

Machine Learning 0 - Intro

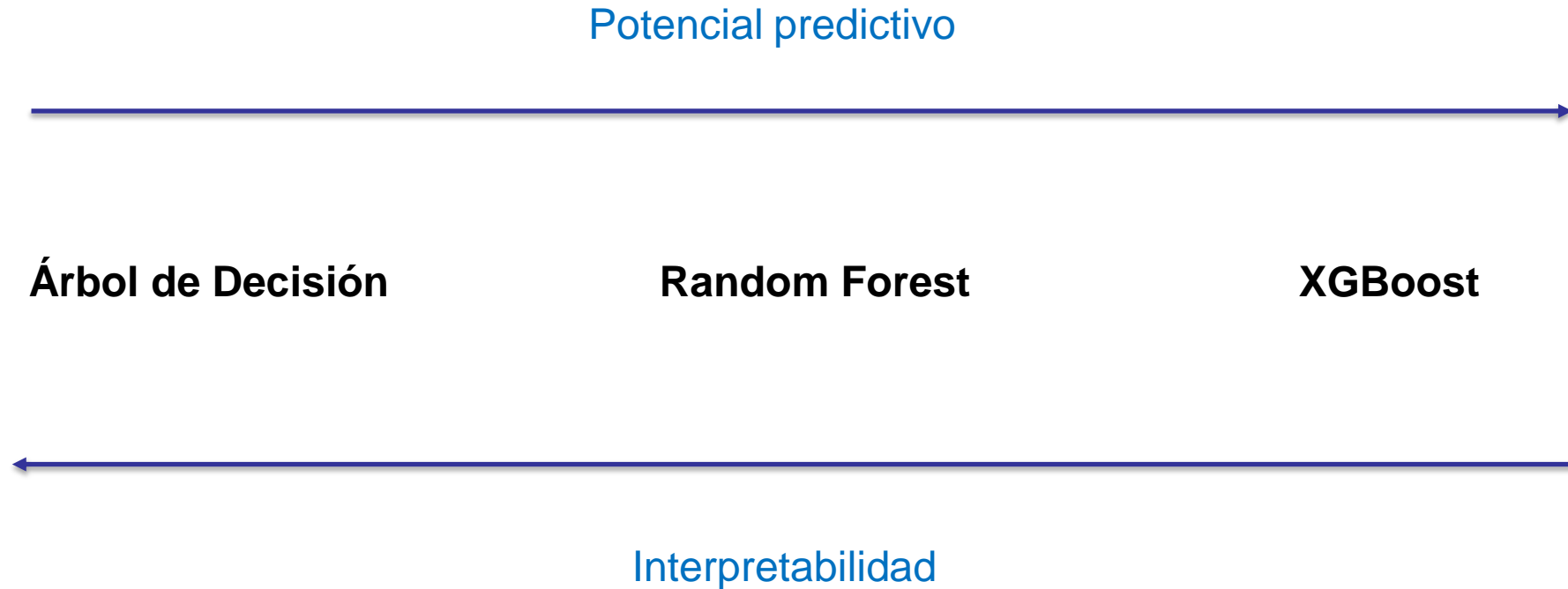
↳ orus.ml

Jesús Prada Alonso - HORUS ML

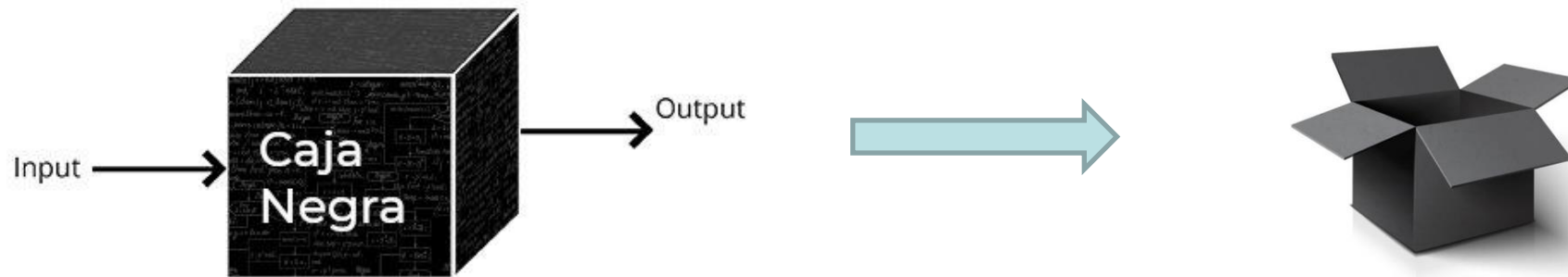
INTERPRETABILIDAD



INTERPRETABILIDAD VS PRECISION (I)



