

# Resum TFG

Francesc Llabrés Massanet

En els models predictius aplicats a dades biomèdiques la incertesa pot tenir conseqüències crítiques. L'objectiu principal d'aquest treball serà el d'analitzar i caracteritzar aquesta incertesa en els nostres models, centrant-nos en la predicció de l'eficàcia d'un tractament per a l'hepatocarcinoma.

Proposam un mètode de quantificació de la incertesa basat en els conjunts creïbles. Analitzarem aquest mètode amb una base de dades senzilla i acadèmica, el *Wisconsin Breast Cancer Dataset*. A continuació, treballarem sobre una base de dades de l'Hospital Universitari de Son Espases, de 144 pacients i 118 variables, que presentarà una major complexitat.

Per a fer prediccions, s'han implementat tècniques com el *bootstrap*, LASSO, LOOCV i SMOTE, i mètodes de classificació com regressió logística i *random forest*. A més, s'ha desenvolupat un mètode d'imputació de valors buits basat en la distància de Gower i els *K-nearest neighbors* i un mètode per a extreure informació sobre quines variables ens aporten més informació mitjançant *bootstrap* i LASSO.

Com a contribucions, el treball ofereix un mètode efectiu per a quantificar la incertesa, una estratègia d'imputació de dades robusta i informació sobre les variables més influents en la resposta al tractament. Hem pogut veure, per exemple, que aspectes en principi irrelevants com la localització tumoral afecten a l'efectivitat del tractament. A més, malgrat que el nostre model final obtingui una alta incertesa degut a la complexitat del problema, hem aconseguit que la capacitat predictiva de l'estudi sigui millor que la que teníem fins al moment, amb una presició del 66%.