



UNIVERSIDAD
DE GRANADA



Facultad de Ciencias



Propuesta TFG_BIOTEC
Curso: 2018-19
DEPARTAMENTO:
Ciencias de la Computación e
Inteligencia Artificial

CÓDIGO DEL TFG: CCIA-3

1. DATOS DEL TFG OFERTADO:

Título: Ciencia de Datos aplicada a problemas Biológicos

Resumen (máx 250 palabras, estructurado en Introducción, Objetivos y Plan de trabajo):

La extracción de conocimiento sobre diferentes problemas en el área de biología según modelos de Machine Learning y Data Mining. Actualmente la dificultad a la hora de modelar nuevos escenarios de trabajo está creciendo. Por un lado, se encuentra el gran volumen de información que se es capaz de extraer de los distintos dominios de aplicación, lo que se conoce como Big Data. Por otro, la representación de esta información necesita ser preprocesada con metodologías robustas para mejorar la calidad de los datos, hasta alcanzar el Smart Data. Por ello se habla de la Ciencia de Datos como el proceso completo necesario para la extracción, interpretación, modificación, generación de modelos, y visualización de resultados. El manejo y comprensión de las herramientas de IA disponibles para realizar esta metodología es de especial interés en el área de la biología, la química, o la medicina.

Se pretende familiarizar al alumno/a con el proceso completo de Ciencia de Datos, mediante el uso e implementación de técnicas (mediante programación en Java, R, Python u otros lenguajes) y su uso en algunos problemas relacionados con la biología. Para ello se propone el siguiente plan de trabajo:

- Estudiar, comprender y resumir el funcionamiento de los algoritmos básicos más conocidos para preprocesamiento de información: selección de características y filtrado de ruido.
- Seleccionar métodos robustos de generación de modelos para clasificación, seleccionados de bibliotecas estándar.
- Familiarizarse con los problemas relativos a su campo de trabajo, en particular basados en micro-arrays. Ha de utilizar y buscar en los repositorios públicos disponibles conjuntos de datos relacionados con la biología.
- Realizar un análisis comparativo de los distintos algoritmos para determinar cuáles son los más interesantes.
- Implementar en Java, R, Python o lenguaje de programación equivalente los algoritmos que considere más interesantes.
- Intentar mejorar los resultados obtenidos por los modelos de clasificación estándar.
- Realizar un informe de visualización para representar los resultados obtenidos.

Tabla de actividades y dedicación estimada:

Planteamiento, orientación, supervisión, y preparación de la memoria	20
Preparación de la memoria	9
Desarrollo del trabajo	120
Exposición del trabajo	1
TOTAL (6 ECTS)	150 horas

OFERTADO POR:

Profesor del Departamento X

Profesor del Departamento junto con Empresa o Institución

Propuesto por alumno (*)

(*) En el caso de TFG propuesto por alumno, por favor completar la siguiente información sobre el mismo:

Apellidos:

Nombre:

e-mail institucional:

2. MODALIDAD: 1

1. Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado
2. Elaboración de un informe o un proyecto de naturaleza profesional
3. Elaboración de un plan de empresas

4. Simulación de encargos profesionales
5. Trabajos experimentales, de toma de datos.
6. Trabajos bibliográficos sobre el estado actual de una temática relacionada con el Grado.
7. Trabajos derivados de la experiencia desarrollada en prácticas externas.

3. COMPETENCIAS Y RESULTADOS DEL APRENDIZAJE:

CG1 - Capacidad para la modelización, simulación y optimización de procesos y productos biotecnológicos.

CB2 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio

CB3 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética

CB4 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado

CB5 - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CT1 - Capacidad de análisis y síntesis

CT2 - Capacidad de organizar y planificar

CT3 - Capacidad de aplicar los conocimientos en la práctica y de resolver problemas

CT4 - Capacidad de comunicar de forma oral y escrita en las lenguas del Grado

CT5 - Razonamiento crítico

CT8 - Capacidad para la toma de decisiones

CT9 - Capacidad de trabajar en equipo y en entornos multidisciplinares

CE2 - Poseer habilidades matemáticas, estadísticas e informáticas para obtener, analizar e interpretar datos, y para entender modelos sencillos.

CE3 - Saber buscar, obtener e interpretar la información de las principales bases de datos biológicos (genómicos, transcriptómicos, proteómicos, metabolómicos y similares, derivados de otros análisis masivos) y de datos bibliográficos

4. BIBLIOGRAFIA BÁSICA:

- Pang-Ning Tan, Michael Steinbach & Vipin Kumar: Introduction to Data Mining Addison-Wesley, 2006
- 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, Fourth Edition, Morgan Kaufmann, 2016
- Margaret H. Dunham. Data Mining: Introductory and Advanced Topics. Prentice Hall, 2003.
- G. Shmueli N R Patel P C P Bruce, Data mining for business intelligence. Wiley 2010 (2nd. edition)
- V. Cherkassky, F.M. Mulier Learning from Data: Concepts, Theory, and Methods, 2nd Edition Wiley-IEEE Press, 2007

5. ACLARACIONES PARA EL ESTUDIANTE:

Debe tener conocimientos previos de programación en Python, R, Java o similar.

3. DATOS DEL TUTOR/A UGR:

Apellidos: FERNÁNDEZ HILARIO
Teléfono: 958 240079

Nombre: ALBERTO
e-mail: alberto@decsai.ugr.es

**En el caso de trabajos desarrollados en Empresas u otras Instituciones ajenas a la Universidad de Granada, por favor completar la siguiente información:

TUTOR/A DE LA EMPRESA O INSTITUCIÓN:

Apellidos:
Empresa/Institución:
Teléfono:

Nombre:

e-mail: