MTAT.03.319

Business Data Analytics

Lecture 1: Introduction

Rajesh Sharma
Email: rajesh.sharma@ut.ee

Warm-up question

- We are a charity. We have a database containing 100K donors who have donated in the past 10 years. We know their basic demographics, address and how much they have donated in past.

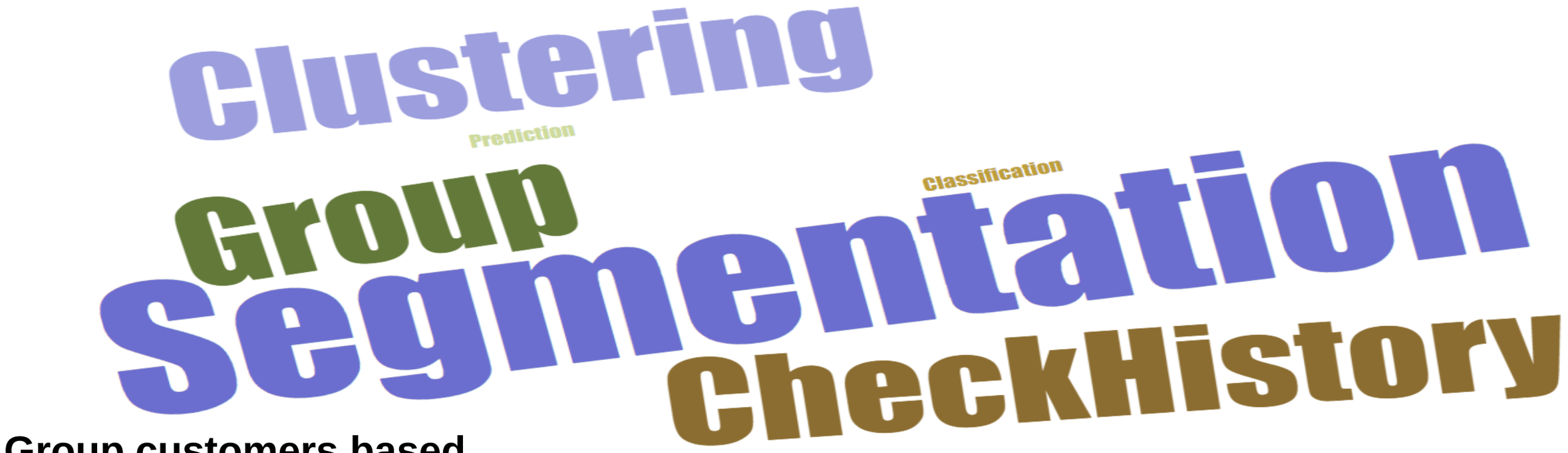
- Sending a snail-mail (using postcards) asking for donation costs 60 cents/piece. When we mail out, the average donation comes at about 80 cents.

- **Questions:**

  - Should we send a (postal) mail to all 100K donors?

  - How can we efficiently solve this problem ?

Suggestions …



Group customers based on localities

Talk to postal team and get cheap postcards or delivery prices

Contact partners of rich husbands and wives

Outline

- What is Business Data Analytics (BDA) ?
- Applications areas of BDA ?
- How BDA works ?
- Course Logistics

What is Business Data Analytics?

Business Analysis

vs

Business **Data** Analytics

BA Vs. BDA

- **Business Analysis**

- Analyzing the problem and mapping it to the solution space (without the data)

- Qualitative in nature

- When data is not available or cannot be get (time constraint)

- Eg: Put adv on street for donation.

- **Business DATA Analytics**

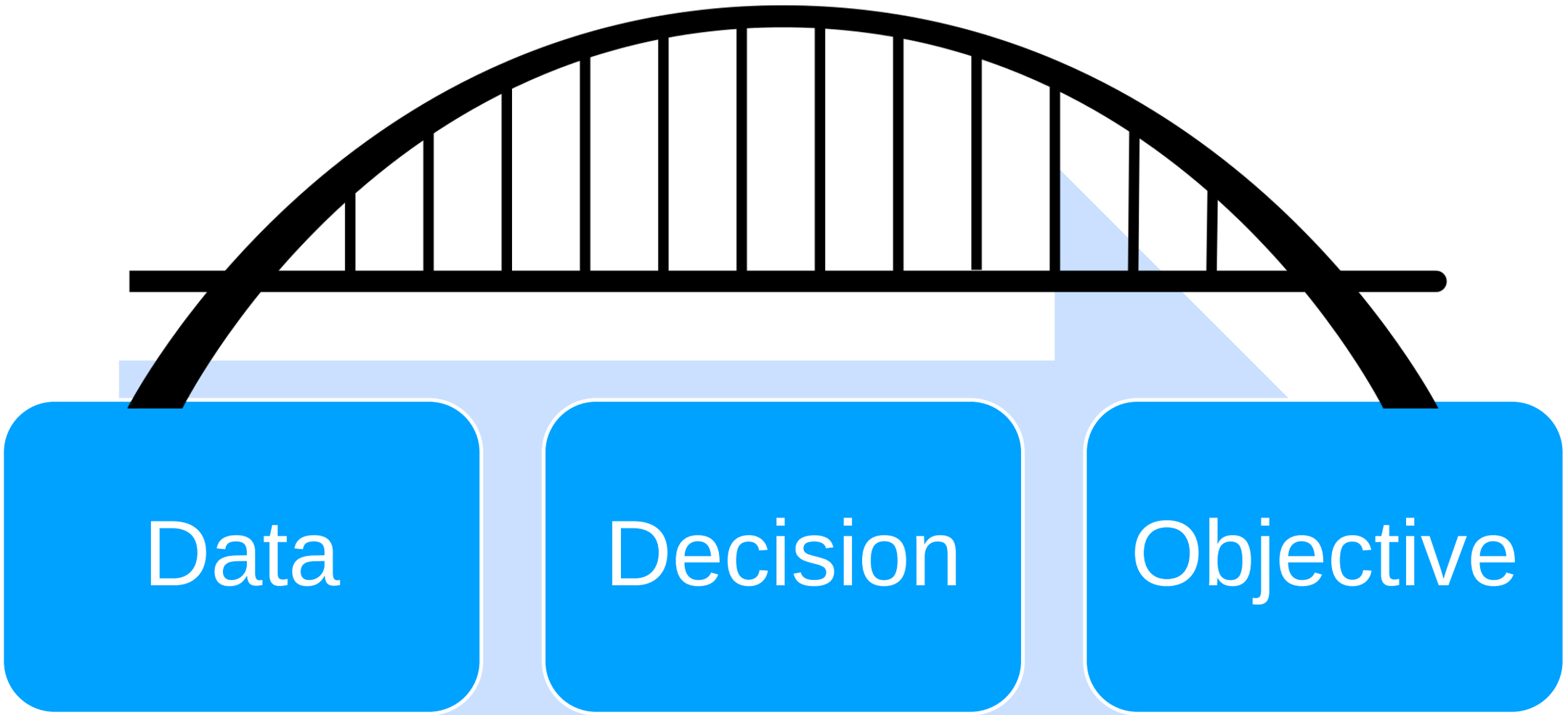
- Mapping of the problem to the solution space based on the data analysis

- Quantitative in nature

- Take actions based on the data.

- Eg: Send mails to people who have donated in the past based on analysis.

Business Data Analytics

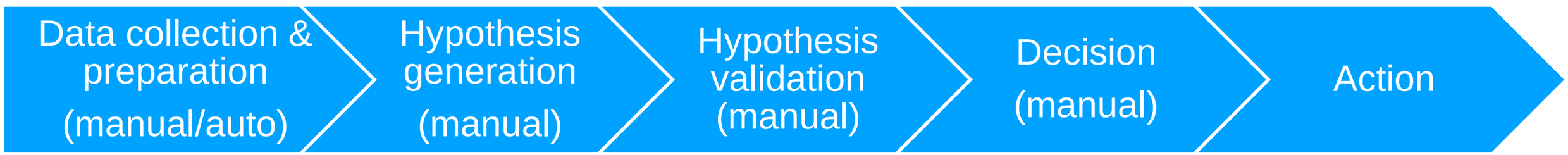


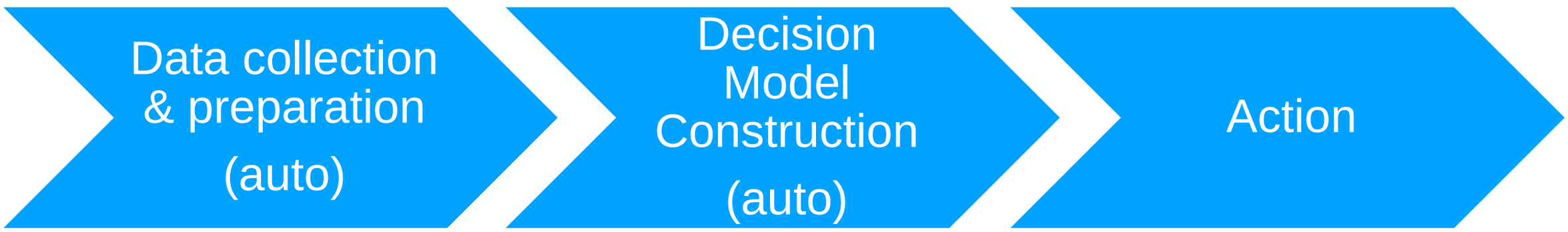
The application of repeatable methods to use data
in order to inform or make decisions
to achieve or maximize a business objective

BA Vs. BDA: Problem perspective

- Do 18-30 y.o. male customers living in cities of 50K+ inhabitants buy more of product X than others?
- In which zip codes have the sales of product X increased the most in the months of June-July?
- If a customer buys product X, which other products might he/she buy?
- Are customers whose number of calls in the past 3 months is lower than the 12-months average more likely to churn in the next 6 months?
- Which customers are most likely to buy product X?
- How much we will sell of products X in each zip code next month?
- For each customer, which product are they most likely to buy given their current shopping cart?
- ???

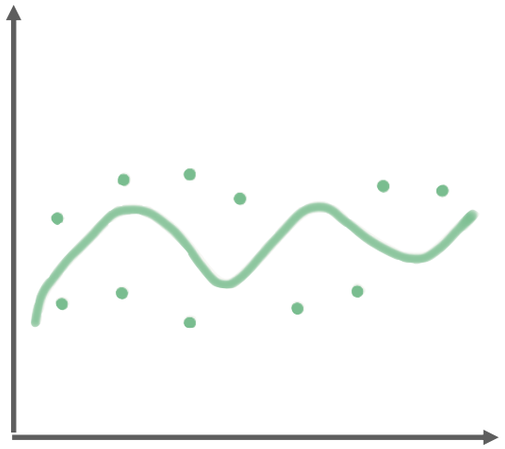
Business Analysis vs. Business Data Analytics





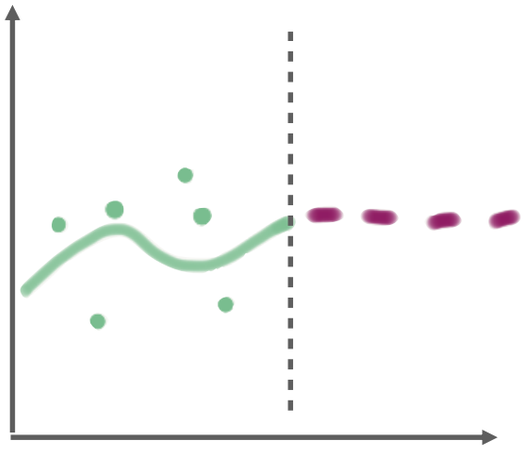
Types of Business Analytics

Descriptive



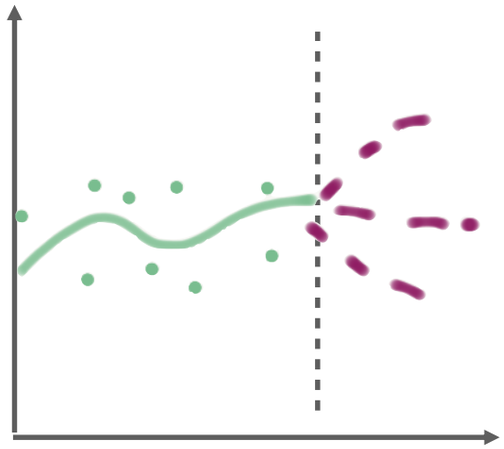
What has happened?
E.g. what top five customer segments we have? Which pairs of products are bought together?

Predictive



What will happen?
E.g. Who will buy? Who will churn?

Prescriptive



What to do to achieve my goals?
When should I make my next customer call, to whom and what should I tell them?

Typical classes of analytics questions

- Clustering: Which objects belong together/are similar?
- Classification: Who likes X? Who will do X? …
- Regression: How much will X spend/buy/earn/donate …?
- Recommendation systems: Which products can be placed together?
- Forecasting: What will be the price/cost of X in future? What will be the average expenditure/income, etc. of a population in future?
- Link Analysis: Which are most important people in a network of customer transactions?
- Opinion mining: What is the brand value of the company?

Other related terms

- Business Intelligence
  - Dashboard and report extraction from enterprise databases for monitoring, data exploration and hypothesis validation
- (Statistical) Data Analysis
  - Very useful for exploration & hypothesis validation; works with “small data”; driven by models
- Data Mining: Extracting patterns from data automatically
  - Business analytics applies data mining to business problems
- Machine learning
  - The algorithms behind data mining
- Data Science
  - Catch-all term for data analysis, analytics, mining, machine learning, deep learning, Big Data processing, network science, opinion mining, and other techniques

Outline

- What is Business Data Analytics (BDA) ?
- Applications areas of BDA ?
- How BDA works ?
- Course Logistics

When & where to use business data analytics?

Application Areas of Business Data Analytics

CRM

Customer Relationship Management

BPM and ERP

Business Process Management

Enterprise Resource Planning

BRM

Business Risk Management

BDA in CRM

Customer segmentation:	Behavioral Product/category based Brand based
Product Recommendations	Up sell Cross sell Next sell
Customer Lifecycle Management	Acquisition Retention Win-back

Customer Lifecycle (Linoff & Berry)

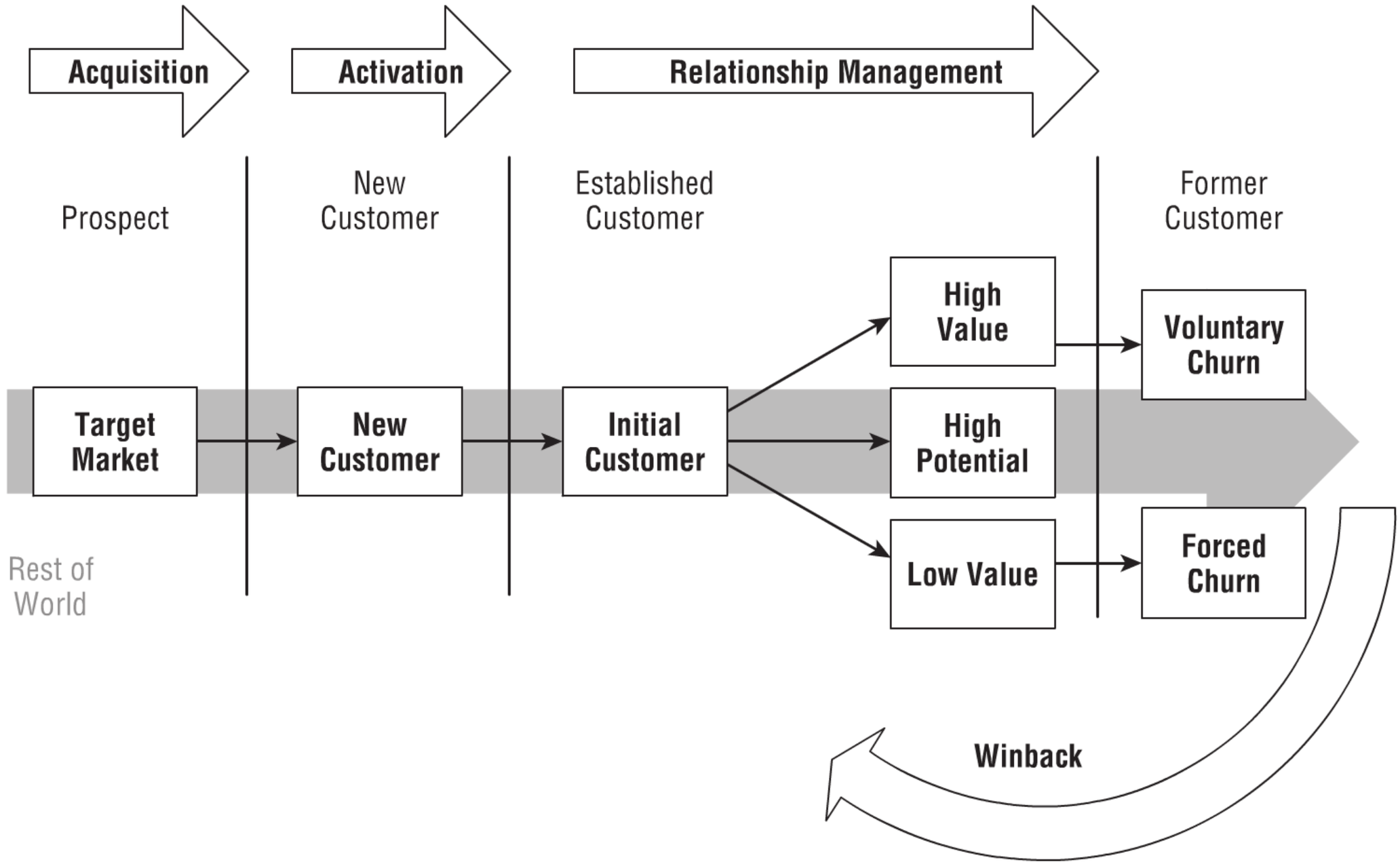


Figure 2-1: The customer lifecycle progresses through different stages.

BDA in BPM and ERP

- Discover how a process is actually executed (BPM)
  - Including deviations with respect to how it should be done
- Identify root causes of poor process performance (BPM)
  - E.g. customer complains, defects, rework, delays, etc.
- Predict that a given execution of a process will end up in an undesirable outcome (BPM)
  - E.g. predict delays at runtime, before they happen
- Enterprise Resource Planning
  - HR/Talent management

Identifying High Potential Talent: A Neural Network based Dynamic Social Profiling Approach , ICDM 2019 (using Graph Neural Networks)

BDA in BRM

Strategic and compliance	industry changes? economic stability merger and acquisition activity
Financial and operational	supply chain **fraud detection** investments, portfolio credit scoring

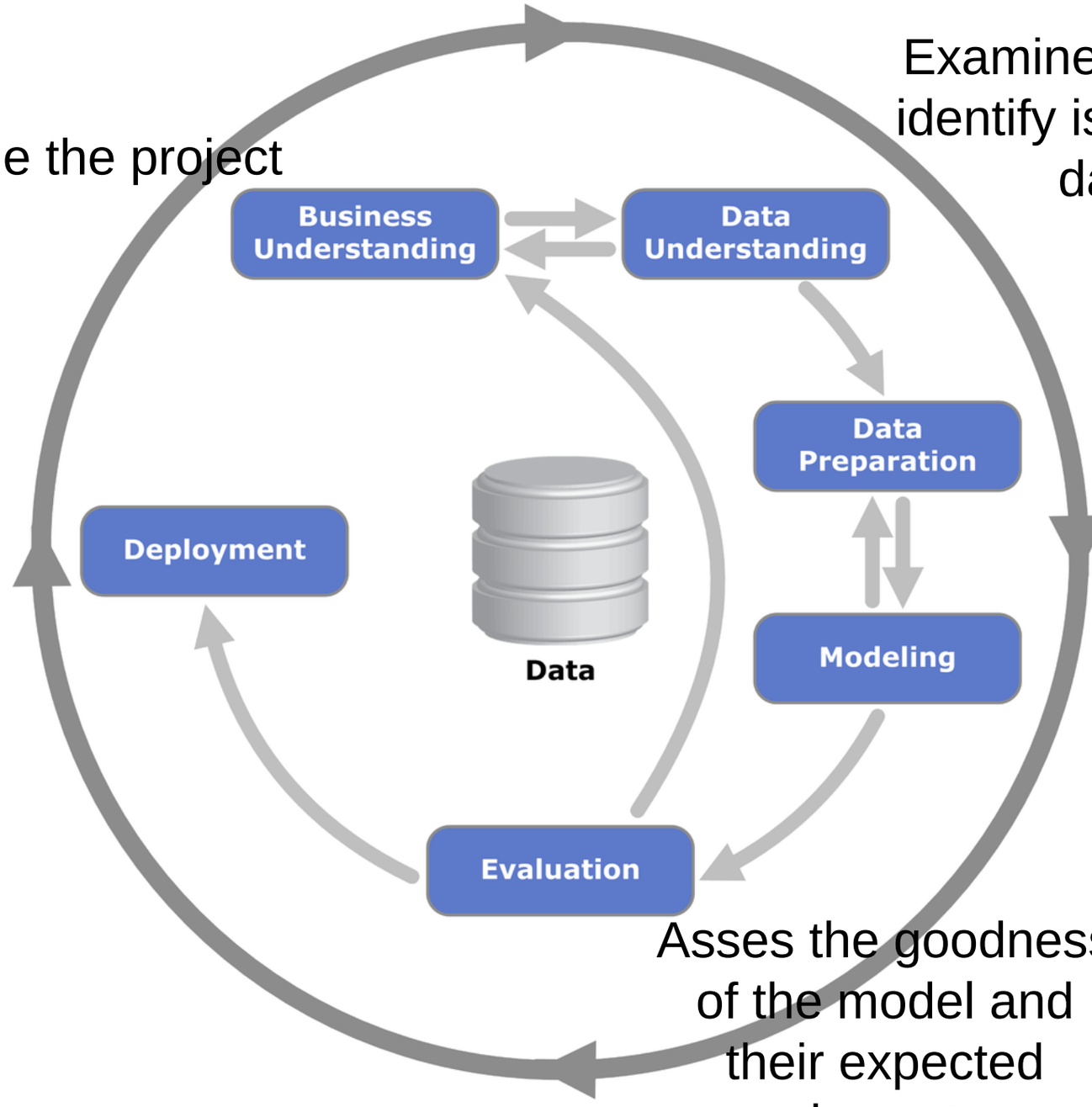
Application Areas of Business Data Analytics

CRM	**BPM and ERP**	**BRM**
Customer Relationship Management	Business Process Management & Enterprise Resource Planning	Business Risk Management
✓ Marketing ✓ Sales, ✓ **Customer Lifecycle Management (CLM)**	✓ Process optimization ✓ Inventory management, ✓ HR/Talent management, ✓ Facilities management	✓ Risk management ✓ Underwriting

Outline

- What is Business Data Analytics (BDA) ?
- Applications areas of BDA ?
- How BDA works ?
- Course Logistics

CRISP-DM
Cross-Industry Standard Process for Data Mining



Define the project

Examine the data; identify issues with data

Fix data issues; create derived variables (features)

Build models

Asses the goodness of the model and their expected impact

Put models and insights into use

Who is involved?

- Business sponsor: CEO, VP
- Domain expert(s): Financial expert, Biomedical expert etc.
- Analytics expert: Data Scientist
- Data steward & database expert: Knows company’s DB very well.

Matching: CRISP-DM Steps and Roles

Domain Experts

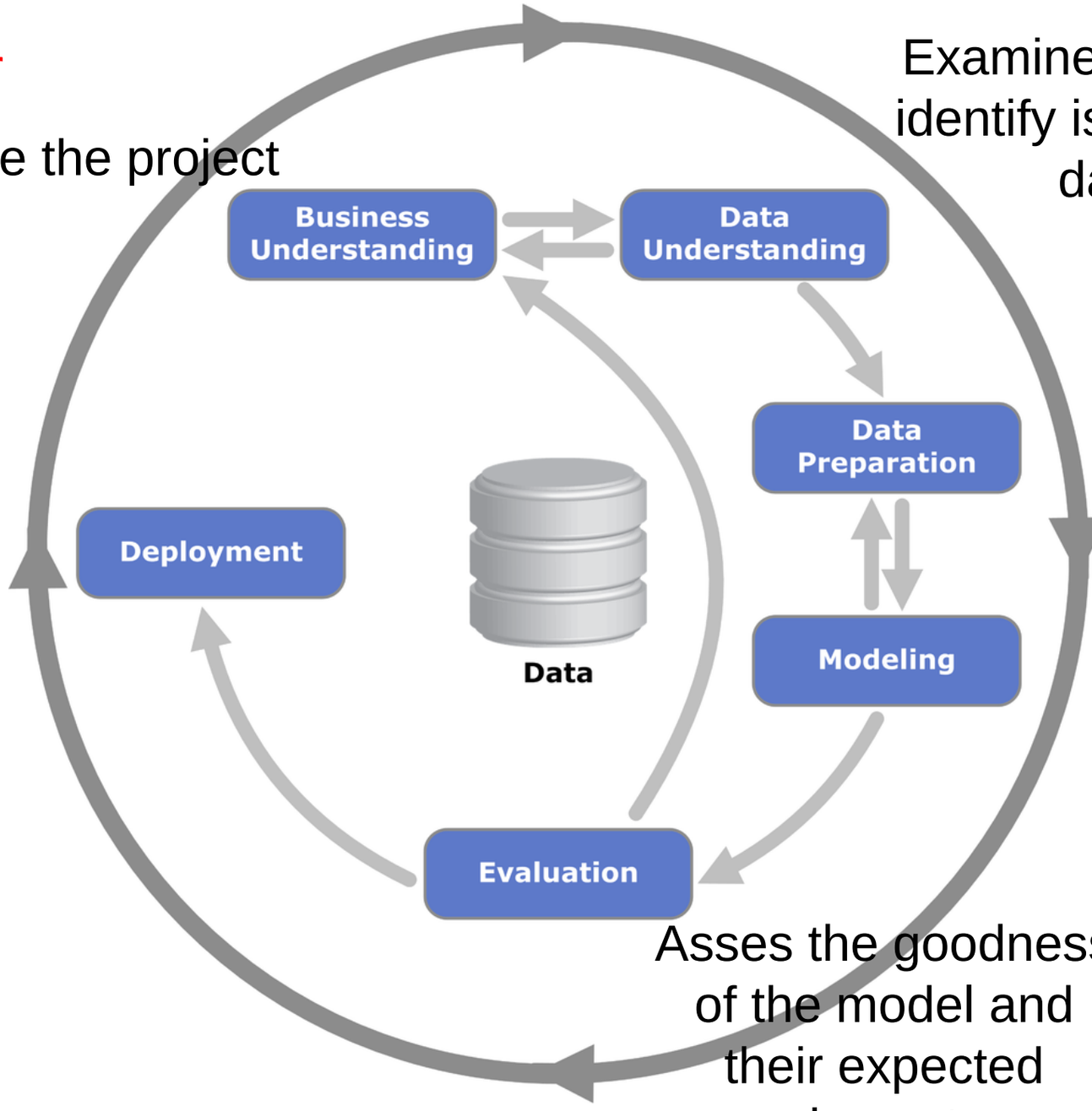
Data steward

Analytics expert

Business sponsor

Examine the data; identify issues with data

Define the project



Data steward

Analytics expert

Fix data issues; create derived variables (features)

