



1. DATOS BÁSICOS DEL TFG:

Título: Recopilación e implementación de las medidas de eficiencia de un clasificador binario

Descripción general (resumen y metodología):

La clasificación binaria de un conjunto de datos consiste en su dicotomización o asignación de estos datos a una categoría entre dos posibles. Este tipo de clasificación es del máximo interés en diversas áreas, como son el diagnóstico clínico o el desarrollo de sistemas de decisión automática. Actualmente hay numerosos índices para caracterizar la eficiencia de un clasificador binario. También son numerosos los métodos que permiten comparar estos índices de cara a discriminar qué clasificador es preferible cuando se presentan diferentes alternativas.

Esta propuesta de Trabajo de Fin de Grado plantea una revisión bibliográfica inicial que permita recopilar las medidas de eficiencia y criterios comparativos considerados actualmente en la literatura y la implementación de las medidas y métodos más relevantes. Es deseable que esta implementación contemple la elaboración de un interfaz de usuario amigable, como puede ser una aplicación Shiny.

Tipología: Trabajos bibliográficos sobre el estado actual de una temática relacionada con el Grado.

Objetivos planteados:

- Realizar una revisión bibliográfica que permita recopilar las medidas de eficiencia relativas a un clasificador binario y aquellas que permiten comparar entre clasificadores alternativos.
- Implementar un interfaz en forma de aplicación Shiny que tome como entrada las frecuencias de matrices de confusión (tablas de contingencia 2x2) y proporcione la estimación de las medidas de eficiencia y criterios comparativos recopilados en la revisión realizada.

Bibliografía básica:

- Martín, A. & Femia, P (2006) Chance-Corrected Measures of Reliability and Validity in 2×2 Tables. Communications in Statistics - Theory and Methods, 37:5, 760 - 772
- Parker, C. (2011) An Analysis of Performance Measures For Binary Classifiers. IEEE 11th International Conference on Data Mining, Vancouver, BC, Canada, 2011, pp. 517-526, doi: 10.1109/ICDM.2011.21.
- Pepe, MS (2003) The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford.
- Wickham, H. (2021) Mastering Shiny. O'Reilly.

Recomendaciones y orientaciones para el estudiante:

Tener conocimientos básicos de programación en R. Tener interés por la programación.

Plazas: 2

2. DATOS DEL TUTOR/A:

Nombre y apellidos: PEDRO JESÚS FEMIA MARZO

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: pfemia@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:

Correo electrónico: