

# Una aplicación de R en el análisis del desempeño y equidad de algoritmos de clasificación con ruido en la clase y atributos

Ana Vargas<sup>1,2,\*</sup>, Rosa Delgadillo<sup>2</sup>

(1)Departamento Académico de Estadística en Informática, Universidad Nacional Agraria La Molina, Perú

(2) Programa de Postgrado en Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos, Perú

**Palabras clave:** clasificación, modelo de ruido, equidad

## Abstract

El ruido en los datos es una fuente de sesgo que ha sido poco explorada. En este estudio, se utiliza R para investigar el impacto del ruido en las métricas de rendimiento y equidad, específicamente la exactitud (ACC) y la diferencia absoluta de probabilidades (AODD). Se analizan tres clasificadores clásicos: random forest (RF), gradient boosting (GB) y regresión logística (RL), empleando tres conjuntos de datos públicos para clasificación binaria. Se realiza un análisis comparativo que examina los efectos de distintos tipos y niveles de ruido introducidos tanto en las variables predictoras (atributos) como en la variable de respuesta (clase).

## Introducción

El ruido en los datos es una fuente de sesgo poco explorada que, según Corbett-Davies et al. (2017), puede aumentar la discriminación contra grupos minoritarios o vulnerables (Barocas et al., 2017). Debido a que el ruido en los datos reales es difícil de cuantificar y sus características son desconocidas, muchos estudios optan por introducir ruido sintético de manera experimental y controlada para extraer conclusiones basadas en su tipo, frecuencia y características. La mayoría de estas investigaciones se han centrado en el efecto del ruido en la clase, mientras que menos estudios han explorado su impacto en los atributos, y pocos han considerado ambas condiciones (Sáez, 2022).

En este trabajo, se utiliza la librería `noisemodel` en R (Sáez, 2023) para introducir ruido de manera controlada en conjuntos de datos de clasificación. Además, se emplean las librerías `caret` (Kunh et al., 2020) para el entrenamiento de modelos y `fairness` (Kozodoi y Varga, 2021) para estimar la métrica de equidad (AODD).

# UNA APLICACIÓN DE R EN EL ANÁLISIS DEL DESEMPEÑO Y EQUITAD DE ALGORITMOS DE CLASIFICACIÓN CON RUIDO EN LA CLASE Y ATRIBUTOS

## Análisis experimental

Se diseñó un experimento para evaluar el desempeño y la equidad de tres clasificadores clásicos: random forest (RF), gradient boosting (GB) y regresión logística (RL), utilizando tres conjuntos de datos públicos (Lsa, Cobertura y Empleo) orientados a tareas de clasificación binaria.

Cada conjunto de datos se dividió aleatoriamente en un 70% para entrenamiento y un 30% para prueba. Se introdujo ruido en niveles del 0%, 10%, 20%, 30% y 40%, considerando su aplicación en las variables de clase (ln), en los atributos (an) o en ambos (cn), mediante un modelo de ruido uniforme y simétrico (Sáez, 2022) de la librería `noisemodel` de R (Sáez, 2023). Cada escenario se repitió 10 veces.

Para evaluar el desempeño, se utilizó la métrica de exactitud (ACC), y para evaluar la equidad, se consideró el sexo (femenino y masculino) como un atributo sensible o protegido, con el fin de determinar si los modelos construidos son justos para ambos grupos. La métrica utilizada fue la Diferencia Absoluta Promedio de Probabilidades (AODD) que se estimó empleando la librería `fairness` (Kozodoi, y Varga, 2021).

Un AODD de 0 indica equidad perfecta, mientras que un valor mayor sugiere diferencias en las tasas de error entre los grupos, señalando sesgo en el modelo.

## Resultados preliminares

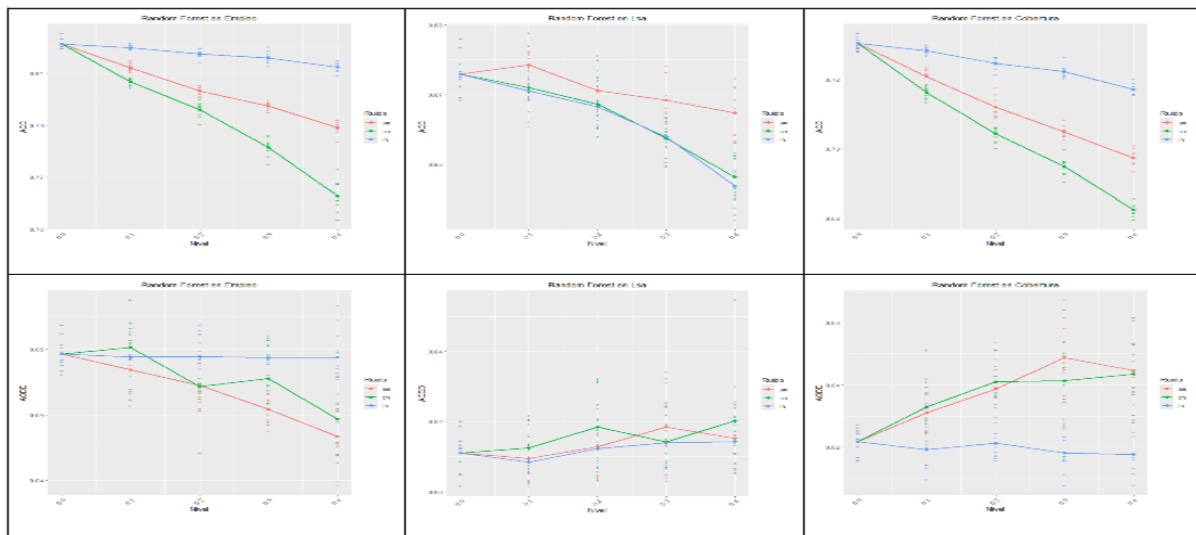


Figura 1. ACC y AODD según el tipo de ruido inyectado y nivel de ruido con RF

# UNA APLICACIÓN DE R EN EL ANÁLISIS DEL DESEMPEÑO Y EQUITAD DE ALGORITMOS DE CLASIFICACIÓN CON RUIDO EN LA CLASE Y ATRIBUTOS

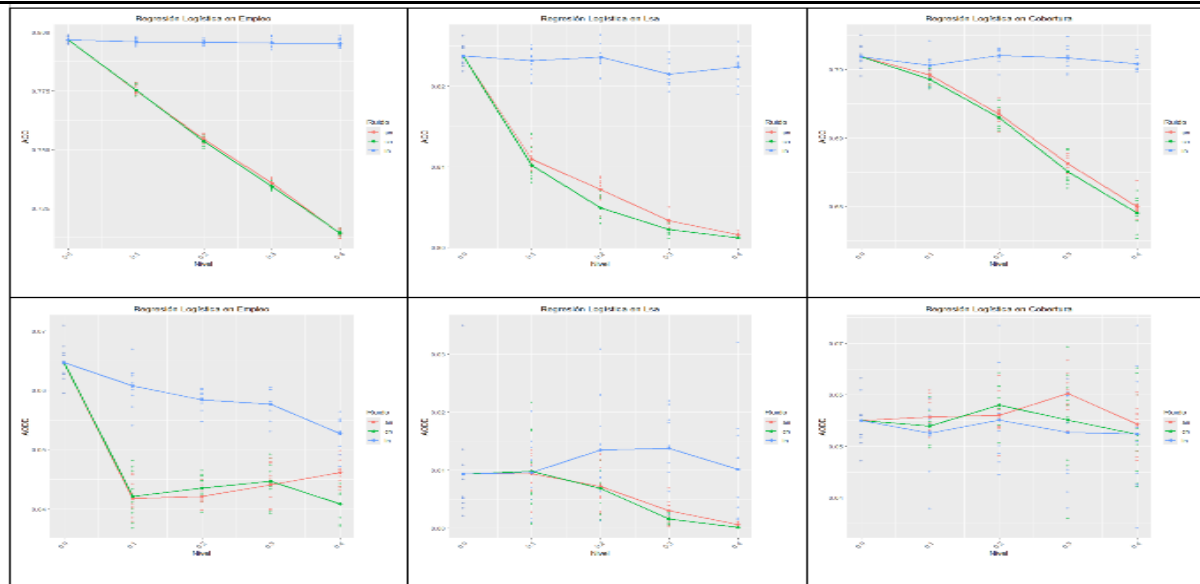


Figura 2. ACC y AODD según el tipo de ruido inyectado y nivel de ruido con RL

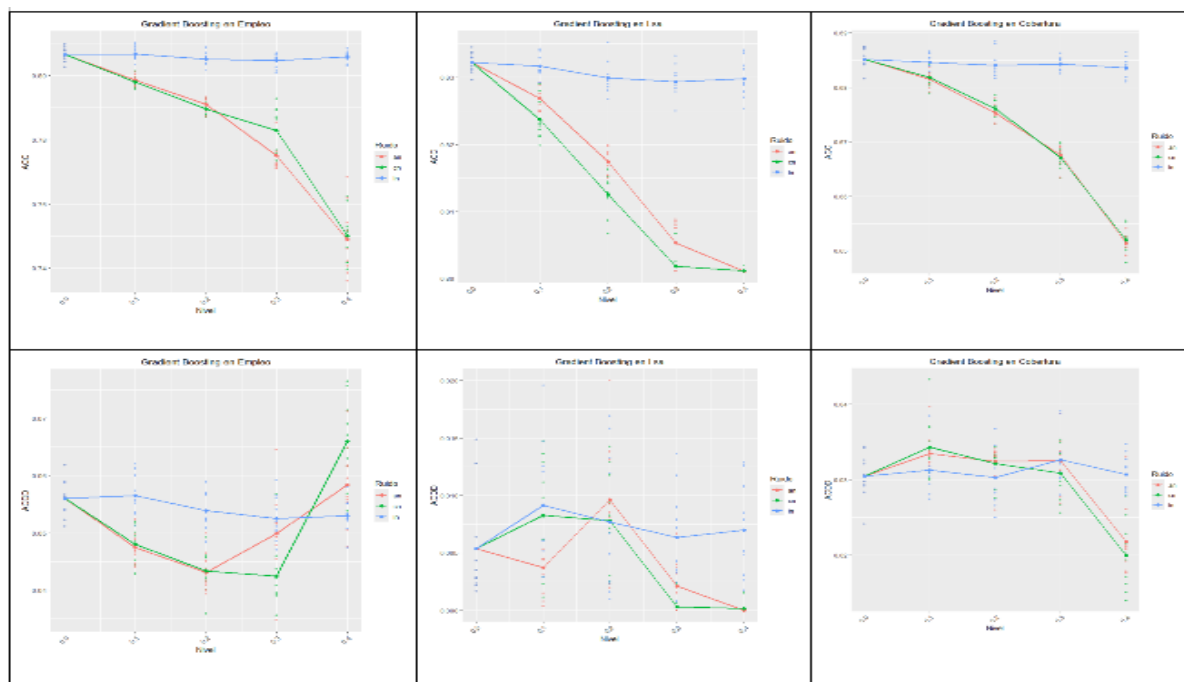


Figura 3. ACC y AODD según el tipo de ruido inyectado y nivel de ruido con GB

- En general, la medida ACC disminuye conforme aumenta el nivel de ruido añadido (10%, 20%, 30%, 40%), ya sea que el ruido sea del tipo an o cn, en cualquiera de los algoritmos. Sin embargo, AODD muestra un comportamiento variable según el algoritmo.
- El efecto del ruido, cuando se introduce en las etiquetas (ln), se muestra generalmente robusto tanto para la métrica de exactitud como para la de equidad a medida que aumenta el porcentaje de datos ruidosos.

## Referencias

Barocas, S., Hardt, M., & Narayanan, A. (2017). Fairness in machine learning. *Nips tutorial*, 1, 2.

Corbett-Davies, S. and Goel, S. (2018). The measure and mismeasure of fairness: A critical review of fair machine learning. arXiv preprint arXiv:1808.00023.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>

Sáez, J. A. (2022). Noise models in classification: Unified nomenclature, extended taxonomy and pragmatic categorization. *Mathematics*, 10(20), 3736.

Kozodoi, N., & Varga, T. V. (2021). *fairness: Algorithmic fairness metrics* (Version 1.2.2) [R package]. CRAN. <https://CRAN.R-project.org/package=fairness>

Sáez, J. A. (2023). Noise simulation in classification with the noisemodel R package: Applications analyzing the impact of errors with chemical data. *Journal of Chemometrics*, 37(5), e3472.