Put models and insights into use

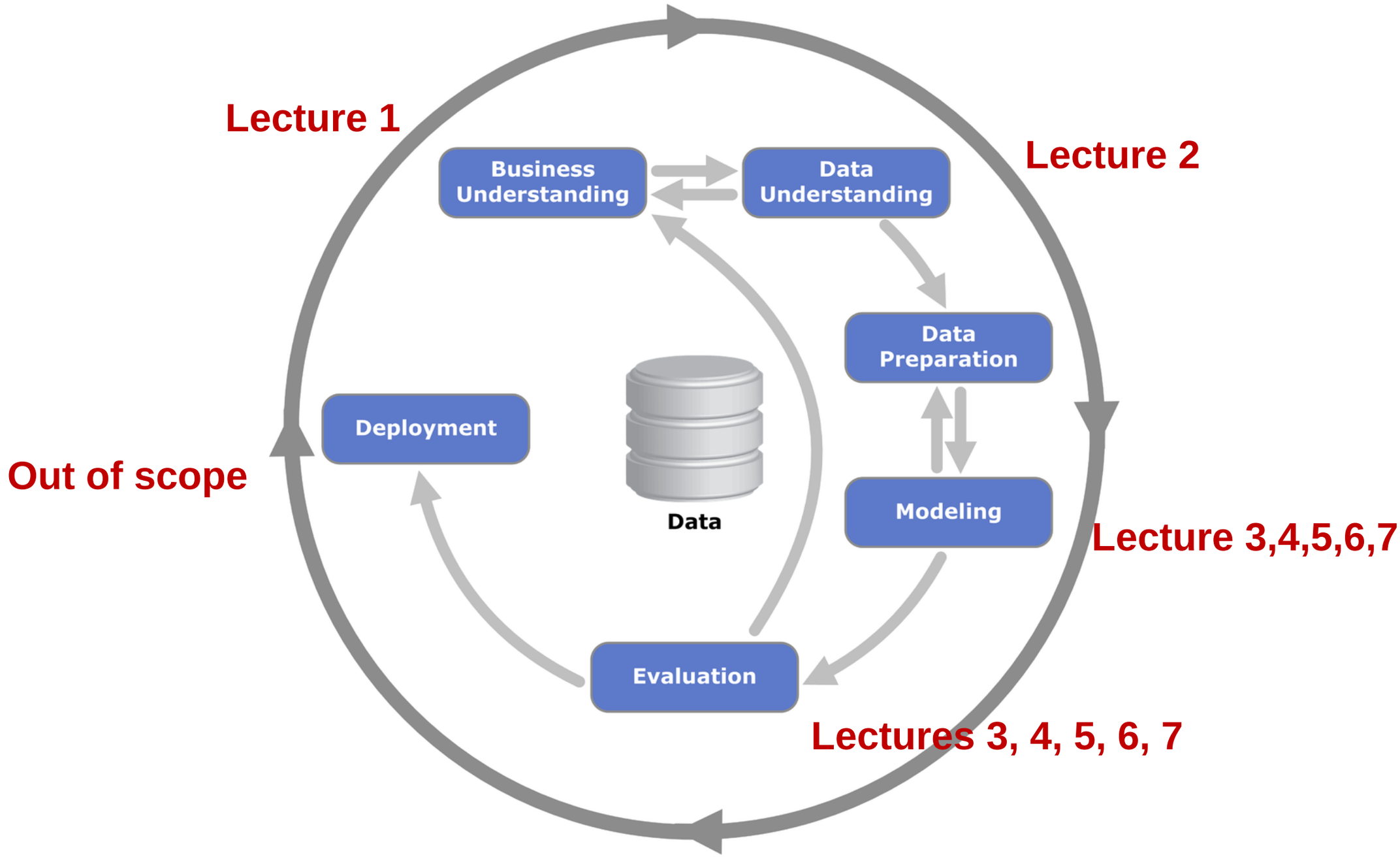
Build models

Analytics expert

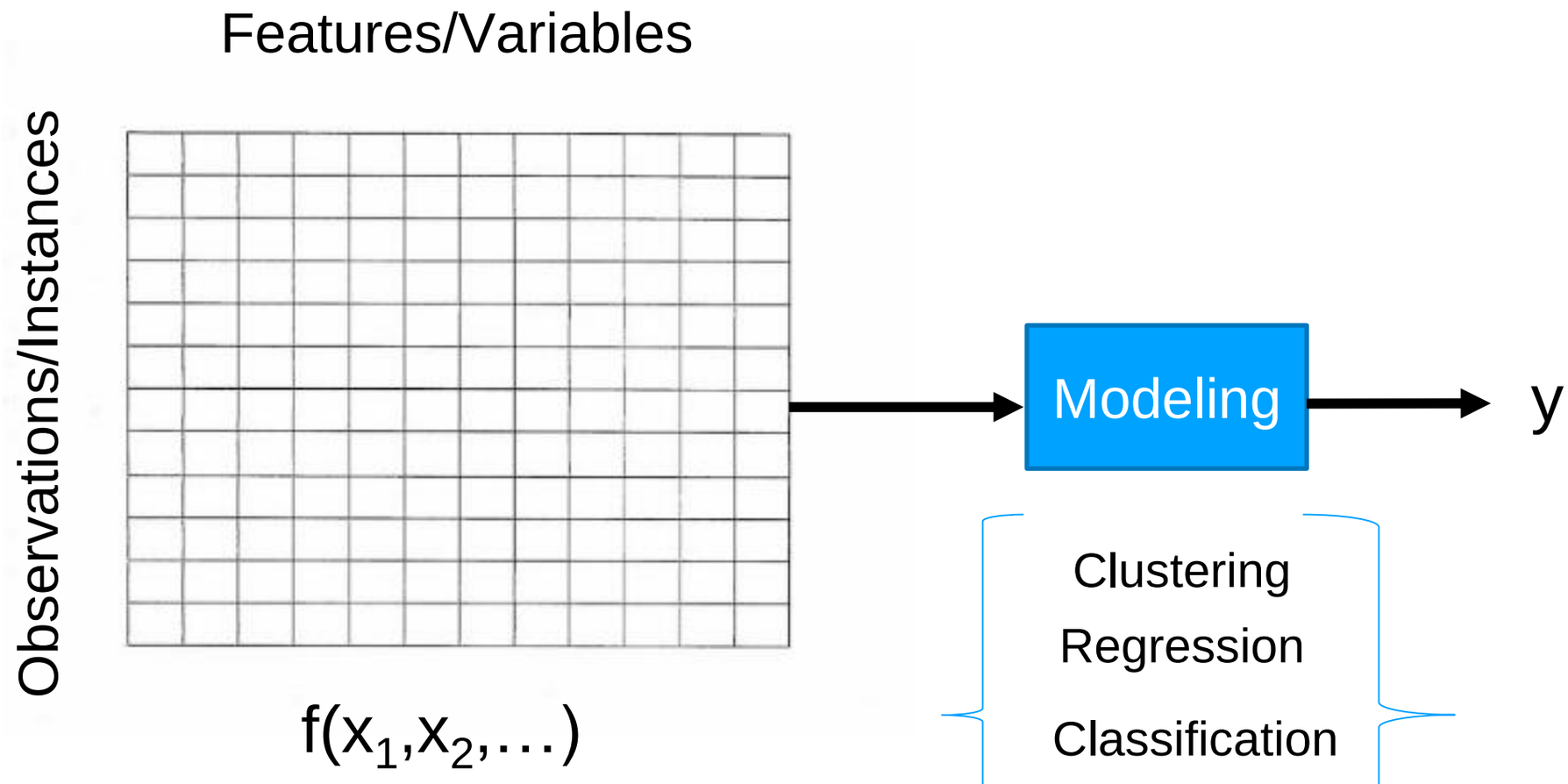
Asses the goodness of the model and their expected impact

