**INFORME ENTREGA INTERMEDIA**

* Enviamos finalmente para la entrega intermedia 2 ficheros, resultados del modelo que ya habíamos realizado XGB + RandomForest con ligeras diferencias en las muestras para el entrenamiento respecto al anterior:
  + Entrega 1 → 6% test y 5250 muestras residential aleatorias. 0.6969
  + Entrega 2 → 20% test y 6000 muestras residential aleatorias. 0.6938
* Ello nos coloca en 3er puesto nacional por el momento. Hemos empeorado nuestros resultados respecto a la entrega de la Fase I (0.6999).
* A continuación comentamos nuestras ideas y planificación cara a la entrega final:
  + Mejora del modelo (no creemos que otro modelo vaya a mejorar los resultados y no vemos factible realizar uno nuevo a estas alturas):
    - Feature engineering y reducción de dimensionalidad:
      * Coordenadas X, Y: Hemos realizado un ‘mapa’ a partir de los K-vecinos de cada muestra, según la clase de estos, o la probabilidad de que sean de esa clase después. Por ahora no parece mejorar nada.
      * Colores: Hemos pensado en:
        + Reducir los 11 valores de cada color en 1 para cada, al representar la densidad de color por décil en una sola.
        + O también, en calcular la densidad total de cada color de RGB (suma de todas las variables) e introducir 3 columnas, una por color, con el valor de posición de 0 a 2 (siendo 0 la menor densidad total de los 3 colores para esa muestra y 2 la mayor).
        + Canal NIR: Variables condicionales?: Por ejemplo, si la densidad final es mayor que un valor x, ponemos un 1, si no un 0.
      * Geometría y área: sabemos que el valor medio del área de la clase Residential es mucho menor que la del resto, pero no sabemos cómo sacarle partido a esto exactamente. Además, las 4 variables geométricas desconocemos qué representan. Hemos pensado que podrían ser 4 lados del terreno, o uno de ellos el perímetro…
      * Media de las variables geométricas en su sustitución: no mejora.
      * Eliminación de deciles 2,3,4,6,7,8: no mejora.
    - Presentación y estructura de la entrega final (muy importante, especialmente al estar todo tan ajustado):
      * Informe al estilo paper que conecte el trabajo de forma general.
      * Jupyter notebooks (aún no queda claro cuántas ni cuáles).
      * Archivos py bien organizados en una carpeta.
      * Carpetas con las visualizaciones y datos que hemos extraido para apoyar nuestro trabajo.
      * Código no empleado finalmente, también, de forma separada.

Muchas gracias!