**Mentoría - Diplomatura en Ciencia de Datos y Aprendizaje Automático**

Modelo de recomendación de profesionales médicos para una mejor experiencia de usuarios

**Objetivos:**

El principal propósito de este informe es generar modelos de clasificación que nos permitan determinar cuáles son las variables más importantes a la hora de definir una categoría de socio y de prestador.

Para ello nos hemos valido de las variables resultantes del análisis exploratorio y de ingeniería de features realizado en los trabajos prácticos anteriores.

**Consigna 1**

En primer lugar, se ha definido la variable objetivo a ser predicha, la cual puede asumir dos valores: Standard y Silver+Gold, los culaes se corresponden con categorías de socio.

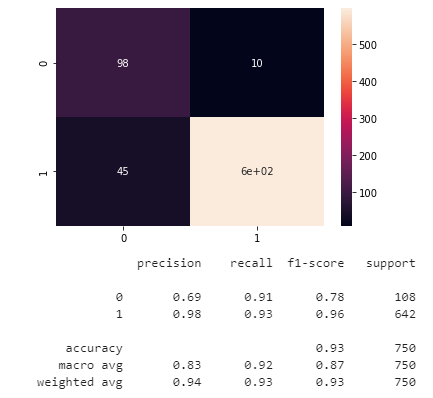
Target: Categoria\_socio\_Standard

Se utilizarán tres algoritmos de clasificación:

* Regresión Logística.
* Decision Tree
* RandomForest

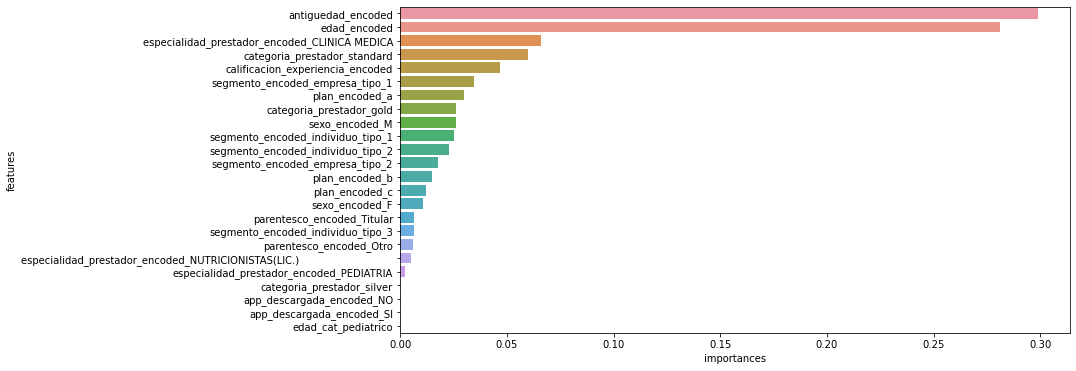
Se deberá encontrar un modelo que devuelva las mejores métricas. Para ello se hizo uso de la técnica de validación cruzada con GridSearch para obtener los mejores hiperparámetros que maximizan una métrica elegida.

El mejor modelo fue un Árbol de Decisión, obteniéndose los siguientes resultados:



Una vez obtenido el mejor modelo, atendiendo a los objetivos de este trabajo, nos valdremos del atributo “feature\_importances”.

Los resultados indican que las variables más importantes son: antigüedad y edad. En menor medida si el tipo de atención recibida es de Clínica Médica y si la categoría del prestador es standard.

****

**Consigna 2**

- Elección del clasificador

Para aplicar los modelos de clasificación propuestos a nuestro conjunto de datos, primero vamos a decidir nuestra variable a predecir, es decir nuestro target.

Target: Categoria\_prestador \_Standard

Se trabajará con este target porque cuando esta variable es 0, significa que el prestador es de categoría Silver+Gold

Se emplea los modelos propuestos decision tree y logistic regression. Para encontrar los valores óptimos de los hiperparametros que emplean estos modelos, además de encontrar cuál de estos dos modelos es el mejor por lo que se empleara en método GridSearchCV, para ello.

Por medio del método mencionado anteriormente podemos acceder cual es el mejor modelo, la mejor puntuación y los mejores parámetros encontrados a través de los atributos best\_estimator,best\_score y best\_params.Por lo que el clasificador elegido es decision tree.

* Feature importance

Por medio del siguiente gráfico, el cual nos da una puntuación que nos indica la importancia relativa de cada feature al realizar la predicción, vemos que la variable clasificación\_experiencia (puntaje que le otorgo el paciente a su médico), tiene el puntaje mas alto.

