

Sistema de Recomendación Turístico

Jhoan S. Moscoso & J. Manuel Agudelo



Índice

1. Introduccion al Recomendador Turistico

2. Analisis y Depuracion de Datos

3. Creacion del perfil del usuariosy lugares

4. Modelos predictivos

5. Implementación

6. Conclusion

Introducción

El turismo es una de las industrias más dinámicas y significativas a nivel global.

La creciente diversidad de atracciones hace necesario un sistema que concilie preferencias individuales



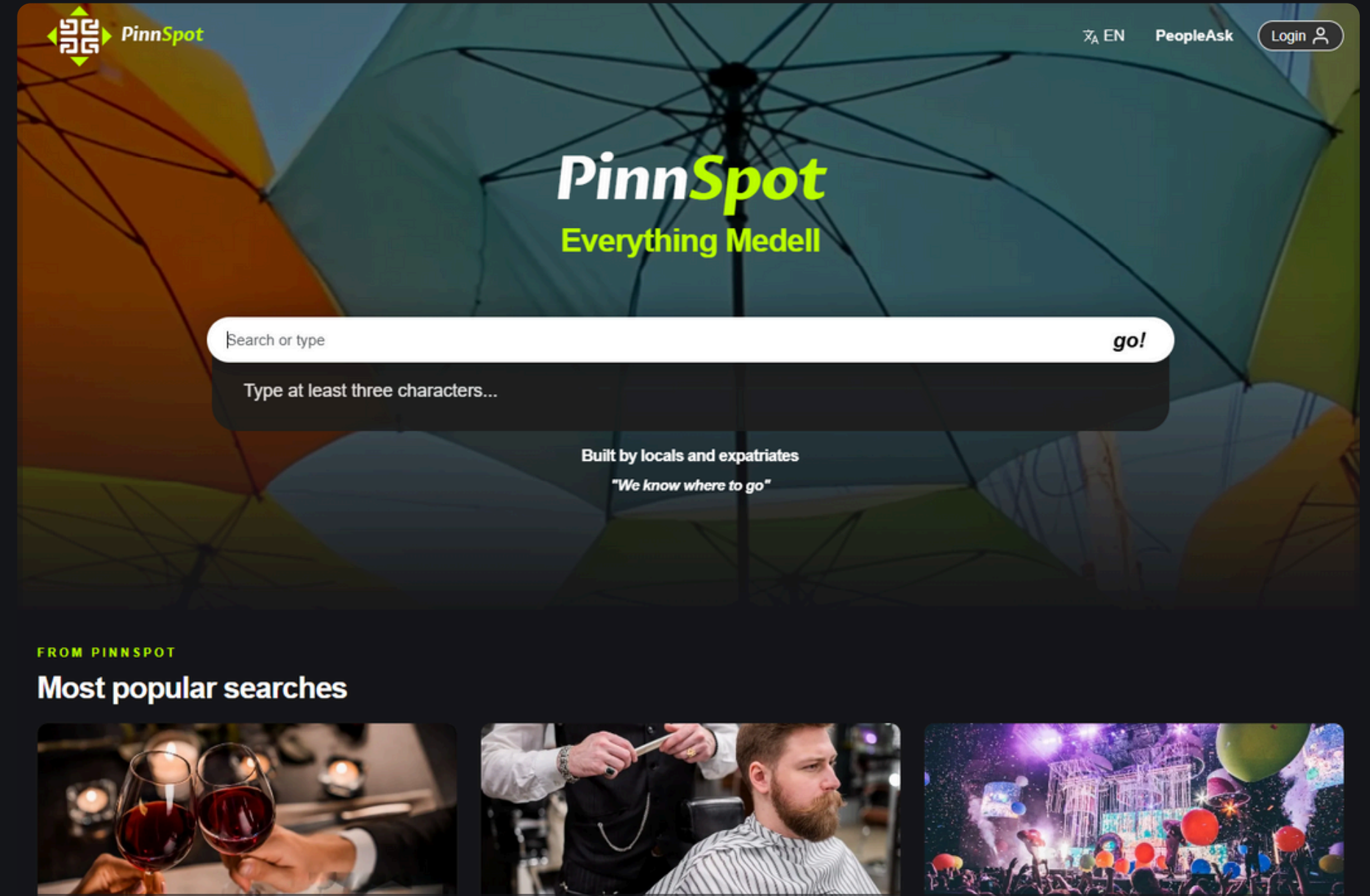


ORIGEN

Nace de la idea de crear una plataforma interactiva para los turistas con recomendaciones genuinas creadas por la comunidad local, expatriotas y nómadas

VISION

Crear una comunidad en línea donde se presenta al mundo los mejores lugares para visitar en Medellín. Ayudar en la creación de experiencias únicas para los visitantes, así como beneficiar a los lugares promovidos.



