GUÍA DOCENTE
2021 - 2022

FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Big Data y Gobierno del dato
Código	DTC-MBD-512
Título	Máster en Big Data. Tecnología y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics
Impartido en	Máster en Big Data. Tec. y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics [Primer Curso]
Nivel	Master
Cuatrimestre	Semestral
Créditos	3,0 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Telemática y Computación
Responsable	Carlos Morrás Ruiz-Falcó
Horario de tutorías	Concertar por email

Datos del profesorado	
Profesor	
Nombre	Carlos Morrás Ruiz-Falcó
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	cmorras@icai.comillas.edu
Profesor	
Nombre	Leticia Catalina López-Lapuente Gutiérrez
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	lclopezlapuente@comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura

Aportación al perfil profesional de la titulación

Big data is a new technology that plays a leading role in all processes where there is a large volume of data or where artificial intelligence, or machine learning algorithms are required. It is allowing to highly increase efficiency and effectiveness and enabling new business models that were previously impossible or unimaginable. Is a challenging to process and analyze in order to support decision-making.
It is necessary to understand the impact that this technology has and on what concepts it is based. Not only in the technical detail - which will be seen in other subjects - but also to understand in global terms its main characteristics, but also the ethical and legal connotations that using this technology implies and how to take care of the new valuable asset that is, and how to manage it with data governance principles.

During the course, students will learn the most relevant aspects of big data technology,

The objective of this course is for students to understand the power of Big Data, its transformative and disruptive value and the value of data and how to manage them in a company properly, from the ethical, legal and data governance aspects.
By the end of the course, students will:

- Understand what Big Data is in a global way. How and why it arises and its main characteristics.
- Understand the value of the data and know how to organize a data governance to maximize its value, and company's value
- Know the ethical and legal aspects (GDPR) that Big Data and Advanced Analytics systems must comply, and who to design in this mode. By the end of the course, students should have enough knowledge of big data technology to understand its potential, and have developed an informed criterion to determine when and how to use it in a professional context.

Prerequisitos

There are not special prerequisites for this course

Competencias - Objetivos

Competencias

General competences
CG1. Have acquired advanced knowledge and demonstrated, in a research and technological or highly specialized context, a detailed and well-founded understanding of the theoretical and practical aspects, as well as of the work methodology in one or more fields of study.
Haber adquirido conocimientos avanzados y demostrado, en un contexto de investigación científica y tecnológica o altamente especializado, una comprensión detallada y fundamentada de los aspectos teóricos y prácticos y de la metodología de trabajo en uno o más campos de estudio.

CG2. Know how to apply and integrate their knowledge, understanding, scientific rationale, and problemsolving skills to new and imprecisely defined environments, including highly specialized multidisciplinary research and professional contexts.
Saber aplicar e integrar sus conocimientos, la comprensión de estos, su fundamentación científica y sus capacidades de resolución de problemas en entornos nuevos y definidos de forma imprecisa, incluyendo contextos de carácter multidisciplinar tanto investigadores como profesionales altamente especializados.

CG3. Know how to evaluate and select the appropriate scientific theory and the precise methodology of their fields of study in order to formulate judgements based on incomplete or limited information, including, when necessary and pertinent, a discussion on the social or ethical responsibility linked to the solution proposed in each case.
Saber evaluar y seleccionar la teoría científica adecuada y la metodología precisa de sus campos de estudio para formular juicios a partir de información incompleta o limitada incluyendo, cuando sea preciso y pertinente, una reflexión sobre la responsabilidad social o ética ligada a la solución que se proponga en cada caso.

CG4. Be able to predict and control the evolution of complex situations through the development of new and innovative work methodologies adapted to the scientific/research, technological or specific professional field, in general multidisciplinary, in which they develop their activity.
Ser capaces de predecir y controlar la evolución de situaciones complejas mediante el desarrollo de nuevas e innovadoras metodologías de trabajo adaptadas al ámbito científico/investigador, tecnológico o profesional concreto, en general multidisciplinar, en el que se

desarrolle su actividad.

CG5. Be able to transmit in a clear and unambiguous manner, to specialist and non-specialist audiences, results from scientific and technological research or state-of-the-art innovation, as well as the most relevant foundations that support them.
Saber transmitir de un modo claro y sin ambigüedades, a un público especializado o no, resultados procedentes de la investigación científica y tecnológica o del ámbito de la innovación más avanzada, así como los fundamentos más relevantes sobre los que se sustentan.

CG6. Have developed sufficient autonomy to participate in research projects and scientific or technological collaborations within their thematic area, in interdisciplinary contexts and, where appropriate, with a high knowledge transfer component.
Haber desarrollado la autonomía suficiente para participar en proyectos de investigación y colaboraciones científicas o tecnológicas dentro de su ámbito temático, en contextos
interdisciplinares y, en su caso, con una alta componente de transferencia del conocimiento.

CG7. Being able to take responsibility for their own professional development and their specialization in one or more fields of study.
Ser capaces de asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más campos de estudio.

