UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERIA

ESCUELA DE ESTUDIOS DE POSTGRADO

MAESTRIA PARA LA INDUSTRIA

CON ESPECIALIZACION EN CIENCIAS DE LA COMPUTACION

CURSO: Minería de Datos.

Por: Denys Fernando Orozco.

EJERCICIO ACERCA DE RANDOM FOREST – Tarea bota nota.

Estimado Ingeniero:

Por favor, sustituir la presente actividad por la tarea #4 acerca de K-Means (clusterización)

Enlace al repositorio a github con actividad completada:

REPOSITORIO PRINCIPAL:

<https://github.com/denysorozco/MineriaDeDatosFIUSAC.git>

REPOSITORIO DE RANDOM FOREST:

<https://github.com/denysorozco/MineriaDeDatosFIUSAC/tree/Random_Forest>

El algoritmo Random Forest es un algoritmo que permite hacer clasificaciones y/o regresiones, a partir de varios árboles de decisión, entrenados por medio de datos aleatorios, los que aportan al resultado final por medio de votación mayoritaria (clasificación) o por un promedio (regresión).

Este tipo de algoritmo permite hacer predicciones basadas sobre la base de un conjunto de datos aleatorios, de los que se tomó un subconjunto para su entrenamiento y otro subconjunto para la predicción. El algoritmo introduce cierta diversidad en su entrenamiento y es bastante robusto para grandes cantidades de información.

Este algoritmo destaca por su capacidad para evitar los sobreajustes, de tener cierta inmunidad a los valores atípicos con características no lineales.

Es un algoritmo que tiene la característica de poder predecir o asignar una calificación a los datos, basándose en argumentos bastante sencillos de ajustar, siempre que se comprenda la naturaleza de la pregunta a completar.

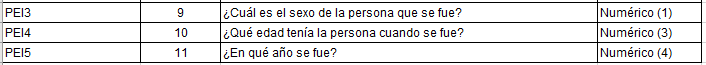
La predicción llevada a cabo, toma información acerca de la “Migración en Guatemala”. La fuente de toma es del instituto nacional de estadística, acerca del archivo nacional de Migración. Esta, como un fenómeno complejo social, se aborda desde diversas perspectivas.

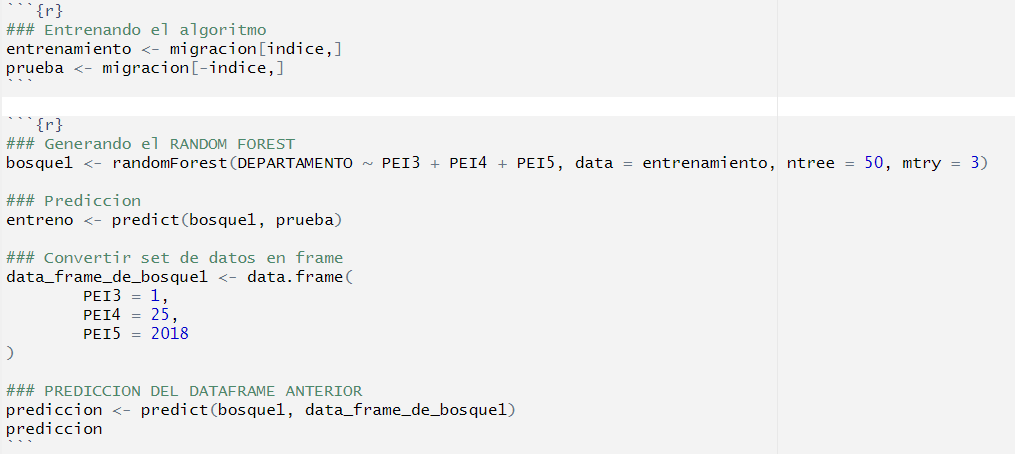
La población guatemalteca, en general de clase baja o media; se pueden ver expuestos a situaciones desfavorables, lo que aumenta la migración.

En el presente archivo, se toma una muestra de los datos del archivo de migración y se procede a su análisis por medio de Random Forest. Se pretende averiguar de forma muy generalizada, de que departamento del país de Guatemala tienen su origen los migrantes que dejan el país; dados los argumentos que indican la edad y el año.

El archivo del INE, tiene ciertas consideraciones que se toman en cuenta para la migración, sin considerar detalles como la escolaridad, si una persona labora en el sector formal o informal, los ingresos promedios mensuales; algunos detalles que se pueden considerar importantes pero que los gobiernos de país omiten, que podrían indicar claramente que las causas en muchas instancias son humanitarias, relacionadas a la calidad de vida que se pueda a través de las políticas ofrecer a la población.

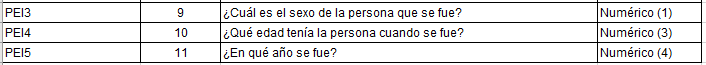
Para el primer análisis, se tomó en cuenta la pregunta que resuelve el tema de que departamento proviene un migrante, dado el género, año de migración y su edad:

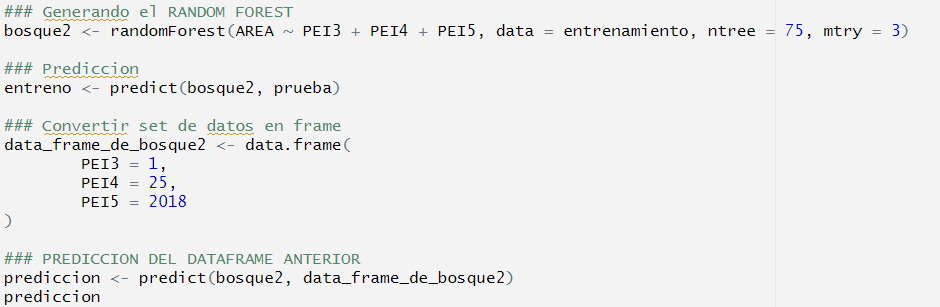




Conclusión: De acuerdo al resultado, se hace la predicción de que el migrante proviene del departamento de Huehuetenango; dadas las características ingresadas (hombre, de 25 años de edad; en el año 2018)

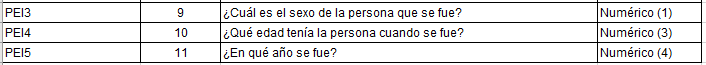
Para el segundo análisis, se tomó en cuenta la pregunta que resuelve el tema de si el migrante es una persona del área urbana o rural, dado el género, año de migración y su edad:

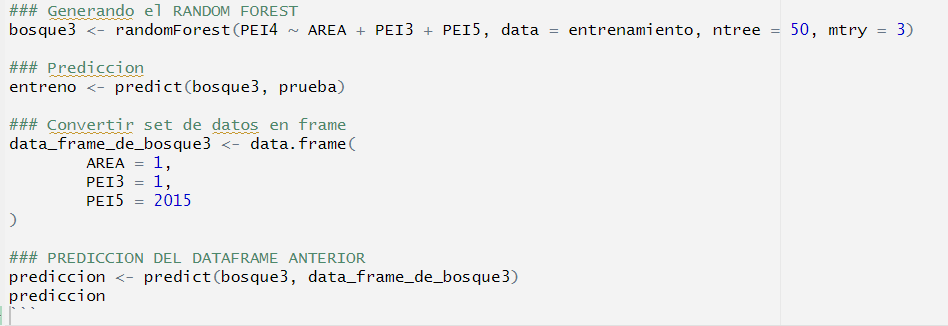




Conclusión: De acuerdo al resultado, se hace la predicción de que el migrante proviene del área urbana; dadas las características ingresadas (hombre, de 25 años de edad; en el año 2018)

Para el tercer análisis, set tomó en cuenta la pregunta que resuelve el tema de que edad tiene el migrante, dado su género, el año de migración y su area (rural o urbana)





Conclusión: De acuerdo al resultado, se hace la predicción de que un caballero migrante del área rural, es dos años más en joven en promedio (23) frente a uno de área urbana, quien se estima tenga un promedio de edad mayor (25).