

PROFESORA:

Valentina Grajales



INTEGRANTES (GRUPO 1):

Diego Felipe Carvajal Lombo (201911910)
 Brenda Catalina Barahona Pinilla (201812721)
 Sergio Julián Zona Moreno (201914936)

Tabla de contenido

1	Introducción.....	1
2	Conocimiento del dataset del trabajo.....	1
3	Definición y construcción de un modelo híbrido de recomendación	2
4	Sintonización y evaluación del modelo	4
5	Construya una aplicación Web sencilla.....	6
6	Análisis de resultados	8
7	Bibliografía y referencias	9

1 Introducción

En este documento se presenta solución al taller 2 del curso Sistemas de recomendación. Realizaremos un sistema de recomendación híbrido de lugares usando la librería *surprise* para *Python*, un backend desarrollado en *FastAPI*, y un frontend desarrollo en *React* + *Material UI* + *Bootstrap*.

A continuación, se encuentra el enlace al sitio web que permite visualizar los resultados:

<https://taller2-sr-202310-app.onrender.com/>

2 Conocimiento del dataset del trabajo

- a. Ubique los datos en <https://www.yelp.com/dataset>. En el link de descarga puede encontrar la documentación sobre la versión del dataset en la versión que escoja.

Al analizar el conjunto total de datos, por temas de limitación computacional, optamos por trabajar con un extracto de los mismos. Particularmente, trabajamos con una muestra de 50000 datos en cada uno de los 3 JSON de interés: Business, Users, Reviews.

Esta muestra de 50000 datos abarca (aproximadamente): el 25% de los negocios, el 2.5% de los usuarios y el 0.007% de las reseñas.

b. Estudie el formato de los datos, en particular identifique la manera de relacionar negocios, usuarios, visitas y reportes.

Para poder hacer un entendimiento de los datos en primer lugar se descargaron los 3 archivos JSON proporcionados por el curso para poder realizar el análisis de los mismos. Al realizar esto pudimos encontrar las siguientes columnas relevantes:

- De los negocios: name, city, stars, review_count.
- De las reseñas: user_id, business_id, stars
- De los usuarios: user_id, name, review_count, friends

Decidimos no hacer uso del conjunto “Tip” debido a que los comentarios de algunos usuarios a los negocios no tenían relación alguna con los mismos. Por ejemplo, se pudo encontrar un tip que decía “Avengers time with the ladies”. En la siguiente imagen se puede ver:



Con estos tres archivos, podemos hacer las siguientes relaciones:

- Buscar los negocios cercanos a un negocio en específico con base en su ubicación geográfica (principalmente la ciudad). Esto nos permitirá realizar recomendaciones sensibles al contexto geográfico del establecimiento.
- Utilizar el conjunto de amigos de un usuario específico para hacer recomendaciones que se basen en sus reseñas.
- Por último, y más importante, la tabla de reseñas conecta usuarios con negocios, y permite otorgarles un puntaje numérico de 1 a 5, que es la base de todas las recomendaciones que vamos a realizar.

3 Definición y construcción de un modelo híbrido de recomendación

- c. Establezca un objetivo de recomendación propuesto bajo el modelo conceptual de Jannach (Jannach & Adomavicius, 2016) visto en clase. Identifique objetivos estratégicos del proveedor del servicio y del usuario del servicio. Una vez establecidos, identifique un propósito de recomendación para el proveedor y el usuario del servicio. En base a los propósitos, identifique las tareas operativas de su sistema y escoja las métricas adecuadas para validar que las tareas operativas del sistema se estén cumpliendo. Describa detalladamente en el informe los objetivos, propósito de recomendación, tareas operativas y métricas que escogió para su solución.***

Utilizando como base el modelo de recomendación de Jannach, podemos implementar su Framework: *From Goals to Metrics* (véase la siguiente página).

Al implementar el Framework, logramos evaluar el propósito de recomendación y comprender ambos puntos de vista (tanto del usuario final, como de los negocios). Siendo entonces, una herramienta muy útil para plantear con claridad el sistema de recomendación.

