



**UNIVERSIDAD
DE GRANADA**



Propuesta TFGB. Curso 2017-18

**DEPARTAMENTO: Ciencias de la
Computación e Inteligencia Artificial**

CÓDIGO DEL TFG: CCIA-08

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):

Introducción

Durante mucho tiempo la extracción de conocimiento sobre diferentes problemas en el área de biología se ha realizado en base a modelos estándar de Machine Learning (aprendizaje automático) y Data Mining (minería de datos). Sin embargo, actualmente la dificultad a la hora de modelar los nuevos escenarios de trabajo está creciendo a pasos agigantados. 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 lado, la propia representación de esta información necesita ser preprocesada con metodologías robustas para mejorar la calidad de los datos, hasta alcanzar el Smart Data. Es por ello que actualmente 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, la medicina, etc....

Objetivos

El objetivo del presente TFG es 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:

- El alumno/a debe estudiar, comprender y resumir el funcionamiento de los algoritmos básicos más conocidos para preprocesamiento de información, en particular selección de características y filtrado de ruido. Ello servirá para preparar los datos para su correcto aprendizaje.
- Además, debe ser capaz de seleccionar métodos robustos de generación de modelos para clasificación. Para ello, hará un repaso completo de las bibliotecas estándar de Machine Learning a este respecto.
- El alumno/a debe familiarizarse con los problemas relativos a su campo de trabajo. Son de especial interés aquellos basados en micro-arrays. En ese sentido, debe ser capaz de utilizar y buscar en los repositorios públicos disponibles conjuntos de datos relacionados con la biología, que le permitan analizar estadísticamente el comportamiento de dichos algoritmos.
- Realizar un análisis comparativo de los distintos algoritmos para determinar cuáles son los más interesantes.
- El alumno/a debe ser capaz de implementar en Java, R, Python o lenguaje de programación equivalente los algoritmos que considere más interesantes con el objetivo de mostrar que es capaz de entender y reproducir el funcionamiento de los métodos básicos para el procedimiento de Ciencia de Datos.
- De entre todos los conjuntos de datos analizados, el alumno/a debe seleccionar uno que le resulte interesante e intentar mejorar los resultados obtenidos por los modelos de clasificación estándar. Para ello, deberá realizar un flujo de procesamiento completo como el indicado al comienzo en la introducción de este documento.
- Finalmente, realizará un informe de visualización para representar los resultados obtenidos de forma que la comparativa sea útil a la hora de extraer conocimiento.

Con todo ello se pretende que el alumno/a sea capaz de enfrentarse a un problema de clasificación que se le pudiese plantear en su futuro desempeño como Biólogo/a, estudiando los algoritmos existentes, seleccionando el/los mejor/es, e incluso mejorándolos específicamente para obtener los mejores resultados posibles.

Conocimientos previos recomendados

Tener conocimientos de programación en Java, R, Python o similar.

Tabla de actividades y dedicación estimada:

Planteamiento, orientación y supervisión	30 horas
Exposición del trabajo	1 horas
Desarrollo del trabajo	209 horas
Preparación de la memoria	60 horas
TOTAL (12 ECTS)	300 horas

OFERTADO POR:

Profesor/a del Departamento

Profesor/a del Departamento junto con Empresa o Institución

Propuesto/Acordado por estudiante y profesor/a (*)

(*) En este caso, por favor completar la siguiente información sobre el estudiante:

Apellidos:

Nombre:

e-mail institucional:

2. MODALIDAD: Trabajo bibliográfico
Trabajo experimental **
Informe o proyecto de naturaleza profesional **



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: