# Modelo Clasificatorio: Intubación COVID-19 en México

# Cargamos los datos de SINAVE al 1ro de julio junto con librerias

# Interfaz de usuario gráfica, Texto, Aplicación Descripción generada automáticamente

# A la base original le modificamos algunas columnas de interés, asegurándonos que solo contengan valores binarios así como la variable objetivo dejarlo en “yes y no”. Después nos deshacemos de las columnas no necesarias y convertimos las columnas de string a factor.

# Interfaz de usuario gráfica, Texto, Aplicación Descripción generada automáticamente

# Creamos una muestra Aleatoria con 10,000 datos y eliminamos la variable de municipio, por que son muchísimas combinaciones no nos puede dar datos significativos para el procesamiento. Nos aseguramos que la variable categórica de entidad sea factor y creamos el modelo bootsrtap

# Interfaz de usuario gráfica, Texto, Aplicación Descripción generada automáticamente

# Preparamos los datos, especificando la variable objetivo, la fuente de datos y super importante el tratamiento a la variable categórica, la cual convertimos a dummies. Tamboen se normaliza y se aplica un filtro near zero

# Imagen que contiene Interfaz de usuario gráfica Descripción generada automáticamente

# Nuestros datos se visualizan así:

# Tabla Descripción generada automáticamenteTabla Descripción generada automáticamente

# Ahora, se creara un Random Forest, para encontrar los parámetros adecuados

# Interfaz de usuario gráfica, Texto, Aplicación Descripción generada automáticamente

# Obtenemos la siguiente grafica donde observamos que en el parámetro mtry entre mas bajo mejor desempeño, sin embargo, en el min\_n no hay una tendencia como tal.

# Gráfico, Gráfico de dispersión Descripción generada automáticamente

# Por lo que procedermos a ahacer un fine\_tuning Imagen que contiene Interfaz de usuario gráfica Descripción generada automáticamente

# Con esto obtenemos el mejor modelo seleccionado de los siguientes;

# Gráfico, Gráfico de líneas Descripción generada automáticamente

# Sin embargo, debido a su aleatoridad en la selección de las muestras y al realizar el modelo, este reporte se hizo con los resultados de un reandom forest con mtry=5 y min\_m=10

# Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico Descripción generada automáticamente

# Este modelo nos da las siguientes medidas, accuracy de 96.26%, PPV de 98.60% y NPV de 28.58% basado en:

# Patrón de fondo Descripción generada automáticamente con confianza bajaTexto Descripción generada automáticamente

# Las variables mas importantes son:

# Interfaz de usuario gráfica, Texto, Aplicación Descripción generada automáticamente

# Gráfico, Gráfico de embudo Descripción generada automáticamente

# Conclusión:

# El modelo creado es bueno en general, con un accuracy mayor al 95%, sin embargo este resultado se puede ver sesgado, ya que la mayoría de los pacientes no fueron intubados, y el modelo es bastante bueno clasificando los no intubados, con un accuracy de mas de 98%. El problema es cuando se busca clasificar a los que si fueron intubados, aquí el accuracy es un poco menor al 30%. En la práctica, este modelo es inservible, ya que se usa la variable de defuncion como entrada, lo cual no tiene lógica, no tiene una aplicación útil, es meramente demostrativo.