

Herramienta Educativa R Shiny para Diagnóstico de Datasets

Palabras clave: R Shiny, análisis exploratorio de datos, educación estadística, ciencia de datos

Abstract

This paper presents a R Shiny application designed as an educational tool for the diagnosis and exploration of datasets. The application allows students from different educational levels to perform exploratory data analysis in an interactive and visual way, facilitating the understanding of fundamental statistical concepts.

Introducción:

En la era actual de la información, la capacidad de analizar y comprender datos es una habilidad crucial para estudiantes de diversas disciplinas. Sin embargo, la complejidad de las herramientas estadísticas tradicionales puede resultar abrumadora para los principiantes. Este trabajo presenta una aplicación R Shiny (Wickham, 2021) diseñada como herramienta educativa interactiva para el diagnóstico y exploración de datasets, con el objetivo de facilitar el aprendizaje de conceptos estadísticos y de ciencia de datos en diversos niveles educativos. Inicialmente, la aplicación se diseñó para las escuelas de habla portuguesa de Brasil, sólo para las pruebas preliminares.

Funcionalidades principales:

Las principales funcionalidades siguen los conceptos estadísticos de Spiegel y Stephens (2008) y las principales consideraciones sobre la calidad de un conjunto de datos de Foxwell (2020). Estas funcionalidades son:

- Carga y visualización de datasets en formato .CSV
- Resumen estadístico detallado de variables
- Detección de valores atípicos, datos faltantes y duplicados
- Generación de gráficos exploratorios (histogramas, diagramas de caja, gráficos de dispersión)
- Análisis de balance de clases para variables categóricas
- Interfaz intuitiva con controles interactivos

Aplicabilidad educativa:

La herramienta está diseñada para su uso en diversos contextos educativos, incluyendo:

- Curso de Análisis y Desarrollo de Sistemas, disciplina Fundamentos de Probabilidad y Estadística
- Introducción a la Ciencia de Datos y Aprendizaje Automático
- Curso Técnico Integrado de Telecomunicaciones

Objetivos pedagógicos:

La herramienta se diseñó para que estuviera centrada en el alumno, utilizará una metodología activa para que los alumnos participaran activamente en el aprendizaje (Prince, 2004) y un aprendizaje interactivo, para que los alumnos pudieran interactuar con la tecnología (Sanusi y Oyelere, 2020).

Herramienta Educativa Shiny para Diagnóstico de Datasets

Como método de instrucción, la herramienta propone la interacción a través de actividades prácticas (Ekwueme et al., 2015). Todo el marco pedagógico de la herramienta busca involucrar a los estudiantes en un pensamiento de orden superior (analizar, sintetizar, evaluar y reflexionar) sobre aquello con lo que están interactuando y aprendiendo (Sanusi y Oyelere, 2020).

Pantalla de inicio:

	Id	SepallengthCm	SepalwidthCm	PetalLengthCm	PetalWidthCm	Species
1	1	5.1	3.5	1.4	0.2	Iris-setosa
2	2	4.9	3.0	1.4	0.2	Iris-setosa
3	3	4.7	3.2	1.3	0.2	Iris-setosa
4	4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	5.0	3.6	1.4	0.2	Iris-setosa
6	6	5.4	3.9	1.7	0.4	Iris-setosa
7	7	4.6	3.4	1.4	0.3	Iris-setosa
8	8	5.0	3.4	1.5	0.2	Iris-setosa
9	9	4.4	2.9	1.4	0.2	Iris-setosa
10	10	4.9	3.1	1.5	0.1	Iris-setosa

```
'data.frame': 150 obs. of 6 variables:  
 $ Id : int 1 2 3 4 5 6 7 8 9 10 ...  
 $ SepallengthCm: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
 $ SepalwidthCm : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
 $ PetallengthCm: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
 $ PetalwidthCm : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
 $ Species : chr "Iris-setosa" "Iris-setosa" "Iris-setosa" "Iris-setosa" ...
```

Trabajos futuros:

El desarrollo futuro de esta herramienta se enfocará en:

- Expandir las funcionalidades para incluir etapas iniciales del proceso de desarrollo de modelos de Machine Learning, como selección de características y partición de datos.
- Que esté disponible en español, francés, guaraní, quechua e inglés.
- Implementar la aplicación en un ambiente educativo real, evaluando tanto la percepción de aprendizaje como la experiencia de uso por parte de los estudiantes.
- Refinar continuamente la herramienta basándose en el feedback de los usuarios, siguiendo una metodología de Investigación-Acción (Kemmis et al., 2014).
- Crear recursos complementarios, como tutoriales y guías de estudio, para maximizar el potencial educativo de la herramienta.

Referencias:

- Ekwueme, C., Ekon, E. E., & Ezenwa-Nebife, D. (2015). The Impact of Hands-On-Approach on Student Academic Performance in Basic Science and Mathematics. *Higher Education Studies*, 47(5).
- Foxwell, H. J. (2020). Creating Good Data: A Guide to Dataset Structure and Data Representation. Apress. <https://doi.org/10.1007/978-1-4842-6103-3>
- Kemmis, S., McTaggart, R., & Nixon, R. (2014). The Action Research Planner: Doing Critical Participatory Action Research. Springer.
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223-231.
- Sanusi, I. T., & Oyelere, S. S. (2020). Pedagogies of Machine Learning in K-12 Context. In *IEEE Frontiers in Education Conference* (pp. 1-8). Uppsala, Sweden.
- Spiegel, M. R., & Stephens, L. J. (2008). Statistics. Theory and Problems of Statistics (4th ed.). McGraw-Hill.
- Wickham, H. (2021). Mastering Shiny. O'Reilly Media.

URL de la aplicación R Shiny:

<https://presentedtolatinr2024anonimous.shinyapps.io/diagnosticdataset/>