Analytics expert

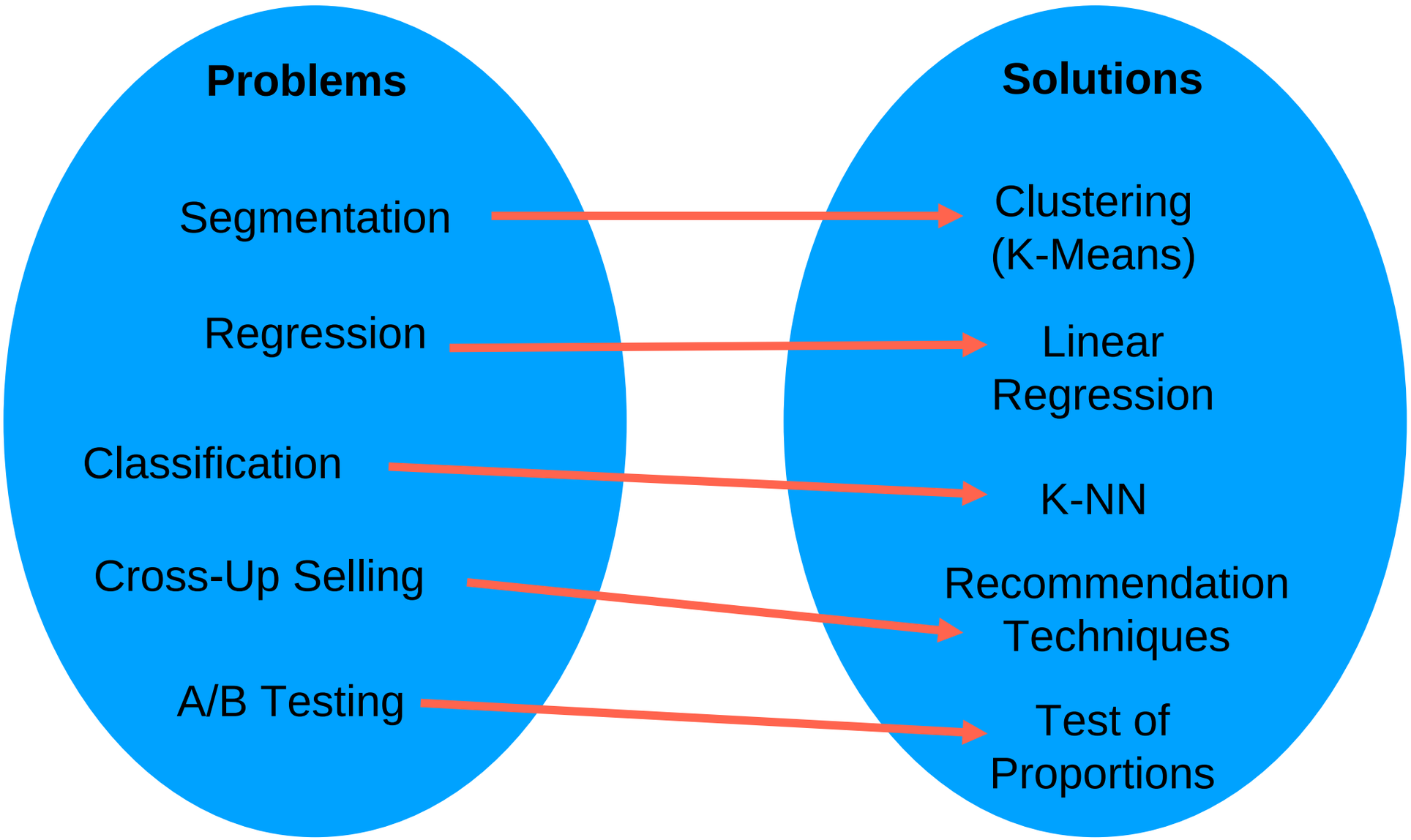
CRISP-DM & Course Structure



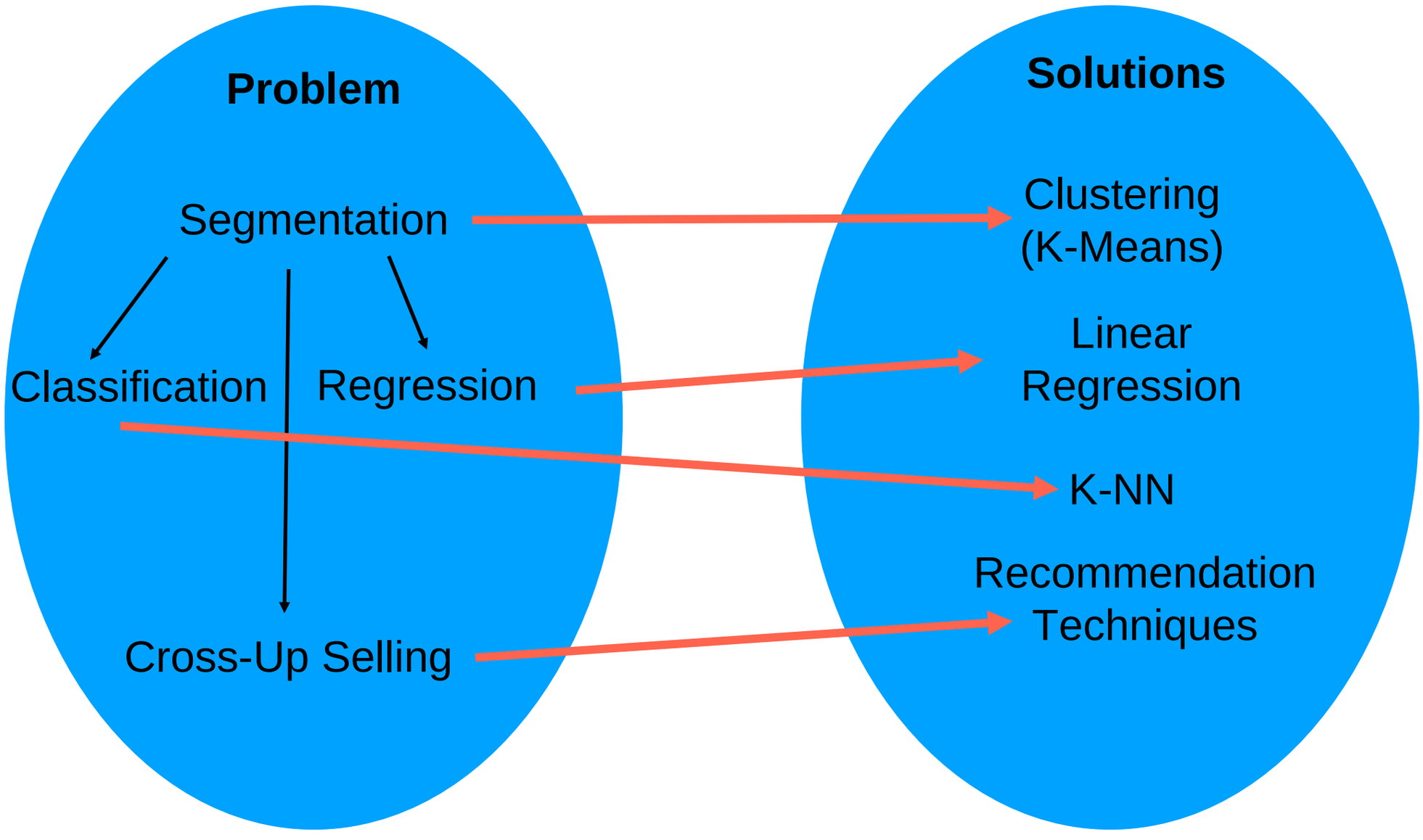
Modeling


Features/Variables
Observations/Instances
$f(x_1,x_2,\ldots)$
Modeling
y
Clustering
Regression
Classification
Recommd.

During the course: Bipartite view



At the end of the course: mesh structure



Back to warm-up question

- We are a charity. We have a database containing 100K donors who have not donated in the past 12 months. We know their basic demographics, address and how much they have donated in past (and when). We do not have emails.

- Sending a mail asking for donation costs 60 cents/piece. When we mail out, the average donation comes at about 80 cents.

- Should we send a (postal) mail to all 100K donors?

Outline

- What is Business Data Analytics (BDA) ?
- Applications areas of BDA ?
- How BDA works ?
- Course Logistics

Thanks

Outline

- What is Business Data Analytics (BDA) ?
- Applications areas of BDA ?
- How BDA works ?
- Course Logistics

Lecture Topics

Introduction
Descriptive analysis and visualization
Customer segmentation
Customer Lifecycle management - regression problems
Customer Lifecycle management - classification problems
Cross-sell/Up-sell recommendations
A/B testing in marketing

Course Logistics

- **Moodle:** https://moodle.ut.ee/
  On moodle, you will get all the information.

- Please do NOT follow https://courses.cs.ut.ee/2021/BDA/fall

- Course will have three important elements:
  - Problem oriented course
  - Theory
  - Practical

- We will upload material before 10.15 am on Thursdays, the actual timing of the Lecture.

- It is an applied course. We will use already existing methods to use them in different problems

- **It is not a programming course**: We compliment theory with Labs, but main intension is not to teach you programming or spoon feed you. We will try our best to help you. For example, during practice session, we demonstrate you coding examples, which will help you in homeworks. We cannot demonstrate everything in practice. Some questions in the homeworks might not be covered in the lab and expect they could be challenging.

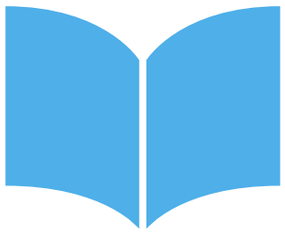
Course Logistics

- Lab sessions:
  - We will discuss enough material for you so that you can work on homeworks. Some questions might be similar to lab sessions and some might challenge.
  - You can ask questions about homework through moodle but do let us know what you did (for example, what Googling you did) before asking the questions. Also be patient for the reply (weekends, night … difficult to reply)
  - Homeworks' Feedback: Limited resources, we cannot give personalized feedback. But we will let you know why we cut the marks. Time in ~ 2 weeks.

- Send an email: rajesh.sharma@ut.ee

- **Exam dates:** Wait and Watch – COVID situation (In January 2022). During every class we will give you enough pointers. Last lecture is about exam preparation, where we will show some sample questions to give you an idea what kind of questions you can expect.

Course Structure & Assessment



Homeworks. 70 points in total

Deadline = At 24.00 ~ 2 weeks.
Example: Thursday 10 September, deadline: 23 September, 24 midnight. (Hard Deadline)
You can submit as Individual or in pair (2)

Theory exam

30 points.
Questions based on discussion in Lecture and Lab (no coding questions will be asked)

Passing condition

Condition 1: at least 12 marks to pass the course irrespective of the marks you receive in homeworks.
Condition 2: You need 51 points at least (theory+homeworks) to pass the course.

Some questions (based on previous years)

- Q: Can we get easy homeworks because I am not from computer science department?
  A: No, for us everyone is equal irrespective of your background.

- Q: Can you provide answers to homeworks in terms of graphs or some values in advance … so it is clear what kind of graphs we have to make.
  A: No, Questions are open ended. This is not a basic programming course.

- Q: I could not submitted homeworks, can I submit a Project.
  A: Sorry, No

- Q: I do not want to give theory exam, can I submit a Project
  A: Sorry, No

- Q: I slept at 11.30 pm and forgot to submit it on time. Can I send you next day by email ?
  A: Sorry, No.

- Q: My total homework marks are low, and I want A, can I do extra assignments to compensate that?
  A: Sorry, No.

- Q: My R is better than Python. Can I submit in R?
  A: Sorry, No.

Some questions (based on previous years)

- Q: I missed video lectures, can I just go through slides and this is enough for Theory exam?
  A: No, Sometimes we discuss elaborately in lectures. So you are strongly advice to go through video lectures.

- Q: I never went through course logistic slides.
  A: We assume you went through these slides and course logistic video as it is part of course.

- Q: I am not getting emails regularly but my course mates are getting it.
  A: We sent emails about course logistics and updates regularly at your email id which you have entered in SIS/OIS (Student Information System). So if you are not getting it enter email which you check often and also check your SPAM folder (we will discuss about it as a use case in one of the lectures)

- Q: What is the response rate on Moodle ?
  A: Once a day and no response on weekends.

- Q: The exam in Course “X” is open book, why not this one (asked 2 days before the exam).
  A: Every course is designed differently.

Lab sessions and homeworks

- Why Python ?
  - Courses in the Institute of Computer science have moved to Python.
  - Production site
  - Taking over R in the market.

Online Platform: Colab (From google)

- You need to have a google account
  Please make one before the lab starts.
- You need to be connected (internet) when you are working
- You don’t have to worry about packages or OS.
- Colab: colab.research.google.com
- Hosted by google, cloud system, interactive session, Jupyter interface

Thanks !