

En general los resultados obtenidos por ranger son inferiores a los obtenidos por Árboles Azarosos. Al ejecutar el punto 3 usando mis semillas, obtuve un score ínfimo en Kaggle de 10569. Para mi sorpresa, luego de haber procesado el script z433 esperando cerca de 8 horas, al ordenar por ganancia y subirlo a Kaggle, el resultado fue espantoso: -67667. Leyendo en Zulip pude observar que a otros compañeros le sucedió lo mismo. La causa mencionada es: "Podría estar sucediendo que por algún motivo el modelo, o el predict() NO estén generando probabilidades bien calibradas, es decir que este ordenando correctamente a los registros pero NO les esté asignando el correcto valor de probabilidad."

Se mencionó dos formas de sortear esto:

- Cortar por cantidad de envíos: pude salir de los negativos, pero no fue suficiente, el mayor score que obtuve en kaggle fue de 22459.
- Utilizar Isotonic Regression: estuve tratando de implementar esto, leyendo diferentes fuentes y no he tenido éxito en la implementación, en algo estoy fallando. @profesor tendrías el código de esto?

A continuación muestro los mejores hiperparámetros para Random Forest y para Árboles Azarosos.

	Random Forest (Ranger)			Arboles azarosos
num.trees	279		arbolitos	500
min_node_size	83		min_bucket	10
max_depth	11		max_depth	10
mtry	50		minsplit	50
kaggle	-67667		kaggle	56417

Nota: dejé corriendo un nuevo script con mayor cantidad de iteraciones (200) y 20 semillas en lugar de 5. Cuando tenga los resultados los compartiré por acá.