- **+300 Lugares registrados**
- **+130 categorías y características**
- **Actividades y tours**
- **Blogs y recomendaciones**

¿Cómo crear recomendaciones personalizadas para el usuario?

Todo lo que
necesitas
son DATOS!

Algoritmo
Machine
Learning

¡Recomendaciones
personalizadas!



Punto de Partida

yelp  data licensing



6,990,280

Reviews



150,346

Businesses



11

Metropolitan areas

- Contiene información en inglés de reviews de usuarios de su plataforma en más de 300.000 negocios comerciales en EE.UU

- Plataforma con más de 10 años de experiencia en el campo, con una gran comunidad que crea herramientas para mejorar el turismo y la calidad de los servicios

- **¿Cómo crear un sistema de recomendaciones para este dataset?**

- **¿Cómo ajustar o extrapolar el resultado a nuestra base de datos para PinnSpot?**

2. Planeación y limpieza

- Después de analizar el dataset, decidimos que los pasos a seguir serían:
 1. Limpieza del dataset.
 2. Creación de perfiles
 - a. Usuarios
 - b. Lugares.
 3. Creación del dataset con el puntaje de la review como target.
 4. Modelos y puntos de vista:
 - a. Aproximación lineal con SGD Regression
 - b. Aproximación con DNN con Keras
 5. Análisis de resultados
 6. Modelo definitivo como servicio en la nube (Inference API)
 7. Ajuste del modelo a nuestra estructura de datos.
 8. Implementación en la web.



- Los perfiles son un array que representan el peso normalizado de las diferentes categorías(tags) que tiene Yelp para determinado Usuario o Lugar

3.a Creacion del Perfil de usuarios

$$\frac{\sum_i (r_{u,i} - \mu_u)}{Count_u(c)} + \mu_{global}$$

Adjusted Rating:

se utiliza para ajustar las calificaciones en funcion de su promedio personal y un promedio global

$$\frac{Count_u(c)}{Count_u(c) + C}$$

Confidence:

Se establece un puntaje de confianza que considera la cantidad de reseñas realizadas por el usuario.

TF-IDF (Term frequency-inverse document frequency)

Este enfoque penaliza las categorías comunes y enfatiza las menos frecuentes, puesto que generan mayor interes en el usuario

$$TF_u(c) \times IDF(c)$$

Influencia Social

Se evaluan las preferencias de los amigos y se le da un valor fijo de 30%

$$\alpha \times \frac{\sum_f W_f(c) \times Similarity(u, f)}{\sum_f Similarity(u, f)}$$

$$W_{c,u} = AdjustedRating_{user}(c) \times Confidence_{user}(c) \times TF - IDF_{user}(c) + SocialBoost_{user}(c)$$

3.b Creacion del Perfil de lugares

$$TF_b(c) \times IDF(c)$$

Especificidad Categorías:

Este enfoque penaliza las categorías comunes y enfatiza las menos frecuentes, puesto que generan mayor interes en el usuario.

$$\frac{\sum (UserRating \times UserConfidence) \alpha \mu}{\sum UserConfidence + \alpha}$$

Bayesian-Adjusted (Credibilidad):

Se implementa un sistema de calificación ajustada que considera la confianza del usuario en sus reseñas

Este ajuste se basa en la calificacion promedio del usuario y el promedio global.

$$\log(1 + ReviewCount)$$

Escala de popularidad

Se aplica una escala de popularidad que utiliza un enfoque logaritmico sobre el conteo de reseñas.

