

Plan Docente

* Asignatura: [Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Créditos: [Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Programa: [Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Módulo: [Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Materia: [Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Código: [Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Abreviatura:[Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Profesor coordinador: [Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Curso académico: [Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Convocatoria:[Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Cuatrimestre:[Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Campus:[Georgia 11, Normal], A cumplimentar por la dirección del programa.
* Referencia SGIC:DA\_PEPD\_ESP
* **Última revisión:**[Georgia 11, Normal], A cumplimentar por la dirección del programa.
* **Fecha de aprobación:**[Georgia 11, Normal], A cumplimentar por la dirección del programa.
* **Fecha de publicación:**[Georgia 11, Normal], A cumplimentar por la dirección del programa.

**Tabla de contenidos**

[1. Profesorado 3](#_Toc526704812)

[2. Presentación 5](#_Toc526704813)

[2.1. Descripción 5](#_Toc526704814)

[2.2. Campos profesionales en los que se proyecta 6](#_Toc526704815)

[3. Competencias 6](#_Toc526704816)

[3.1. Resultados del aprendizaje 6](#_Toc526704817)

[4. Programa 6](#_Toc526704818)

[5. Metodologías docentes 7](#_Toc526704819)

[6. Sistemas de evaluación 8](#_Toc526704820)

[7. Bibliografía 8](#_Toc526704821)

[7.2. Recomendada **¡Error! Marcador no definido.**](#_Toc526704822)

# Profesorado

**Montse Llos i Bombardó**

(montse.llos@campus.eae.es)



Montse Llos i Bombardó, Ingeniera Superior en Informática por La Salle – Universidad Ramon Llull, ha centrado su actividad profesional principalmente en el desarrollo de herramientas que ayuden en la toma de decisiones empresarial.

Aficionada al deporte de montaña y a descubrir el porqué de la existencia, entiende que al igual que uno crece personalmente a partir de la interacción con su entorno y con uno mismo, en una organización empresarial la información del entorno que le rodea es necesaria procesarla debidamente para llevar a cabo las acciones oportunas para su supervivencia y crecimiento. De ahí su interés por el mundo del Data Science y Machine Learning.

Con varias certificaciones de RapidMiner tanto a nivel de Use Cases, Data Engineering y Machine Learning, actualmente trabaja en Mind Analytics como Head of Data Science, donde desarrolla proyectos de Machine Learning para clientes que necesitan del uso de estas nuevas tecnologías para la obtención automática de conocimiento.

# Profesorado

**Miguel Rooney i Pelazzini**

(miguel.rooney@campus.eae.es)

A person wearing a suit and tie smiling at the camera

Description automatically generated

Miguel Rooney Pelazzini, Ingeniero Superior en Informática por la Universidad de Buenos Aires, cuenta con más de 30 años de experiencia en la Consultoría tecnológica aplicada a las diversas áreas funcionales de la empresa. Ha prestado sus servicios en sectores tales como: banca, seguros, automoción, logística industrial, química y telecomunicaciones. Tiene una amplia experiencia profesional en áreas: comercial, financiera, supply chain, calidad, ingeniería y logística. En todos ellos, con el objetivo de aplicar la tecnología como una herramienta de soporte a la actividad empresarial mediante la mejora de procesos de gestión.

Aficionado a las dos ruedas: bicicleta de montaña y motos de gran cilindrada. Desmontar y montar motores (sin que sobren piezas) y que funcionen mejor que antes.

Actualmente desarrolla funciones de consultoría de negocio y tecnológica, gestión y ejecución de proyectos en Mind Analytics.

# Presentación

## Descripción

Año tras año los gobiernos, empresas, colectivos e incluso las personas, generan cada vez más datos. Durante los últimos años se ha ido resolviendo la forma en como almacenar esta cantidad ingente de datos. Pero albergar muchos datos no tiene sentido sin una explotación adecuada de los mismos para la obtención de conocimiento para la empresa (o incluso terceros). Conocimiento que ayude en la toma de decisiones de forma objetiva, y no subjetiva como se hacía anteriormente.

La entrada del Business Intelligence ha ayudado muchísimo en plasmar toda la información relativa de la empresa. Y de la cual, después de analizarla, nos ayuda a saber en qué punto estamos y de dónde venimos. Pero, ¿y si pudiéramos actuar de forma proactiva? Determinar un punto crítico sobre el que se podría obtener una mejoría importante por el hecho de conocer de antemano y de forma automática lo más probable que suceda, nos puede ayudar en decidir qué hacer para evitarlo o, al menos, disminuirlo en grado de impacto negativo. Ahora es cuando el Machine Learning entra en acción.

De hecho, aunque surgió hacia los años 50 del siglo pasado, no ha sido hasta ahora con la mejora del hardware y el aumento de cantidad de datos, que ha empezado a usarse más ampliamente en el ámbito empresarial. De todas formas, aunque gran parte de las empresas comprenden que deben entrar en este mundo (desconociendo en qué ámbito), sólo un 25% de las empresas de todo el mundo tienen implementadas soluciones de Advanced Analytics.

En esta asignatura el alumno será guiado por la historia del aprendizaje automático desde sus inicios hasta cómo se pueden aprovechar todas las tecnologías de sistemas distribuidos BigData para aplicar modernos algoritmos.

Al finalizar el curso el alumno habrá adquirido los conocimientos suficientes para identificar qué algoritmos son los que se deben utilizar en cada caso de uso real, así como la capacidad de implementar y ajustar dichos algoritmos de manera óptima, consiguiendo con ello generar desde segmentaciones de grupos dentro de un conjunto de datos hasta complejos sistemas predictivos.

## Campos profesionales en los que se proyecta

La asignatura se proyecta en campos como:

* Business Intelligence, explotación de datos
* Customer Service, estudio de sentimientos, propensión de abandonos, etc
* Marketing, segmentación de mercados, campañas de publicidad orientada
* Finanzas, control de fraude, predicción de valores
* Industria y sensorización, detección de anomalías, etc.

# Competencias

## Resultados del aprendizaje

Los participantes desarrollarán habilidades técnicas y teóricas que les permitirán usar diferentes tipos de algoritmos de Machine Learning así como ajustarlos para su correcto funcionamiento utilizando diferentes herramientas como R, Python, Keras, Microsoft Azure ML o RapidMiner.

# Programa

* **Tema 1. Introducción a Data Science y Machine Learning:** Se verán las diferencias entre la programación clásica y el aprendizaje automático, casos de uso y herramientas actuales disponibles en el mercado.
* **Tema 2. Algoritmos supervisados de clasificación:** Durante este tema conoceremos los distintos tipos de algoritmos de clasificación más conocidos, cómo parametrizarlos y cómo evaluar su rendimiento mediante Confusion Matrix y ROC-Curve.
* **Tema 3. Metodología de trabajo de un proyecto de DS:** Hablaremos del Feature Engineering y de cómo mejorar un proceso de aprendizaje: Overfitting, Underfitting y Unbalanced Data.
* **Tema 4. Introducción al Python:** Trabajaremos dos aspectos complementarios. En primer lugar, entender el proceso mental requerido para desarrollar un programa de ordenador en cualquier lenguaje. Paralelamente, y mediante ejercicios prácticos, aprenderemos y aplicaremos los conceptos básicos que nos permitirán desarrollar y probar programas de nivel inicial en Python que se utilizarán posteriormente en otras unidades del programa.
* **Tema 5. Algoritmos supervisados de regresión:** Veremos cómo resolver problemas de predicción numérica, qué algoritmos usar y las diversas formas de evaluar su rendimiento. Hablaremos también de cuando es necesario normalizar los datos.
* **Tema 6. Algoritmos no supervisados:** Estudiaremos qué casos resolver con algoritmos no supervisados y veremos en detalle cómo implementarlos.
* **Tema 7. Reinforcement Learning, Deep Learning y otros:** Nos adentraremos en las últimas novedades en Machine Learning. Hablaremos de cuando es necesario usar Deep Learning y su implementación con Tensorflow, Keras,… También daremos una pincelada al Reinforcement Learning y al Natural Language Processing (NLP) o Text Mining.
* **Tema 8. Series temporales:** Las series temporales, como método tradicional de predicción también han evolucionado con el Machine Learning. Haremos un repaso a todas ellas.
* **Tema 9.** **Tips&Tricks:** Trucos y consejos para ser un buen data scientist. En este apartado veremos cómo crear un sistema de aprendizaje automático en tiempo real (Streaming)

En la mayoría de temas se realizarán prácticas en diferentes plataformas para afianzar los conocimientos.

