Diseño de un modelo para la predicción del ausentismo a citas médicas.



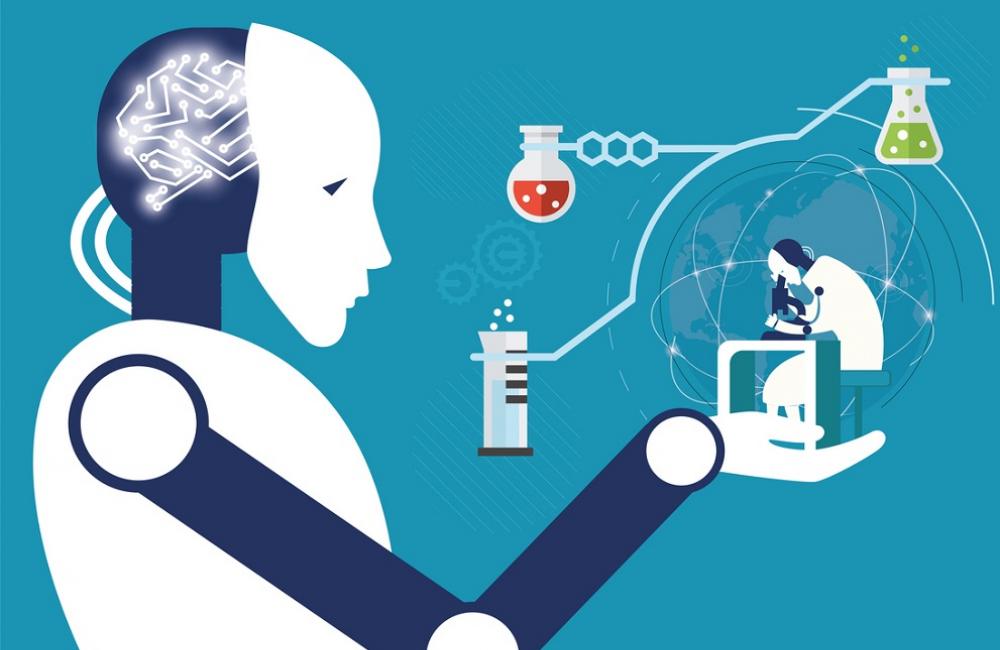
## **Introducción**

Cuando un paciente no se presenta a su cita médica sin previo aviso, genera una serie de contratiempos a las organizaciones sanitarias.

Se sabe que entre el 10 y 30 % de las personas que programan turnos médicos externos no se presentarán a la cita. Las investigaciones sobre este aspecto problemático de la atención sanitaria estiman que, por ejemplo, un 25 % de ausentismo:

* Aumenta notablemente los tiempos de espera para la asignación de nuevos turnos
* Disminuye las oportunidades para el diagnóstico y tratamiento en tiempo y forma oportuna
* Reduce la eficiencia de los sistemas de salud.
* Supone una infrautilización de los recursos humanos y materiales

La secretaría de Salud de Bogotá ha alertado que en el año 2017 los **pacientes incumplieron más de 569.044 citas médicas** en red pública lo cual ha generado pérdidas por más de 7,5 millones de dólares.



