Stichting NIOC en de NIOC kennisbank

Stichting NIOC (www.nioc.nl) stelt zich conform zijn statuten tot doel: het realiseren van congressen over informatica onderwijs en voorts al hetgeen met een en ander rechtstreeks of zijdelings verband houdt of daartoe bevorderlijk kan zijn, alles in de ruimste zin des woords.

De stichting NIOC neemt de archivering van de resultaten van de congressen voor zijn rekening. De website www.nioc.nl ontsluit onder "Eerdere congressen" de gearchiveerde websites van eerdere congressen. De vele afzonderlijke congresbijdragen zijn opgenomen in een kennisbank die via dezelfde website onder "NIOC kennisbank" ontsloten wordt.

Op dit moment bevat de NIOC kennisbank alle bijdragen, incl. die van het laatste congres (NIOC2023, gehouden op donderdag 30 maart 2023 jl. en georganiseerd door NHL Stenden Hogeschool). Bij elkaar bijna 1500 bijdragen!

We roepen je op, na het lezen van het document dat door jou is gedownload, de auteur(s) feedback te geven. Dit kan door je te registreren als gebruiker van de NIOC kennisbank. Na registratie krijg je bericht hoe in te loggen op de NIOC kennisbank.

Het eerstvolgende NIOC vindt plaats op donderdag 27 maart 2025 in Zwolle en wordt dan georganiseerd door Hogeschool Windesheim. Kijk op www.nioc2025.nl voor meer informatie.

Wil je op de hoogte blijven van de ontwikkeling rond Stichting NIOC en de NIOC kennisbank, schrijf je dan in op de nieuwsbrief via

www.nioc.nl/nioc-kennisbank/aanmelden nieuwsbrief

Reacties over de NIOC kennisbank en de inhoud daarvan kun je richten aan de beheerder:

R. Smedinga kennisbank@nioc.nl.

Vermeld bij reacties jouw naam en telefoonnummer voor nader contact.

ACADEMIE IT & MEDIA DESIGN

APPLIED DATA SCIENCE

DOOR MAYA SAPPELLI & TILMAN TODT

HAN_UNIVERSITY
OF APPLIED SCIENCES

What we've got:

Data lakes
Proof-of-concepts
Nice stories & screenshots

Infrastructural headaches
Privacy and security scares
Few actual profits

What organizations need:

Stuff that actually works
Is scalable
Related to business & research goals

THAT REQUIRES SKILLS AND KNOWLEDGE OTHER THAN TAUGHT IN REGULAR DATA SCIENCE STUDIES

APPLIED DATA SCIENCE

TARGET GROUP

Professionals and bachelor's students can apply if they meet the following admission requirements:

- A bachelor's degrees,
- Knowledge in mathematics and statistics as well as programming skills, Proficient mastery of English (Level English: B2 or higher).

Degree
Master of Science

Study load
60 ECTS credits

Mode of study
Part-time only

Program duration
2 years (part-time)

Program Start
September

Language
English

Location
Arnhem

HAN_UNIVERSITY
OF APPLIED SCIENCES

DEVELOPMENT

- Multidisciplinary team
  - BioInformatics (Christof Francke)
  - Lean Manufacturing (Vincent Wiegel)
  - Data & Knowledge Engineering (Stijn Hoppenbrouwers)

- In collaboration with organizations such as Machine2Learn

MASTER APPLIED DATA SCIENCE

- Training professionals in the use of data-driven solutions to solve problems and create added value.

- Target groups: professionals with different backgrounds and levels of knowledge, different levels of experience with programming and data science.

- Why so broad? Because all sectors of the economy are facing the challenge of innovating with data and becoming more efficient in an increasingly digital environment.

MASTER APPLIED DATA SCIENCE

- How can we train students in our degree programme to use Data Science effectively?

1. train professionals to understand the actual problem of the problem owner and translate it into a Data Science approach in a structured way. For this purpose, we have developed our own Machine Learning Canvas that covers important and generalised features of a Data Science problem.

2. providing a schema of a Data Science Life Cycle, which makes the phases of Data Science tasks transparent and gives the student guidance on how to proceed. Here we have adapted the CRISP-DM for our purposes.

3. The development of Data Science competencies as well as communication and research skills through project work on real problems.

CRISP-DM


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

Machine Learning Canvas

Designed for: Designed by: Date: Iteration: .

Business Objectives
What business objectives are we serving? What KPI will be improved and by how much? What is the business value and how can it be measured?

End-User Persona
Who are the end-users of the predictive model? What are the jobs they need to get done? What are their immediate pains and potential gains in trying to do so?

Value Proposition
How do we help the end-users get the job done and satisfy their needs? How do we alleviate or eliminate their pains and create value through gains?

Model Evaluation
What methods and metrics can be used to evaluate the 'offline' performance? What is the cost of being wrong?

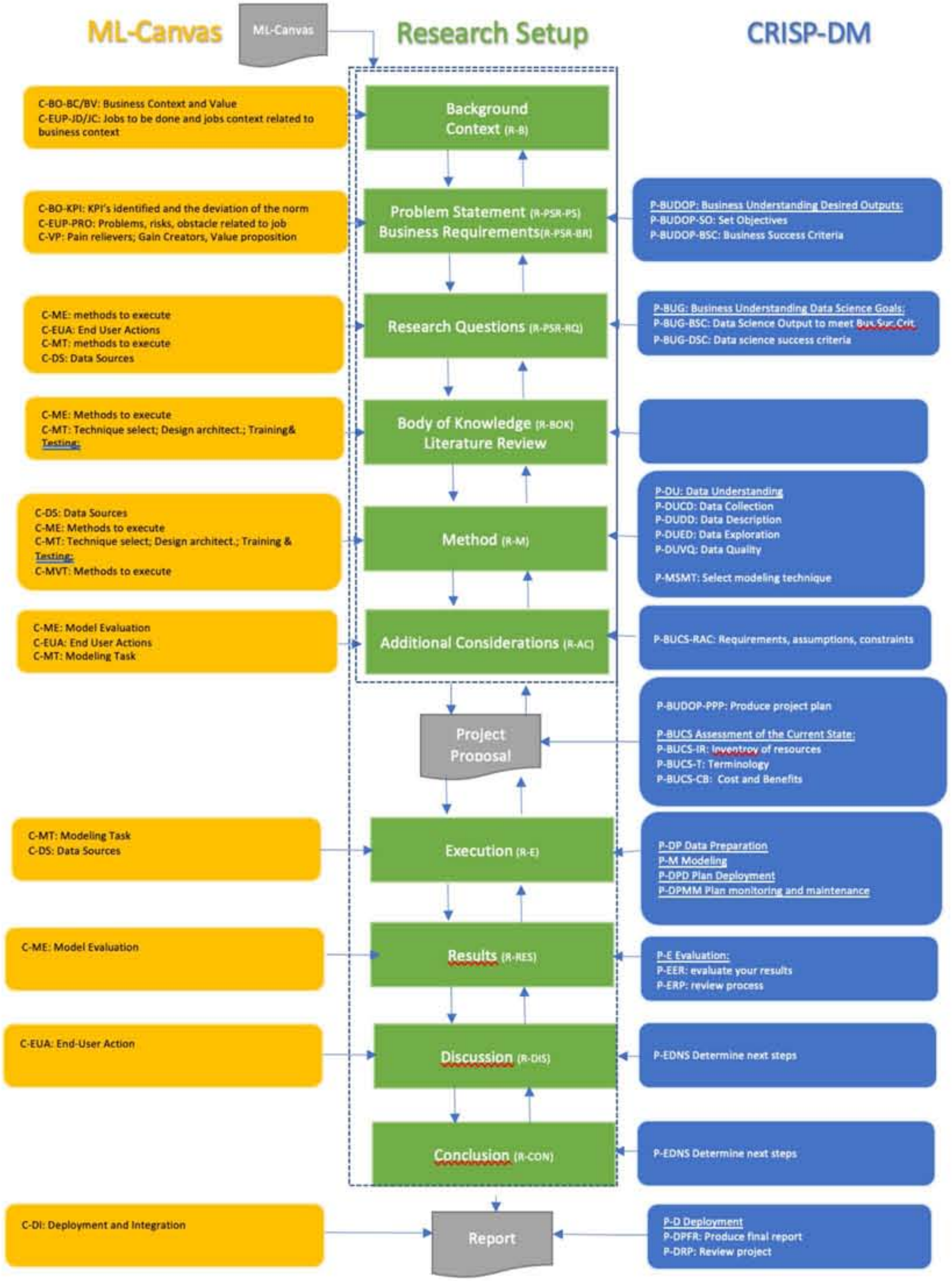
End-User Actions
How do predictions elicit actions (or decisions) that result in the proposed value for the end-user?

