



1. DATOS BÁSICOS DEL TFG:

Título: Modelos de clasificación y análisis discriminante para identificar perfiles de riesgo en salud

Descripción general (resumen y metodología):

Las herramientas de inteligencia artificial han surgido en los últimos años como una ayuda para los profesionales de la salud a la hora de identificar enfermedades o pacientes en riesgo de sufrir una enfermedad o de recaer en ella. Los algoritmos utilizados para esta tarea reciben un conjunto de datos del cual “aprenden” las reglas y patrones que existen entre las distintas variables de este, elaborando así modelos de clasificación automática en una u otra categoría. Dichos algoritmos han ido sofisticándose con los años, a medida que ha mejorado la calidad de los datos de entrenamiento y la potencia predictiva de los métodos de aprendizaje automático. Estos algoritmos conviven con los modelos desarrollados por la estadística clásica muchos años antes, y que todavía siguen ofreciendo un rendimiento bastante aceptable cuando se comparan con los algoritmos de aprendizaje automático más modernos.

En este Trabajo Fin de Grado se realizará una aplicación de los modelos de clasificación en el contexto de la salud, con el objetivo de estudiar su eficacia a la hora de identificar pacientes en riesgo de sufrir una enfermedad. Se introducirá el problema de clasificación desde un punto de vista estadístico, así como los principales métodos y algoritmos de clasificación existentes a día de hoy, para posteriormente realizar una aplicación utilizando un conjunto de datos real.

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

El objetivo principal de este trabajo es el estudio y aplicación de técnicas de clasificación y análisis discriminante en el contexto de la salud. Los objetivos específicos son los siguientes:

OE1: Estudio de técnicas de clasificación: análisis discriminante, regresión logística, clasificación mediante vecinos más cercanos, modelos basados en árboles, métodos de boosting, máquinas de soporte vectorial.

OE2: Estudio de los métodos de entrenamiento y validación de los modelos de clasificación: selección de variables, técnicas de regularización, validación cruzada, bootstrap.

OE3: Aplicación de las técnicas en un conjunto de datos real para identificar pacientes en riesgo de sufrir una enfermedad. Evaluación de la precisión de los modelos de clasificación e interpretación de los resultados.

Bibliografía básica:

- Agresti, A. (2013). Categorical data analysis. John Wiley & Sons.
- Barber, D. (2012). Bayesian reasoning and machine learning. Cambridge University Press.
- Bramer, M. (2007). Principles of data mining. London: Springer London.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with applications in R. New York: Springer.
- Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. New York: Springer.

Recomendaciones y orientaciones para el estudiante:

Se recomienda que el estudiante haya cursado las materias Modelos Lineales, Análisis Multivariante, Análisis de Datos y Áreas de Aplicación del Grado en Estadística.

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: RAMÓN FERRI GARCÍA

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: rferri@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos:

Ámbito de conocimiento/Departamento:

Correo electrónico:

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: Belén Rivero Rodríguez

Correo electrónico: belenrivero@correo.ugr.es