

## TALLER 2

Modelos de Recomendación Híbridos y Evaluación.

Leidy Yinet Ferro Lara,  
Daniel Felipe Barón Espitia,  
Julio Alejandro Morales Buitrago

Maestría en Ingeniería de Información, Facultad de Ingeniería

Universidad de los Andes

Bogotá

2022

## Tabla de contenido

I.	Definición y construcción de modelo híbrido de recomendación	3
II.	Sintonización y evaluación del modelo	6
III.	Aplicación Web	9
IV.	Análisis de resultados	9

## I. Definición y construcción de modelo híbrido de recomendación

### a. Objetivo de recomendación propuesto

#### Objetivos estratégicos del proveedor de servicio y del usuario de servicio

Como proveedor de servicio, ofrecerles a los usuarios un top 10 de nuevas opciones para visitar, a partir de sus características de perfil y hábitos de consumo, así mismo, promover las visitas a nuevos sitios, a partir de las características más significativas que el usuario potencial busca a la hora de seleccionar un lugar para visitar.

Como usuario del servicio, contar con una opción que le permita contar con la recomendación de nuevos lugares a visitar, de acuerdo a los lugares previamente visitados.

#### Propósito de recomendación para el proveedor y el usuario de servicio

Como propósito general de esta solución tecnológica, se busca ofrecer tanto para el proveedor, como el usuario de servicio, una opción que le permita ofrecer opciones a partir tanto de las posibles similitudes que tienen los usuarios en sus opciones de visita, como de las características propias de los lugares, realizar posibles recomendaciones que les permitan conocer nuevos lugares, en la ciudades que son frecuentes en sus actividades diarias.

#### Tareas operativas del sistema

El sistema se encuentra operando bajo dos esquemas de selección, el primero, basado en usuario – usuario, y el segundo basado en contenido.

En la primera etapa, donde se usa la opción de usuario – usuario, se parte de la premisa de realizar el proceso a partir de un usuario, dado el volumen de datos. A partir de esto, se determinan:

- Las similitudes con los otros usuarios,
- Se realiza el cálculo de ratings para este usuario
- Se identifican los usuarios más similares a él
- Se genera una recomendación de los posibles sitios a visitar

Se ha manejado una muestra de usuarios para este entrenamiento, los cuales se relacionan a continuación:

	name	review_count	yelping_since
173	Archie	1022	2008-11-25 08:46:24
234	Howard	157	2010-05-20 20:54:52
538	Sarah	647	2008-06-12 03:53:41
569	Eric	5887	2007-03-28 19:08:35
718	Vicky	808	2007-12-27 18:26:42
897	Sophia	205	2010-11-04 03:23:05
922	Paul	769	2010-05-21 22:56:03
1047	Nick	178	2008-01-05 01:12:46
1220	Vicki	406	2010-06-03 03:11:57
1333	brook	83	2009-10-29 17:05:11

En la segunda etapa, donde se usa el sistema de recomendación basado en contenido, partimos de la premisa de usar la información relacionada de los restaurantes, a la cual realizamos previamente un ejercicio de tratamiento de información, tal que nos permitiera usar estos datos de la siguiente manera:

- Nombre del sitio
- Dirección
- Coordenadas
- Ciudad
- Estado
- Calificación
- Las características de cada sitio

Adicionalmente, se realiza una priorización de las características más relevantes, seleccionando los siguientes elementos:

```
['Electronics',  
'Professional Services',  
'Mobile Phones',  
'Caterers',  
'Barbeque',  
'Internet Service Providers',  
'Television Service Providers',  
'Alcohol Free',  
'Smokehouse',  
'Airports']
```

Para el proceso de entrenamiento, se ha tomado entonces como información para caracterizar al usuario lo siguiente:

- Los sitios visitados por el usuario
- Los sitios visitados por el top 10 de los usuarios más similares

El modelo entrenado es un K-means, donde de acuerdo a las evaluaciones de desempeño, se identifica como parámetro de clasificación a  $K=2$ .

Una vez se tiene los resultados, se toman los sitios que no ha visitado, pero combinando la información relacionada a las ciudades que ha visitado el usuario, de acuerdo a su contexto de ubicación.

Finalmente, se muestra al usuario, los sitios que no ha visitado, en orden de mejor calificación.