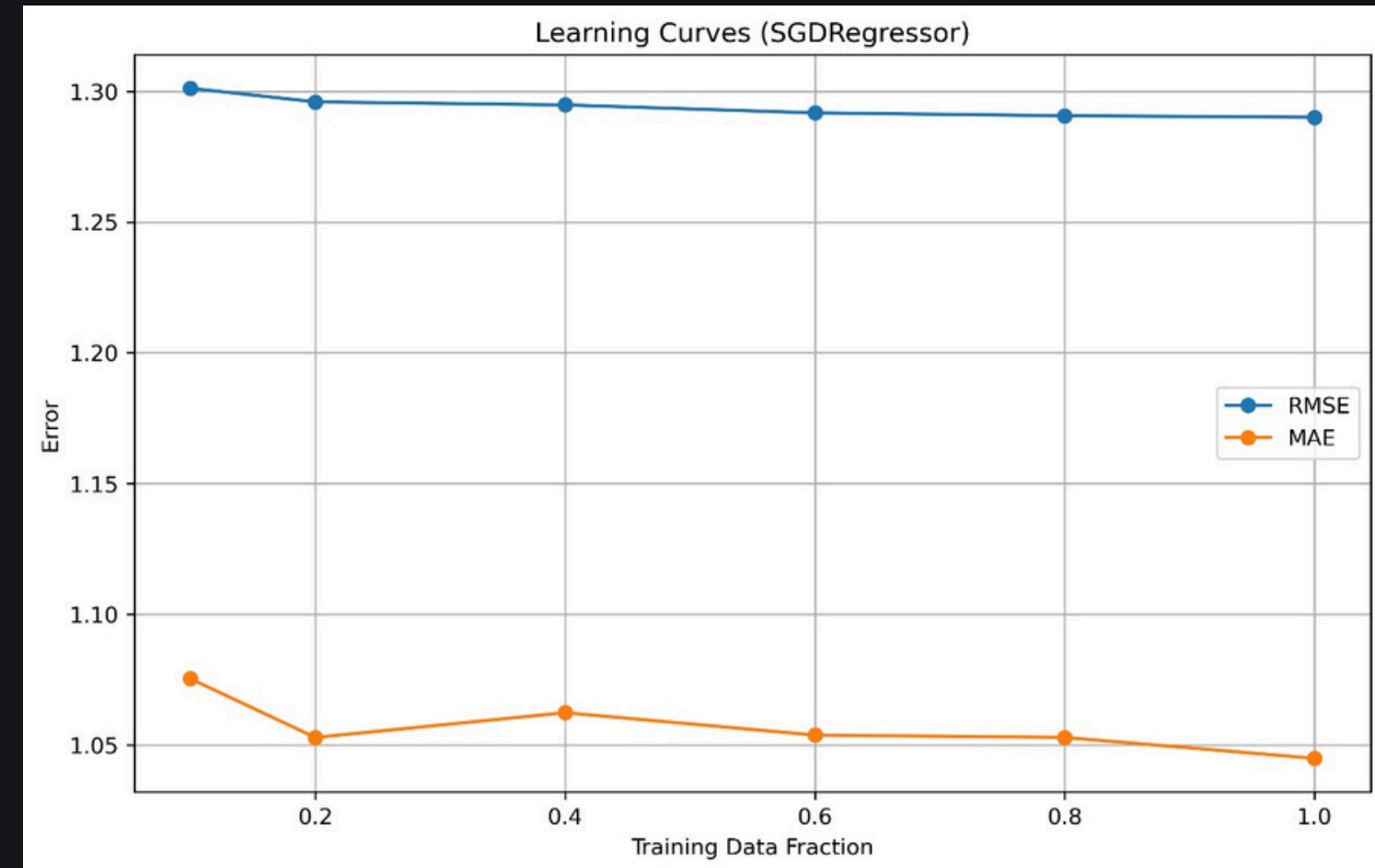
Esto ayuda a moderar la influencia de los lugares con un gran numero de reseñas

