

CMU Portugal
Advanced Training Program
Foundations of Data Science

DAVID SEMEDO
RAFAEL FERREIRA
NOVA SCHOOL OF SCIENCE AND TECHNOLOGY

Today's Topics

1. USE CASE – FINAL TOUCHES

2. USE CASE PRESENTATIONS

3. MODULE OVERVIEW

4. WHAT'S NEXT?

01

Use Cases – Final Touches

02

Use Case - Presentations

Use-Case Discussion Session

Let's simulate a Data Science team discussion:

- Bring your expertise and point of view!



Use-Case Details

Requirements - One operation of each of the following:

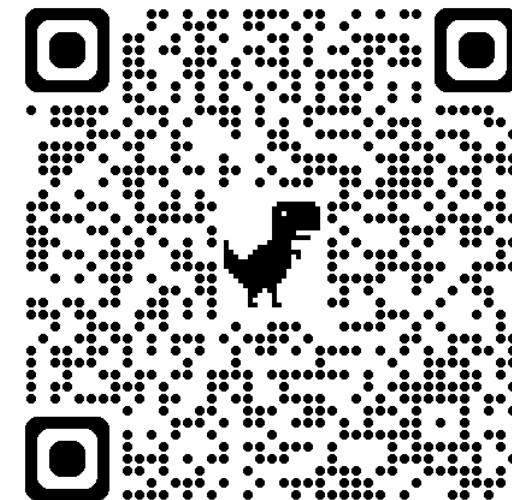
- Dataset Descriptive Statistics
- Data Cleaning (e.g. checking for NaNs, column removal, etc.)
- Model Selection, Feature Engineering, and Normalization
- Plotting (frequency, correlation between feature pairs)
- Supervised Learning:
 - Training a linear classifier
 - Evaluate its performance over multiple metrics

Use Case – Teams

- Groups and topics are [here](#).
- Presentations on the 18th of October.

Presentation:

- Show and discuss your notebook to the class
- 7 to 10 minutes per presentation
- 5 to 10 minutes discussion



Presentations!



Use Case – Final Submission

Send us an email with your notebooks and materials:

- df.semedo@fct.unl.pt
- rah.ferreira@campus.fct.unl.pt

Deadline: **24 October 2024** (next Thursday by the end of the day)

