UNIVERSIDAD DE GRANADA

Propuesta TFGB. Curso: 2020-2021

DEPARTAMENTO: CCIA

CÓDIGO DEL TFG: CCIA-07

1. DATOS DEL TFG OFERTADO:

Título: Big Data en problemas de aprendizaje para biología

Resumen (Introducción, Objetivos y Plan de trabajo; máx. 2.460 caracteres con espacios y fuente Arial 9):

Nos encontramos en el mundo del Big Data, donde el volumen de información existente en diferentes aplicaciones se ha incrementado exponencialmente. Cada vez resulta más sencillo generar y almacenar datos con el objetivo de extraer conocimiento interesante de los mismos. Por tanto, debemos tener en consideración los requisitos de escalabilidad en el desarrollo de modelos de Machine Learning. Las soluciones tradicionales deben adaptarse a entornos distribuidos como MapReduce para poder resolver los nuevos problemas de forma satisfactoria. El área de la biología es uno de estos contextos en los que la cantidad de información se considerar como Big Data. Por ello, para abordar con éxito este tipo de aplicaciones, será necesario el manejo de herramientas tales como Hadoop o Spark.

Se quiere familiarizar al alumno/a con el área de Ciencia de Datos y Big Data, utilizando problemas de carácter biológico. Para ello, deberá usar e implementar de tecnicas (mediante programación en Scala, Python, Java, u otros lenguajes) y su uso en algunos problemas de gran volumen relacionados con la biología. Para ello se propone el siguiente plan de trabajo:

- Estudiar, comprender y resumir el funcionamiento de las tecnologías asociadas con Big Data y Cloud Computing: Hadoop, Spark, etc.
- Seleccionar métodos robustos de generación de modelos en Machine Learning que sean escalables. Para ello, hará un repaso completo de las bibliotecas de Big Data a este respecto.
- Familiarizarse con los problemas relativos a su campo de trabajo, por ejemplo aquellos basados en matrices de expresión genética.
- Realizar un análisis comparativo de los distintos algoritmos para determinar cuales son los más interesantes.
- Implementar en Scala, Python, Java o lenguaje de programación equivalente los algoritmos que considere más interesantes.

Tabla de actividades y dedicación estimada:

Planteamiento y desarrollo del trabajo	200
Elaboración de la memoria	80
Preparación y ejecución de la exposición	20
TOTAL (12 ECTS)	**300 horas**

2. MODALIDAD (*): Trabajo Experimental / de Investigación

(*) En el caso de trabajos experimentales, el tutor considera conveniente que el estudiante realice el taller "Prevención de riesgos y eliminación de residuos en el laboratorio" ☐

3. DATOS DEL TUTOR/A UGR (**):

Apellidos: Fernández Hilario **Nombre**: Alberto

Teléfono: 958240079 ***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: **Nombre:**

Empresa / Institución:

Teléfono: ***e-mail:***

4. DATOS DEL ESTUDIANTE (***):

(***) Si ha sido acordado por el estudiante y profesor/a, por favor completar la siguiente información sobre el estudiante:

Apellidos: **Nombre:**

e-mail institucional: