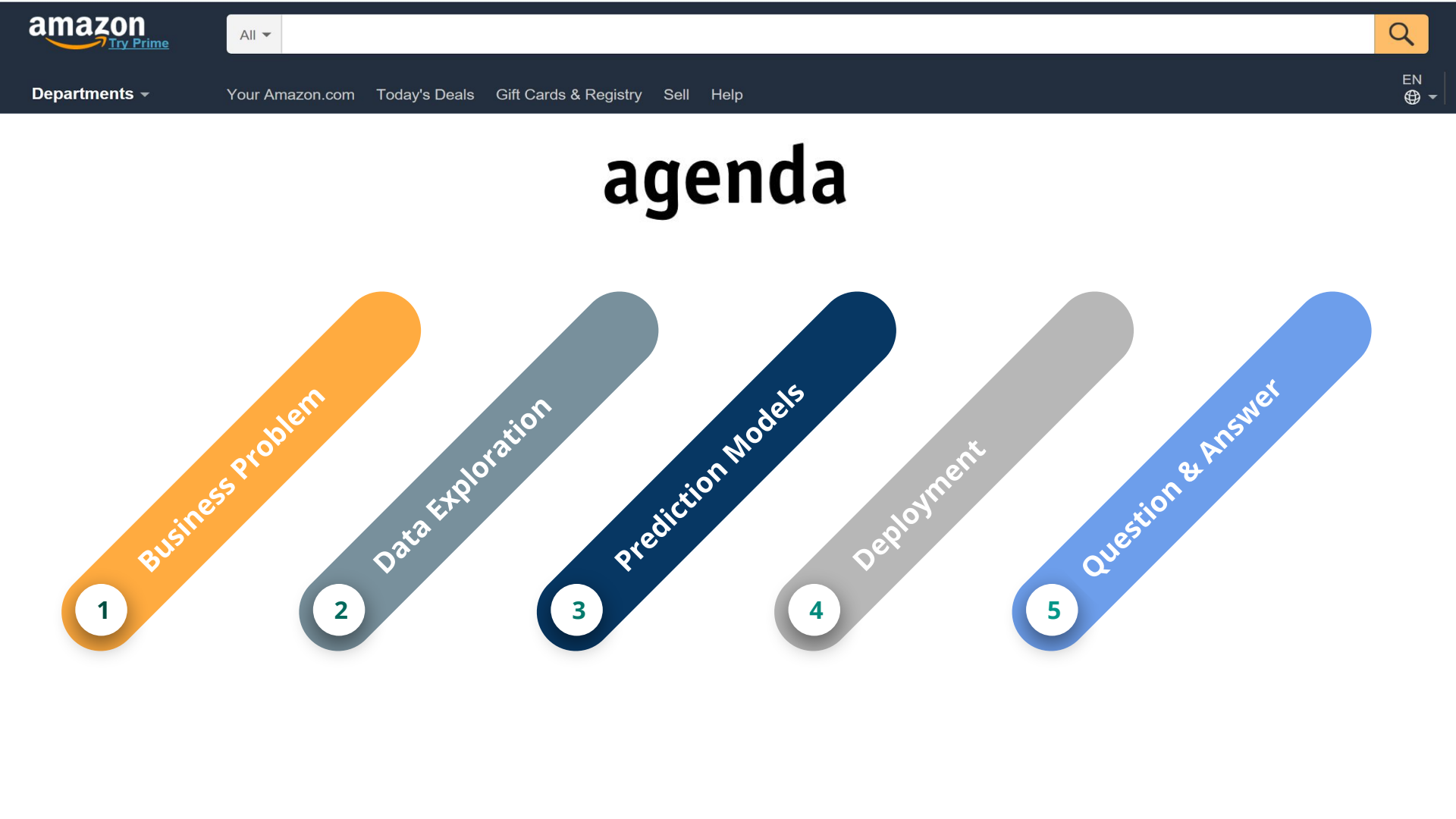
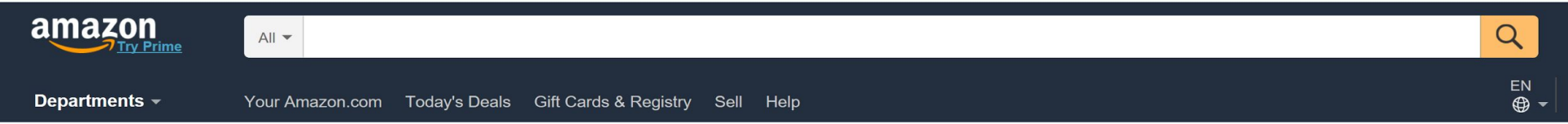
Will Amazon be interested in Elena's biography?

Athena Li
Lancy Mao
Elena Lopez



agenda

1. Business Problem
2. Data Exploration
3. Prediction Models
4. Deployment
5. Question & Answer

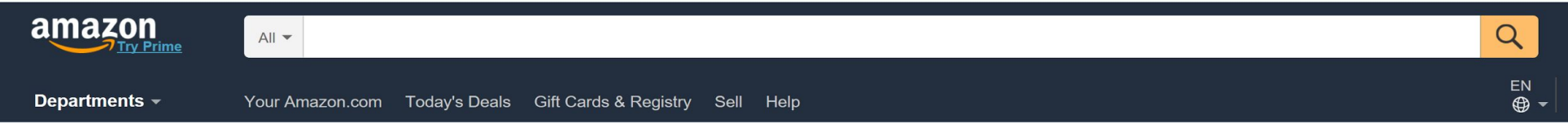


business background

Customer Reviews and Rankings

Add Business Value

Higher Revenue



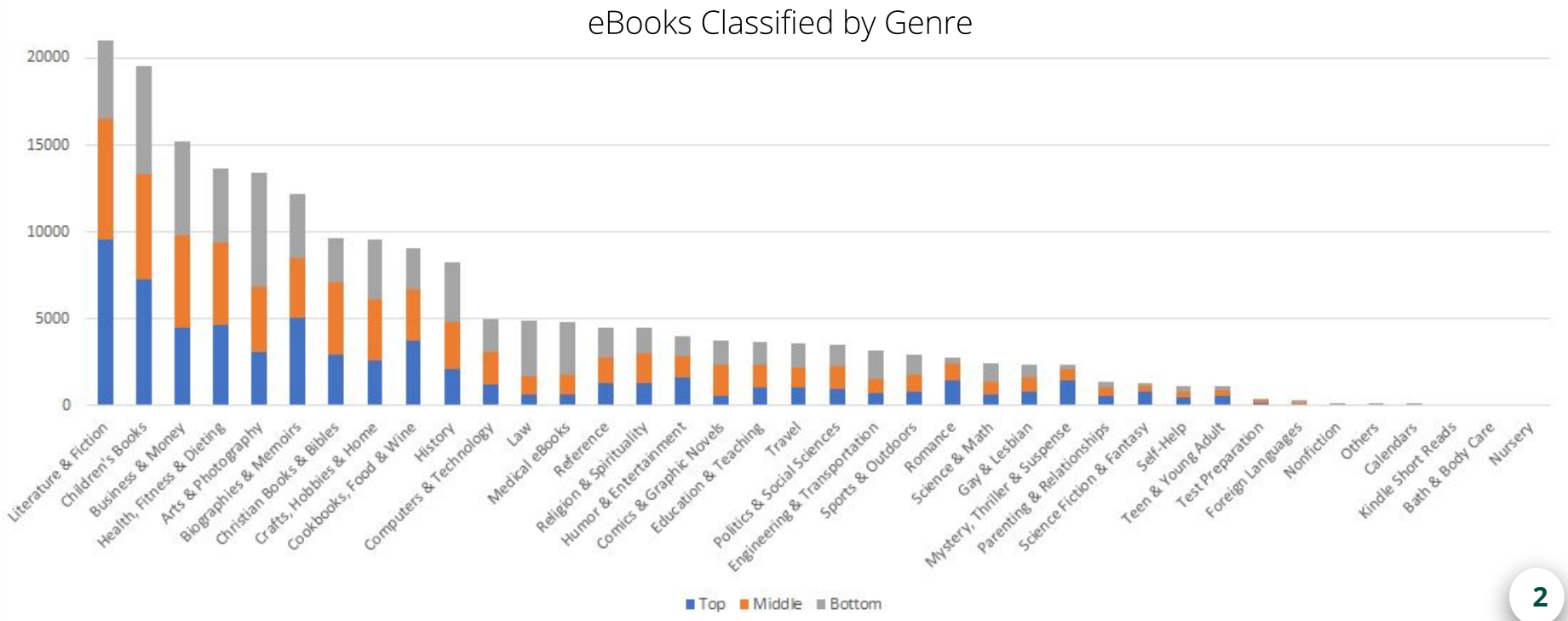
data exploration

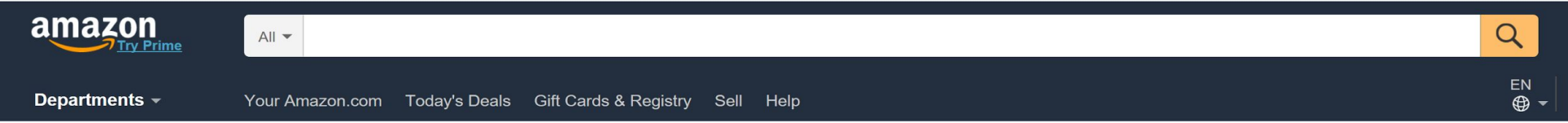
Attribute Selection → Cleansing → Classification → Filter

Sales Price	Preorder
Title	Date Published
Author	Categories (Genre)
Sales Rank	Digital Rights Management (DRM)
Unit Rank	Kindle Unlimited (KU)
Total Reviews	Type of Publisher
Average Rating	Page Count
Daily Units Sold	Daily Author Revenue