El modelo está planteado en los siguientes archivos de configuración:

**Preprocesamiento.py** : En esta capa realizamos las siguientes tareas,

- Cargue de datos
- Filtro por método usuario – usuario
- Similitud tipo coseno y la similitud entre los usuarios

**Content\_Based\_RS.py**: En esta capa realizamos las siguientes tareas,

- Transformación de datos, en formato columnas, dando prioridad las características de cada sitio, limpieza, unificar formatos.
- Se realiza priorización de variables, para dejar solo las que representaban mayor relevancia en los resultados.
- Se realiza segmentación de datos para entrenamiento y prueba

Una vez se realiza el entrenamiento, se realiza el proceso de predicción de la siguiente manera:

- Se toman los datos de los sitios que no ha visitado el usuario, para que a partir de la predicción, le pueda definir un posible rating , se ordena la predicción y se toman las K recomendaciones.
- Finalmente se arma la estructura de la información adicional de las recomendaciones:
  - o Nombre
  - o Dirección
  - o Ciudad
  - o Estado
  - o La calificación de acuerdo a la predicción

**Hybrid\_RS.py:** En esta capa se realiza la combinación de los resultados,

- Se recopilan los metodos genrados en las capas previas.
- Se procede a la combinación de resultados, para generar la recomendación final.

### Métricas para validación de tareas

Para la solución, se usan las siguientes métricas:

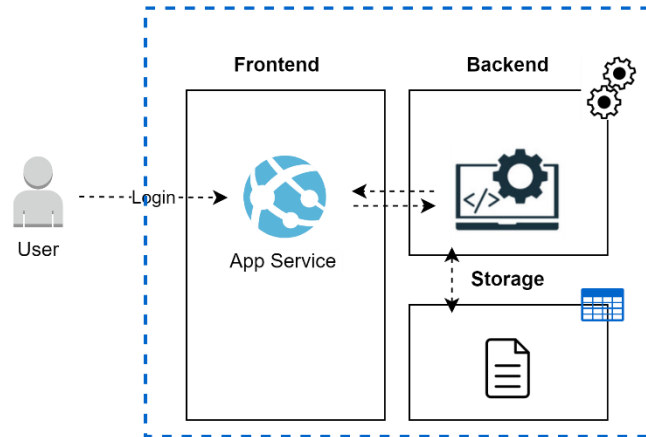
- Precisión de predicción
- Cobertura
- Medida F del modelo

A continuación relacionamos los objetivos y la estrategia para el sistema de recomendación propuesto:

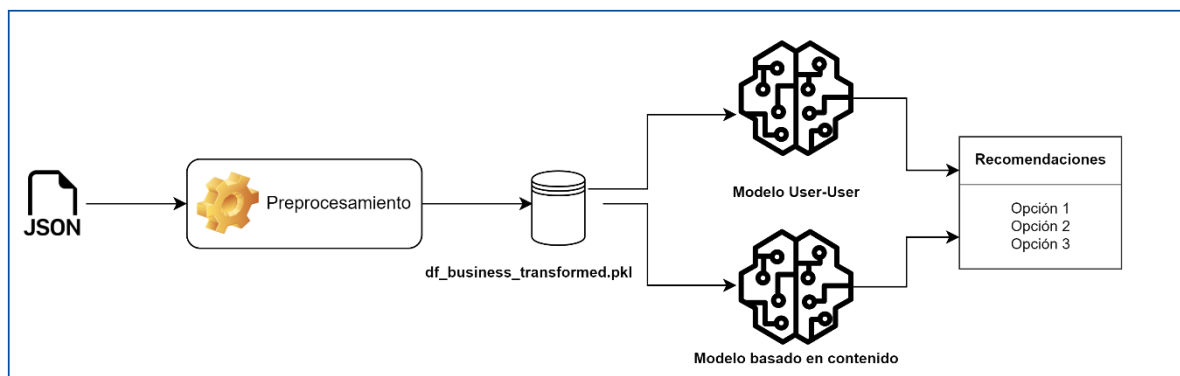
		<b>Usuario</b>	<b>Proveedor</b>
<b>Perspectiva Estratégica</b>	<b>Objetivo General</b>	Proponer nuevos lugares a visitar, de acuerdo con gustos y ubicación	Promover visitas de sitios
	<b>Propósito de recomendación</b>	Proponer nuevos lugares Explorar opciones no conocidas en ciudades visitadas	Aumento de visitas Proponer visitas de lugares no visitados Aumentar la demanda
<b>Perspectiva Operacionanl</b>	<b>Tareas del sistema</b>	Seleccionar lugares, de acuerdo con características más relevantes Recomendar lugares, teniendo en cuenta elementos de contexto del usuario.	
	<b>Métricas computacionales</b>	Precisión de predicción Cobertura Medida F del modelo	

