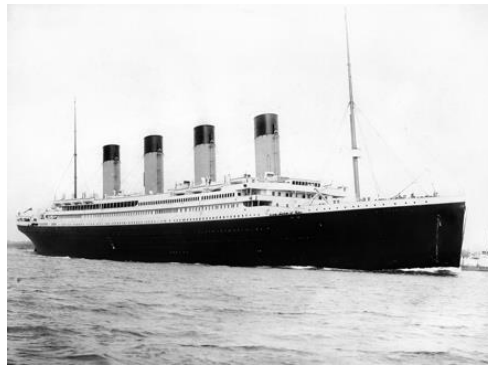


EXERCICI DE LLIURAMENT OBLIGATORI

Objectiu: aplicar diferents classificadors (regressió logística, perceptró i arbres de decisió) a un problema real. Analitzar les característiques del problema, estudiar i ajustar els paràmetres dels classificadors.

Materials: el problema es realitzarà amb les llibreries emprades a classes de pràctica i especialment amb Python Notebooks. L'entrega serà el *notebook* i una versió convertida en pdf.



Problema

L'enfonsament del Titànic és un dels naufragis més importants de la història. El 15 d'abril de 1912, durant el seu viatge inaugural, el considerat "insumergible" RMS Titànic es va enfonsar després de xocar amb un iceberg. Desafortunadament, no hi havia suficients bots salvavides per a tots a bord, la qual cosa va resultar en la mort de 1502 de 2224 passatgers i tripulants.

Si bé va haver-hi algun element de sort involucrat en la supervivència, sembla que alguns grups de persones tenien més probabilitats de sobreviure que uns altres. En aquest problema, utilitzant dades reals dels passatgers del Titànic, heu de realitzar un model predictiu que respongui a la pregunta: "quin tipus de persones tenien més probabilitats de sobreviure?"

La pràctica té dues parts diferents.

1. La primera part consisteix a entrenar un conjunt de predictors i comparar els seu rendiment. Aquests predictors han de ser la regressió logística, perceptró i arbres de decisió. S'ha de dur a terme el procés complet que hem vist a classe, és a dir, s'han de fer: anàlisi de dades, entrenament i extracció de resultats. Totes les decisions preses i resultats s'han d'explicar detalladament. Teniu en compte que les dades són "brutes".

AJUDA: Aquest problema és un problema típic dels concursos de ML[1]. Trobareu el més conegut a Kaggle i una de les millors solucions a [2].

2. La segona part consisteix en per una part obtenir i analitzar la importància de cada característica per cada un dels models entrenats. Aquesta tasca és clàssica al camp de l'aprenentatge automàtic i rep el nom de *feature importance*. Per l'altra seleccionar els millors paràmetres de cada model per l'entrenament de forma exhaustiva.



Entrega: La pràctica s'ha de realitzar en parella. Hi haurà dues dates d'entrega diferent. El primer dia 13 de gener permeten obtenir una nota fins a 10. La segona data d'entrega serà dia 7 de febrer amb una nota màxima de 7. No s'admetran entregues posteriors. En ambdós casos, els dies posteriors de l'entrega tots els membres del grup hauran de fer una entrevista presencial amb el professor. **Si qualche grup no fa aquesta entrevista la pràctica quedarà suspesa.**

Referències:

[1] <https://www.kaggle.com/c/titanic/overview>

[2] <https://www.kaggle.com/vinothan/titanic-model-with-90-accuracy>