



UGR | Universidad  
de Granada



Propuesta TFG. Curso 2015-16

Departamento  
Ciencias de la Computación e  
Inteligencia Artificial

## 1. DATOS DEL TFG OFERTADO

**Título del trabajo: Estudio y Análisis Comparativo de Algoritmos de Regresión y Clasificación aplicados sobre Conjuntos de Datos relacionados con la Biología**

**Resumen** (máx 250 palabras) estructurado en Objetivos y Plan de trabajo. Se debe incluir en folio adjunto.

**Palabras clave:** Minería de datos, Sistemas de regresión, Sistemas de clasificación.

**Número de alumnos por trabajo ofertado (máximo 3): 1**

**Ofertado por:**

1. Profesor del Departamento
2. Profesor del Departamento junto con Empresa ó Institución
3. Propuesto por alumno (  )

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

Nombre y apellidos del alumno:  
e-mail institucional:

## 2. MODALIDAD

1. Trabajo bibliográfico
2. Trabajo experimental (  )
3. Informe o proyecto de naturaleza profesional (  )

(  ) En el caso de trabajos experimentales e informes o proyectos de naturaleza profesional desarrollados en empresas u otras instituciones ajenas a la Universidad de Granada, por favor, completar la siguiente información

Nombre de la empresa/institución:  
Domicilio social:  
Teléfono/ e-mail de contacto:

## 3. DATOS DEL TUTOR DE LA UGR Y TUTOR DE LA EMPRESA O INSTITUCIÓN (en su caso) DEL TFG OFERTADO

**Nombre y apellidos del tutor/a UGR: Rafael Alcalá Fernández**

Teléfono: 958 240427

e-mail: [alcala@decsai.ugr.es](mailto:alcala@decsai.ugr.es)

**Nombre y apellidos del tutor/a de la empresa o institución:**

Empresa o Institución:

Teléfono:

e-mail:

## Resumen

### Introducción

El diseño de un modelo de regresión o clasificación a partir de datos, desde el punto de vista del aprendizaje supervisado, consiste en la obtención de modelos (sistemas basados en reglas difusas, sistemas basados en reglas intervalares, sistemas basados en árboles, sistemas basados en redes neuronales, etc.) que permitan predecir valores futuros de una variable objetivo a partir de mediciones o valores disponibles para otras variables de un sistema (regresión) o determinar la clase de un nuevo ejemplo dentro de un conjunto de clases conocido (clasificación). En algunos casos, caso de los modelos basados en reglas, el modelo puede ser interpretado desde el punto de vista del usuario final que pretende obtener un sistema regresor o clasificador para un problema dado, y en cualquier caso estos modelos pueden ser analizados y utilizados para inferir el posible valor o clase de la variable objetivo (normalmente desconocida y/o difícil de medir para ciertos problemas). El manejo y comprensión de éstas herramientas de la IA 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 con dicho tipo de técnicas mediante el uso de las mismas en algunos problemas relacionados con la biología. Para ello se propone el siguiente plan de trabajo:

- El alumno debe estudiar, comprender y resumir el funcionamiento de algunos de los algoritmos básicos más conocidos para problemas de regresión. Se proponen los siguientes algoritmos a estudiar: M5Rules-R, CART-R, Ensemble-R, EPSILON\_SVR-R, NU\_SVR-R y WM+GFS-LLRS-T.
- El alumno debe estudiar, comprender y resumir el funcionamiento de algunos de los algoritmos básicos más conocidos para problemas de clasificación. Se proponen los siguientes algoritmos a estudiar: C4.5Rules, FURIA, y FARC-HD-C.
- El alumno debe familiarizarse con la herramienta KEEL (<http://www.keel.es/>), para aplicar los algoritmos nombrados en los siguientes conjuntos de datos (o similares) para regresión: Relative location of CT slices, KEGG Metabolic Relation Network (directed and undirected), Breast Cancer Wisconsin (Prognostic), Physicochemical Properties of Protein Tertiary Structure, Fertility, etc. Y en los siguientes conjuntos de datos (o similares) para clasificación: Ecoli, Zoo, Abalone, Yeast, Breast, Bupa, newThyroid e Iris.
- El alumno debe realizar un análisis comparativo de los distintos algoritmos para determinar cuál es el más adecuado para cada tipo de problema. Adicionalmente, puesto que cualquier algoritmo de regresión puede ser adaptado fácilmente para abordar problemas de clasificación, deberá considerar el o los mejores algoritmos de regresión adaptados al problema de la clasificación para comparar con el o los mejores de clasificación, sacando las conclusiones oportunas respecto a cuáles son los más indicados para cada tipo de problema.

### Plan de trabajo

Tabla de actividades y dedicación horaria estimada	
Planteamiento, orientación y supervisión	30 horas
Formación en el uso de la herramienta Keel – Estudio de la bibliografía asociada	20 horas
Estudio de la bibliografía asociada a los algoritmos propuestos	40 horas
Búsqueda y preparación de los conjuntos de datos para los experimentos (al menos 15 para cada tipo problema)	40 horas
Ejecución de los algoritmos sobre los conjuntos de datos	40 horas
Aplicación de los tests estadísticos (Friedman, Holms...) para comparación de algoritmos y adaptación de los algoritmos de regresión para aplicarlos y compararlos con los de clasificación	40 horas
Preparación de la memoria	60 horas
Preparación de la exposición del trabajo	29 horas
Exposición del trabajo	1 horas
<b>TOTAL (12 ECTS)</b>	<b>300 horas</b>

### Requisitos

Capacidad de leer textos en inglés a nivel medio, Conocimientos de programación básicos (lenguaje JAVA a ser posible, Matlab en su defecto)

## Referencias

- Kuncheva L (2000) Fuzzy classifier design, Springer, Berlin
- Ishibuchi H, Nakashima T, Nii M (2004) Classification and modeling with linguistic information granules: Advanced approaches to linguistic data mining, Springer, Berlin
- J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, F. Herrera. KEEL: A Software Tool to Assess Evolutionary Algorithms to Data Mining Problems. *Soft Computing* 13:3 (2009) 307-318.
- J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera. KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. *Journal of Multiple-Valued Logic and Soft Computing* 17:2-3 (2011) 255-287.