



Seattle Public Library

James Cameron, Pradyut Karmakar, Kaitlynn Lilly, Danielle Reagan, and Josephina Wright

Introduction

Goal: develop predictive insights into library book checkouts using machine learning and time series analysis

Questions:

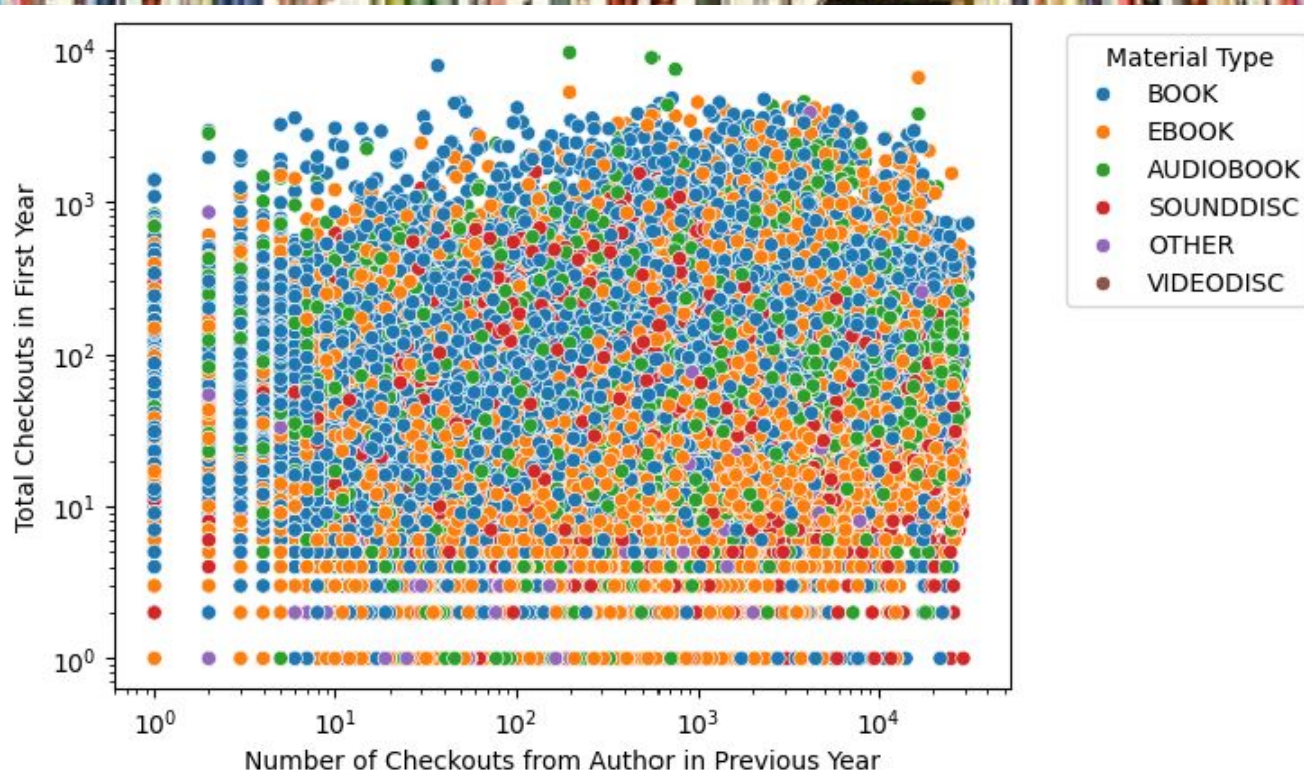
- 1) Can we predict the number of checkouts in the first year for a new book based on a number of features?
- 2) Can we forecast future checkouts over several months by analyzing past checkout patterns?

