

# Modelos predictivos y selección de variables en el monitoreo de enfermedades agrícolas: un enfoque con R

**Palabras clave:** Patosistemas, variables biometeorológicas, reducción de dimensionalidad, regresión logística.

## Abstract

El monitoreo de cultivos es una labor intensiva que consume tiempo y a menudo requiere de métodos destructivos. Por ello, contar con modelos predictivos de la probabilidad de desarrollo de una enfermedad según el ambiente en el cual se desarrollará, permite planificar técnicas de manejo, selección de cultivares o líneas tolerantes/resistentes para evitar pérdidas por enfermedades. Para modelar las variaciones en el desarrollo de enfermedades de cultivos agrícolas debida a realizaciones aleatorias de las temperaturas, precipitaciones y otras variables meteorológicas en determinados periodos de tiempo previo y durante el ciclo del cultivo, mediante modelos interpretables y parsimoniosos resulta indispensable recurrir a herramientas analíticas que permitan reducir la dimensionalidad seleccionando un subconjunto de variables con capacidad predictiva. Este trabajo tiene por objetivo ilustrar el uso de R para la descarga y selección de variables para construcción y validación de modelos logísticos para la predicción de probabilidad de presencia de enfermedad en un patosistema. En una primera instancia a través del paquete *rgee* pudimos conectarnos a la plataforma de google earth engine para descargar las variables climáticas de interés desde ERA 5 Land, que es la quinta generación de reanálisis atmosférico del Centro Europeo de Previsiones Meteorológicas a Medio Plazo Medio (ECMWM), que proporciona estimaciones por hora de una gran cantidad de variables climáticas atmosféricas, terrestres y oceánicas. En una segunda instancia, nos enfrentamos a el desafío de seleccionar las variables más relevantes para construir y validar modelos logísticos. Para ello, utilizamos diferentes métodos de selección de variables disponibles en varios paquetes de R, la versatilidad del lenguaje nos permito generar una función en la cual podemos comparar métodos como *Boruta*, *LASSO*, *Algoritmo genético* y *Stepwise* simultáneamente en términos de eficiencia y eficacia. Esta función abarca las etapas cuatro etapas de la selección de variables: generación de subconjuntos de variables predictoras, evaluación de los subconjuntos, detención del proceso y validación de resultados logrados con los conjuntos de predictoras. Esta función incorpora las variables seleccionadas por cada método a un modelo de regresión logística. A partir de los modelos calcula las métricas como precisión, sensibilidad, especificidad y área bajo la curva que se obtiene por cada método. En resumen, La comparación de métodos de selección de variables nos ha permitido identificar las variables más relevantes y obtener modelos con un alto rendimiento predictivo. Estos resultados nos brindan información valiosa para la toma de decisiones en el manejo de cultivos y la mitigación de enfermedades, contribuyendo así a la optimización de la producción agrícola y la reducción de pérdidas. R es una poderosa herramienta que nos permite acceder a la descarga, selección y construcción de modelos logísticos que poseen la capacidad predictiva para la presencia de enfermedades.