



Universidade de Vigo

Trabajo Fin de Máster

---

# Recomendación de productos

---

Ángel Padín Cores

**Máster en Técnicas Estadísticas**

**Curso 2019-2020**

Se tomó la decisión, de mutuo acuerdo con FINSA, de la no publicación del trabajo original para preservar la privacidad de los datos proporcionados por la empresa para la realización del mismo. En su defecto, se proporciona este resumen introductorio al proyecto original:

## **RESUMEN TFM**

FINSA es una empresa que produce y comercializa productos derivados de la madera y dispone de un catálogo de miles de productos que se comercializan en la actualidad entre centenares de clientes. Estos productos están organizados en familias con diferentes materiales y acabados.

Además, la información de estos productos se muestra en un portal comercial a los clientes en función de su perfil.

El objetivo de este proyecto es analizar las ventas históricas de los productos de FINSA a sus clientes para obtener las cestas de productos más habituales y crear un sistema de recomendación que sugiera productos relacionados con los productos que han comprado los clientes.

Dentro del proyecto están definidos los siguientes objetivos:

- Estudiar los datos disponibles en la base de datos de FINSA, creada con tecnología de Google Cloud. Se crearán dashboards para poder visualizarlos y analizar la relación entre los diferentes datos disponibles sobre la infraestructura de Google Cloud.
- Obtener los datos de las entregas realizadas a los clientes de FINSA desde la base de datos creada en Google Cloud.
- Trabajar con los datos disponibles para procesarlos en R atendiendo a sus características de volumen y dimensionalidad.
- Estudiar los principales algoritmos disponibles para realizar un recomendador de productos con los datos disponibles.
- Desarrollar diferentes recomendadores de productos con los datos existentes.
- Definir las métricas asociadas a los recomendadores y analizar el rendimiento de cada uno de ellos.
- Crear una herramienta dinámica en Shiny para visualizar el resultado del recomendador de productos que permita hacer simulaciones de nuevas cestas de productos y sus recomendaciones.

- Presentar los resultados obtenidos durante el proyecto a diferentes departamentos de FINSA para que los evalúen.

En este trabajo de fin de máster se intentará dar solución a las cuestiones propuestas en los objetivos teniendo en cuenta las limitaciones temporales de las que se disponen en una empresa.

## ¿Que es un sistema de recomendación?

Un sistema de recomendación se podría definir como una herramienta diseñada para interactuar con conjuntos de datos con el fin de proporcionar al usuario información que sea de su interés de forma automatizada, normalmente en forma de recomendaciones que probablemente desconozca, basándose en técnicas estadísticas aplicadas sobre un conjunto de datos tal como un histórico de ventas o de valoraciones. Este proceso ayuda al cliente a encontrar productos que podrían adaptarse a sus necesidades más rápidamente, promoviendo la venta cruzada, que consiste en vender productos complementarios a los que consume o pretende consumir un cliente, y creando valor añadido a la relación negocio-cliente.

La mayoría de los sistemas de recomendación se pueden clasificar en tres grupos: Basados en contenido, filtrado colaborativo y los mixtos, combinación de los anteriores. Los sistemas basados en contenido se enfocan en la idea de “muéstrame más cosas como las que me han gustado”. Para desarrollar esta idea se necesita disponer de una lista de valoraciones de cada usuario acerca de ciertos productos y una cuantificación de las características de dichos productos. A partir de los productos que han sido valorados positivamente por el usuario, ya sea en forma de puntuación o simplemente en términos de si lo ha consumido o no, se pueden recomendar otros cuyas características sean parecidas a las que tenían los productos anteriores.

Los principales inconvenientes de este tipo de sistemas de recomendación son:

- Requiere establecer cuáles son los atributos principales de los productos.
- Las recomendaciones suelen ser productos muy similares a los ya consumido por el usuario.
- Los valores de las características de los productos no aportan información acerca de su calidad.

En cuanto a los sistemas de filtrado colaborativo podemos decir que se fundamentan en la idea de “muéstrame más cosas que le hayan gustado a gente parecida a mí”. En este caso necesitaremos un listado con las valoraciones de los usuarios a los diferentes productos. Al conjunto de las valoraciones de un usuario a los diferentes productos se le conoce como perfil del usuario. La facilidad para recolectar información del usuario

en contraposición a las dificultades que aparecen a la hora de valorar las características de los productos hace que los sistemas de filtrado colaborativo estén mucho más extendidos que los sistemas basados en contenido, sobre todo en grandes negocios donde se manejan miles de productos. Adaptándonos a las características de dimensionalidad de los datos, así como a la falta de una categorización de características de los productos, nos decantaremos por implementar este tipo de sistema de recomendación, más allá de recomendar los productos más vendidos en general. En cuanto a los inconvenientes, la principal desventaja de este tipo de sistemas es el denominado *cold start*, fenómeno que surge cuando aparece un nuevo usuario en el sistema y no se dispone de ninguna información sobre su comportamiento, por lo que se hace difícil hacer alguna recomendación.

Otro de los principales problemas que surgen cuando se desea utilizar un sistema de recomendación basado en el filtrado colaborativo reside en la construcción del perfil de los usuarios. Esta información se puede recolectar de forma directa mediante encuestas o testes en los que el usuario debe valorar los productos. Existen muchas investigaciones y son varios los algoritmos desarrollados para crear recomendaciones en los casos en que tenemos unas valoraciones de los ítems de forma explícita. Sin embargo, en algunas situaciones no es práctico hacer una valoración de los productos que compramos, por ejemplo, en los supermercados o en los centros comerciales, y por lo tanto no se dispone de forma explícita de una lista de valoraciones por parte de los usuarios. Cuando no disponemos explícitamente de dichas valoraciones podemos obtener el perfil de un usuario de forma indirecta, por ejemplo, según el número de sus visualizaciones a los productos en una web o en función de las compras que realiza.

En el caso de los datos de los que dispone FINSA nos encontramos en esta tesitura, ya que no se dispone de una valoración directa por parte de los usuarios a los productos. Los datos se corresponden con las entregas realizadas a los diferentes clientes, que a su vez son empresas que se dedican a la distribución, o bien los utilizan para fabricación de nuevos productos para su posterior venta. Esta situación se puede modelizar interpretando los comportamientos de los clientes y traduciéndolos a datos binarios (es decir utilizando 0 y 1) de la siguiente forma: 0 será la valoración que dará un cliente en los casos en que no haya consumido un producto y 1 en el caso contrario. Notemos que existe la complicación de que en el 0 están incluidos los casos de productos que podrían interesar al usuario y los que no le interesan definitivamente, ya que la razón por la que un cliente no ha consumido un cierto producto es desconocida, pero debería ser una de las siguientes: El cliente no necesita el producto en este momento; el cliente no conoce el producto, en cuyo caso es un candidato para una recomendación o bien al cliente no le gusta ese producto, con lo que ese producto no debe ser recomendado a dicho cliente. Dada esta situación, el objetivo del sistema de recomendación será valorar los productos que un cliente nunca ha consumido (es decir, los valorados con un 0) en

función de los que sí sabemos que valora positivamente (los que se codifican con el 1) y clasificarlos en productos recomendados o no recomendados.

Desde el punto de vista estadístico, esta situación se corresponde con un problema de clasificación, perteneciente al campo del aprendizaje estadístico, un ámbito relacionado con el aprendizaje automatizado de la ciencia de los datos.

Se considera aprendizaje estadístico a un conjunto de técnicas que permiten entender y modelar problemas que involucren conjuntos de datos complejos. Estas técnicas se engloban principalmente en dos grupos: el aprendizaje supervisado y el no supervisado. Los problemas del aprendizaje supervisado se engloban los problemas en los que para cada observación de unas ciertas variables predictoras existe asociada una variable respuesta. En esta situación se desea obtener un modelo que relaciones las variables predictoras con la respuesta, con el objetivo de hacer predicciones de la variable respuesta para futuras observaciones o entender mejor la relación entre las predictoras y la respuesta. En este marco podemos encontrar técnicas como la regresión o los propios clasificadores.

El aprendizaje no supervisado, en contraste con el anterior, abarca los problemas en que disponemos de una serie de variables, pero no se dispone de una respuesta asociada a ellas. La solución consiste en entender las relaciones entre las distintas variables. Entorno a este problema encontramos la formación de *clusters* como principal mecanismo de solución.

## **Desarrollo de trabajo**

Tras la introducción, en el capítulo 2 se describe el proceso de modelado de este problema, estudiando los principales tipos de filtrado colaborativo para la creación de sistemas de recomendación y todas sus métricas asociadas. Además, se exponen también los métodos de evaluación que se pueden llevar a cabo y la forma de implementar todas estas ideas en un software como R.

En el capítulo 3 se aplican de una forma práctica todo lo visto en el capítulo 2 a los datos de los que dispone FINSA, tratando de resolver todas las tareas asignadas al proyecto, desde el análisis la base de datos, descarga y tratamiento de los datos hasta el desarrollo e implementación de los distintos algoritmos de recomendación, comprobando el rendimiento de cada uno para los datos existentes.

Por último, en el capítulo 4 se muestran el producto principal del proyecto. Se presenta un recomendador de productos en forma de aplicación interactiva, en la cual podremos analizar los perfiles de los clientes ya existentes en la base de datos y lanzar sugerencias de productos personalizadas y también podremos simular una cesta de la compra,

añadiendo los productos que queremos de todo el catálogo y recibiendo instantáneamente recomendaciones de productos para dicha cesta.