## b. Arquitectura del sistema de recomendación

A continuación relacionamos la arquitectura del sistema de recomendación:



A nivel de modelo, se presentan las etapas realizadas para la transformación, procesamiento e implementación de los modelos:



Como se observa en la figura anterior, para esta solución, se ha realizado una etapa de preprocesamiento, de la cual obtenemos una base de información intermedia (df\_business\_transformed.pkl), con la cual se han realizado las actividades de entrenamiento y predicción de los dos modelos.

## II. Sintonización y evaluación del modelo

### a. Esquema de experimentación y evaluación

Dada la cantidad de datos y el formato en que se encuentra la información, previamente se ha requerido realizar una actividad de preprocesamiento tal que nos permitiera contar con una estructura de datos para posteriormente ser usada en los modelos. La base de información objeto de este taller, se ha usado de Yelp (<https://www.yelp.com/dataset/download>), puntualmente la relacionada a Review.json y Business.json. Para el entrenamiento, se ha tomado como base una muestra de 11 usuarios, con sus respectivas valoraciones y histórico de visitas.

Para la evaluación del modelo, se tomarán como criterios de aceptación, los siguientes elementos:

- Los resultados presentados de recomendación, de acuerdo a la proporción de sitios visitados por ciudad, y las posibles opciones que no han sido visitadas y que tienen una alta valoración, tal como se presenta en el ejemplo a continuación:

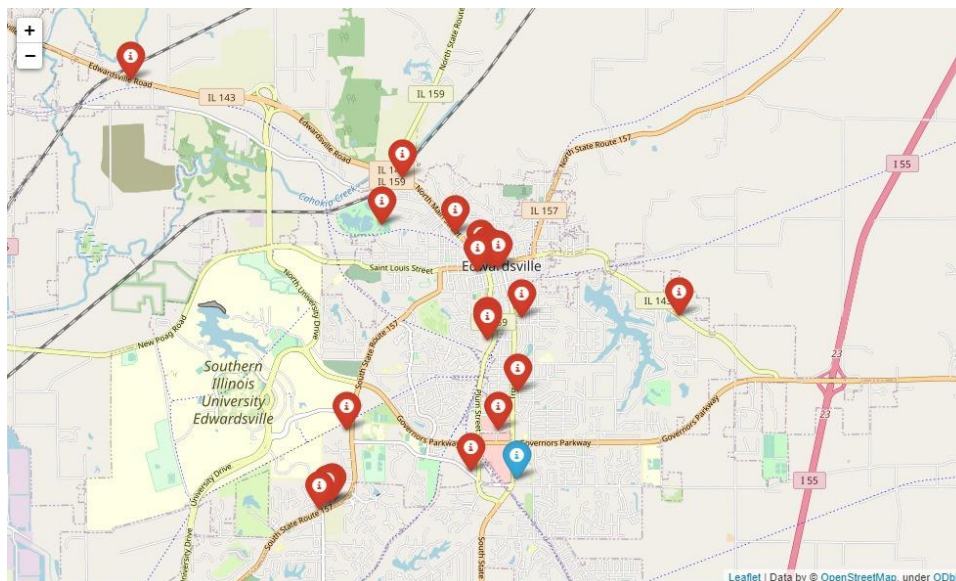
### Usuario ejemplo 1:

	name	review_count	yelping_since
1984695	Bill	1	2021-06-13 23:26:52

Resultados obtenidos, vista detalle sitios:

name	address	city	state	review_stars_predicted	latitude	longitude
Edwardsville Children's Museum	722 Holyoake Rd	Edwardsville	IL	5.000000	38.804395	-89.949733
Mike Shannon's Grill	871 S Arbor Vitae	Edwardsville	IL	5.000000	38.801945	-89.955650
Joe's Pizza & Pasta	4 Club Centre Ct	Edwardsville	IL	5.000000	38.779284	-89.983001
Fazoli's	6663 Edwardsville Crossing Dr, Ste A	Edwardsville	IL	5.000000	38.789029	-89.953745
El Maguey	1534 Troy Rd	Edwardsville	IL	5.000000	38.794259	-89.950390
Coyote's Ranch Grill & Saloon	4 Club Center Ct	Edwardsville	IL	5.000000	38.779004	-89.983799
Watershed Nature Center	1591 Tower Ave	Edwardsville	IL	5.000000	38.817067	-89.974173
Eclipse Car Wash	62025-2812	Edwardsville	IL	5.000000	38.783439	-89.958672
Edwardsville Public Library	112 S Kansas St	Edwardsville	IL	5.000000	38.811135	-89.953855
Lucy n' Ricardo's	3624 Edwardsville Rd	Edwardsville	IL	5.000000	38.836850	-90.018415
Bella Milano	1063 S State Route 157	Edwardsville	IL	7.000000	38.789031	-89.980409
Northside Dairy Haven	1902 N Main St	Edwardsville	IL	6.000000	38.823463	-89.970683
Jacks 66 Auto Repair	601 N Main St	Edwardsville	IL	6.000000	38.815861	-89.961313
Cleveland-Heath	106 N Main St	Edwardsville	IL	5.666667	38.811610	-89.955697
Dewey's Pizza	112 E Vandalia St	Edwardsville	IL	5.500000	38.811217	-89.954867
Andria's Countryside	7415 State Rt 143	Edwardsville	IL	5.500000	38.804629	-89.922006
Peel Wood Fired Pizza	921 S Arbor Vitae	Edwardsville	IL	5.400000	38.801353	-89.955765
Jerry's Tire Sales	104 W Vandalia St	Edwardsville	IL	5.250000	38.810691	-89.957465
Bigelo's Bistro	140 N Main St	Edwardsville	IL	5.000000	38.812598	-89.956829
Treo Salon	2117 South Rte 157	Edwardsville	IL	5.000000	38.778213	-89.985281

Resultados obtenidos, vista ubicación geográfica:





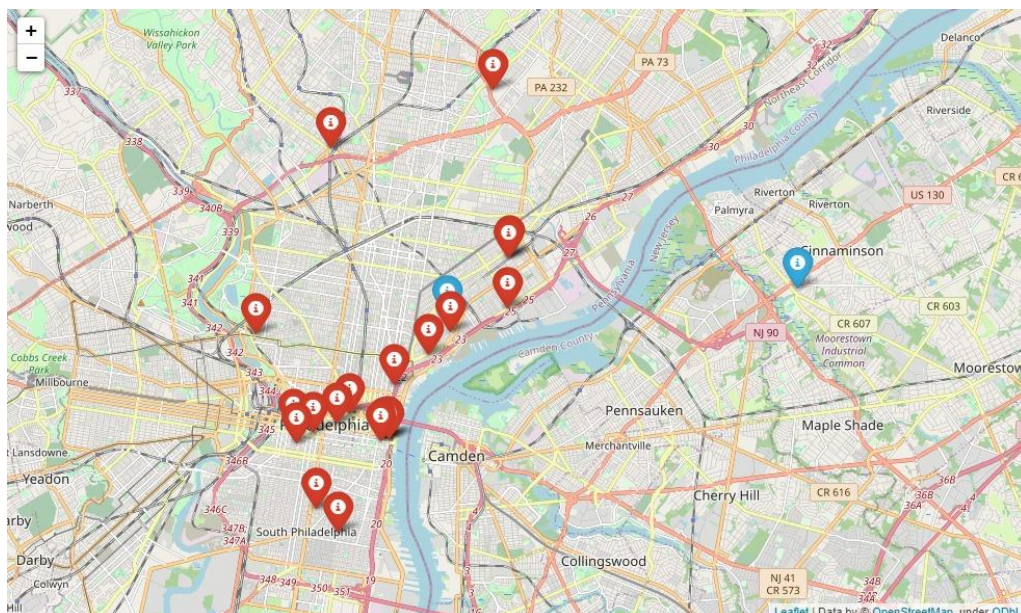
## Usuario ejemplo 2:

	name	review_count	yelping_since
1067051	Anthony	2	2012-11-03 01:48:27

Resultados obtenidos, vista detalle sitios:

	name	address	city	state	review_stars_predicted	latitude	longitude
0	St Honore Pastries	935 Race St	Philadelphia	PA	5.0	39.955505	-75.155564
1	Chroma Hair Salon	108 Market St	Philadelphia	PA	5.0	39.949643	-75.142662
2	honeygrow	110 S 16th St	Philadelphia	PA	5.0	39.950723	-75.167501
3	Tinto	114 S 20th St	Philadelphia	PA	5.0	39.951501	-75.173845
4	Swanky Bubbles	10 S Front St	Philadelphia	PA	5.0	39.949346	-75.142349
5	iMunch Cafe	1233 N 31st St	Philadelphia	PA	5.0	39.975918	-75.186330
6	Nam Vang Restaurant	571 Adams Ave	Philadelphia	PA	5.0	40.037561	-75.108069
7	Vista Peru	20 S 2nd St	Philadelphia	PA	5.0	39.949418	-75.143959
8	The Victor Cafe	1303 Dickinson St	Philadelphia	PA	5.0	39.931653	-75.166327
9	Metropolitan Bakery	262 S 19th St	Philadelphia	PA	5.0	39.948295	-75.172815
9584	Caliber Collision	2609 E York St	Philadelphia	PA	7.0	39.976358	-75.122156
10173	Olde City Escape Games	221 Chestnut St, Ste 300	Philadelphia	PA	7.0	39.948628	-75.144984
4137	Deke's BBQ	137 W Berkley St	Philadelphia	PA	7.0	40.022987	-75.161698
12098	Luhv Vegan Deli	51 N 12th St	Philadelphia	PA	7.0	39.953210	-75.159409
5494	Panther Pro Audio	3475 Collins St	Philadelphia	PA	7.0	39.995002	-75.102785
10990	La Virgen De Las Nieves	1839 S 8th St	Philadelphia	PA	7.0	39.925589	-75.159266
8943	Kurant Brew & Brew	436 E Girard Ave	Philadelphia	PA	7.0	39.970645	-75.129318
110	The Lunar Inn	3124 Richmond St	Philadelphia	PA	7.0	39.982293	-75.103160
13433	Art Plus Science	801 N 2nd St, 2nd Fl	Philadelphia	PA	7.0	39.963035	-75.140628
13429	Kensington Pub	2116 E Tioga St	Philadelphia	PA	7.0	39.995567	-75.102530

Resultados obtenidos, vista ubicación geográfica:





### III. Aplicación Web

La solución web propuesta para esta solución, presenta un flujo de trabajo, tal que al consultar un usuario particular, nos presenta:

- **Datos del usuario:** ID, Nombre, cantidad de revisiones, Fecha de asociación a la base de información.
- **Recomendaciones:** Se presentan las características más relevantes, y las posibles recomendaciones, de acuerdo a su historial.

### IV. Análisis de resultados

Cuando se plantean soluciones de sistemas de recomendación, es importante manejar como punto de partida, cuales son los objetivos tanto del usuario potencial, como del mismo proveedor, para que los esfuerzos de desarrollo de estos sistemas sean encaminados a resolver las necesidades del negocio de manera integral.

Esto no solo permite aplicar la combinación de técnicas tanto para la segmentación de los datos, la priorización de los datos que serían más relevantes para el caso de uso, y permitir dar respuesta a la posible recomendación.

A lo largo de la generación de la propuesta de solución, se han identificado aspectos que requieren ser considerados para aprovechar los datos de la base de información. Algunos de estos elementos son:

- Técnicas para extracción, limpieza y calidad de la base de información.

Dado el volumen de datos y la estructura de almacenamiento, se requieren evaluar técnicas tales que permitan no solo disminuir los tiempos de procesamiento requeridos, sino también su aprovechamiento, dado que para este caso de uso, fue necesario trabajar solo con una muestra pequeña, que permitiera ser analizada y tratada para aplicar los modelos propuestos.