**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 practico es genera un modelo de recomendación para cada especialidad. Para ello nos hemos valido de las variables resultantes del análisis exploratorio y de ingeniería de features realizado en los trabajos prácticos anteriores.

**Desarrollo:**

1. Preprocesado del dataset.

* Se verifica que no se repitan prestadores entre las distintas especialidades.
* Se hace un filtrado del dataset, para obtener los registros de los pacientes que fueron atendidos por prestadores, al cual puntuaron con calificaciones altas. Este filtrado se lo hace por medio de los cuantiles.
* Se elimina las variables que no son necesarias para elegir el modelo de recomendación.

1. Elección del Modelo
2. Target: id\_prestador

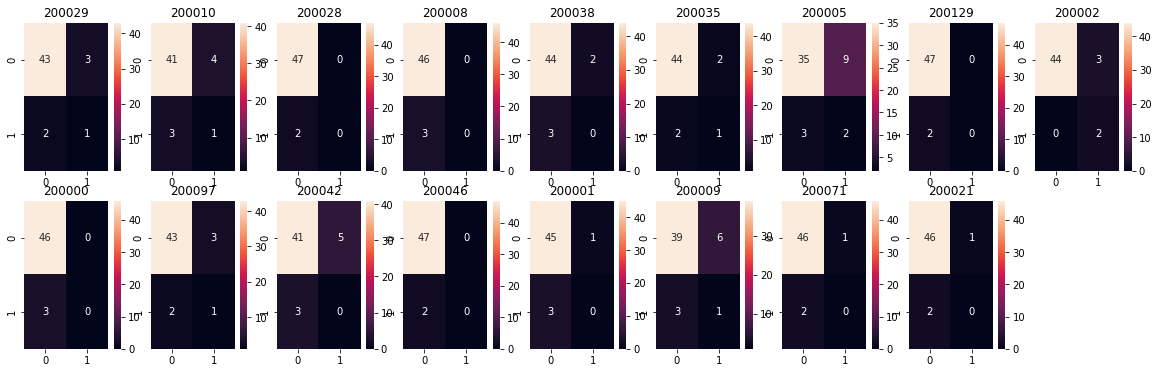
Se utilizarán tres algoritmos de clasificación:

* Decision Tree
* RandomForest
* k Nearest Neighbours

Para encontrar un modelo que devuelva las mejores métricas, se usó de la técnica de validación cruzada con GridSearch para obtener los mejores hiperparámetros que maximizan una métrica elegida.

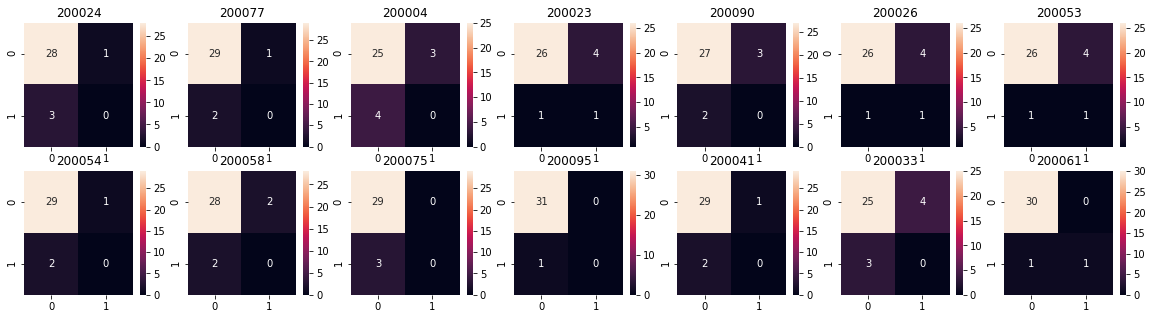
Para todas las especialidades, el mejor modelo fue RandomForest con lo cual se obtuvieron los siguientes resultados.

Médicos Clínicos



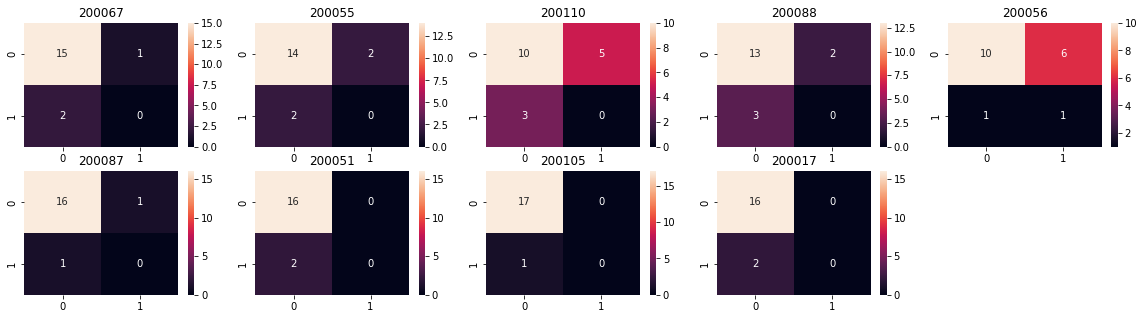
Luego de analizar las métricas globales y las matrices de confusión de cada una de las clases, podemos concluir con que los resultados de la clasificación no son buenos. El mejor modelo ha logrado clasificar de manera correcta únicamente para las clases 200010,2000029,200035,200005,200002,200097,200009. Por ejemplo, para el prestador 200010 solamente se clasificó bien una vez, hay 3 falsos negativos y cuatro falsos positivos. El mejor f1-score se obtuvo para la clase 200002.

Nutricionistas



Con las matrices de confusión podemos observar que los resultados de la clasificación no son los mejores en casi todas las clases. Solamente se ha clasificado correctamente para 4 clases (200023,200053,200061,200026), habiendo en casi todas ellas varios FN y FP, como lo indican las métricas de precisión y sensibilidad. Donde se tiene mejor F1-score es con la clase 200061, que es donde mejor se ha clasificado teniendo en cuenta por supuesto el número de registros disponibles para esta clase.

Médicos Pediatras



Con las matrices de confusión podemos observar que los resultados de la clasificación no son los mejores en casi todas las clases. Solamente se ha clasificado correctamente 1 clase (200056) habiendo en casi todas ellas varios FN y FP, como lo indican las métricas de precisión y sensibilidad. Donde se tiene mejor F1-score es con la clase 200056.

**Conclusión:**

Los resultados que se obtuvieron al utilizar un clasificador Randon Forest no fueron los mejores. Por lo que se tendrá que buscar otro algoritmo/modelo que nos sirva para armar nuestro modelo de recomendación, se plantea buscar técnicas de aprendizaje no supervisado.