Modeling Task
What are we going to predict and with what information? What is the type of machine learning problem?

Data Sources
Which raw data sources can we use (internal and external)? How and with what frequency should the data be collected?

Monitoring and Value Testing
What methods and metrics can be used to monitor 'online' model performance? How is the actual business value measured?

Deployment and Integration
How can the model be deployed and brought into production? How is it exposed to the end-user and integrated in a product or service?

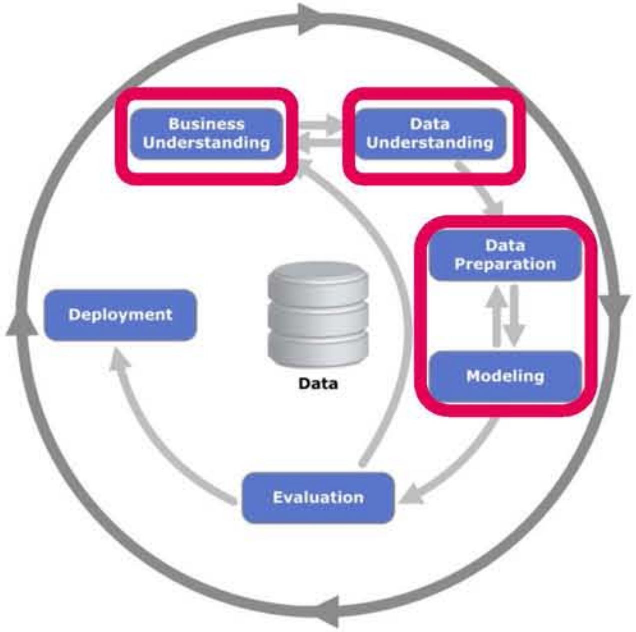


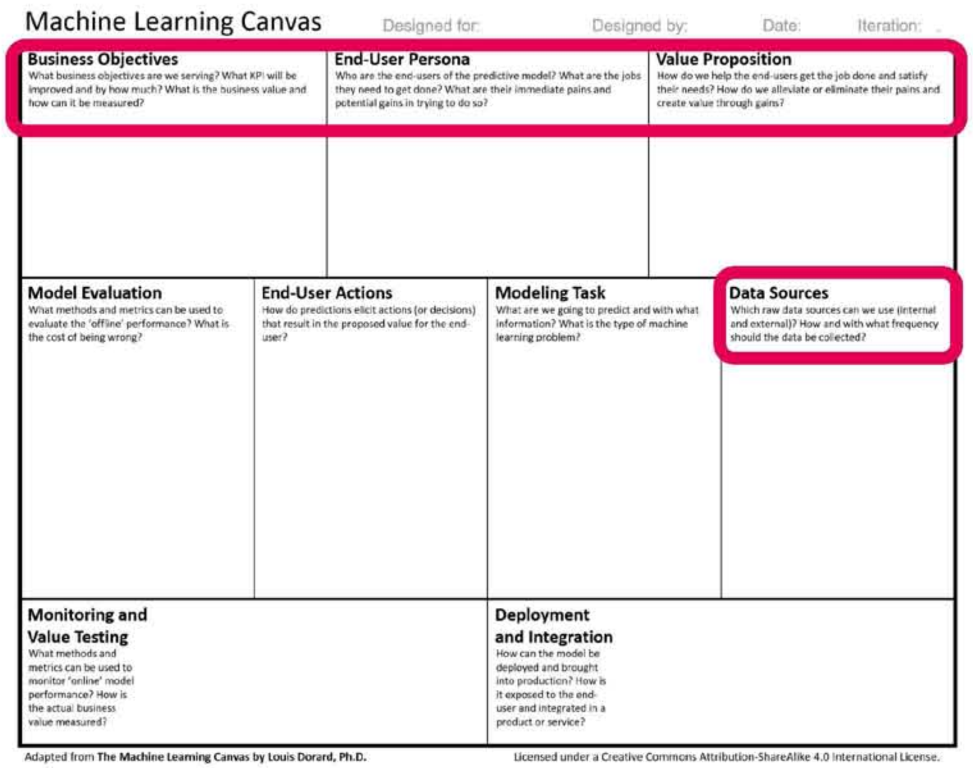
HAN_UNIVERSITY OF APPLIED SCIENCES

APPLIED DATA SCIENCE

FIRST YEAR		SECOND YEAR	
