

## APRENDIZAJE AUTOMÁTICO APLICADO A LA PREDICCIÓN DE CARACTERES COMPLEJOS

Reinoso-Peláez<sup>1</sup>, E.L., Gianola<sup>2</sup>, D. y González-Recio<sup>1</sup>, O.

<sup>1</sup>Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria. Ctra. de La Coruña, km 7,5, 28040 Madrid, España. <sup>2</sup>University of Wisconsin–Madison, Department of Animal and Dairy Sciences. 1675 Observatory Dr, Madison, WI 53706, Estados Unidos; edgar.reinoso@inia.es.

### INTRODUCCIÓN

El crecimiento de la inteligencia artificial y el aprendizaje automático (ML) ha sido exponencial en los últimos años. Las aplicaciones de la predicción usando datos genómicos masivos (GWP) se han desarrollado enormemente, con aplicaciones variadas en agricultura y ganadería. Pero ¿cuál es el mejor modelo de ML?, ¿son los algoritmos de ML mejores que los modelos lineales clásicos o bayesianos? Este trabajo realiza un meta-análisis involucrando modelos de predicción o clasificación para identificar aquellos con mayor capacidad predictiva para caracteres complejos aplicados a la mejora genética animal.

### MATERIAL Y MÉTODOS

Se realizó una revisión sistemática de artículos publicados entre 2010 y 2020 que usaban al menos un modelo de ML. De un total de 34 estudios se obtuvieron 244 comparaciones con 19 modelos estadísticos. Se realizó un meta-análisis usando un modelo bayesiano tipo Thurston para rankings con el software GIBBSTHUR. El modelo fue:  $l_m = X_{bm}b + Z_{pm}p + e_m$ , donde  $l$  corresponde al ranking del modelo;  $b$  corresponde al efecto del tipo de estudio (planta o animal);  $p$  corresponde al efecto del modelo estadístico con distribución  $p \sim N(0, \sigma_p^2)$ , donde  $\sigma_p^2$  es la varianza asociada al efecto del método;  $e_m \sim N(0, \sigma_e^2)$  fue el residuo con varianza residual  $\sigma_e^2 = 1$ ; y  $X_{bm}$  y  $Z_{pm}$  fueron las respectivas matrices de incidencias.

### RESULTADOS Y DISCUSIÓN

Los modelos con mejor capacidad predictiva en términos de correlación de Pearson en el set de predicción fueron: espacios de Hilbert con núcleo reproductor (RKHS), Naive Bayes (NB) y redes neuronales Bayesianas (BNN); a estos les siguen Bayes B (BB), Bayes A (BA), Regresión ponderada de contracción bayesiana (wBSR), Bayes C (BC), Bagging (BAGG), Redes neuronales artificiales (ANN), GBLUP, Bayes LASSO (BL), bosques aleatorios (RF), Máquinas de vectores de soporte (SVM), redes neuronales convolucionales (CNN), selección de variable de búsqueda estocástica (SSVS), LASSO, regresión de cresta (RR), algoritmo de Boosting (BOOST), y K vecinos más cercanos (KNN). Cuando se usó el error cuadrático medio de la predicción como métrica de comparación, los algoritmos BOOST y BNN se situaron en los primeros puestos, seguido de BL, SVM, RF, CNN, BB, GBLUP, LASSO, BC, BA, RR, y ANN. Se debe tener en cuenta que las desviaciones estándar de la distribución posterior de las soluciones fueron amplias, y los resultados deben ser interpretados con cautela.

### CONCLUSIÓN

Los modelos Bayesianos (BNN, NB, BA, BB) y de kernel (RKHS) mostraron las mejores capacidades predictivas. Los métodos de ensamblaje (BOOST, RF, BAGG) han presentado puntajes variados, aunque se ha demostrado que son modelos robustos y ampliamente usados. Cada modelo tiene cualidades que pueden ser efectivas o no, dependiendo de los datos, el ambiente, y el objetivo de estudio.

### REFERENCIAS BIBLIOGRÁFICAS

- Abdollahi-Arpanahi, R. *et al.* 2020. Genet. Sel. Evol. 52: 12.
- Abdollahi-Arpanahi, R. *et al.* 2015. J. Anim. Breed. Genet. 132: 218–228.
- Gentzmittel, L. *et al.* 2019. Genome Biol. 20: 1–20.
- González-Recio, O., and S. Forni. 2011. Genet. Sel. Evol. 43: 7.
- González-Recio, O. *et al.* 2010. Genet. Res. (Camb). 92: 227–237.
- Grinberg, N.F. *et al.* 2016. Front. Plant Sci. 7: 133.
- Grinberg, N.F. *et al.* 2020. Mach. Learn. 109: 251–277.
- Van der Heide, E.M.M. *et al.* 2019. J. Dairy Sci. 102: 9409–9421.
- Heslot, N. *et al.* 2012. Crop Sci. 52:146–160.
- Jiménez-Montero, J.A. *et al.* 2013. J. Dairy Sci. 96: 625–634.
- Kim, K. *et al.* 2015. PLoS One 10: e0139685.
- Kwong *et al.* 2017. BMC Genet. 18:107.
- Liu, Y. *et al.* 2019. Front. Genet. 10: 1091.
- Long, N. *et al.* 2011. Theor. Appl. Genet. 123: 1065–1074.
- Ma, W. *et al.* 2018. Planta 248 :1307–1318.
- Sarkar, R. *et al.* 2015. Genet. 94: 187–192.
- Schulz-Streeck, T. *et al.* 2013. Appl. Genet. 126: 69–82.
- Waldmann, P. 2018. Genet. Sel. Evol. 50: 70.
- Xavier, A. *et al.* 2016. G3 Genes, Genomes, Genet. 6: 2611–2616.
- Zingaretti, L.M., *et al.* 2020. Front. Plant Sci. 11: 1.