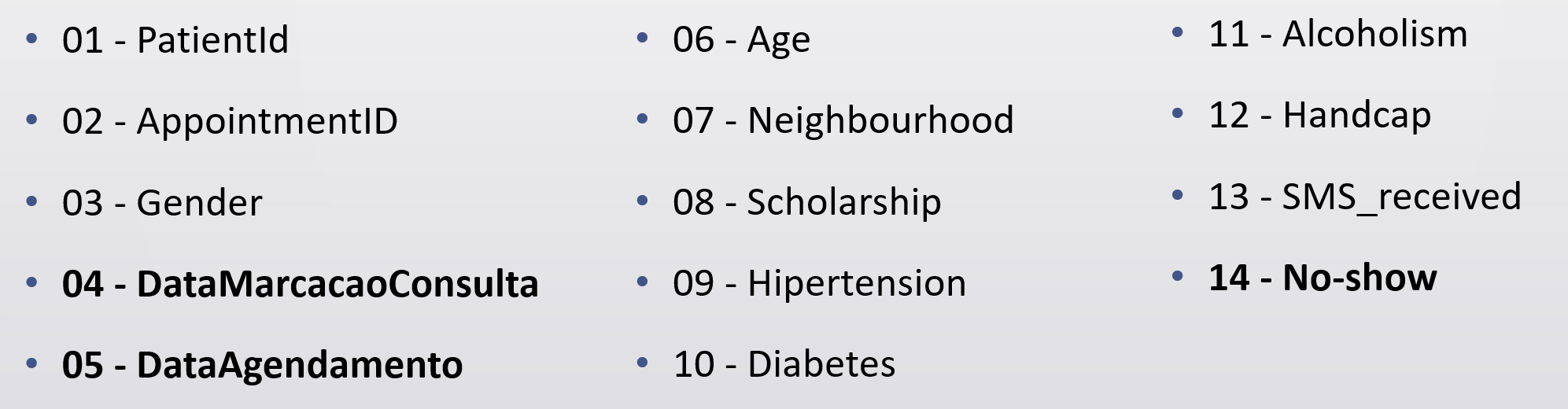
## **Objetivos**

La inteligencia artificial tiene numerosas aplicaciones en sanidad y farmacia, que consiste Procesar cantidades ingentes de información clínica en poco tiempo.

Sin embargo, muchos de estos datos no son utilizados. Por lo que mediante este modelo buscamos presentar las posibilidades de la aplicación de AI y sus ventajas.

Diseñando un modelo para la predicción de posibles casos de ausentismo a las citas médicas.

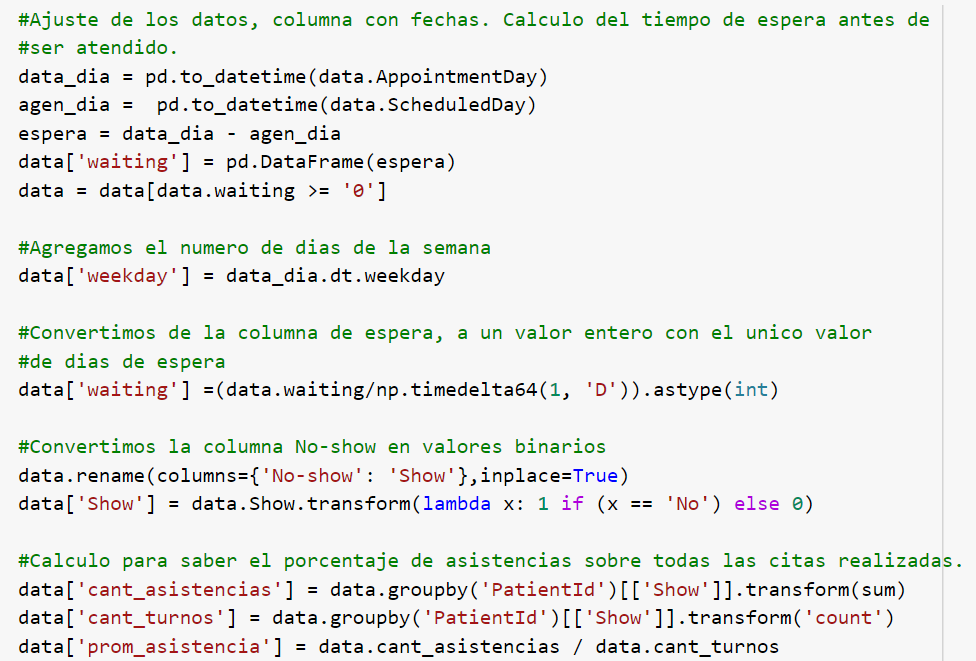
## **Dataset**

Los datos utilizados en el proyecto provienen de un estudio subido a Kaggle En el cual se recolecto información de **110.527** citas médicas de Brasil el año 2016, a las cuales se asocian 14 variables. 

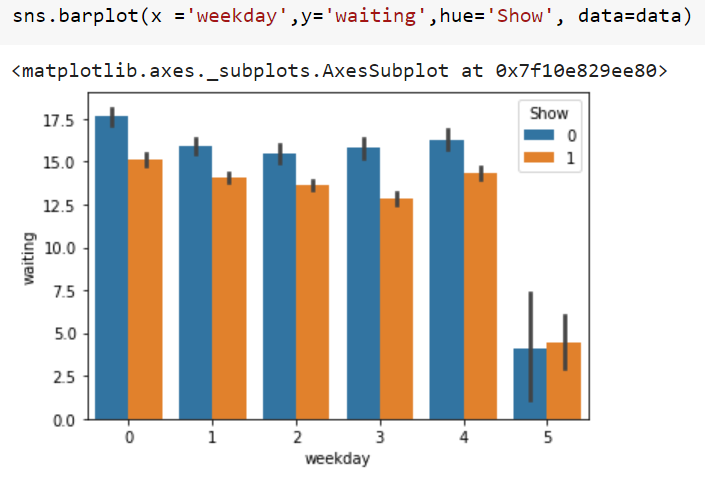
## **Procesamiento de los datos**

El primer paso fue ajustar las fechas desde que se realizó la cita médica hasta la presentación de la misma y el tiempo de espera antes de ser atendidos y si faltaron a la cita médica o asistieron a la misma.

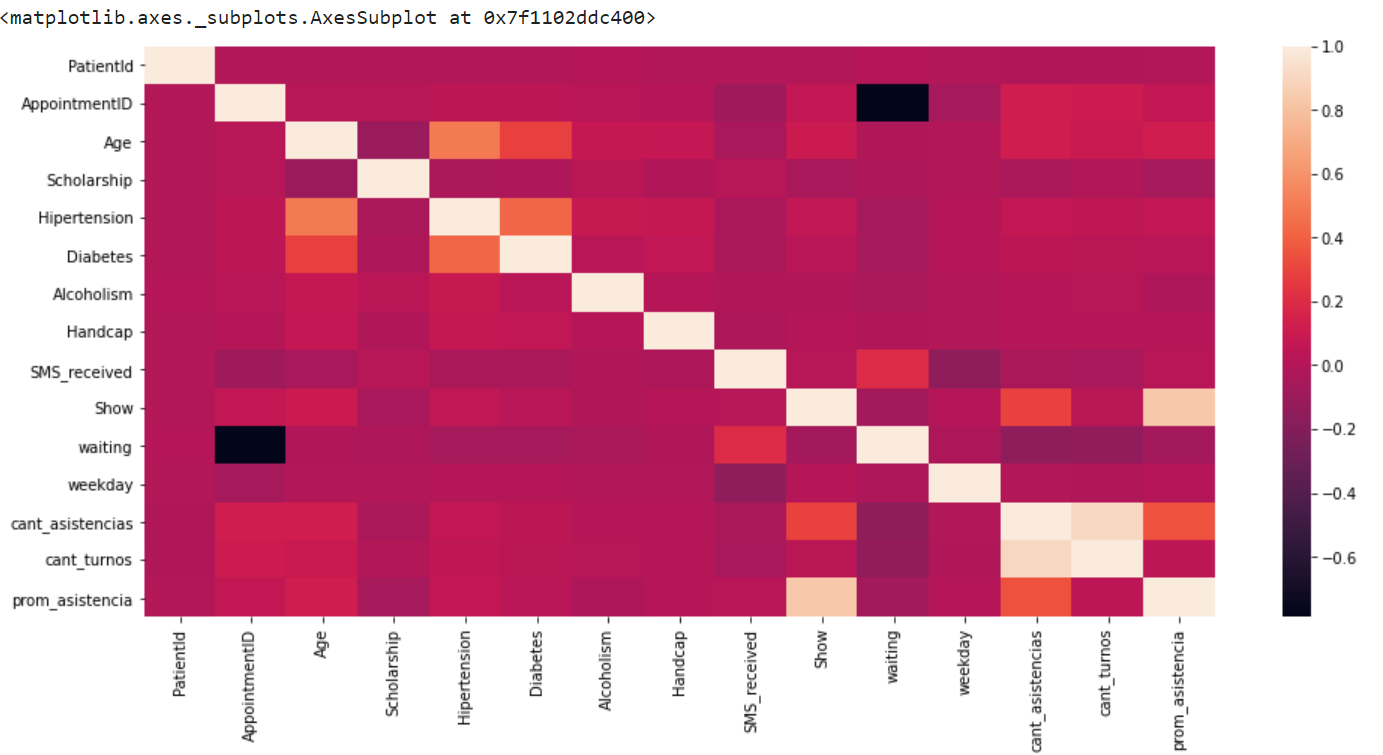
Además de crear dos columnas, día que se realizó la cita médica y tiempo de espera, que nos permitirían tener un mejor vistazo a todo el conjunto de datos obtenidos



Con los datos preparados para ser analizados, el siguiente paso fue elaborar graficas que nos permitan entender mejor el comportamiento de los datos y que estos se encuentren bien distribuidos en las diferentes columnas, para ello se elaboraron las siguientes tablas.



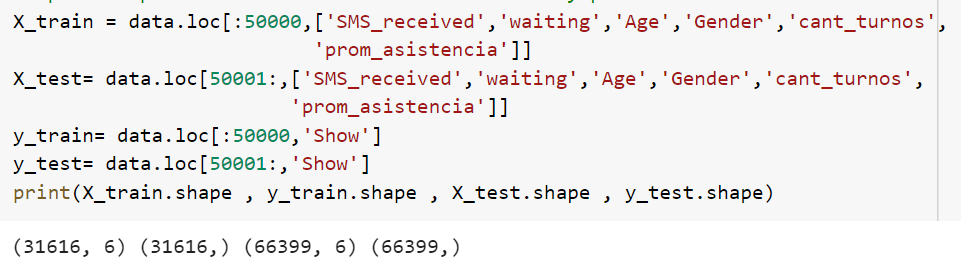
Esta tabla separa los datos por día de la semana (Citas médicas de lunes a viernes y algunos casos especiales de sábado) y al mismo tiempo muestra si el paciente se presento a su cita medica en base al tiempo que estuvo esperando para ser atendido. Por los datos se encuentran bien distribuidos en base al día de la semana y los tiempos de espera son muy similares en estos.



Un mapa de calor para poder identificar si existe alguna correlación entre las diferentes variables, en la cual se puede ver que la edad esta muy relacionada con la Hipertensión y la Diabetes. Sin embargo, el ausentismo no se encuentra fuertemente relacionado a todas estas variables.

## **Elaboración de los modelos**

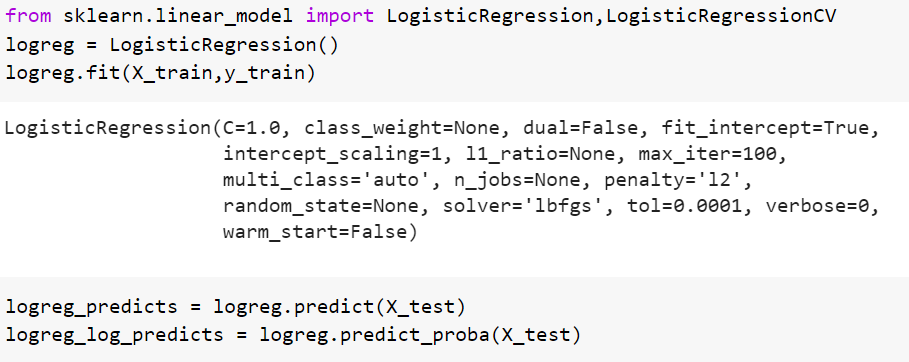
Después de tener el Dataset preparado, realizamos la distribución de los datos en un conjunto de entrenamiento y otro de prueba utilizando CrossValidation para evitar las dependencias de los datos.

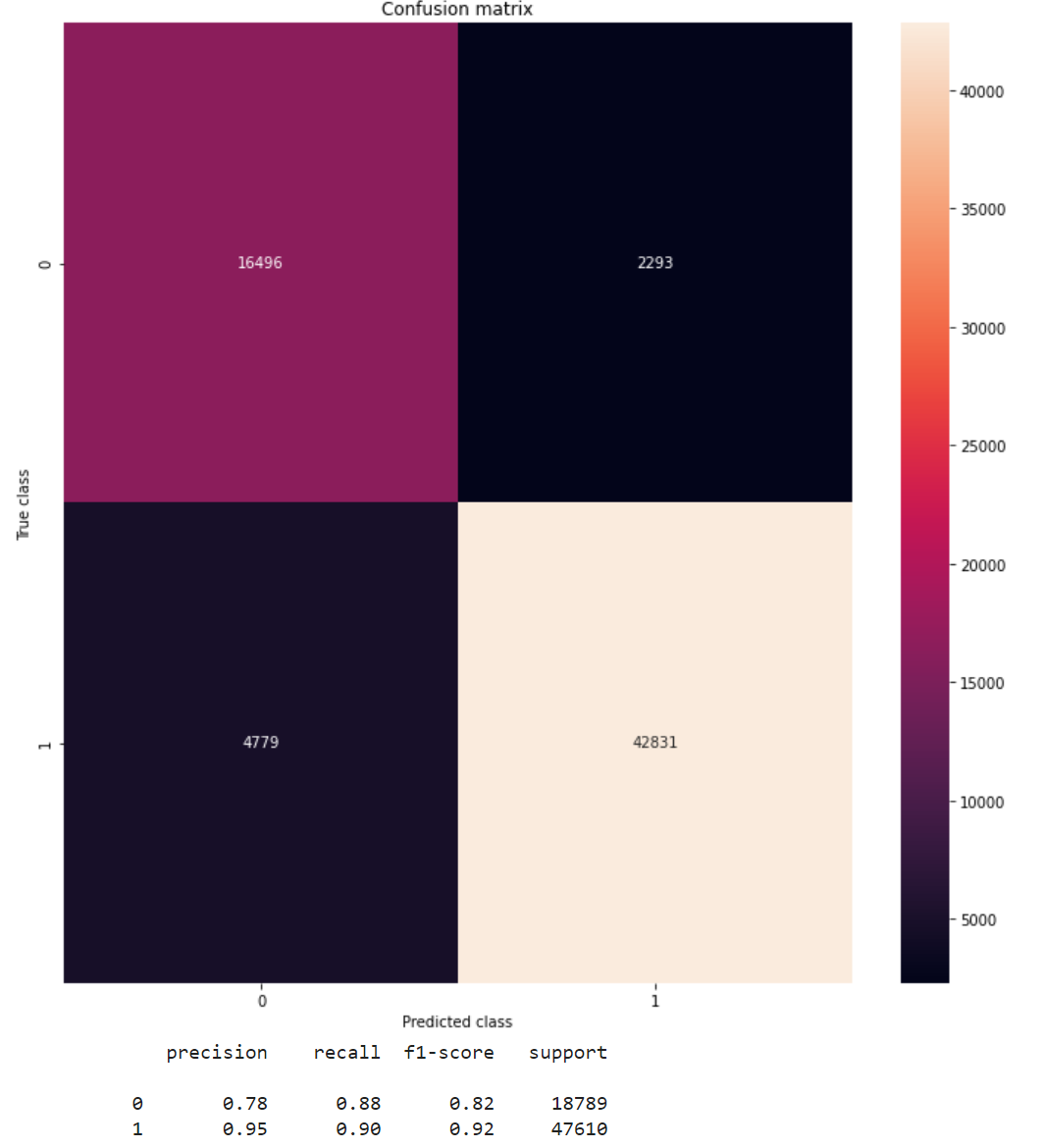


Debido a que tenemos un problema de clasificación realizamos tres modelos diferentes que se ajusten a nuestras necesidades, para poder comparar los resultados finales y tener un mejor modelo.