DATA EXPLORATION (UNIT 1, 15 EC)	**PREDICTIVE MODELLING (UNIT 2, 15 EC)**	**MODEL DEPLOYMENT (UNIT 3, 15 EC)**	**GRADUATION (UNIT 4, 15 EC)**
Applied data science (1) Machine learning (1): regression and clustering Big data, data wrangling & data exploration	Applied data science (2) Machine learning (2): concepts and methods of predictive modelling Data architecture Data ethics (1)	Model deployment Machine learning (3): advanced methods Capita selecta: Deep learning, Text mining, & AI Data privacy & ethics (2)	Graduation project
Projects on real-life cases			
Research skills & professional skills			

APPLIED DATA SCIENCE – UOS 1

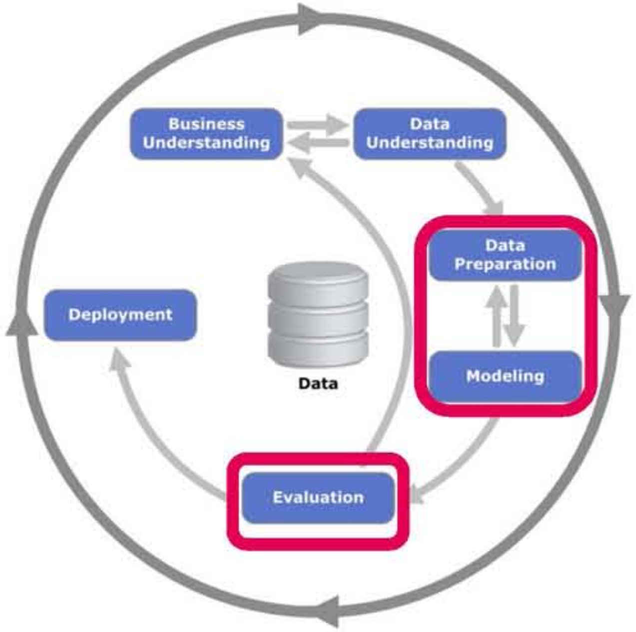


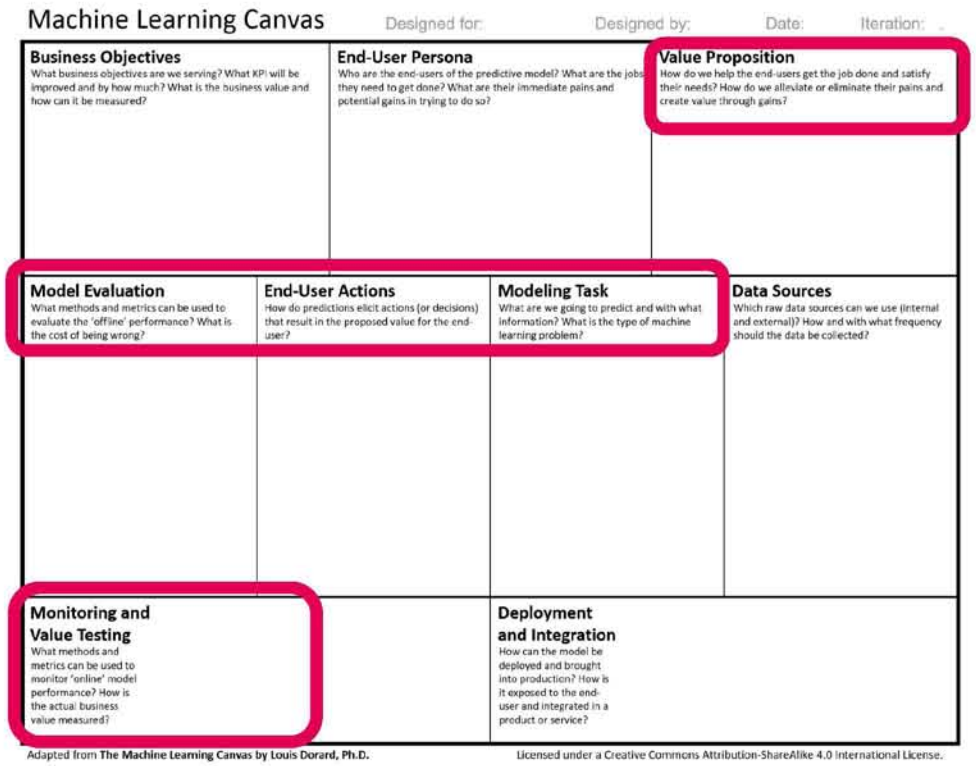


DATA EXPLORATION (UNIT 1, 15 EC)