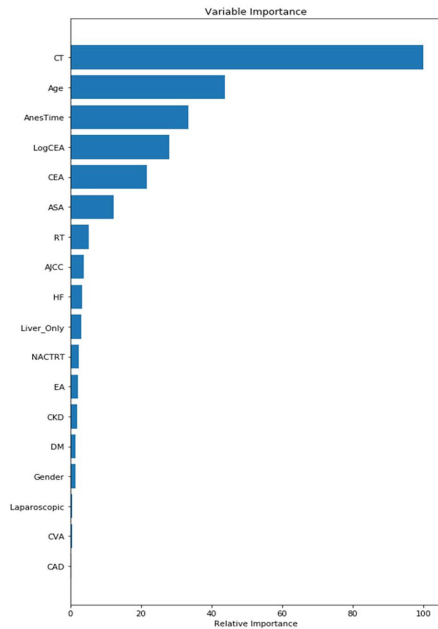
- En aprendizaje automático habitualmente los **modelos más precisos son los más complejos**.
- Por esta razón, los modelos que dan mejores resultados predictivos son normalmente los **menos interpretables**.
- Hasta ahora, había que tomar una decisión entre un **modelo interpretable** o un **modelo que funcionara bien** a nivel predictivo.
- Sin embargo, este **enfoque ha empezado a cambiar** recientemente.



- Algunos factores que han influido en el cambio de enfoque son:
 - Nuevas leyes de protección de datos, **GDPR**.
 - Uso en **sanidad**.
- La extracción de interpretabilidad es una línea de investigación popular en estos momentos, pero aún es un problema sin resolver por completo.
- Sin embargo, ya se han desarrollado algunos algoritmos:
 - **Arboles de decisión** para explicar las decisiones del modelo (atributo *feature_importances_*).
 - Predicción basada en **ensembles de reglas**.
 - **Métodos estadísticos** basados en insertar ruido a los inputs y analizar variaciones en el output.

INTERPRETABILIDAD VS PRECISION (IV)

Interpretabilidad Global

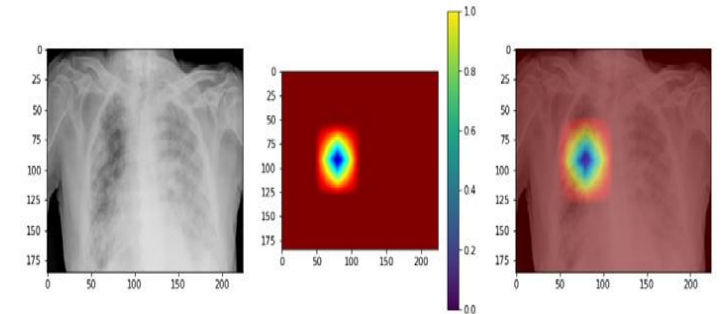


Interpretabilidad Predicción

- ID Paciente: 0000001
- Predicción de probabilidad de Exitus: 0.9997221

1. des_d1: Infarto agudo de miocardio con elevación de ST
2. des_grd: OXIGENACIÓN POR MEMBRANA EXTRACORPÓREA
3. des_d2: Insuficiencia cardíaca, no especificada
4. flag_uci: 1
5. des_p1: Derivación de mamaria interna
6. edad: 78
7. des_grd_ent: OXIGENACIÓN POR MEMBRANA EXTRACORPÓREA
8. des_d7: Bloqueo auriculoventricular, completo
9. numproc: 15
10. previous_diag: 14

Interpretabilidad Imagen



Jesús Prada Alonso
jesus.prada@horusml.com