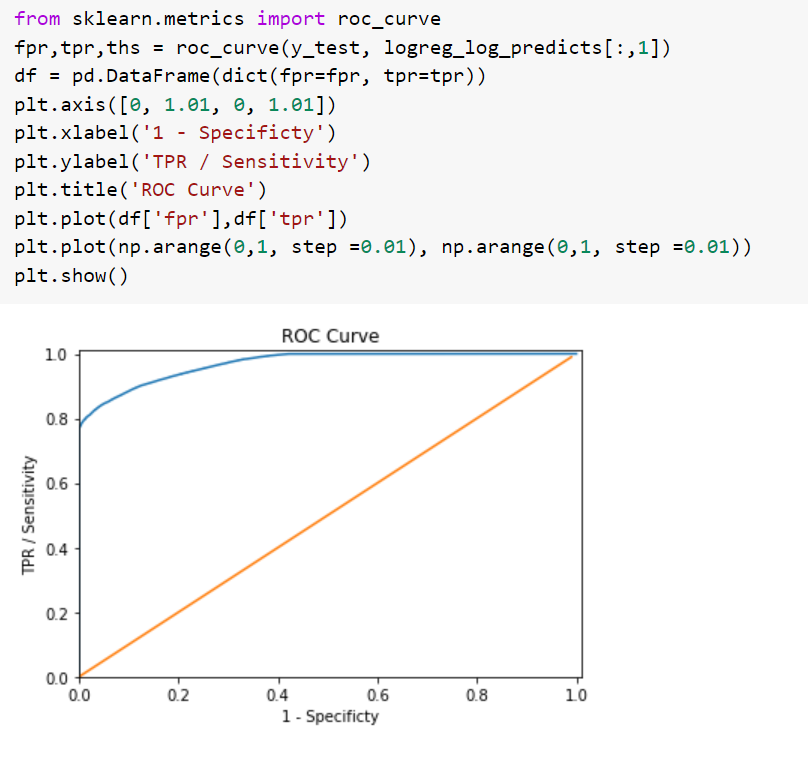
## **Regresión logística**

El primer modelo que utilizamos fue la Regresión Logística para el cual aplicamos el siguiente código:





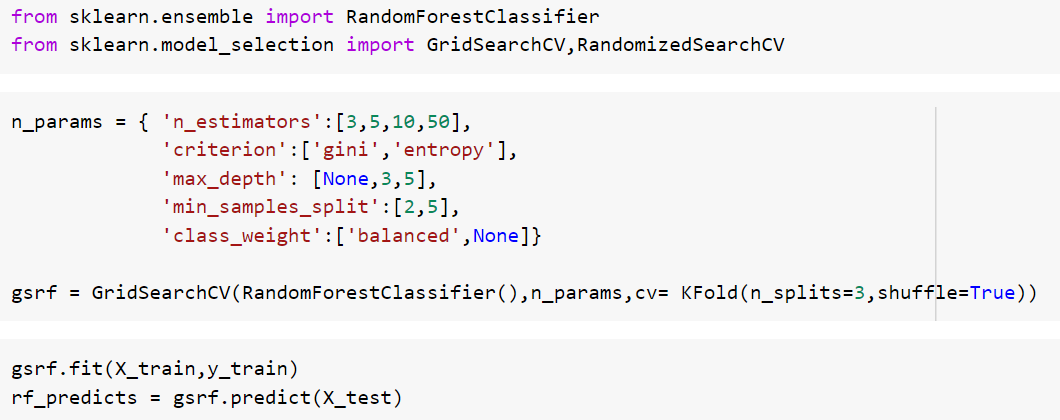
Este modelo dio como resultado una precisión del 89% en la predicción de ausentismo utilizando el conjunto de prueba. Sin embargo, para estar mas seguros de los resultados también se elaboro una Curva de ROC :