Applied data science (1)
Machine learning (1): regression and clustering
Big data, data wrangling & data exploration

APPLIED DATA SCIENCE – UOS 2

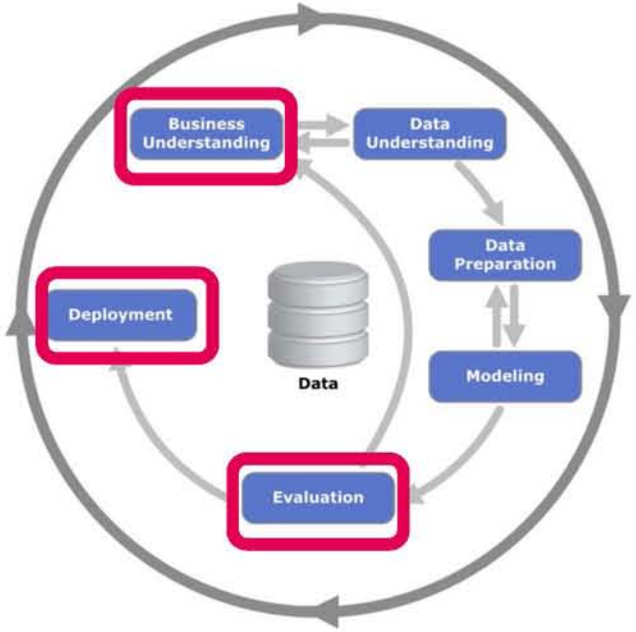


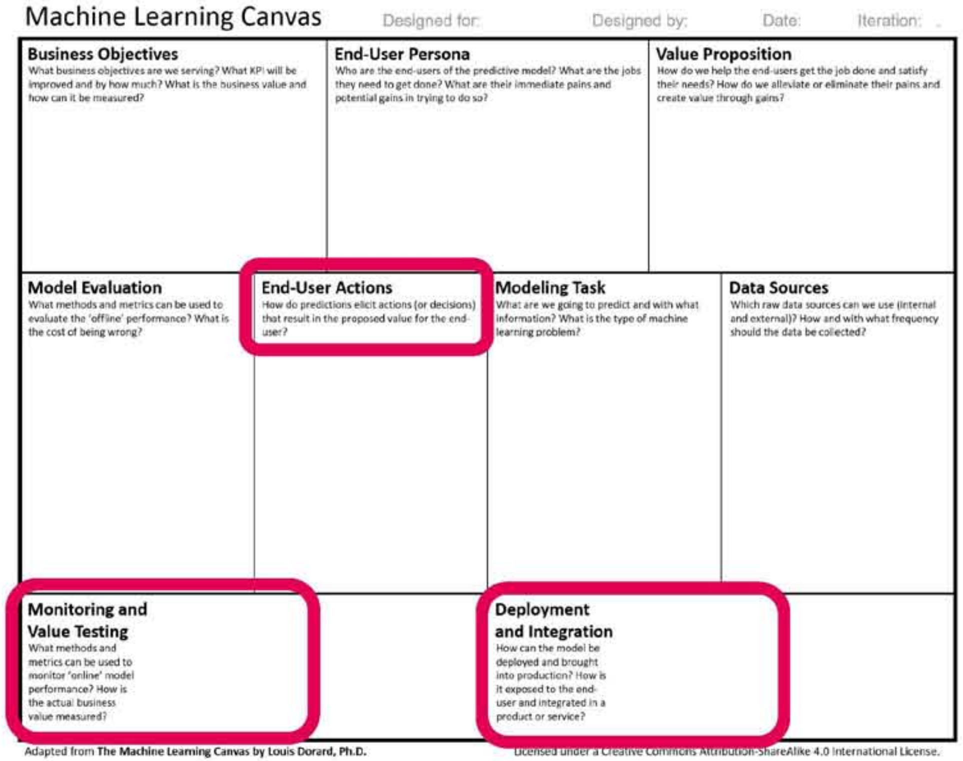


PREDICTIVE MODELLING (UNIT 2, 15 EC)