03

Module Overview

Foundations of Data Science

Introduction to Programming

Statistics and Probability

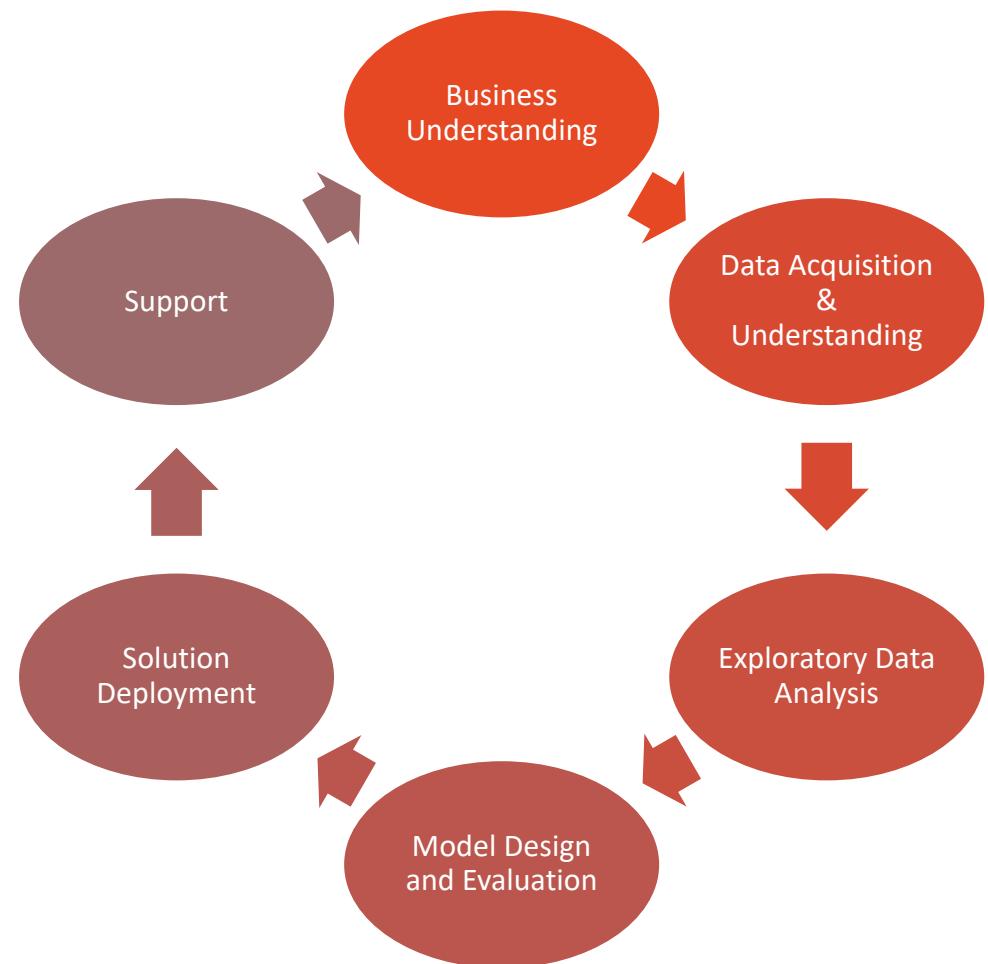
Data Preparation

Intro to Machine Learning

Model Evaluation

Data Visualization

Data Science Lifecycle



Module Progression



python™



matplotlib
seaborn



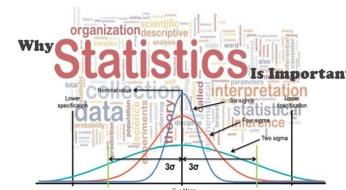
NumPy

pandas

Machine Learning

scikit
learn

Module Progression – From Zero to Hero



Leveraging data!



Machine
Learning



Use-Case Skills



Collaboration and Teamwork

Gain experience working in teams.



Problem-Solving

Adapt to new and complex data situations.



Effective Communication

Create interpretable visualizations.

Explain findings in a way that's easy to understand.



Hands-On Tool Usage

Build practical skills for real-world applications.

Assess the success of your solutions.

What's Next? – Core Course Structure

We covered the very fundamental concepts and tools!

Course	Lecturer	Teaching hours	ECTS
Foundations of Data Science	David Semedo & Rafael Ferreira	30	2
Machine Learning	Chryssa Zerva & Sweta Agrawal	30	2
Data Collection and Pre-Processing	Cátia Pesquita	30	2

Data Collection and Pre-Processing



Main Instructor:
Prof. Cátia Pesquita

- **Topic:** Obtaining and preparing data for analysis.
 - **Data Sources & Types:** Understanding various data origins and categories
 - **Collection Methods:** Exploring different techniques for data gathering
 - **Challenges with Datasets:** Addressing issues like incompleteness, noise, inconsistency, and bias
 - **Emphasis on Data Quality:** Data cleaning, Integration, Transformation, Reduction, Discretization.
- **Goal:** Gain foundational skills in collecting, cleaning, integrating, exploring, and sharing data.

Machine Learning

- **Topic:** Fundamental concepts and tools of machine learning.
 - **Learning settings:** Supervised, Unsupervised and Reinforcement Learning
 - **Feature Engineering and Selection:** Pre-processing, feature selection strategies.
 - **Model Evaluation and Validation:** Evaluation and generalization assessment.
 - **Models:** Regression, SVMs, Associate Rules, Decision Trees, Neural Networks, etc.
- **Goal:** Gain foundational skills in collecting, cleaning, integrating, exploring, and sharing data.



Main Instructor:
Prof. Chryssa Zerva

From Zero to Hero – Ok, but what else?



- **Through the optional courses, you can specialize on:**

- Vision and Language data
- Complex data
- AI system engineering

- **Capstone Project:**

- Lectured by all lecturers
- Create and test a real system

Course	Lecturer	Teaching hours	ECTS
Deep Learning	Chryssa Zerva & David Semedo	18	1
Vision and Language	Bruno Martins & João Magalhães	18	1
Complex Data Analysis	André Falcão	18	1
Cloud-based Data Processing	Nuno Preguiça & Rodrigo Rodrigues	18	1
CapStone	All	48	3



David Semedo – Short bio

df.semedo@fct.unl.pt

Assistant Professor @ NOVA FCT, Integrated Researcher @ NOVA LINCS

AI for vision and language. Neural approaches to conversational and contextualized media understanding.

Rafael Ferreira – Short bio

rah.ferreira@campus.fct.unl.pt



4th year PhD Student @ NOVA FCT

Conversational AI. Team leader of the award-winning TWIZ in the Alexa TaskBot Challenge.



Thank you!!