Specific competences
CE5. Know the techniques used to extract information from large datasets, as well as the different platforms, tools, and languages that make it possible.
Conocer las técnicas para extraer información de grandes conjuntos de datos, así como las diferentes plataformas, herramientas y lenguajes que lo hacen posible.

Resultados de Aprendizaje

By the end of the course students should:
RA1. Have familiarity with Bigdata common language and analytics that allows them to communicate with specialists.

RA2 List the characteristics and advantages of Big data systems.
RA3 Understand and propose uses of Bigdata in general and in the industry in particular.
RA4 Calculate the impact on the bigdata and analytical business
RA5 Capable of addressing simple analytical projects.
RA7 Develop analytical algorithms.
RA8 Properly assess the impact on people, business and society of the Bigdata, including its ethical and moral aspects.
RA8 Know the legal limitations of data use and privacy legislation.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

Theory

Unit 1. Introduction to big data: The value of the data in decision making

1.1 Data value

1.2 Is real or buble?

1.3 Why now? The digital society

Unit 2. Big data origin, motivation and history

2.1 Exponential and disruptive technologies

2.2 The big data revolution: The new paradigm of problem solving, from program-oriented to data-oriented

2.3 The problem of Google and Yahoo!

2.4 Open source and the Nutch project

2.5 Hadoop and evolution

2.6 Current trends: Internet of Things (IoT)

Unit 3. How does big data work?

3.1 Distribution and parallelism. Programming features

3.2 HDFS file system

3.3 MapReduce and Yarn

3.4 Redundancy and fault tolerance

Unit . The big data revolution:

.1 Revolutions around the history.

.2 Digital transformation

.3 The new paradigm of problem solving, from program-oriented to data-oriented

Unit 4. Fundamental characteristics of big data

4.1 The four V's

4.2 Structured and unstructured information: Impact of big data

Unit 5. Ethical aspects of data, big data and artificial intelligence

5.1 Ethics, privacy and data protection

5.2 Data compliance and other ethical aspects

5.3 Principles of GDPR

Unit 6 Privacy and data protection. Data compliance and other ethical aspects.

6.1 Basic principles. Constitution, human rights other

6.2 GDPR

6.3 Consents and legitimate use

6.4 Privacy by design and by default

6.5 Rights of the individual. Transparency, access, rectification, forgetting, portability, no treatment. obligations of the companies

6.6 Anonymization and other design criteria.

6.7 Fines and infractions

6.8 Practical case of GDPR implementation

Unit 7 Data governance and data Stewardship
7.1 Principles of data governance.
7.2 Management models. Figures of data management.
7.3 Data cycle
7.4 Data Management, quality, Security and compliance
7.5 Overview and main actors (AWS, Azure and Google)

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

To ensure useful and practical learning, theoretical classes will be combined with master classes that reflect the reality of the market. Real case examples with also be studied from business and technical perspectives, some of which will be used in practical sessions.

Metodología Presencial: Actividades

In-class activities
▪ Lectures: The lecturer will introduce the fundamental concepts of each unit, along with some practical recommendations, and will go through worked examples to support the explanation. and by proposing quizzes and short application exercises to be solved in class.
Competences: CG1, CG3, CG4, CG7, CE5
▪ Active participation and class discussion: will be encouraged by raising open questions to foster discussion., With the teacher as moderator, discussion in class and in electronic media will be encouraged by the students of topics and readings delivered in advance so that the student faces real situations
Competences: CG1, CG2, CG3, CG5, CG6, CG7, CE5
▪ Tutoring for groups or individual students will be organized upon request. –

Metodología No presencial: Actividades

Personal study of the course material and resolution of the proposed exercises.
Competences: CG1, CG3, CG4, CG7, CE5
▪ Preparation of work at home. Reading articles and watching videos to prepare the discussions
Competences: CG2, CG5, CE5

RESUMEN HORAS DE TRABAJO DEL ALUMNO

STUDENT WORK-TIME SUMMARY

IN-CLASS HOURS

Lectures	Quizzes and class participation
25	5

OUT-OF-CLASS HOURS

Self-study	Practice preparation	Report writing	Homework assignments
41,5	9.5	7.5	11

ECTS credits: 3 (90 hours)

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Assessment activities Grading criteria Weight

Quizzes and class participation 40%
▪ Understanding of the theoretical concepts.
▪ intensity, frequency, critical sense and adequacy of the participations

Final exam 60%
▪ Understanding of the theoretical concepts.
▪ Application of these concepts to problem-solving.
▪ Critical analysis of numerical exercises' results.

