Propuesta de Trabajo Fin de Máster
Año académico 2026-2027
MÁSTER EN CIENCIA DE DATOS PARA CIENCIAS EXPERIMENTALES

Proyecto Nº 8
Título: Artificial Intelligence (AI) to define accionable genetic risk predictors
Departamento/ Laboratorio: Biología Computacional
Director: Carlos Ruiz Arenas Correo electrónico: cruizarenas@unav.es Codirector: Mikel Hernáez Correo electrónico: mhernaez@unav.es
Resumen: Although an individual's genetic variants affect their predisposition to different common diseases, their clinical utility for prevention and treatment is currently very limited. The main reason is that Polygenic Risk Scores (PRS), the best predictors of an individual's genetic risk, are constrained to a reference population (limiting their application) and provide a global measure of risk without specifying which regulatory pathways may be affected and in which tissues to propose specific treatment. The shortcomings of PRS can be addressed by predicting the effect of variants on gene expression and combining them into Gene Set Activity Scores (GSAS). GSAS provide a measure of how the gene expression of different molecular processes is affected in various tissues, offering a much more specific and actionable measure of genetic risk than current PRS. In this project, we propose to develop tissue-specific genetic risk predictors based on GSAS and molecular functions. To achieve this, the predicted expression by different variants of an individual will be combined into GSAS using NetActivity, a tool developed by our group. In this way, GSAS will be obtained for each tissue and molecular pathway. These GSAS will be correlated with the risk of breast cancer in public datasets, in addition to comparing their prediction with established PRS. Thus, specific markers of risk for molecular pathways and tissues will be generated. These new markers may enhance the genetic prediction of breast cancer and provide actionable genetic markers for common variants.

OPTATIVAS RECOMENDADAS
1.   Advance topics in machine learning 2.   Deep learning 3.   Análisis e interpretación de datos de alto rendimiento 4.