Universidad de Alcalá

TEACHING GUIDE

Fundamentals of Data Science

Degree in
Computer Science Engineering

Universidad de Alcalá

Academic Year 2024/2025

4th Year - 1st Semester

TEACHING GUIDE

Course Name:	**Fundamentals of Data Science**
Code:	**781006**
Degree in:	**Computer Science Engineering**
Department and area:	**Ciencias de la Computación** **Computer Science and Artifical Intelligence**
Type:	**Compulsory**
ECTS Credits:	**6.0**
Year and semester:	**4ᵗʰ Year, 1ˢᵗ Semester**
Teachers:	Juan J. Cuadrado Gallego The updated full list of professors for the 2021/2022 academic year can be can be consulted in the website of the subject: https://www.uah.es/es/estudios/estudios-oficiales/grados/asignatura/Fundamentos-de-la-Ciencia-de-Datos-781006/
Tutoring schedule:	Consultar al comienzo de la asignatura
Language:	Spanish/English Friendly

1. COURSE SUMMARY

In the study plans, the degree in Ingeniería Informática is structured in the three formative blocks: Basic, Mandatory and Optative. The basic and mandatory formative blocks cover the main body of knowledge proposed by the reports or curricular guides: Computing Curricula: Computer Science 2001, Computer Curricula Software Engineering 2004, Computing Curricula 2005 from ACM-IEEE, Computing Engineering of 2004 and the 2002 IS Guide from the AIS. The matter Fundamentals of Data Science has 6 ECTS credits and all of them are taught in a single subject with the same name as the matter. The subject has by objective that the student acquires the knowledge corresponding to the body of knowledge corresponding to Data Science.

The body of knowledge used to establish the contents of the subject Fundamentals of Data Science has been developed in the EDISON project, said project was founded by the European Union in 2015 to establish the foundations of Data Science, including the definition of the body of knowledge and its curriculum, the results of which were published in 2020 in the book from Springer publisher The Data Science Framework: A View from the EDISON project. To introduce what these contents are about, the definitions given in the book and Data Science can be defined as "Data Science is a complex discipline that uses conceptual and mathematical abstractions and models, statistical methods, together with modern computational tools to obtain knowledge/derive insight from data to (uncover correlations and causations in business data) support decision making in scientific research and business activity" and if it was be defined only with one sentence: "Science that studies how to obtain knowledge from data."

The knowledge that students obtain attending the course Fundamentals of Data Science are of the most current and interest since it will provide them, through the development and deepening of the same, to become a Data Scientist, a profile for which, according to the most prestigious consulting firms in the sector, it is expected 1.2 million jobs will be created in the coming years.

Prerrequisites and Recomendations

For the subject Fundamentals on Data Science it is recommended to have passed the subjects basic and mandatory of the degree on Statistics, Mathematics, Algorithm and Programming.

2. SKILLS

Basic, Generic and Cross Curricular Skills.

This course contributes to acquire the following basic, generic and cross curricular skills:

en_CG8 - Knowledge of the basic subjects and technologies, which enable them to learn and develop new methods and technologies, as well as those that provide them with great versatility to adapt to new situations.

en_CG9 - Ability to solve problems with initiative, decision making, autonomy and creativity. Ability to know how to communicate and transmit the knowledge, skills and abilities of the profession of Computer Engineering Engineer.

en_CB1 - That students have demonstrated to possess and understand knowledge in an area of study that is based on general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that involve knowledge from the forefront of their field of study.

en_CB2 - That the students know how to apply their knowledge to their work or vocation in a professional manner and possess the competencies that are usually demonstrated through the elaboration and defense of arguments and the resolution of problems within their area of study.

en_CB3 - That students have the ability to gather and interpret relevant data (usually within their area of study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

en_CB4 - That students can transmit information, ideas, problems and solutions to both a specialized and non-specialized public.

en_CB5 - That the students have developed those learning skills necessary to undertake further studies with a high degree of autonomy.

en_TRU1 - Capacity of analysis and synthesis.

en_TRU2 - Oral and written competencies.

en_TRU3 - Ability to manage information.

en_TRU4 - Autonomous learning skills.

en_TRU5 - Team work.

Specific Skills

This course contributes to acquire the following specific skills:

en_CC4 - Ability to know the fundamentals, paradigms and techniques of intelligent systems and analyze, design and build systems, services and computer applications that use these techniques in any field of application.

en_CC5 - Ability to acquire, obtain, formalize and represent human knowledge in a computable way to solve problems through a computer system in any field of application, particularly those related to aspects of computing, perception and performance in intelligent environments or environments.

en_CC7 - Ability to know and develop computational learning techniques and design and implement applications and systems that use them, including those dedicated to automatic extraction of information and knowledge from large volumes of data.

Learning Outcomes

After succeeding in this subject the students will be able to:

RA1. Know the differences between supervised and unsupervised learning as well as its application context.

RA2. Being able to model and solve learning problems using computational paradigms (supervised and unsupervised).

RA3. Being able to explain the applied models and learning results obtained, including their limitations and benefits.

RA4. Being able to understand different types of data, problems and possible transformations for their application both in learning and in the evaluation of results.

3. CONTENTS

Contents Blocks	Total number of hours
Lesson 1. Introduction to Data Science	6 hours
Lesson 2. Data	10 hours
Lesson 3. Association Analysis	10 hours
Lesson 4. Anomaly Detection	10 hours
Lesson 5. No Supervised Classification	10 hours
Lesson 6. Supervised Classification	10 hours

4. TEACHING - LEARNING METHODOLOGIES. FORMATIVE ACTIVITIES.

4.1. Credits Distribution

Number of on-site hours:	58 hours (50 hours on-site +8 exams hours)
Number of hours of student work:	92
Total hours	150

4.2. Methodological strategies, teaching materials and resources

Sessions of large group in the classroom in combination with practices in the laboratory:

Lessons in person will be held in the classroom and laboratory. It will be two kind of lessons:

- Without use of software: these lessons will be held in the classroom and the theoretical concepts of the subject will be presented through the use of presentation tools, also practical cases and exercises will be solved, complementing the theoretical contents.
- With use of software: these lessons will be held in the laboratory and will be solved, thorough the use of software for Data Science, the same practical cases and exercises that had been solved in the previous lesson in the classroom without the use of software.

Group and collaborative work

The practical teaching, through the the solution of questions, exercises and practices (laboratory exams) will have a fundamental role, as methodological strategy, in the teaching of the subject. In consequence will be asked to the student a high degree of involvement and participation in the solution of questions and exercises, individually, but also in group, specially in the realization of the practices o laboratory exams, that must be solved in group. Moreover, as a part of that work, oral presentation could be made by the students groups about the work made, to the rest of the students in the course, about the work done.

Personal work and study

Inside personal work and study the following activities will be performed:

- Study of the theoretical concepts imparted.
- Resolution of exercises. This will be the fundamental activity that the student must realize as personal work and study. It will be very important that the student dedicates enough time to personal work and study because of this dedication will depend the use of the lessons in person.
- Resolution of questions with the software environment used.
- Realization of textual summaries and schemes about the concept imparted.
- Consultation of bibliography sources and resources, physical and electronic.

Both the lessons in person and the autonomous work will be supported by tutorials, in official scheduled hours established for the subject, that will be held individually or in group. Moreover, the students can ask for tutorials out of the established schedule for them, in the case that they can't attend them.

5. ASSESSMENT: procedures, evaluation and grading criteria

Preferably, students will be offered a continuous assessment model that has characteristics of formative assessment in a way that serves as feedback in the teaching-learning process.

5.1. PROCEDURES

The evaluation must be inspired by the criteria of continuous evaluation (Learning Assesment Guidelines, LAG, art 3). However, in compliance with the regulations of the University of Alcalá, an alternative process of final evaluation is made available to the student in accordance with the Learning Assesment Guidelines as indicated in Article 10, students will have a period of fifteen days from the start of the course to request in writing to the Director of the Polytechnic School their intention to take the non-continuous evaluation model adducing the reasons that they deem convenient. The evaluation of the learning process of all students who do not apply for it or are denied it will be done, by default, according to the continuous assessment model. The student has two calls to pass the subject, one ordinary and one extraordinary.

Ordinary **Call**

Continous Assessment:

The continuous assessment process will use different strategies and will collect, during the term, evidences related with all the teaching-learning process during the teaching of the subject. These evidences will consist in the following assessment exams:

1. **Continuous Assessment Exam 1 (PEC1)**. It will consist in written questions and exercises about the contents of lessons 1, 2 and 3; and it could contain also questions about the software environment used to address those lessons. This written exam will be performed without computer.
2. **Continuous Assessment Exam 2 (PEC2)**. It will consist in written questions and exercises about the contents of lessons 4, 5 and 6; and it could contain also questions about the software environment used to address those lessons. This written exam will be performed without computer.
3. **Laboratory Exam 1 (PL1)**. It will consist in the realization of a practice, that will be done with the software for Data Science that had been studied in the subject, and that will be based in the contents of lessons 1,2 and 3. The PL1 may have an associated test, oral or written, on the contents covered in it, which will be carried out individually. If said test is not passed, the qualification obtained in PL1 will not be considered.
4. **Laboratory Exam 2 (PL2)**. It will consist in the realization of a practice, that will be done with the

software for Data Science that had been studied in the subject, and that will be based in the contents of lessons 4,5 and 6. The PL2 may have an associated test, oral or written, on the contents covered in it, which will be carried out individually. If said test is not passed, the qualification obtained in PL1 will not be considered.

About the realization of the exams, they will be the following:

- PEC1 and PEC2 will have a maximum duration of 1 hour and 55 minutes and will be in the classroom, in two dates that will be published during the presentation of the subject.
- PL1 and PL2 will be done during the weeks in which are being studied the contents in each one of them and they will be delivered in the following week to the teaching of the lessons 3 and 6 respectively.

Assessment through final exam:

In the case of evaluation by means of a final exam, the evaluation elements to be used will be the following:

Final Assessment Exam (PEF). It will consist in questions about the contents of the subject and the software for Data Science used. It will take place in the date officially set by the direction of the studies for this exam and it will have a maximum duration of 2 hours and 50 minutes.

Extraordinary Call

The extraordinary call will have the same format and duration as the final assessment exam and it will take place in the date officially set by the direction of the studies.

5.2. EVALUATION

EVALUATION CRITERIA

Evaluation criteria that will be applied to the subject Fundamentals of Data Science are linked with the Learning Outcomes defined for it section 2 and will be oriented to collect both the level of understanding of the contents and the capacity to apply the acquired knowledges to problems resolution, with and without the help of software tools. Moreover, it will be taken into account the positive and collaborative attitude in the realization of the tasks, both individually and in group, and the degree of interest shown in the realization and accomplishment of the classroom and laboratory activities and in the autonomous and collaborative work, so it will by essential to deliver in time and with correct format all the practices activities.

It will be used the following percentages for each one of the criteria of assessment for the subject, related with the learning outcomes:

CE1. Knowledge and applications of the contents: 70%

CE2. Solution using software of practical problems: 25%

CE3. Interest and behavior in the subject: 5%

GRADING TOOLS

This section summarizes the assessment tools that will be applied in each one of the evaluation criteria:

1. **CE1**. PEC1 and PEC2
2. **CE2**. PL1 and PL2
3. **CE3**. Interest and behavior in the lessons in the classroom and in the laboratory

GRADING CRITERIA

In the ordinary call-continuous assessment the relationship between the competences, learning outcomes, criteria and evaluation instruments is as follows.

Skill	Learning Outcomes	Evaluation criteria	Grading Tool	Contribution to the final mark
en_CG8, en_CG9, en_CB1-en_CB5, en_TRU1, en_TRU5, en_CC4-en_CC7	RA1, RA2, RA3, RA7	CE1, CE2, CE3	PEC1	35%
en_CG8, en_CG9, en_CB1-en_CB5, en_TRU1, en_TRU5, en_CC4-en_CC7	RA4, RA5, RA6, RA7	CE2, CE3, CE6	PEC2	35%
en_CG8, en_CG9, en_CB1-en_CB5, en_TRU1, en_TRU5, en_CC4-en_CC7	RA1, RA2, RA3, RA7	CE2, CE3	PL1	12,5%
en_CG8, en_CG9, en_CB1-en_CB5, en_TRU1, en_TRU5, en_CC4-en_CC7	RA4, RA5, RA6, RA7	CE2, CE3	PL2	12,5%
en_TR4, en_TR5		CE3	Interest and behavior	5%

The minumum rating to pass the subject is 5 points.

If the joint rating of the two PECs is less than 3.25 out of 7, the rating of the PECs cannot be compensated with the rest of the ratings.

If the joint rating of the two PECs and the two PLs in less than 4.25 the rating cannot be compensed with the rating of interes and behaviour.

In the ordinary call-final evaluation, the relationship between the competences, learning outcomes, criteria and evaluation instruments is as follows.

Skill	Learning Outcomes	Evaluation criteria	Grading Tool	Contribution to the final mark
en_CG8, en_CG9, en_CB1-en_CB5, en_TRU1, en_TRU5, en_CC4-en_CC7	RA1, RA2, RA3, RA4, RA5, RA6	CE1, CE2, CE3	PEF	100%

The grade of "Not Presented" will be granted to the student who, having opted for the continuous evaluation procedure, meets any of the following requirements:

- When the student has failed to attend 70% of the classes.
- When the student has not taken the continuous assessment tests.
- When the student has not submitted the laboratory tests.

Extraordinary call

In the case of the extraordinary call, the same percentages that have been established in the case of the evaluation by means of a final exam will be maintained.

6. BIBLIOGRAPHY

6.1. Basic Bibliography

- J.J. Cuadrado-Gallego and Y. Demchenko, Data Analytics: A Theoretical and Practical View from the EDISON Project. Springer, 2023. ISBN: 978-3-031-39129-3. Introduce in a theoretical and practical view the main techniques and algoritms for data analytics.
- J.J. Cuadrado-Gallego and Y. Demchenko, The Data Science Framework: A View from the EDISON Project. Springer, 2020. ISBN: 978-3-030-51023-7. Introduce the Data Science framework and allows the reader to know from the Data Science Body of Knowledge to the Data Science professional profiles.
- J. Han et al. Data Mining: Concepts and Techniques. Morgan Kaufman, 2012. ISBN: 978-0-12-381479-1. It is one of the reference books more used in the learning of Data Analytics, and allow to the reader to get into the concepts of the subject.

-
- M. Crawley, The R Book. John Wiley & Sons, 2007. An interesting book to get in deep the concepts of R. Can be access on-line from the library of the University of Alcalá.

6.2. Additional Bibliography

- J. Leskovec, Mining of Massive Datasets. Cambridge University Press, 2011. A very interesting book to get in deep in the concepts of Data Analytics.
- P. Teetor, R Cookbook. O'Reilly Media, 2011. ISBN: 978-0-596-80915-7. Very useful book to get into the R environment. Can be access on-line from the library of the University of Alcalá.

Disclosure Note

During the evaluation tests, the guidelines set out in the Regulations establishing the Rules of Coexistence of the University of Alcalá must be followed, as well as the possible implications of the irregularities committed during said tests, including the consequences for committing academic fraud according to the Regulation of Disciplinary Regime of the Students of the University of Alcalá.