NOVA IMS
Information Management School

0

COURSE OVERVIEW

Data Science for Marketing

Acreditações e Certificações

Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa

Theory classes + Course coordination

HELLO!

Nuno António

nantonio@novaims.unl.pt

www.linkedin.com/in/nunoantonio

www.researchgate.net/profile/nuno_antonio

Practical classes

Ana Edmundo

aedmundo@novaims.unl.pt

www.linkedin.com/in/ana-edmundo

Diogo Diogo

ddiogo@novaims.unl.pt

www.linkedin.com/in/diogodiogo

Intended learning outcomes

- Understand the fundamental concepts associated with Data Science
- Recognize and identify Data Science applications in Marketing
- Understand a Data Science project lifecycle
- Identify the most common types of data and how to collect data from SQL databases
- Use Excel PowerPivot to do basic and intermediate data description/characterization tasks
- Use Python as an analytical tool
- Be able to execute basic and intermediate data preparation and preprocessing tasks
- Be able to use data visualization as both a data exploration tool and a communication tool
- Calculate and explain the most relevant performance measures used in association rules and describe the Apriori algorithm
- Calculate and explain data similarity and dissimilarity measures
- Calculate and interpret the RFM model

Tentative schedule

W1 06 SEP Theory	W2 13 SEP Theory	W3 20 SEP Practice	W4 27 SEP Practice	W5 04 OCT Theory	W6 11 OCT Theory	W7 18 OCT Practice
Course overview Introduction to DS CRISP-DM proc. model	Common data types and SQL Data characterization/ description	SQL exercises Data characterization/ description exercises	Data characterization /description exercises	Data understanding Communication and Data visualization	Data preparation	Python quiz Data understanding and visualization
W8 **05 NOV** **Theory**	**W9** **08 NOV** **Practice**	**W10** **15 NOV** **Theory**	**W11** **22 NOV** **Theory**	**W12** **29 NOV** **Practice**	**W13** **06 DEC** **Practice**	**W14** **Practice**
Association rules	Association rules (including data underst. and preparation)	Data similarity and dissimilarity measures	RFM model	Data similarity and dissimilarity measures exercise	RFM model exercise	Group project Q&A

Extra classes

For those not familiar with Python

- Saturday, 24 Sep. 15h – 17h (online)
  - Tools setup
- Saturday, 01 Oct. 14h30 – 18h30 (hybrid)
  - Python Bootcamp 1/2
- Saturday, 08 Oct. 14h30 – 18h30 (hybrid)
  - Python Bootcamp 2/2

Pedagogical model Fall 2022

- For now, all classes will be in hybrid mode (presential and online):
  - In presential classes:
    - A minimum of 5 students is required
    - If the minimum is not reached in one class, the class goes to fully online mode after that class
  - In online classes:
    - Students should have their webcams turned on or could be moved to the Zoom “waiting room”
    - As the classes will have a high practical component, it is recommended to use two screens or a large screen (one for zoom, one for running the exercises)
- Python introductory online courses:
  - If you are not familiarized Python, enroll and complete Datacamp’s online courses announced on Moodle until week 3

Evaluation method

Python quiz	10%
Group project: • The minimum grade is 8.0 out of 20 • To be delivered in the exam season (to be defined according to the 1st season exam date)	50%
Exam: • The minimum grade is 8.0 out of 20 • 1st season and 2nd season have the same weight	40%
TOTAL	**100%**

Evaluation will be continuous and must include a practical component.
Therefore, **there will be no exam that alone corresponds to 100% of the grade**

Quiz and exam details

- Individual, with materials consultation (paper or device)
- Quizzes and exams are made online, in Moodle
- If done remotely (depends on the rules at the moment of the quiz/exam), uses proctoring software (Respondus Lockdown browser)
- Multiple choice questions:
  - Single answer questions (e.g., 1 point – 4 possible answers):
    - Correct answer: full score of the question (e.g., 1 point)
    - Incorrect answer: negative (-1/possible answers- e.g.,-1/4 = -0.25)
  - Multiple answers questions:
    - Each correct answer: number correct questions/full score of the question
    - Each incorrect answer: -number correct questions/full score of the question
  - If no answer is given or if “I do not want to answer” is selected, the score will be 0 (zero)
- Typically, a quiz has 10 to 20 questions and an exam, 40 to 50 questions. The time to answer each question is 1 min. 30 sec.

Group project (1/2)

Objective:
To develop student's ability to build a Data Science project employing the CRISP-DM process model – more details in the respective notebook

- Students must organize themselves into groups of 2 to 4 students
- Students can be from different theoretical and practical classes

Group project (2/2)

- **Deliverables:**
  - Python source code (Jupyter notebook or .py files) and/or Excel files. Python code should be commented to facilitate comprehension
  - Powerpoint presentation
  - Report:
    - Maximum of 20 pages (excluding appendixes)
    - The minimum font size is 10
    - Should describe the main outputs according to CRISP-DM, including a brief description of the problem, methods, results, and their discussion
- **Presentation:**
  - To be done in the exam season with all group members present
  - Slots of 20 minutes per group
  - 10 minutes of presentation, another 10 for discussion

Bibliography

[A] Miller, T. W. (2015). Marketing Data Science: Modeling Techniques in Predictive Analytics with R and Python, Pearson

[B] Keller, G. and Gaciu, N. (2020). Statistics for Management and Economics (2nd edition), Cengage Learning

[C] Han, J., Kamber, M., Pei, J. (2012). Data Mining - Concepts and Techniques (Third edition), Morgan Kaufmann

[D] Linoff, G. S., and Berry, M.J.A (2011). Data Mining Techniques for marketing, sales, and customer support (Third edition). Wiley Publishing, Inc.

[E] Provost, F., and Fawcett, T. (2013). Data Science for Business, O'Reilly

[F] Materials and references URLs provided in class by the instructors

Communications

- Course instructors to students – materials and logistics:
  - Announcements via Moodle (followed by email)
  - Moodle: publication of course materials
- Students to course instructors – questions – by priority order:
  - Check course materials
  - Post a question in Moodle (Forum) – anyone can post and answer/help
  - Email
  - Meeting (available Monday and Thursday, 18h30-20h) (requires scheduling by email)
- Submissions (project deliverables, groups memberships, presentations' time slots selection):
  - Must be made via Moodle in the appropriate slots
  - Must be made before the deadlines

Questions?

Data Science for Marketing

Acreditações e Certificações

Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa