



Título del trabajo: Aplicación de técnicas estadísticas en resolución de problemas de ciencia de datos
Tutor/a: Alberto Fernández Hilario
Cotutor/a:
Departamento responsable: Ciencias de la Computación e I.A.
Perfil y número de estudiantes al que va dirigido (máximo 2): <i>(Rellenar sólo en caso de que la propuesta esté realizada a través de estudiante)</i>
Estudiante que propone el trabajo (Nombre, Apellidos, DNI):
Tipo de trabajo <i>Debido a las materias que se van a tratar podría ser considerado tanto un trabajo de investigación como una simulación de encargo profesional</i>
Competencias <i>(estas son las mínimas; consultar (**) si se considera añadir otras)</i> <i>Competencias generales:</i> G02, G03, G04, G05, G06, G07, G08, G09. <i>Competencias específicas:</i> E01, E02, E03, E04, E05, E08, E09, E10.
Resultados de aprendizaje <i>(estos son los mínimos; añadir otros si se considera)</i> <ul style="list-style-type: none">• Adquirir competencias globales ligadas al desarrollo y aplicación de los conocimientos del Grado.• Adquirir competencias ligadas a la búsqueda y organización de información y documentación relevante sobre el tema objeto de estudio.• Aplicar el “pensamiento estadístico” y tener capacidad para enfrentarse a las distintas etapas de un estudio estadístico (desde el planteamiento del problema hasta la exposición de resultados).• Saber presentar, de forma escrita y oral, la memoria, los resultados y las conclusiones del trabajo realizado.
Antecedentes y resumen del tema propuesto: <p>La Ciencia de Datos es una de las ramas de conocimiento más importantes en la actualidad. Se trata de la extracción y presentación de conocimiento en grandes volúmenes sobre problemas estándar (clasificación, regresión) y no estándar (minería social, aprendizaje multi-vista), por lo que está íntimamente relacionado con el uso de herramientas para Big Data.</p> <p>En este trabajo, se pretende resolver algún problema de esta índole, preferentemente alguno de los disponibles en la competición kaggle. Para ello, se deberá tener en cuenta todo el proceso asociado a la gestión de este tipo de proyectos:</p> <ol style="list-style-type: none">1. Análisis descriptivo de los datos: qué características tiene el problema a resolver2. Preprocesamiento de datos: preparar la información para realizar un aprendizaje efectivo, incluyendo limpieza, selección, y transformación de los mismos.3. Aplicación de Técnicas de Minería de Datos: Técnicas de extracción de patrones y medidas de interés4. Presentación de Resultados: Técnicas de Visualización y de Representación del conocimiento <p>Es de especial relevancia resaltar que en función de la dimensión del problema a tratar, algunas</p>



de las herramientas o soluciones a aplicar se encontrarán en el contexto de Big Data, como pueda ser uso de bases de datos NoSQL, o las bibliotecas MLlib y Mahout cuyos modelos están implementados en el marco de trabajo MapReduce.

Breve descripción de las actividades presenciales y no presenciales a realizar:

Actividades presenciales (15-30%)	Planteamiento, orientación y supervisión	90
	Exposición del trabajo	1
	Otras:	9
Actividades no presenciales (70-85%)	Preparación del trabajo	250
	Elaboración de la memoria	50
	Otras:	
Total (12 ECTS)		300 horas

Objetivos que se pretenden alcanzar:

- Estudio y descripción de los algoritmos a aplicar según el problema tratado en cada uno de los pasos del trabajo.
- Estudio y aplicación de diferentes programas de minería de datos que cubran los diferentes algoritmos a aplicar.
- Realizar un estudio, descripción y presentación de los diferentes resultados obtenidos en el trabajo realizado.

Bibliografía básica para la puesta en marcha del trabajo:

- Pang-Ning Tan, Michael Steinbach & Vipin Kumar: Introduction to Data Mining Addison-Wesley, 2006.
- Cathy O'Neil and Rachel Schutt. Doing Data Science . O'Reilly Media, 2013
- Jiawei Han & Micheline Kamber: Data Mining: Concepts and Techniques Morgan Kaufmann, 2006.
- Michael R. Berthold, Christian Borgelt, Frank Hppner, and Frank Klawonn. 2010. *Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data* (1st ed.). Springer Publishing Company
- I.H. Witten, E. Frank. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, Morgan Kaufmann, 2005.
- Margaret H. Dunham. Data Mining: Introductory and Advanced Topics. Prentice Hall, 2003.
- G. Shmueli N R Patel P C P Bruce Prentice Hall, 2003 Data mining for business intelligence. Wiley 2010 (2nd. edition)
- V. Cherkassky, F.M. Mulier Learning from Data: Concepts, Theory, and Methods, 2nd Edition Wiley-IEE Prees, 2007



Tipo de trabajo (*):

1. Estudio de profundización en algún tema concreto de Estadística, o como proyecto de aplicación de la misma a estudios o problemas de otros ámbitos científicos o sociales.
2. Realización completa de todas las fases de un proyecto estadístico, bien con auxilio de prácticas en empresas o con prácticas propuestas y dirigidas por el tutor.
3. Estudio de casos, teóricos o prácticos, relacionados con la Estadística.
4. Elaboración de un informe o un proyecto de naturaleza profesional.
5. Elaboración de un plan de empresa.
6. Simulación de encargos profesionales.
7. Trabajos bibliográficos sobre el estado actual de una temática relacionada con la Estadística.
8. Creación y/o empleo de herramientas informáticas para su uso en Estadística.
9. Trabajos de inicio a la investigación.
10. Trabajos cuya finalidad sea la divulgación de la Estadística en diversos contextos.
11. Trabajos sobre Historia de la Estadística.
12. Trabajos relacionados con la docencia de la Estadística.

Competencias (**)

Competencias generales:

G01. Poseer los conocimientos básicos de los distintos módulos que, partiendo de la base de la educación secundaria general, y apoyándose en libros de texto avanzados, se desarrollan en la propuesta de título de Grado en Estadística que se presenta.

G02. Saber aplicar los conocimientos básicos de cada módulo a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de la Estadística y ámbitos en que esta se aplica directamente.

G03. Saber reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

G04. Poder transmitir información, ideas, problemas y sus soluciones, de forma escrita u oral, a un público tanto especializado como no especializado.

G05. Haber desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.

G06. Saber utilizar herramientas de búsqueda de recursos bibliográficos.

G07. Poder comunicarse en otra lengua de relevancia en el ámbito científico.

G08. Poseer habilidades y aptitudes que favorezcan el espíritu emprendedor en el ámbito de aplicación y desarrollo de su formación académica.

G09. Fomentar y garantizar el respeto a los Derechos Humanos, a los principios de accesibilidad universal, igualdad, y no discriminación; y los valores democráticos, de la cultura de la paz y de igualdad de género.



Competencias específicas:

E01. Conocer los fundamentos básicos del razonamiento estadístico, en el diseño de estudios, en la recogida de información, en el análisis de datos y en la extracción de conclusiones.

E02. Conocer, saber seleccionar y saber aplicar, técnicas de adquisición de datos para su tratamiento estadístico.

E03. Conocer los fundamentos teóricos y saber aplicar modelos y técnicas estadísticas en estudios y problemas reales en diversos ámbitos científicos y sociales.

E04. Saber seleccionar los modelos o técnicas estadísticas para su aplicación en estudios y problemas reales en diversos ámbitos científicos y sociales, así como conocer herramientas de validación de los mismos.

E05. Comprender la importancia de la Investigación Operativa como metodología de optimización, toma de decisiones y diseño de modelos particulares para la resolución de problemas en situaciones específicas.

E06. Comprender y utilizar básicamente el lenguaje matemático.

E07. Conocer los conceptos y herramientas matemáticas necesarias para el estudio de los aspectos teóricos y prácticos de la Probabilidad, la Estadística y la Investigación Operativa.

E08. Conocer y saber utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, bases de datos, visualización gráfica y optimización, que sean útiles para la aplicación y desarrollo de las técnicas estadísticas.

E09. Conocer los conceptos básicos y habilidades propias de un ámbito científico o social en el que la Estadística o la Investigación operativa sean una herramienta fundamental.

E10. Tomar conciencia de la necesidad de asumir las normas de ética profesional y las relativas a la protección de datos y del secreto estadístico, como premisas que deben guiar la actividad profesional como profesionales de la Estadística.