



# Amazon Book Recommender System

---

Project Overview and Insights

By Kyle Rodriguez

# Table of Contents

<b>Intro to Recommender Systems</b>	<b>03</b>
<b>Project Goal</b>	<b>04</b>
<b>Data Overview &amp; Wrangling</b>	<b>05</b>
<b>Modeling &amp; Evaluation</b>	<b>06</b>
<b>Model Optimization</b>	<b>07</b>
<b>Final Model Selection</b>	<b>08</b>
<b>Conclusion &amp; Next Steps</b>	<b>09</b>

# INTRODUCTION TO RECOMMENDER SYSTEMS

## The Role of Recommender Systems

---

- Widely used in user-facing industries such as Netflix, Amazon, LinkedIn and YouTube.
- Helps manage large volumes of content by providing relevant suggestions to users.
- For example, over 70% of watch time on YouTube is driven by its recommender algorithm.



# Project Goal

---

## Objective

- Develop a recommender system for Amazon book shoppers
- Recommend books users are likely to enjoy based on the reviews of others.
- Encourage users to discover books they might not find on their own

## Problem Statement

*How can we accurately recommend 5 new books to a user using an item-item approach to encourage expanding user preferences to books they may not have otherwise found on their own?*

# Data Overview & Wrangling

---

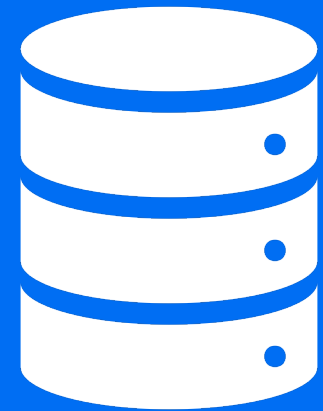
## Dataset

Two datasets:

- the first consisting of >51M rows containing user ratings on books
- the second containing metadata for each book in the initial dataset.

## Wrangling

- Null values were dropped from the dataset (13%).
- Data was downsampled to about 32K rows to manage computational resources.



# Modeling & Evaluation

---

## Baseline Models

Several algorithms were tested including SVD, NMF, and KNN-inspired for best baseline results.

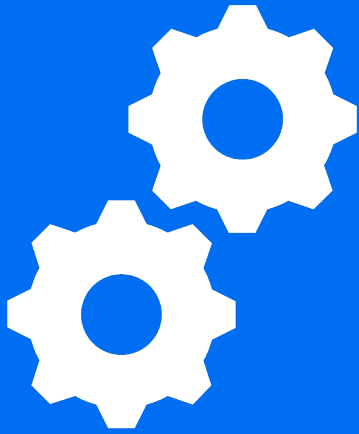
## Performance Metrics

Models were evaluated for accuracy using MAE and RMSE.

## Top Performing Models

The SVD and KNNBaseline models had the best MAE and RMSE scores from the ones tested.





# Model Optimization

---

## **GridSearch CV**

An exhaustive search was done to determine the best hyperparameters for both models.

## **SVD Model Tuning**

Optimized number of epochs, learning rate and regularization parameter.

## **KNN Model Tuning**

Options for similarity measures like MSD and Cosine were optimized as well as the number of neighbors,  $k$ .



# Final Model Selection

---

## Chosen Model

The tuned SVD model was selected as a basis for the recommendation system since it has the best results.

It was chosen over the KNN model since those tend to become more computationally expensive as the size of data grows.

## Predict Top-N Books for Users

Using the recommender system, the top 5 books were predicted for a random user. The metadata was used to provide useful information on each book that was predicted.



	title	price	description	books	predicted_ratings
0	Army Field Manual FM 22-100 (The U.S. Army Lea...	Price Not Available.	No Description Available.	1420928244	5.0
1	The Essences	\$23.99	Writing this book was a very fun process for m...	1514491877	5.0
2	Winning the Staffing Sales Game: The Definitiv...	\$19.99	No Description Available.	1543461468	5.0
3	Wolfbaene	\$7.81	My name is Michelle Dennis. I live in Ellenbro...	1849633436	5.0

Figure: A pandas DataFrame showing an example of the top 4 recommended books for a user along with the system's predicted user ratings and metadata about each book.



# Conclusion and Next Steps

---

## Project Outcomes

- Successfully predict the top 5 recommended books for a user using underlying collaborative filtering algorithm.
- The chosen algorithm was the SVD which had an MAE of 0.71 stars.

## Next Steps

- This system may not scale well as the dataset increases and may need to switch to a more robust recommender system library for handling.
- Design an API endpoint for user testing and interfacing via a Flask web app.