Department of Mathematics
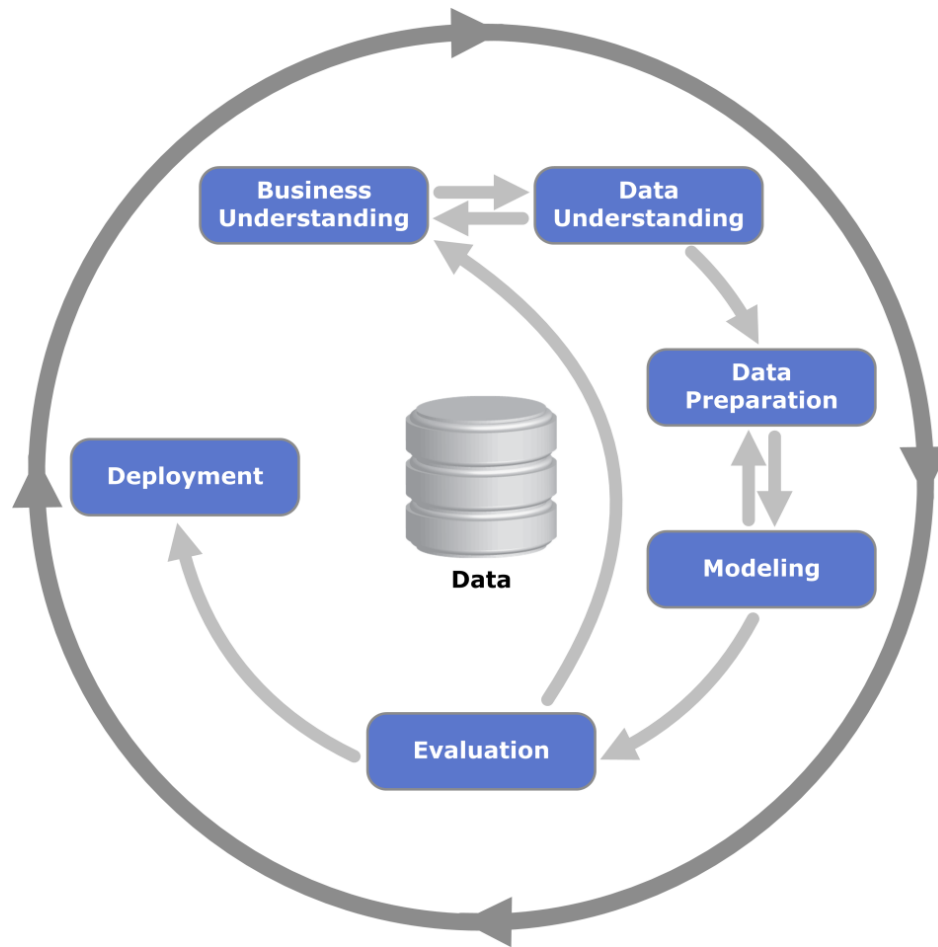University of Calabria

Business Intelligence and Analytics

Data Mining

Case study: Churn Analysis

Ph.D. Ettore Ritacco

The Knowledge Discovery Process (CRISP-DM)

Business Understanding

○ We are interested in modeling the customer attrition of a US mobile network operator (mobile phone company)

- ○ Customer attrition, also known as customer churn, is a business term describing the rate at which customers leave or cease paying for a product or service.

- ○ It's a critical figure in many businesses, as it's often the case that acquiring new customers is a lot more costly than retaining existing ones.

Data Understanding and Manipulation

- The dataset:

  - contains 3333 tuples

  - has 21 attributes

  - is unbalanced (w.r.t. the class attribute):

    - 2850 tuples are labeled as False. (85.5%)
    - 483 tuples are labeled True. (14.5%)

Data Understanding and Manipulation

- Data Schema (1/3):

  - State: categorical, for the 50 states and the District of Columbia

  - Account length: integer-valued, how long account has been active

  - Area code: categorical, regions within the states

  - Phone number: essentially a surrogate for customer ID

  - International Plan: dichotomous categorical, yes or no

  - VoiceMail Plan: dichotomous categorical, yes or no

  - Number of voice mail messages: integer-valued

Data Understanding and Manipulation

- Data Schema (2/3):

  - Total day minutes: continuous, minutes customer used service during the day

  - Total day calls: integer-valued

  - Total day charge: continuous, perhaps based on foregoing two variables

  - Total evening minutes: continuous, minutes customer used service during the evening

  - Total evening calls: integer-valued

  - Total evening charge: continuous, perhaps based on foregoing two variables

  - Total night minutes: continuous, minutes customer used service during the night

Data Understanding and Manipulation

- Data Schema (3/3):

  - Total night calls: integer-valued

  - Total night charge: continuous, perhaps based on foregoing two variables

  - Total international minutes: continuous, minutes customer used service to make international calls

  - Total international calls: integer-valued

  - Total international charge: continuous, perhaps based on foregoing two variables

  - Number of calls to customer service: integer-valued

  - **Churn: dichotomous categorical, yes or no, CLASS ATTRIBUTE**