# Metodologías docentes

El curso utiliza diferentes metodologías de aprendizaje, en la cual se combinan:

* Sesiones plenarias
* Prácticas a realizar en clase
* Trabajos individuales y en equipo

Las sesiones plenarias tienen como objetivo presentar los conceptos relativos a los temas del curso. En este sentido se espera que el alumno tenga una participación activa, a partir de lecturas previas que serán asignadas por el profesor en algunas de las sesiones. En estas sesiones se espera la asistencia de todo el grupo. Para ampliar los temas expuestos se realizarán prácticas técnicas con las herramientas descritas. Para su desarrollo se contará con la asistencia del profesor y de profesores adjuntos cuando proceda.

Por otro lado para el desarrollo de los trabajos finales se formarán grupos de trabajo pequeños o individuales que desarrollarán conjuntamente un caso presentado por el profesor.

# Sistemas de evaluación

|  |  |
| --- | --- |
| **Actividad de evaluación** | **Ponderación** |
| Examen teórico | 15% |
| Trabajo práctico | 60% |
| Participación y ejercicios desarrollados en clase | 25% |

# Bibliografía

* [Comunidad Kaggle](https://www.kaggle.com/)
  + Para Datasets
  + Competiciones de DS
* Estar al día sobre últimas noticias de Data Science
  + [Kdnuggets](https://www.kdnuggets.com/)
  + [Towards Data Science](https://towardsdatascience.com/)
  + [Medium](https://medium.com/)
  + [Data Science Central](https://www.datasciencecentral.com/)
* Libros
  + [An introduction to Statistical Learning](http://www-bcf.usc.edu/~gareth/ISL/) - Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani (free pdf)
  + Python Data Science HandBook – Jake VanderPlas (free)
  + [Predictive Analytics and Data Mining](https://www.amazon.es/Predictive-Analytics-Data-Mining-RapidMiner/dp/0128014601/ref=sr_1_2?ie=UTF8&qid=1535994186&sr=8-2&keywords=data+mining+and+predictive+analytics) – Kotu&Beshpande
  + [Deep Learning with R](https://www.manning.com/books/deep-learning-with-r) (o Python) – François Chollet
  + [Inteligencia Artificial](https://www.casadellibro.com/libro-inteligencia-artificial/9788423429448/6442598?gclid=eaiaiqobchmiwbaq0zof3qiv2ej3ch22ewkteayyasabegkqcpd_bwe&utm_source=google&utm_medium=cpc&utm_campaign=19438) – Pablo Rodríguez
  + [Forecasting. Principles and Practice](https://otexts.com/fpp2/index.html) – Rob J Hyndman & George Athanasopoulus (free online book)
* Manuales
  + [RapidMiner Training Videos](https://rapidminer.com/training/videos/)
  + [RapidMiner Academy](https://academy.rapidminer.com/pages/certification)
  + [Azure Machine Learning Studio Documentation](https://docs.microsoft.com/en-us/azure/machine-learning/studio/)
  + [Machine Learning With Spark](http://shop.oreilly.com/product/9781783288519.do)
  + [Pyhton](https://www.python.org/)
  + [The 15 Python libraries in Data Science](https://medium.com/activewizards-machine-learning-company/top-15-python-libraries-for-data-science-in-in-2017-ab61b4f9b4a7)
  + [Useful R packages](https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages)
  + [Neural Netwoks and Deep Learning](http://neuralnetworksanddeeplearning.com/)
  + [Convolutional Networks to Image Recognition](http://cs231n.github.io/)
* Canales YouTube
  + [Canal StatQuest](https://www.youtube.com/user/joshstarmer)
* Herramientas de apoyo a la programación y test
  + Online Live Programming: [http://pythontutor.com/visualize.html#mode=edit](http://pythontutor.com/visualize.html)
  + Python Cheat Sheets: <http://ehmatthes.github.io/pcc/cheatsheets/README.html>
  + Random data file generator: [https://mockaroo.com](https://mockaroo.com/)
  + Python documentation: <https://docs.python.org/3.7/index.html>
  + Python Official site: <https://www.python.org/>