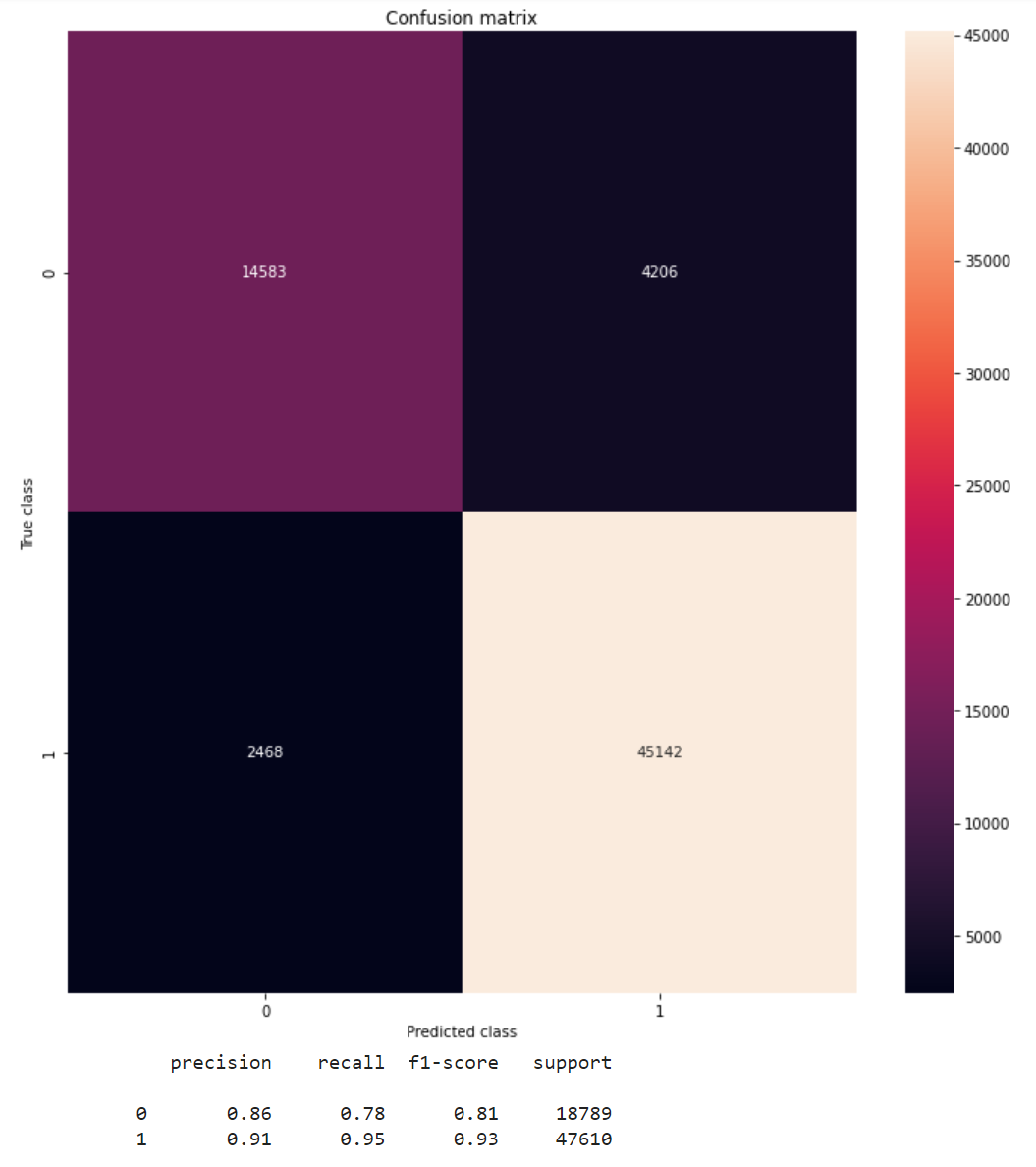
La Curva de ROC nos permite evaluar la predicción de un test, entre mas cerca se encuentra del 1 en TPR/Sensibilidad mejor es el resultado.

