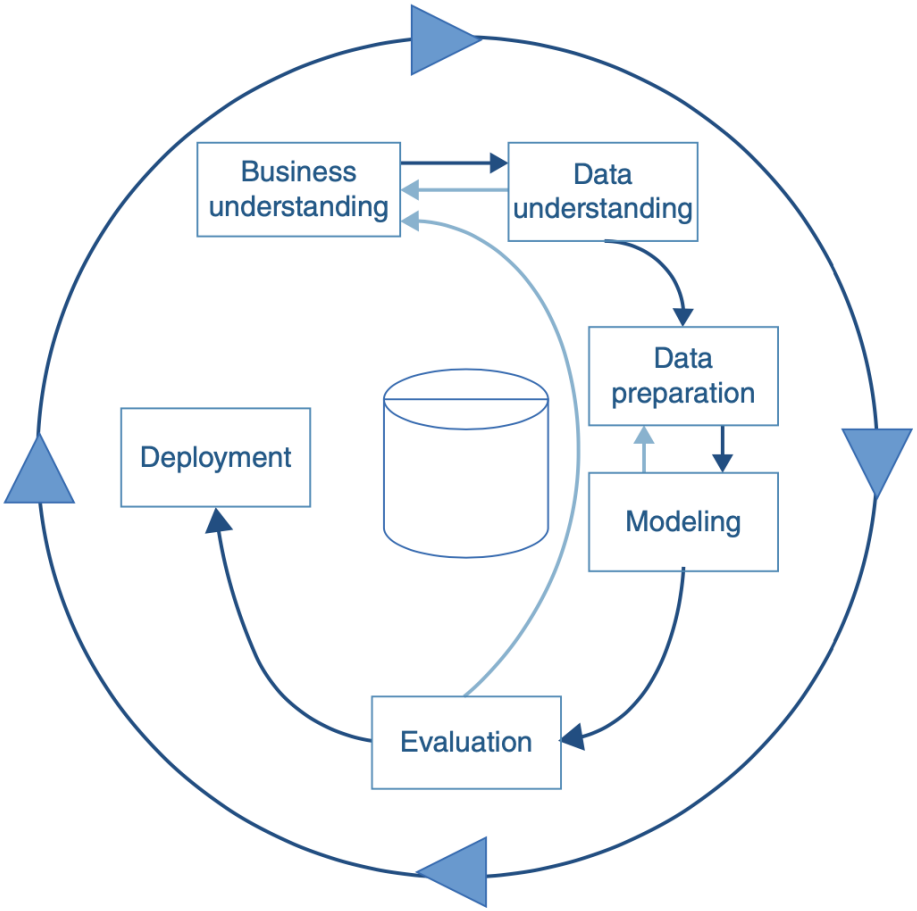
KDD Process (Academia) [1]

Knowledge **D**iscovery in **D**atabases

1. Learning the application domain
2. Creating a target dataset
3. Data cleaning and preprocessing
4. Data reduction and projection
5. Choosing the data mining function
6. Choosing the data mining algorithm(s)
7. Data mining
8. Interpretation
9. Using discovered knowledge

CRISP-DM Methodology (Industry) [1]

CRoss-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining



Outline

Types of Data Analytics [3]

Descriptive, Predictive, and Prescriptive Analytics

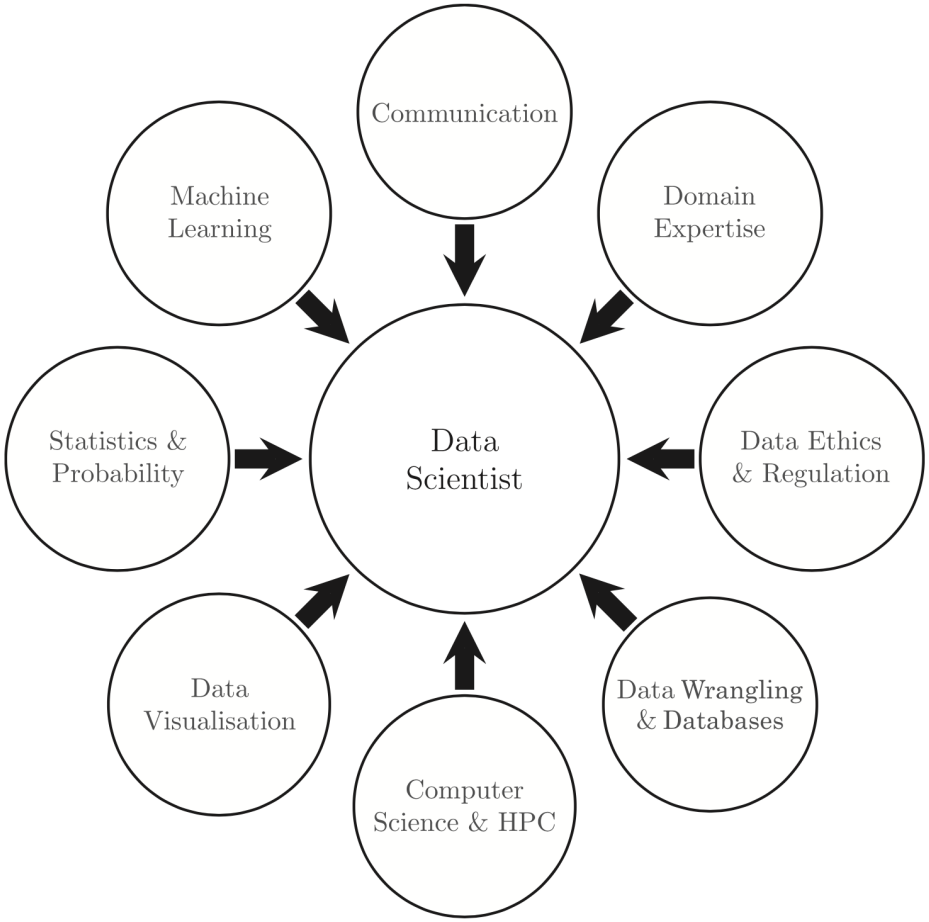
	Descriptive	Predictive	Prescriptive
	What HAS happened?	What COULD happen?	What SHOULD happen?
What the user needs to DO	• Increase asset reliability • Reduce labor and inventory costs	• Predict infrastructure failures • Forecast facilities space demands	• Increase asset utilization • Optimize resource schedules
What the user needs to KNOW	• The number and types of asset failures • Why maintenance costs are high • The value of the materials inventory	• How to anticipate failures for specific asset types • When to consolidate underutilized facilities • How to determine costs to improve service levels	• How to increase asset production • Where to optimally route service technicians • Which strategic facilities plan provides the highest long-term utilization
How analytics gets ANSWERS	• Standard reporting - What happened? • Query/drill down - Where exactly is the problem? • Ad hoc reporting - How many, how often, where?	• Predictive modeling - What will happen next? • Forecasting - What if these trends continue? • Simulation - What could happen? • Alerts - What actions are needed?	• Optimization - What is the best possible outcome? • Random variable optimization - What is the best outcome given the variability in specified areas?
What makes this analysis POSSIBLE	• Alerts, reports, dashboards, business intelligence	• Predictive models, forecasts, statistical analysis, scoring	• Business rules, organization models, comparisons, optimization