Applied data science (2)
Machine learning (2): concepts and methods of predictive modelling
Data architecture
Data ethics (1)

APPLIED DATA SCIENCE – UOS 3





MODEL DEPLOYMENT (UNIT 3, 15 EC)

Model deployment
Machine learning (3): advanced methods
Capita selecta: Deep learning, Text mining, & AI
Data privacy & ethics (2)

VALUE CREATION

- Buy-in from stakeholders (intended users) → requires trust
- Selection of appropriate techniques → not obvious
- Implementation of models and pipelines in process → data quality & reliability
- Proper usage of the models → understanding

REAL CASES

Wat is uw leeftijd?	32
In de laatste 2 weken heb ik wel eens pijn in mijn schouder of nek gehad?	Nee
Is uw werk zwaar of eentoning?	Niet erg — Heel erg
Verzuimt u momenteel (geheel of gedeeltelijk) van uw werk vanwege uw rugklachten?	Nee
Hoe vaak heeft u gemiddeld in de laatste drie maanden periodes van rugpijn gehad?	Niet vaak — Heel vaak
Welk cijfer geeft het best de ernst van uw rugpijn gedurende de afgelopen week (7 dagen) weer?	0 — 10
Hoe waarschijnlijk is het volgens u dat uw rugpijn over 12 weken helemaal over is?	Niet waarschijnlijk — Zeer waarschijnlijk
Als u rekening houdt met alles wat u doet om met de pijn om te gaan, in welke mate bent u dan op een gemiddelde dag in staat om de pijn te verminderen?	Bijna niet — Bijna altijd
Denk bij het beantwoorden van deze vraag aan de situatie in de laatste 2 weken. Ongeruste gedachten gingen vaak door mijn hoofd.	Oneens
Na een moeilijke periode ben ik snel weer de oude.	Bijna altijd

RESULTATEN

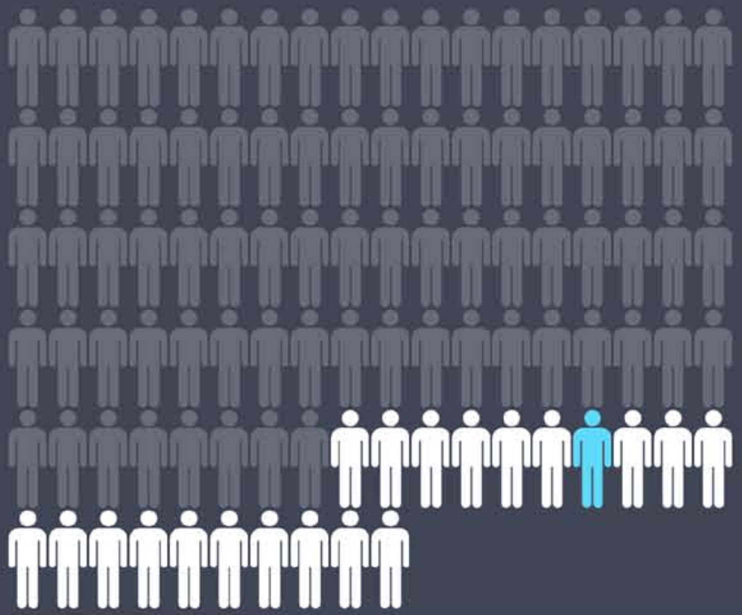
Beoordeling chronische klachten:

Uw rugpijn zal waarschijnlijk snel herstellen. Zelfstandig in beweging blijven is waarschijnlijk voldoende om zo snel mogelijk te herstellen.