Calificaciones

GRADING AND COURSE RULES

Grading

Regular assessment

• Quizzes and class participation: 40%
• Final exam: 40% Mid term exam 20%
In order to pass the course, the weighted average mark must be greater or equal to 5 out of 10 points, the mark of the final exam must be greater or equal to 4 out of 10 points, Otherwise, the final grade will be the lower of the three marks.
Additionally, there may be an optional GDPR and privacy work that gives a maximum of 1 point on the final grade (only in case it is approved)

Retake

Case the part of quizzes and participation in class
• is greater than or equal to 7, then: It will be kept and will make a weighted average (40% / 60%) with a new final exam.
• is less than 7,then The grade will be that of the final exam (100%), having to obtain at least a 5 in it

Course rules

▪ Class attendance is mandatory according to Article 93 of the General Regulations (Reglamento General) of Comillas Pontifical University and Article 6 of the Academic Rules (Normas Académicas) of the ICAI School of Engineering. Not complying with this requirement may have the following consequences:
- Students who fail to attend more than 15% of the lectures may be denied the right to take the final exam

during the regular assessment period.
- Regarding practice, absence to more than 15% of the sessions can result in losing the right to take the final exam of the regular assessment period and the retake. Missed sessions must be made up for credit.
▪ Students who commit an irregularity in any graded activity will receive a mark of zero in the activity and disciplinary procedure will follow (cf. Article 168 of the General Regulations (Reglamento General) of Comillas Pontifical University).

PLAN DE TRABAJO Y CRONOGRAMA

Actividades

WORK PLAN AND SCHEDULE

In and out-of-class activities	Date/Periodicity	Deadline
Final exam	After the lecture period	
Videos and articles, homework	From week 1	

STUDENT WORK-TIME SUMMARY

IN-CLASS HOURS

Lectures	Quizzes and class participation
25	5

OUT-OF-CLASS HOURS

Self-study	Practice preparation	Report writing	Homework assignments
41,5	9.5	7.5	11

ECTS credits: 3 (90 hours)

	In-class activities				Out-of-class activities				Learning outcomes
Week	Time [h]	Lecture	Laboratory	Assessment	Time [h]	Self-study	Practice preparation and report writing	Other activities	Code
1	2	Course overview (0.5h) Introduction to big data (1.5h)			2	Review and self-study (2h)			
		Big data origin,				Review and		Film	

	2	motivation and history (2h)			3.5	self-study (2h)		watching (1.5h)	
2	2	How does big data work? (1.8h)		Quiz (0.2 h)	3	Review and self-study (2h)		Paper homework (1h)	
	2	Big data case 1 (1h.)	Practice 0 (1h)		4.5	Review and self-study (2h)	Practice preparation (2.5h)		
3	2	Fundamental characteristics of big data (2h)			2.5		Report writing (2.5h)		
	2	Ethical aspects of data, big data and AI (1.8h)		Quiz (0.2 h)	4	Review and self-study (2h)		Film watching (2h)	
4	2	RGPD and Privacy 1(2h)			6	Review and self-study (2h)		Paper homework (4h)	
	2	RGPD and Privacy 2(2h) Privacy by design and by default			2	Review and self-study (2h)			
5	2	RGPD and Privacy 1(2h). Rights of the individual. Some real cases			2		Practice preparation (2h)		
	2	Big data ecosystem and basic architecture (2h)	Practice 1 (2h)		2.5	Review and self-study (2h)		Quiz (0.5h)	
6	2	How to develop a big data case (2h)			4	Review and self-study (2h)		Paper homework (2h)	
		Big data case	Practice 2			Review and	Practice		

	2	Big data case 2 (2h)	Practice 2 (2h)		4.5	self-study (2h)	preparation (2.5h)		
7	2	Data governance. Principles and motivation.			2.5		Report writing (2.5h)		
	2	Organization of data governance, data management and data steward.			4.5	Review and self-study (2h)	Practice preparation (2.5h)		
8	2	Tools and real cases	Practice 3 (2h)		2.5		Report writing (2.5h)		
				Final exam[1]	10	Final exam preparation (10h)			

[1] The final exam will be held on the first week of March.

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

Slides prepared by the lecturer (available in Moodlerooms)

• Big Data: A Revolution That Will Transform How We Live, Work, and Think. Viktor Mayer-Schonberger (Autor), Kenneth Cukier (Autor)

• Creating Value with Big Data Analytics (Inglés) Tapa blanda – 14 ene 2016 de Peter Verhoef (Autor), Edwin Kooge (Colaborador) https://www.amazon.es/Creating-Value-Big-Data-Analytics/dp/1138837970

• GDPR: REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016. https://www.boe.es/doue/2016/119/L00001-00088.pdf y http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

• Una ética para Big data. Introducción a la gestión ética de datos masivos. 2018 Rosa Colmenarejo Fernández.

• Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance . ISBN:978-0124103894. David Plotkin.

Bibliografía Complementaria

• George Orwell. 1984 . ISBN: 9788499890944

• Michael Lewis. Moneyball: The Art of Winning an Unfair Game Paperback – March 17, 2004. ISBN: 978-0393324815 or the film 'Moneyball

(2011)'. Sony Pictures Director: Bennett Miller (Netflix, Movistar, otras)

• Facebooksitan: (varias plataformas, Movistar+, Netflix, Dirección: Jakob Gottschau Duración: 58 min

• El gran Hackeo. (The Great Hack 2019). Netflix. Dirección Karim Amer, Jehane Noujaim. 135 min.