## **Random Forest**

El segundo modelo utilizado fue el Random Forest para el cual se aplicó el siguiente código:

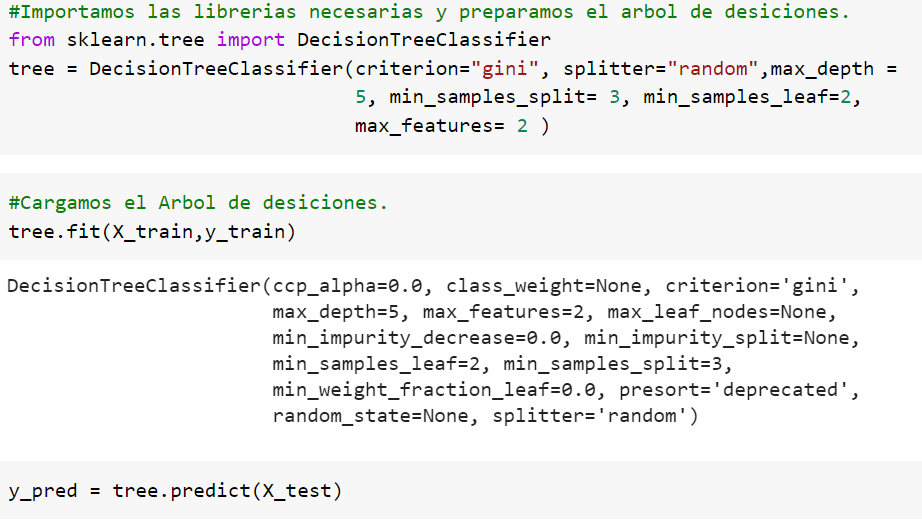


Este modelo tiene una precisión muy parecida a la Regresión logística con un valor del 90%, dando la siguiente matriz de confusión como resultado:

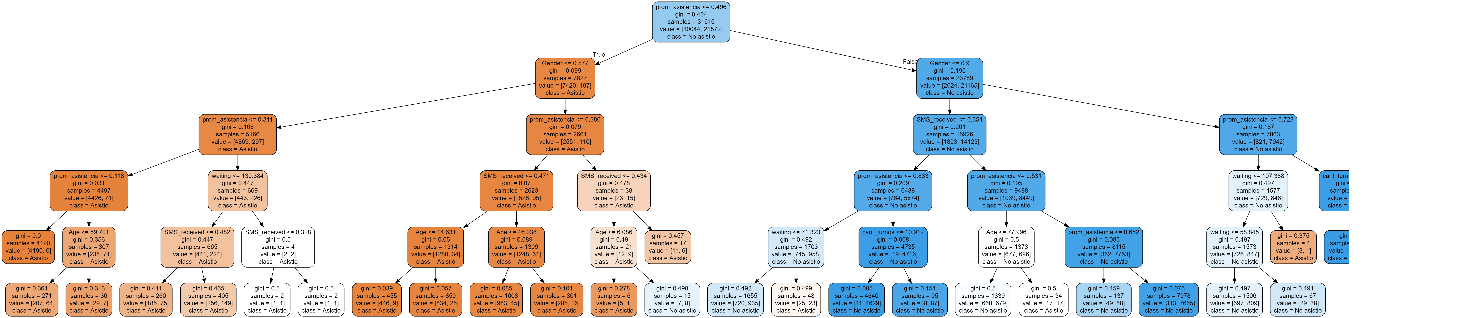


## **Desicion Tree**

Finalmente, el ultimo modelo aplicado fue el Árbol de decisiones aplicando el criterio “Gini” con una profundidad de 5 capas, como se puede ver en el siguiente código:



Este modelo es muy parecido a los anteriores en cuanto a resultado, ya que tiene una precisión del 89% dando como resultado el siguiente Árbol de Decisiones



## **Conclusiones y Recomendaciones**

Los modelos realizados llegan a tener un 90% de precisión. Demostrando que es posible realizar una predicción de la asistencia de un paciente a su cita médica.

Debido a que los datos son del 2016 y están ubicados solamente en Brasil, este trabajo puede utilizarse como una base a futuro para el desarrollo de diferentes investigaciones o proyectos más específicos en el sector de salud como ser:

* Reducir las tasas de ausentismo.
* Habilitar más espacios para la atención médica.
* Proyectos para seguimiento de pacientes.
* Herramienta para el desarrollo de proyectos de investigación e inversión.

Para los cuales seria necesario un estudio diferente para agregar o modificar algunas variables en el Dataset, debido a que cada país, regios o ciudad tienen diferentes culturas, se encuentran en lugares geográficos variados, entre otros.