Business value →

Outline

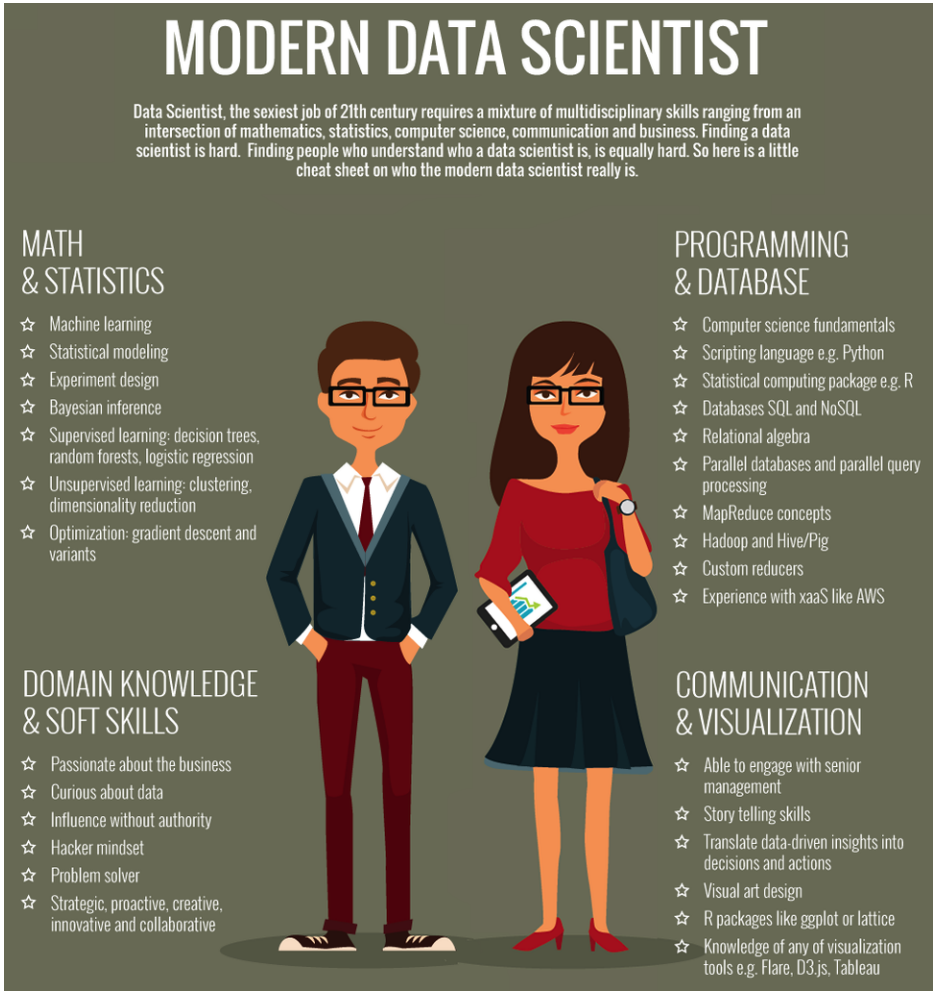
Data Scientist [4]

A Skills-Set Desideratum for a Data Scientist



Data Scientist [5]

Who is a Data Scientist?



References I

[1] Joao Mendes Moreira, Andre C. P. L. F. de Carvalho, and Tomas Horvath. *A General Introduction to Data Analytics*. John Wiley & Sons, 2019. ISBN 978-1-119-29625-6.

[2] Hadley Wickham and Garrett Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, 2017. URL https://r4ds.had.co.nz.

[3] IBM Watson. Descriptive, predictive, and prescriptive analytics, 2015. URL https://gemba.nl/wp-content/uploads/2015/10/watsonbusinessvalue.png.

[4] John D. Kelleher and Brendan Tierney. *Data Science*. The MIT Press, 2018. ISBN 978-0-262-53543-4.

[5] Marketing Distellery. Who is a data scientist?, 2018. URL https://miro.medium.com/max/1040/1*T5GfsoZ-IWK3rcVkZ7R2bw.png.