Enterprise Data Science Bootcamp

Week 1

Syllabus

Projects

Class outline

- Syllabus
- Calendar
- Groups
- Case Studies

Syllabus

- Introduction to enterprise data science project: overview of a data science project, end-to-end
- Data storytelling and call-to-action: how to present your findings and insights to an audience of enterprise decision makers e drive for action
- The enterprise data science technology stack: what technologies are being currently used for data science projects in organizations (including Git and containers)
- Introduction to data science operations and production: how to move your project from prototype to production and keep them up-to-date
- The role of LLM's in data science projects

Key dates

- On a separate document in Moodle
  - Substitution classes schedule for public holidays
  - Project status reports schedule
  - Intermediate presentations schedule
  - Final presentations schedule

Evaluation and contact

- Evaluation:
  - First assessment period:
    - Project status reports (2): 2x5%
    - Interim presentation of the project: 20%
    - Final presentation of the project: 40%
    - Exam: 30%
  - Second assessment period:
    - Presentation of the project: 70%
    - Second Exam: 30%
- Office hours:
  - Tuesdays, 18h00-18h30, by appointment
- Contact:
  - hcarreiro@novaims.unl.pt
  - jazambuja@novaims.unl.pt

Status reports

- Form in Moodle
- The questions will be as follows:
  - Briefly characterize the status of your project
  - What have been your main challenges in the project, so far?
  - What insights/learnings are you expecting to achieve by concluding this project?
  - What do you expect will be the next steps in the project?

Bibliography

- Class slides, scientific papers, analysts reports.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. Sebastopol, CA: O'Reilly Media, Inc.
- Godsey, B. (2017). *Think like a data scientist : tackle the data science process step-by-step*. Shelter Island: Manning.

Technology stack

- The technology stack to use in the projects will be based on the Python stack
- Students will also use Git (GitHub) for code repository, version control and sharing with the course professors
- Preferred technologies are, of course, the ones studied in the program

Approach

- The subject of Enterprise Data Science Bootcamp aims to expose students of the Post-Graduate Program in Enterprise Data Science and Analytics to Data Science projects in situations of real application of knowledge acquired.
- Throughout the school term, students will carry out a data science project from the initial stages of planning to the presentation of the final results.

Expectations for the final presentation

- The expectations for the final presentation are that it will be made to business decision makers — C-Level executives
- **In this sense, it is expected that the presentations will not be merely analytical but will present suggestions for actionable paths**
- The slides should have a professional look, such as the slides left at the end of a board meeting, to support decisions
- A balance should be sought between analytical rigor and accessibility for audiences from different areas and different backgrounds

How *technical* should you be?

- The expectation is that the final presentation will be based on a rigorous analysis of the data and that the description of the methods used will be accurate
- As the project is to be presented to C-level executives, analytical rigor must never be neglected
- In addition, the presentation must be intelligible even to those who don't necessarily have a background in data science

Your experience in Data Science projects and your expectations for the course?

Deliverables

Deliverables (1)

- Create a group of up to 4 members (no less than 3). It is expected that the groups should be defined (in Moodle) after the first week
- It is strongly advisable that the groups are created with the same "T" because of the schedule for the presentations
- **You should present your assessment and conclusions in a set of slides, planned to be delivered in 10 minutes (with further 10 minutes for discussion)**
- You can also make a live demo if you feel comfortable doing it. Try to plan the time you will need for the demo and the presentation

Deliverables (2)

- All the complementary information that you feel it is important, should be put into background slides
- The total number of slides, the ones used for presentation and the ones used for background info should not exceed 30 (hopefully, less than that)
- The slides should be delivered in PDF format, through Moodle
- **(Don’t forget to identify every member of the team with name and number in each deliverable)**

A few words about the intermediate presentation

- The objective of the intermediate presentations is for you the present your findings so far and test the delivery of them to a live and interested audience
- In a word: as close as you can get it to the final presentation, the better feedback you will receive from the overall audience
- Strong recommendation: think of the intermediate presentation as a **rehearsal** for the final presentation

A few words about the exam

- The exam will assess the class subjects not covered in the project
- The exam is based on true/false and multiple-choice questions, and is no longer than 60 minutes
- The base subjects for the exam are the slides of the expositive classes and recommended readings

Descriptors for evaluating the project

	EVALUATION ITEM	DESCRIPTORS	PERCENTAGE OF THE TOTAL CLASSIFICATION
1	Quality and comprehensiveness of the questions in order to explore the case	The questions: a) explore the possibilities presented by the dataset b) give rise to relevant insights in the context.	15%
2	Quality and scalability of the presented solution.	The solution: a) is in accordance with the conditions of the proposed case, b) gives a sustained answer to the questions referred to in point 1,	40%
3	Quality of the presentation.	The presentation: a) is fluid and clear, b) uses language understandable to non-specialists,	20%
4	Quality of the report (in the form of the presentation text and complementary slides).	The report: a) presents the solution clearly, b) indicates the main options taken, c) indicates the limitations and possibilities for future work.	15%
5	Respect for presentation time.	The time allocated to the group for the presentation was respected.	10%

Logistics for delivery

- The delivery will be through Moodle.
- You should upload your Presentation as PDF, well as any scripts and .PBIX files, in one .ZIP file, if possible.
- The name of the .ZIP file should include the code of the group and the number of one member of the group (Mxxxxxx):
  - TP#_GROUPNUMBER_Mxxxxxx_FREETITLE.ZIP

Groups

Groups

- Groups created through Moodle
- The groups should be defined after the first week
- Any changes on the group that might happen afterwards should be communicated to the professors of the course

Project Proposals

Project proposals

- Analysis of customer churn in the telecommunications industry
- Analysis of employee attrition in a consulting company
- Analysis of the effectiveness of a marketing campaign in the financial sector
- BYOD (Bring Your Own Dataset: need to be approved by the professors of the course and closed before the second week)

Class takeaways

- Syllabus
- Calendar
- Groups
- Case Studies