Who is this helpful for? libraries, bookstores, and publishers

Data Cleaning

- Filtered out non-english titles and missing data
- Standardized author names and book titles
- Categorized books by genre using the subjects provided
- Collated the same books together
- Collected first-year checkouts

Data Exploration

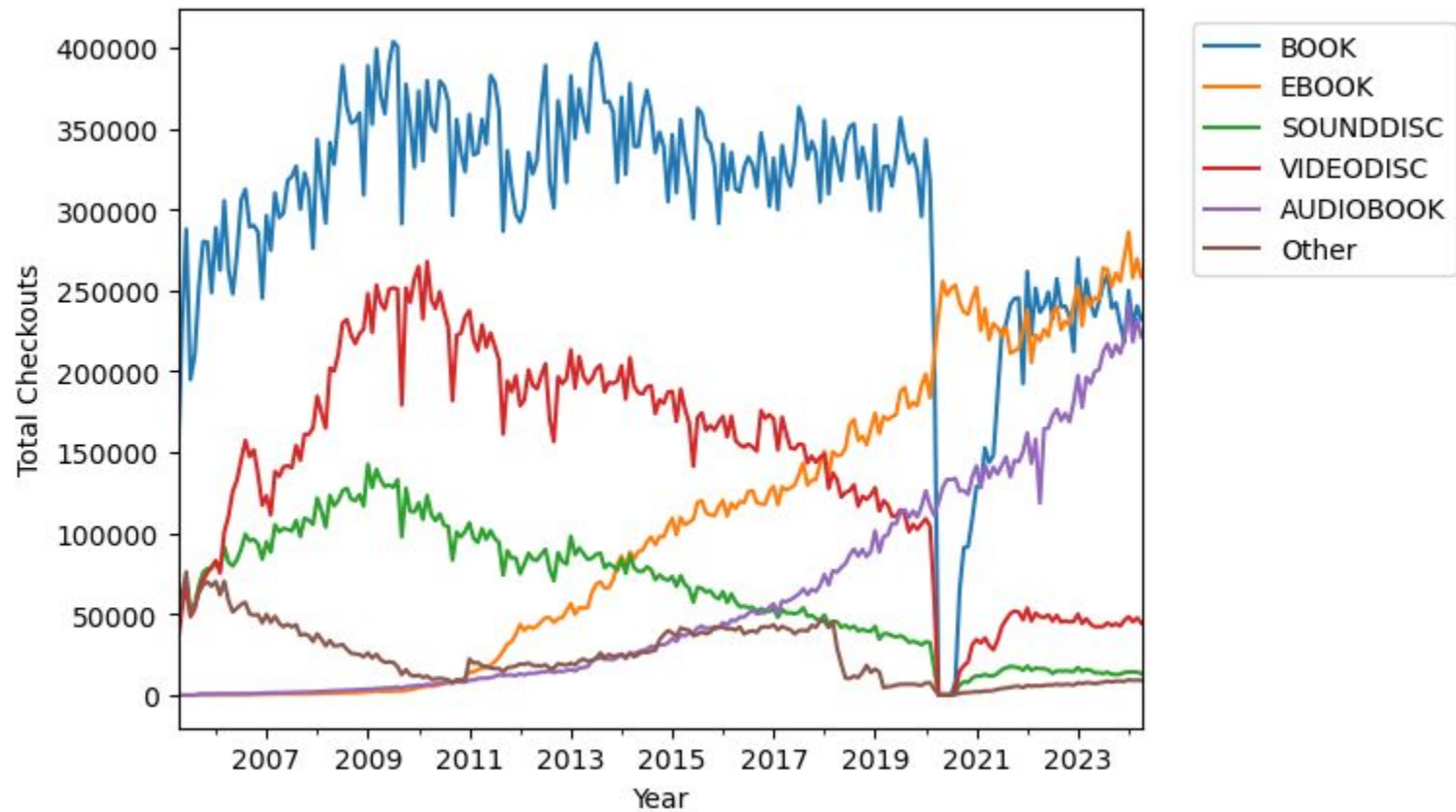


First Year Checkouts

Model	Baseline	Linear Regression	Lasso	Random Forest	Extra Trees	KNN	XGBoost	Neural Network
RMSE	115.980232	111.1058	111.1702	103.9528	108.4143	109.4348	101.8269	106.0051