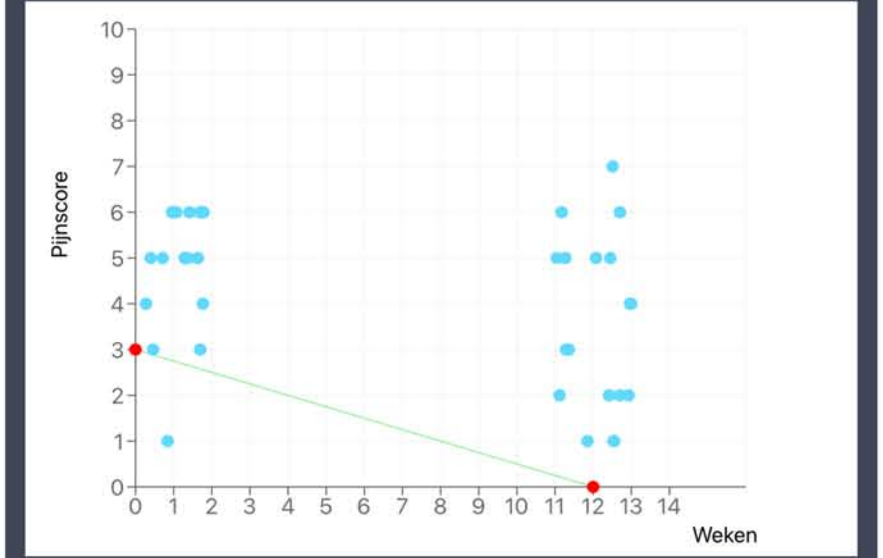
Vragen die het meeste invloed hebben gehad op de beoordeling:

1. Welk cijfer geeft het best de ernst van uw rugpijn gedurende de afgelopen week (7 dagen) weer? (3)
2. Na een moeilijke periode ben ik snel weer de oude. (Bijna altijd)

Pijnscore ten opzichte van anderen:



Verwacht verloop op basis van profiel:



HAN_UNIVERSITY OF APPLIED SCIENCES

ACADEMIE IT & MEDIA DESIGN

SHARING TEACHING MATERIALS WITH HSRW

DOOR MAYA SAPPELLI & TILMAN TODT

HAN_UNIVERSITY OF APPLIED SCIENCES

COLLABORATION – LEARNING NUGGETS

- Can we share teaching materials?
- Can we design our teaching in such a way that students prepare by themselves and classses are for interaction/collaboration/practice?
- How can we do this effectively & efficiently?

COLLABORATION – LEARNING NUGGETS

- **Findable**: Descibe material with metadata & identifier
- **Accesible**: Put the metadata (and later the material) in a format/database that we can access
- **Interopable**: (Meta)data must be interopable with both our universities workflows
- **Reusable**: (Meta)data is rich and nuggets are self contained so they can easily be reused

OBSIDIAN

semester: UOS1
year: 2023
contact: [[👤[redacted]
learning_outcome :

Lecture 2023_UOS1_wk11_Regression_1_of_2

Learning Objectives

- Student is familiar with the concept of correlation and knows how to apply these in data exploration.
- Student understands Pearson linear correlation, its relation to linear regression, and its limitations.
- Student is aware that there exist other variants of correlations, such as rank correlations and distance correlations, and understands that their can be advantages and disadvantages.
- Student is aware of the notion of distance, including the mathematical notion of metric, and similarity, and understands its applicability.
- Student is aware of the notion entropy, cross entropy and relative entropy (KL-divergence).
- Student is aware of the notion of mutual information, its advantages and disadvanteges and understand its applicability.

Lecture Outline

Outline:
- Correlation (intro, usage)
- Pearson correlation (relation to Simple Linear Regression)
- More advanced Correlation (Rank and distance correlations)
- Similarity or Distance Measures
- Entropy and Cross Entropy and Kullback-Leibler Divergence
- Mutual Information

^outline

Additional Reading

📚 Introduction to Data Mining (2nd edition, global edition) Chapter 2.4

Simple Linear Regression

Machine Learning Model → Linear Model for Regression
Types of Learning → Supervised Learning → Regression → Linear Regression
All

---
creation_date: 2022-11-28 16:04
status: todo

id:
type: content
title: Simple Linear Regression

time_necessary:
learning_objectives:
level_of_content: medium
requirements:
contact: 👤[redacted]
scene_path:

up:
- Linear Regression
- Linear Model for Regression
next: Multiple Linear Regression

Simple Linear Regression

tags:: #slope #intercept #parameters

Description

- Is a 1-d form of Linear Regression and Multiple Linear Regression
- The goodness of fit R^2 is given by the (square of the) Pearson correlation
- It fits a 1-d Linear Model for Regression

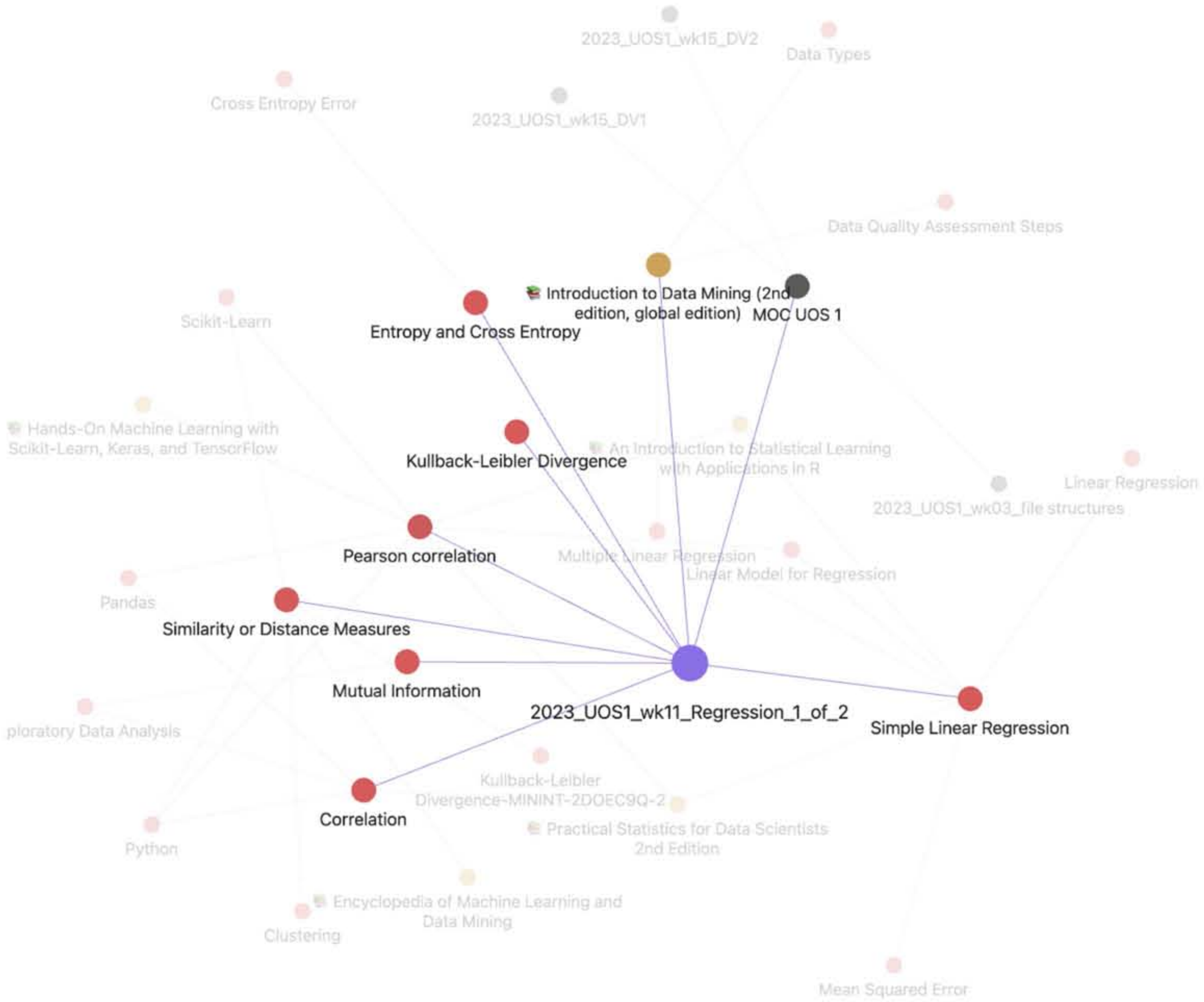
Up Real
1. Linear Regression
2. Linear Model for Regression

Same Real
1. Linear Regression - Statistical Assumptions

Next Real
1. Multiple Linear Regression

OBSIDIAN

- Simple format
  - Markdown
  - Yaml
- Good community
- Many relevant plugins to personalize needs



OUTLOOK

We are interested in sharing and exchanging

- educational material on data science and machine learning
- projects and/or data