Denotamos que nuestro sistema de recomendación (híbrido), cuenta con cuatro submodelos que permiten darle un componente de filtrado colaborativo, y, adicionalmente, sensible al contexto (*Context-Aware*).

$$SR_{Híbrido} = \gamma_1 SR_{Colaborativo_1} + \gamma_2 SR_{Colaborativo_2} + \gamma_3 SR_{Contexto_1} + \gamma_4 SR_{Contexto_2}$$

Donde γ_i es el peso correspondiente a cada submodelo y se optimiza con las métricas del RMSE y MAE.

		Punto de vista del consumidor	Punto de vista del proveedor
Perspectiva estratégica	Objetivo general	Utilidad personal: Sensación de satisfacción/felicidad al momento consumir bienes y servicios de negocios.	Utilidad organizacional: Crecimiento como organización, atracción de más clientes y adquisición de mayores ganancias.
	Propósito de recomendación	<ul style="list-style-type: none"> Ayudar al usuario a encontrar nuevos lugares (de su agrado) donde pueda consumir bienes y servicios. Mostrarle alternativas al usuario diferentes a las que normalmente consume. Recomendar al usuario de acuerdo con su contexto: especialmente, a las preferencias y gustos de sus amigos. Provocar serendipia y nuevas experiencias en el usuario. 	<ul style="list-style-type: none"> Aumentar la demanda del negocio. Conocer métricas de satisfacción en la ciudad donde se ofrece el servicio. Conocer el posicionamiento de marca respecto a otros competidores dentro de la ciudad.
Perspectiva operacional	Tarea del sistema	<ul style="list-style-type: none"> El sistema debe recomendar negocios que se ajusten a las características del usuario y cuyos productos sean de su interés. Recomendar negocios de manera variada según el contexto del usuario, evitando la sobrespecialización. Recomendar negocios teniendo en cuenta las opiniones y reseñas de sus amigos. Recomendar negocios teniendo en cuenta las opiniones y reseñas de otras personas en la ciudad. Recomendar negocios teniendo en cuenta las opiniones y reseñas del conjunto total de usuarios. Otorgar un top de negocios a los cuáles el usuario debería ir, y el porcentaje de similitud (por ejemplo, “<i>este lugar tiene un 93% de probabilidades que te guste</i>”). 	
	Métricas computacionales	Principalmente, se utilizan dos métricas computacionales para desarrollar el modelo híbrido: RMSE y MAE. Ambas permiten calcular el nivel de error de nuestro modelo respecto al mundo real, entre más cercano a cero este sea, nuestro modelo logra explicar mejor las preferencias de los individuos. Los pesos del modelo híbrido se optimizan de acuerdo con estas métricas.	

- d. Para cumplir con su objetivo de recomendación, defina la arquitectura de un sistema de recomendación. En este modelo debe consistir en un modelo híbrido de recomendación. El modelo híbrido debe hacer uso de al menos un modelo de filtrado colaborativo por factorización y un modelo de recomendación sensible al contexto. Puede incluir también modelos de contenido o colaborativo clásico (vecindario) dependiendo de su objetivo de recomendación. Debe justificar en su informe porqué el sistema planteado está alineado con el modelo conceptual propuesto.***

Se decidió realizar un sistema de recomendación híbrido compuesto de 4 criterios.

En primer lugar, se implementaron dos sistemas de recomendación *Context-Aware*. El primero de ellos se basa en la lista de amigos del usuario al que se le van a hacer las recomendaciones. Es decir, obtenemos la lista de los amigos de nuestro usuario. Y revisamos las reseñas y calificaciones les han otorgado a los diferentes negocios. Se concluye que este método es sensible al contexto debido a que se hace la suposición que los amigos del usuario tienen gustos potencialmente similares, por lo que, las reseñas deberían ser parecidas. Para ello se usó el algoritmo de KNNWithZScore.

El segundo *Context-Aware Recommendation System* es basado en el lugar en que se encuentra el negocio. Para ello, obtenemos la ubicación geográfica (la ciudad) y extraemos todas las reseñas que se han dejado a negocios en dicha locación. Concluimos que es sensible al contexto, ya que, entendemos que los usuarios de una misma ciudad son los que con mayor frecuencia han calificado los negocios de la misma, entonces si bien se está sesgando la predicción, permite ajustar los datos al contexto del mundo real. Para este se utilizó el algoritmo de KNNWithZScore.