Permite una evaluación mas equilibrada de todos los destinos turísticos

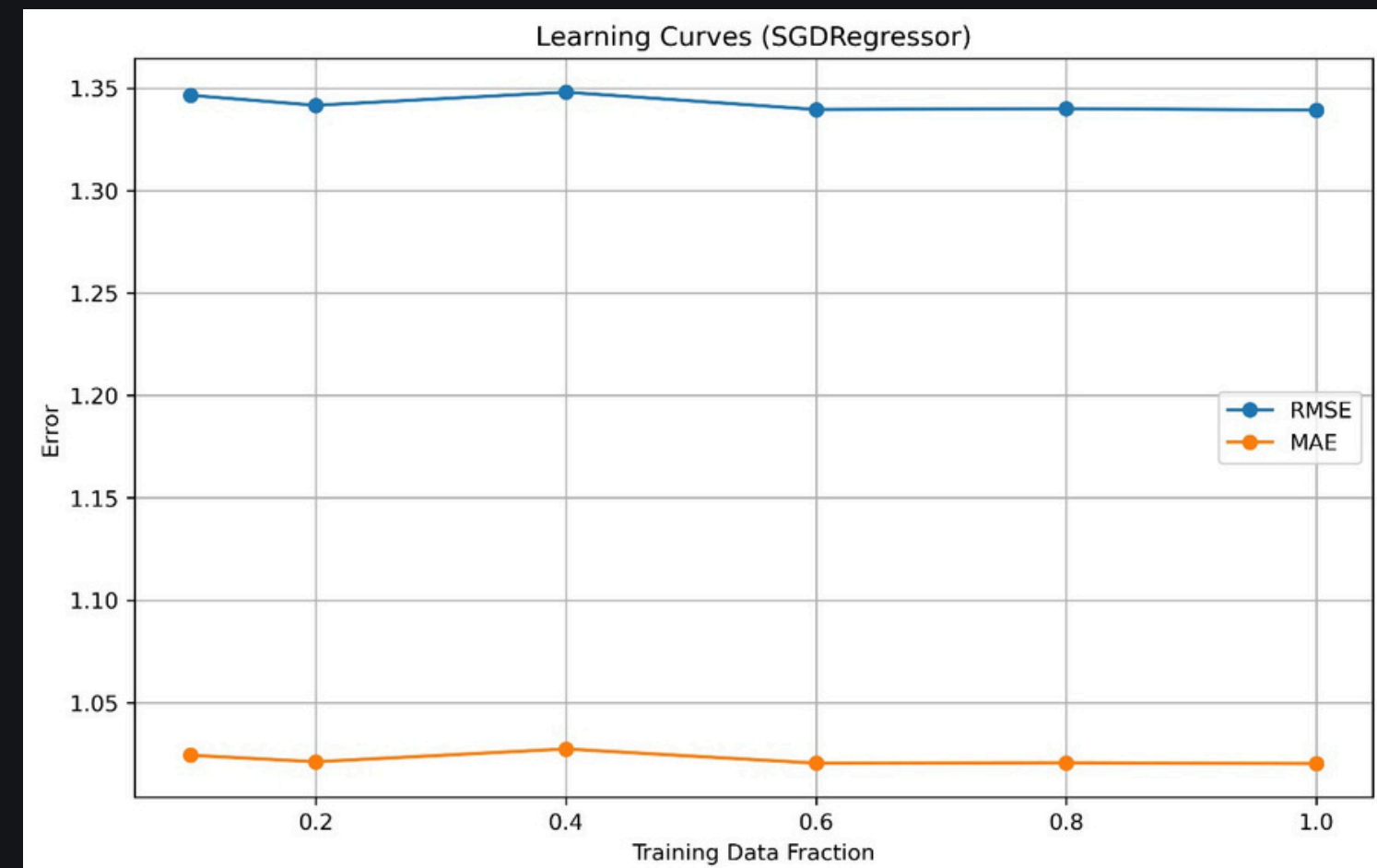
$$W_{c,b} = TF - IDF(c) \times \left(\frac{\sum (UserRating \times UserConfidence) + \alpha \mu}{\sum UserConfidence + \alpha} \right) \times \log(1 + ReviewCount)$$

4.a Aproximación con SGDRegressor

Penalty	l2
Alpha	0.0001
Max Iterations	1000
Loss	squared error
Learning rate	invscaling