Baseline RMSE:	119.1516
XGBoost RMSE:	115.6487

Time Series Aggregation



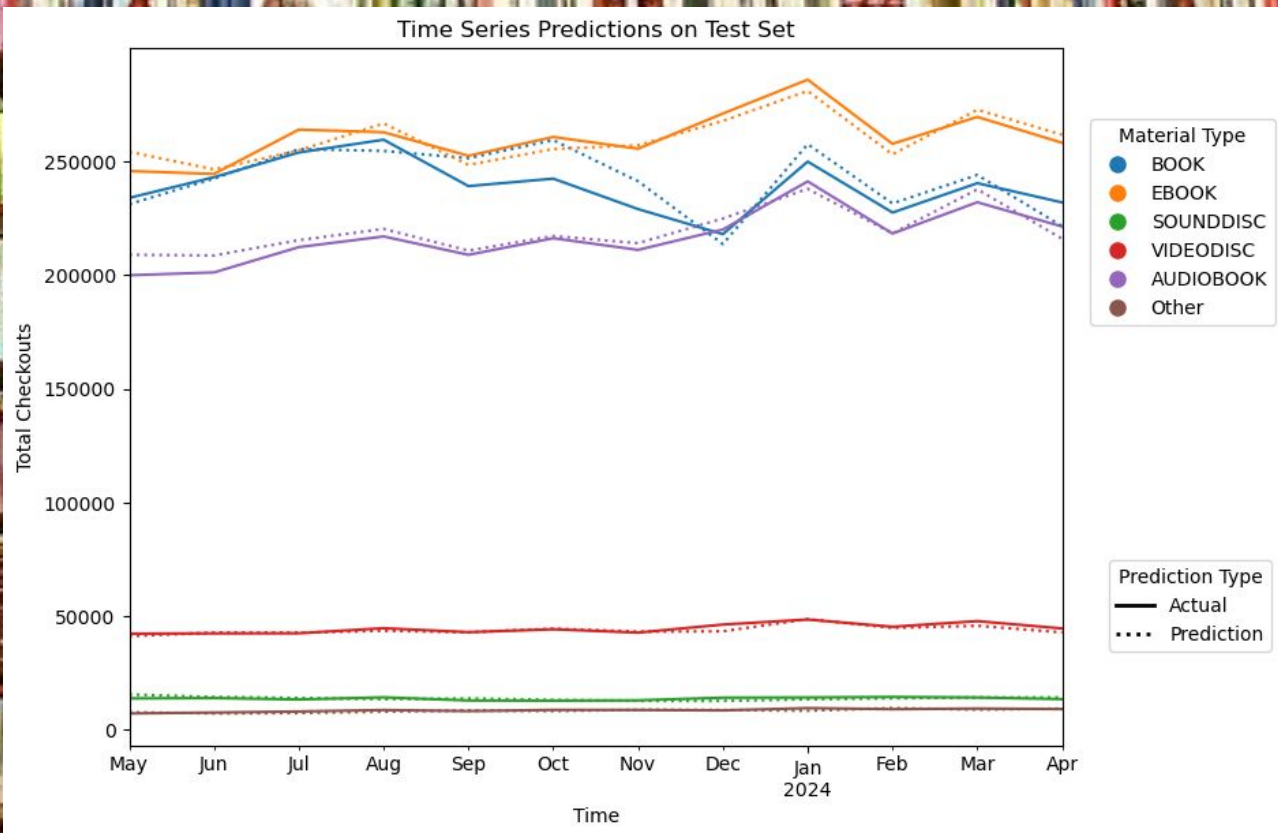
Time Series Models

Model	Books	Ebooks	Sounddisc	Videodisc	Audiobooks	Other
Baseline	0.080167	0.030745	0.184153	0.962654	0.278838	0.211655
Seasonal naive with trend	0.0893	0.064146	0.123656	0.129363	0.191755	0.287499
STL Forecasting	0.099766	0.022056	0.138596	0.155424	0.0517	0.093647
AutoARIMA	0.065118	0.036403	0.059609	0.018656	0.069619	0.07947
Seasonal AutoARIMA	0.058838	0.02226	0.069123	0.048841	0.056387	0.189952

Time Series Results

	Books	Ebooks	Audiobooks	Sounddisc	Videodisc	Other
Baseline Error	0.056188	0.051566	0.151631	0.713263	0.303994	0.293932
Model Error	0.02854	0.017212	0.018635	0.052908	0.028916	0.05525

Time Series Results





Thank you