**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 desarrollar un modelo de recomendación tipo collaborative filtering. Para ello nos valdremos de las variables resultantes del análisis exploratorio y de ingeniería de features realizado en los trabajos prácticos anteriores. En línea con el objetivo planteado para el proyecto, el dataset de entrenamiento y validación del modelo serán aquellos prestadores que muestran altas calificaciones y un nivel de demanda medio y alto. Las altas calificaciones nos aseguran que los prestadores a recomendar brindan un buen servicio. El nivel de demanda medio/alto nos asegura que las calificaciones son representativas estadísticamente.

**Desarrollo:**

1. Preprocesado y Análisis del dataset.

En esta sección se comenta un resumen del Preprocesado y análisis que se realizó al dataset.

* Se verifica que no se repitan los consumos correspondientes a los mismos prestadores entre las distintas especialidades.
* Del análisis realizado de los trabajos prácticos anteriores, se tienen muchos prestadores (médicos) que han realizado pocas consultas, mientras que existen una minoría de prestadores que son los reciben una gran cantidad de consultas.
* 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 del percentil 0.42.
* Se hace un filtrado por aquellos prestadores con una alta demanda, superior al percentil 0.42.
* Se elimina las variables que no son necesarias para el desarrollo del filtro colaborativo.

2- El filtro colaborativo

Para poder hacer un filtro colaborativo, es necesario encontrar similitudes en los consumos de los diferentes socios, por lo que primero se analiza los consumos de los socios/pacientes. En la Fig. 1 se observa que la mayoría de los pacientes consumió un solo médico, esto nos causa un problema para la implementación de un filtro colaborativo que recomiende un médico a un paciente, ya que hay que encontrar un usuario similar que haya consumido más de un médico. En la mayoría de los casos, tenemos que un paciente consume siempre el mismo médico.

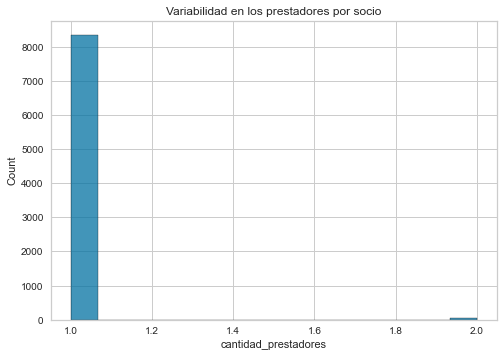


Fig.1: Variabilidad de los prestadores

Con lo mencionando anteriormente probamos que ir por un enfoque de filtro colaborativo no es posible y plantear implementar un filtro basado en contenido tampoco es una opción por los escases y poca variabilidad de los datos que se dispone.

3- Clustering SOM

Si bien en el intento de implementar un filtro colaborativo solo se consideró los consumos de los pacientes. Ahora se utilizarán los features que están relacionadas a las características de los pacientes. La idea es en esencia encontrar los pacientes cercanos y recomendar en función a las características en común. Para ello se define una serie de clusters y se recomienda en función al cluster que pertenezca un determinado cliente/paciente.

Se utilizarán Self Organizing maps (SOM) para hacer clustering. Primero se determina el numero óptimo de cluster utilizando el método del codo y Silhouette. Según las Fig. 2 y Fig. 3, 4 es el número óptimo de clústeres según ambas técnicas elegidas. Sin embargo, al hacer un análisis de segmentación de clientes, es esperable que los métodos no nos proporcionen una clara definición de las fronteras entre los socios. Probablemente sea mejor conseguir variables representativas del negocio para hacer una diferenciación según algún objetivo o característica.

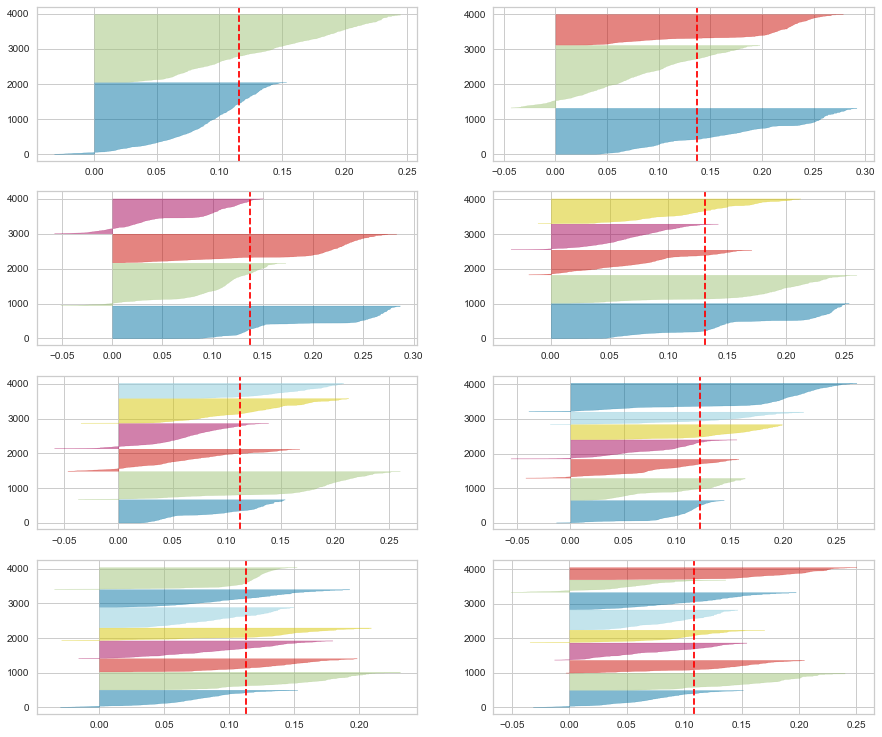


Fig.2: Silhouette

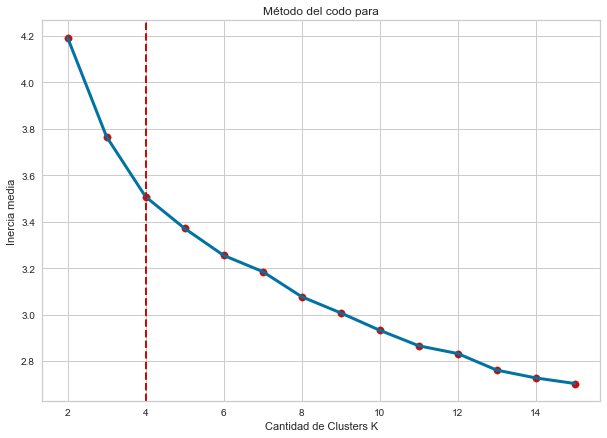


Fig.2: Método del Codo.

Para evaluar la calidad de nuestro modelo de clustering se procede predecir el cluster asignado a cada uno de los socios. Para esto para cada paciente, se obtendra una lista de prestadores que dependerán del cluster asignado. A su vez, dentro de cada cluster se encontrarán los socios más cercanos al buscado para encontrar una lista de *n* prestadores a recomendar.

A continuación, se evalúo la distancia del socio a los socios más cercados y se obtuvo una métrica (mean reciprocal ranking) que cuantifica la distancia al prestador que fue realmente seleccionado por el socio consultado. En la siguiente lista se muestra algunos ejemplos, el cual nos indica que el socio 103487 fue atendido por el medico 200005, y por medio de nuestro modelo este médico se encuentra en la posición 5 de nuestra lista a recomendar, esto no es lo mejor, ya que se esperaría que este médico se encuentre en la posición 1. Además, cabe aclarar que en muy pocos casos se acierta la recomendación, tal es el caso del médico 104510

El socio 103487 tuvo como prestador real: 200005

El prestador esta en el lugar número 5 - Mean reciprocal ranking= 0.2

El socio 106832 tuvo como prestador real: 200005

El prestador esta en el lugar número 5 - Mean reciprocal ranking= 0.2

El socio 108064 tuvo como prestador real: 200110

El prestador esta en el lugar número 38 - Mean reciprocal ranking= 0.02631578947368421

El socio 106540 tuvo como prestador real: 200097

El prestador esta en el lugar número 36 - Mean reciprocal ranking= 0.027777777777777776

El socio 101577 tuvo como prestador real: 200023

El prestador esta en el lugar número 12 - Mean reciprocal ranking= 0.08333333333333333

El socio 104510 tuvo como prestador real: 200000

El prestador esta en el lugar número 1 - Mean reciprocal ranking= 1.0

**Conclusión:**

La mejor manera de desarrollar un sistema de recomendación es por medio de técnicas de aprendizaje no supervisado, tales como Filtro Colaborativo, Filtro Basado en Contenido y Clustering. Para la mayoría de estas técnicas se requiere disponer de una cantidad suficiente de datos que nos permitan implementar este tipo de técnicas. Con lo desarrollado en trabajo practico de mentoría, se pudo observar que la predicción en muy pocos casos acertaba, uno de los motivos al cual se sospecha puede deberse al formato de los datos, es decir la poca diversidad de prestadores en los consumos. Mas allá de que los resultados obtenidos no son los mejores, con la realización de este trabajo se logró comprender el proceso que se debe llevar a cabo para poder implementar este tipo de sistema, los cuales van desde un análisis descriptivo, un preprocesamiento de los datos hasta la búsqueda de un modelo que se adecue a nuestras necesidades y que nos den los mejores resultados.