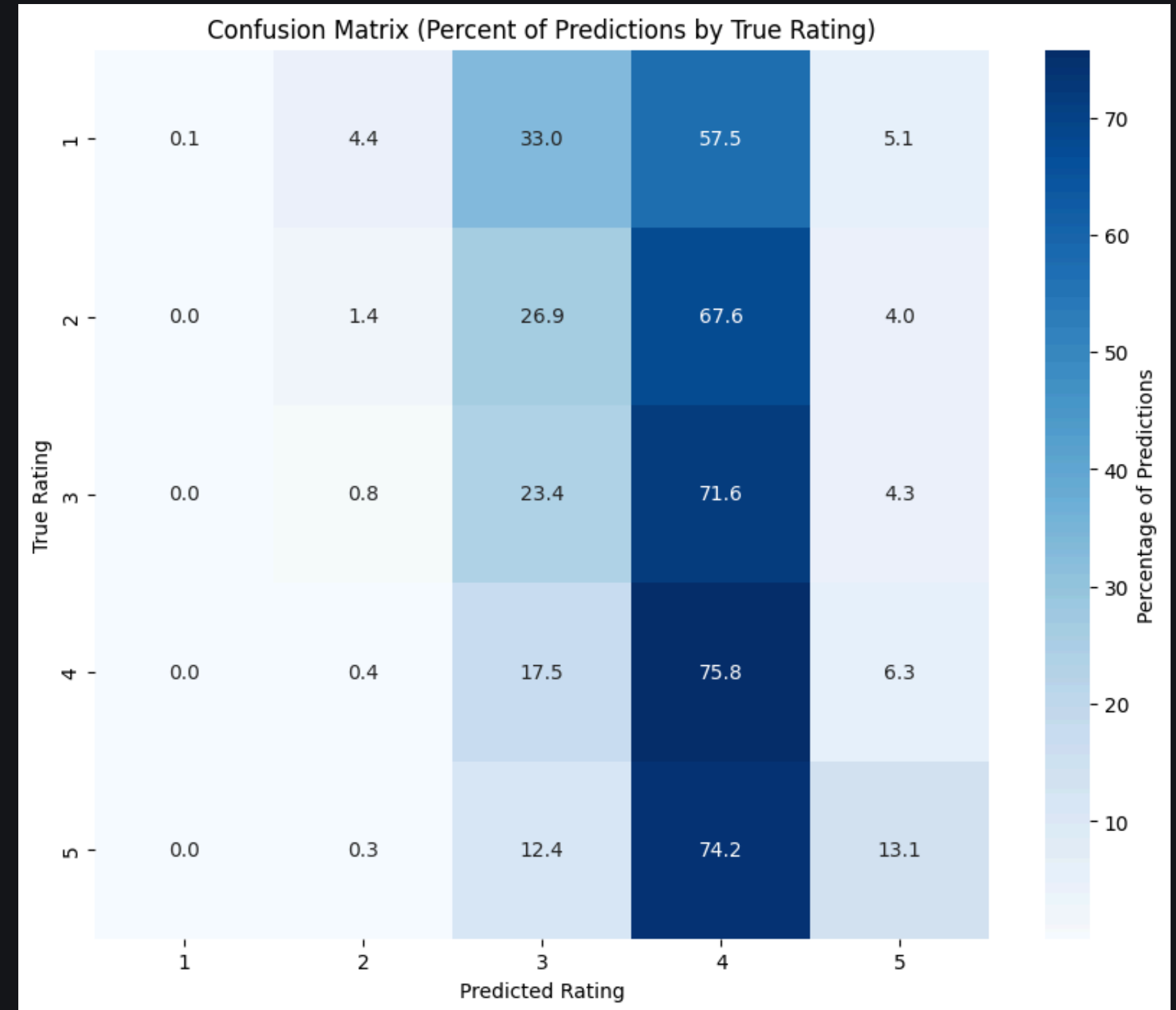
Penalty	elasticnet
Alpha	0.001
Max Iterations	1000
Loss	Huber
eta0	0.01



4.b Aproximación con DNN y Keras

Arquitectura probada

Epocas	5
Capas densas	4
Neuronas	512, 256, 128, 1
Funciones de activacion	[relu], lineal
Optimizador	Adam



5 Análisis de resultados

Metrica	SGDRegressor	DNN (Keras)	Interpretación
MAE	1.05 stars	1.03 stars	Ambos modelos tienen problemas con predecir lugares con un promedio de 1 estrella
RMSE	1.29 stars	1.29 stars	Rendimiento muy similar. Errores grandes penalizados en ambos modelos
R ² Score	Not Provided	0.055	Bajo, modelo solo explica ~5.5% de la variación del puntaje
Within ±0.5 Stars	Not Provided	28.7%	Solo el 28.7% de las predicciones son altamente precisas
Within ±1.0 Stars	~56% (estimated)	56.0%	La mitad de las predicciones tienen errores 'aceptables' (≤1 estrella)
Within ±1.5 Stars	~78% (estimated)	78.4%	Most predictions are within 1.5 stars of true ratings for both models.
Precision (Avg)	N/A	22.7%	Baja precisión, en especial para ratings 1-3.

Implementación futura

- **Formulario en la web: ¿ Cuáles son los intereses del usuario?**
 - Preguntar por categorías de intereses
 - Preguntar por lugares de su interés
- **Cruce de categorías entre Yelp y Pinnsport**
- **Organizar recomendaciones según otros parámetros:**
 - Distancia
 - Interés actual
 - Rating promedio del lugar
 - Clima

Conclusiones

La integración de técnicas de machine learning en sistemas de recomendación turística ha demostrado ser una herramienta poderosa para personalizar la experiencia del viajero. Aunque los resultados pueden presentar un margen de error considerable, este se minimiza al enfocarse en sugerir un conjunto de destinos de interés para el usuario, optimizando así tanto la planificación como la ejecución del viaje.

thank you