Por último, se implementaron dos sistemas de recomendación basados en todos los usuarios y todas las reviews de cada negocio. Para el primero, se implementó un modelo de filtrado colaborativo por factorización (SVD) y el segundo se desarrolló con KNNWithZScore. Nótese que estos dos modelos, al tener más datos (ya que se usa al público general) al optimizar los pesos se les asignó un peso mayor que los modelos sensibles del contexto.

4 Sintonización y evaluación del modelo

- a. Establezca un esquema de experimentación y de evaluación del modelo de recomendación propuesto que permita identificar qué tan bien se cumple el objetivo propuesto por su sistema.***

Para el esquema de experimentación y de evaluación del modelo recomendación se hizo uso de las métricas estadísticas de RMSE y MAE. El RMSE se calcula tomando la raíz cuadrada de la media de los errores al cuadrado entre los valores reales y las predicciones del modelo. Básicamente, el RMSE mide qué tan cerca están las predicciones del modelo de los valores reales. Cuanto menor sea el valor del RMSE, mejor será la precisión del modelo. El RMSE

es útil porque penaliza los errores más grandes de manera más significativa que los errores más pequeños.

El MAE se calcula tomando la media de los errores absolutos entre los valores reales y las predicciones del modelo. El MAE mide el promedio de las diferencias absolutas entre las predicciones y los valores reales. Al igual que el RMSE, el MAE también es una medida de qué tan cerca están las predicciones del modelo de los valores reales.

Para optimizar estas métricas lo que se implementó una selección por hiperparámetros a cada uno de los modelos (utilizando validación cruzada). Con esto, podíamos hacer diferentes búsquedas simultáneas de la mejor “configuración” de cada uno de los modelos para los datos obtenidos. Repitiendo este proceso indefinidamente hasta mejorar sustancialmente los resultados del modelo. Esto se realizó para cada uno de los modelos.



```

SVD

from surprise import SVD
from surprise.model_selection import GridSearchCV, PredefinedKFold

# SVD
param_grid = {'n_factors': [50, 100, 200],
              'n_epochs': [10, 20, 50],
              'biased': [True, False],
              'lr_all': [0.002, 0.005, 0.01],
              'reg_all': [0.02, 0.04, 0.06],
              }

# Set up the grid search
gs = GridSearchCV(SVD, param_grid, measures=['rmse', 'mae'], cv=5, n_jobs=-1)

# Fit the grid search on the dataset
gs.fit(data)

# Print the best score and parameters
print('Best score:', gs.best_score['rmse'])
print('Best parameters:', gs.best_params['rmse'])

[16] Python
... Best score: 1.133144251152205
Best parameters: {'n_factors': 50, 'n_epochs': 20, 'biased': True, 'lr_all': 0.01, '

```

Imagen 1. Evaluación de uno de los modelos (SVD)

Cuando se obtuvieron los mejores parámetros de cada uno de los modelos se procedió a crear el modelo híbrido. Para ello, se definieron los pesos de cada uno de los modelos dependiendo de la cantidad de datos y las métricas previamente obtenidas. Posteriormente, se creó el modelo híbrido cumpliendo las especificaciones del Taller (mínimo un modelo colaborativo por factorización y uno sensible al contexto). Luego, volvimos a calcular métricas de RMSE y MAE del modelo híbrido y se ajustaron los pesos de los submodelos de manera que el modelo se pudiese ajustar mejor a los datos (minimizando el error).

```

[25] # Get the predictions for the test set
test_preds = [hybrid_predict(uid, iid) for uid, iid, _ in testset]

[26] len(test_preds)

... 3308

[30] test_values = []
for _, _, rating in testset:
    test_values.append(rating)

> v
[33] from sklearn.metrics import mean_absolute_error, mean_squared_error

print("RMSE", mean_squared_error(test_values, test_preds))
print("MAE", mean_absolute_error(test_values, test_preds))

... RMSE 0.984242036071063
MAE 0.7684077262758652

```

Imagen 2. Evaluación del modelo híbrido

- b.** *En su entrega debe entregar los scripts o código que le permitió realizar la experimentación, esta también podrá ser evaluada durante la sustentación de la entrega.*

Se adjunta en el .zip entregado el archivo. ipynb, notebook ejecutado con el código realizado.

5 Construya una aplicación Web sencilla

Para la construcción de la aplicación utilizamos la librería de *JavaScript*, *React*. Donde separamos la interfaz del usuario por componentes. Para el desarrollo de esta aplicación, se separaron las carpetas de la siguiente manera:

- **Assets:** para guardar todo lo correspondiente a imágenes estáticas de la aplicación.
- **Pages:** Se encuentran las páginas que se utilizan en la aplicación, esta a su vez está separada por una subcarpeta “Componentes”, donde están los componentes que componen la vista de la página en *React*, cada uno de estos con su correspondiente archivo de diseño *css*.
- **Routes:** Se encuentra el archivo donde se especifican las rutas que se manejan en la página web.
- **Services:** En esta carpeta se encuentran los archivos correspondientes para realizar la conexión con los endpoints del back y obtener la información de los lugares, calificaciones, comentarios, entre otros.

Adicionalmente, al ser una página web sencilla, nos enfocamos en realizar las funcionales más importantes para mostrar e interactuar con el sistema de recomendación. Con lo anterior, se tienen las siguientes funcionalidades en la página web:

- Inicio y cierre de la sesión del usuario: en esta ocasión decidimos realizar un login para los usuarios existentes. En caso de que el id del usuario no exista, se creará un nuevo usuario de manera automática.



Imagen 3 y 4. Log In y Manejo de errores en esta

- Listado de lugares en general
 - Información de cada uno de los lugares tales como nombre, ubicación, con las opciones para realizar una calificación (junto con un comentario) y ver los comentarios anteriores.

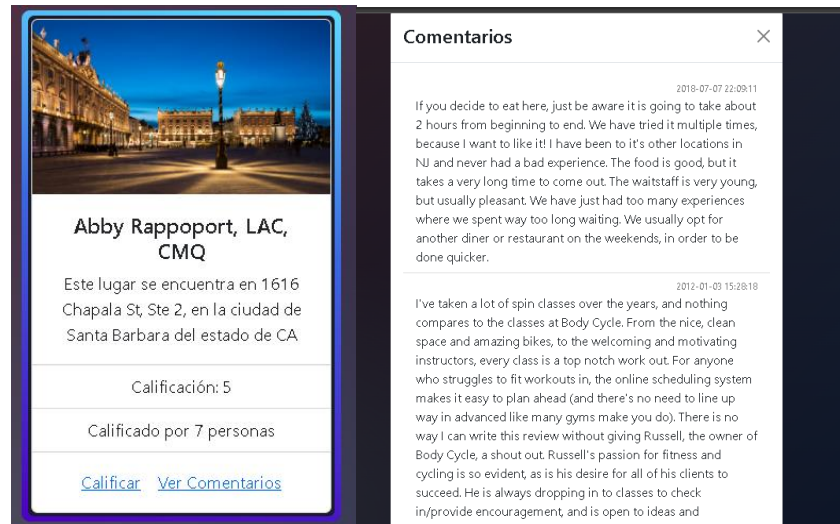


Imagen 4. Vista de uno de los lugares, junto con su calificación.

Imagen 5. Comentarios para uno de los lugares

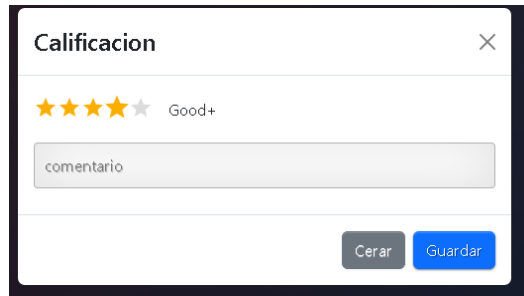


Imagen 6. Opción de calificación para uno de los lugares.

- Recomendación de lugares: Es parecida a la funcionalidad que se mostró anteriormente. Para esta se listan los lugares con su respectiva información además de sus posibilidades de calificar y ver comentarios, la información se presenta en tarjetas, similares a las mostradas anteriormente.

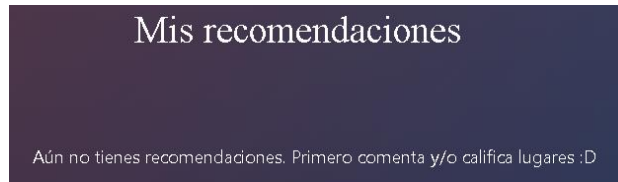


Imagen 7. Vista de Recomendaciones en caso de que sea un nuevo usuario.

Finalmente, también nos apoyamos de otras librerías para facilitar la construcción de los componentes, estas fueron “*react-bootstrap*” y “*Material UP*”

6 Análisis de resultados

- Elabore un informe en el que incluya el análisis solicitado en el punto 2 del enunciado. En el informe deben aparecer además los resultados de la experimentación offline sobre las métricas escogidas para encontrar los parámetros de sintonización, la descripción de la técnica de explicación utilizada y los resultados obtenidos con la aplicación. No olvide incluir conclusiones acerca de los resultados obtenidos***

En la experimentación offline, luego de construir el modelo híbrido, fue posible realizar recomendaciones al usuario de negocios que nunca había reseñado. Al contrastar estos datos, con los de sus amigos, podemos ver similitudes; de igual manera, el ajuste que realiza el modelo de reseñas por ciudad nos permite penalizar el modelo y obtener un mejor ajuste. Siendo entonces el resultado final, un modelo mejorado comparativamente que si se hubiesen utilizado los submodelos de manera individual.

A continuación, presentamos una lista con el TOP 10 negocios recomendados para un usuario, esto también puede ser visualizado en la página web.

--- User: 1M#RwQeZ0ZDhP2Qyry4g
Total business: 4719

	business_id	prediction	percentage	name	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open	attributes
0	m9XUd08b6XULDXs-ZGvQ	4.043075	80.86%	Tips & Toes	531 York Rd	Warminster	PA	18974	40.204633	-75.099462	3.5	13	1	['ByAppointmentOnly': 'False', 'BusinessAccept...
1	8zTCu7guUervileKya0AXg	4.018613	80.37%	Bimmerworks	721 E Nields St	West Chester	PA	19382	39.959157	-75.588299	4.5	13	1	['BusinessAcceptsCreditCards': 'True']
2	KJKKc51Toi7qppOxcvHfg	3.975193	79.5%	La Encantada	2905 E Skyline Dr	Tucson	AZ	85718	32.324742	-110.929997	4.0	88	1	['RestaurantsPriceRange2': '3', 'BusinessAccept...
3	fWbQISAEfNcznozzkCdvj	3.973602	79.47%	Ferrel's Chiropractic	533 E Micheltorena St, Ste 200	Santa Barbara	CA	93103	34.432568	-119.701846	4.5	8	1	['ByAppointmentOnly': 'True', 'AcceptsInsurance...
4	3puk4beQqK6zz8N4nIdhtQ	3.972262	79.45%	Hello Hair Salon	1905 S 13th St	Philadelphia	PA	19148	39.926343	-75.166891	4.0	39	1	['RestaurantsPriceRange2': '2', 'ByAppointment...
5	GJnYfVd9kbM182bKMfSAvg	3.963347	79.27%	Penco	7600 Wydown Blvd	Clayton	MO	63105	38.642721	-90.333404	4.5	152	1	['Alcohol': 'full_bar', 'RestaurantsReserva...
6	sDqoFbuuGPbQyFLmdyiw	3.959181	79.18%	YouFit Gyms	3066 Holiday Dr	New Orleans	LA	70131	29.921225	-90.011517	3.0	14	1	['GoodForKids': 'False', 'DogsAllowed': 'False...

Imagen X. Top recomendaciones usuario específico

También es posible concluir gracias a las métricas de evaluación, que las recomendaciones que hace nuestro modelo híbrido están muy bien rankeadas, gracias a que estas métricas dan valores bajos (0,76 aproximadamente). Por lo podemos tener un gran nivel de confiabilidad en los algoritmos realizados.

7 Bibliografía y referencias

- [1] Universidad de los Andes. (2020). *Material del curso: Sistemas de recomendación*. Disponible en BLOQUE NEÓN para estudiantes de la Universidad
- [2] Kumar, A. (2021, 13 diciembre). *Wilson Lower bound Score and Bayesian Approximation for K star scale rating to Rate products*. Medium. <https://medium.com/tech-that-works/wilson-lower-bound-score-and-bayesian-approximation-for-k-star-scale-rating-to-rate-products-c67ec6e30060>
- [3] Jannach, D., & Adomavicius, G. (2016). Recommendations with a Purpose. Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16) (pp. 7-10). Boston: ACM, New York.