Sales Rank (high, medium, low)
Total Reviews
Average Rating
Categories (Genre)
Digital Rights Management (DRM)
Kindle Unlimited (KU)
Type of Publisher (Indie, Small/Medium, Amazon, Big 5)
Page Count
Sales Price

data exploration

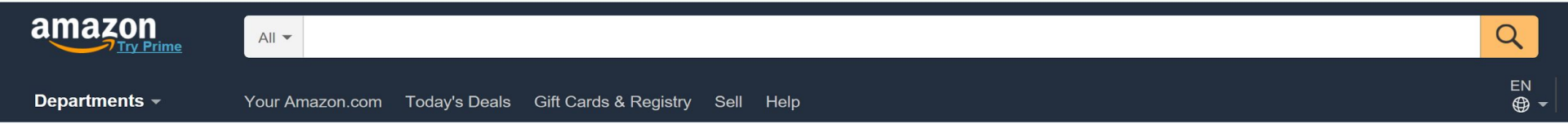




data exploration

Genres Ranked by Customer Ratings

	Highest	Lowest	Average
Top	Christian Books & Bibles (4.3)	Test Prep (2.2)	3.6
Middle	Christian Books & Bibles (4.2)	Foreign Language (1.0)	3.2
Bottom	Nonfiction (4.1)	Kindle Short Reads (0.2)	2.6



data exploration

Type of Publisher

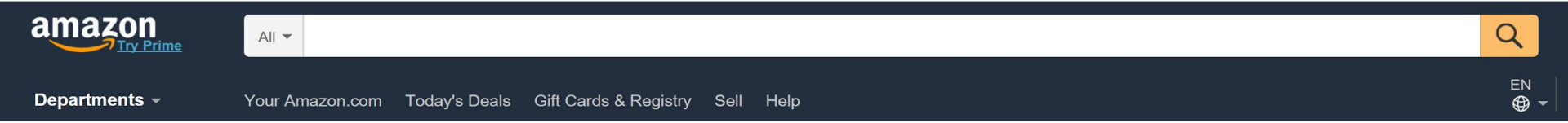
Rank split	*Indie Publisher*	*Small or Medium Publisher*	*Amazon Publisher*	*Big Five Publisher*
Top	38337	1354	1197	25077
Middle	15319	38969	1	582
Bottom	0	53477	0	0

22% of top ranked eBooks use an Indie Publisher

55% of all eBooks are published with a Small/Medium sized publisher

All low ranked books use a Small/Medium sized publisher

Only 0.6% of authors opt in to use Amazon's publishing service



key question

How to maximize Amazon's revenue and customer satisfaction from Kindle eBooks

Model 1: Revenue ranking prediction

- **Target Variable:** Amazon revenue ranking (high, medium, low)
- **Predictors:** Categories (genre), DRM, KU, Type of Publisher, Page Count

Model 2: Customer rating prediction

- **Target Variable:** Customer star rating (Six classes: 0, 1, 2, 3, 4, 5)
- **Predictors:** Categories (genre), DRM, KU, Type of Publisher, Page Count, Sales Price

amazon Try Prime

All

Departments Your Amazon.com Today's Deals Gift Cards & Registry Sell Help EN

model 1: revenue ranking

grid search

Find the optimal combination of parameter values for each model via grid search and 10-fold cross validation.

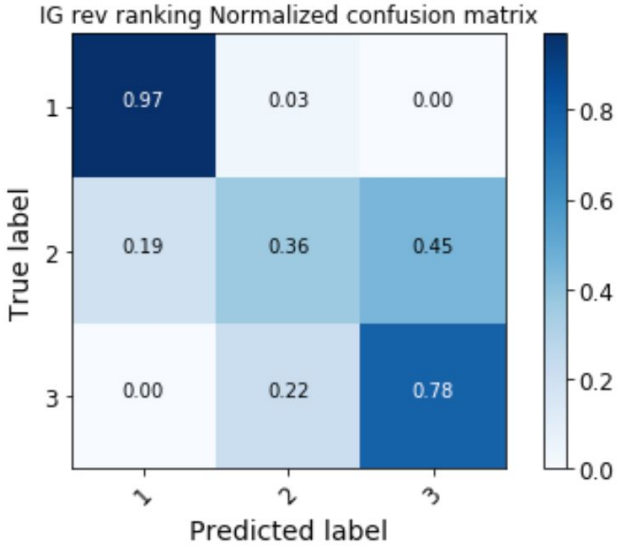
Decision Tree		Logistic Regression		KNN	
Criterion	entropy	**C**	0.1	**P**	2
Max depth	8	**Penalty**	l2	**Metric**	minkowski
Min samples leaf	88	**--**	--	**N neighbors**	147
Min samples split	2	**--**	--	**Weights**	uniform
Random state	0	**Random state**	0	**Random state**	0

model comparison

In-sample accuracy shows how well the models fits in training set, but we focus on out-of-sample accuracy which shows generalized performance.

Accuracy	Out-of-sample	In-sample
KNN (k=147)	0.706	0.711
Decision Tree (IG)	0.707	0.709
Logistic Regression	0.686	0.683

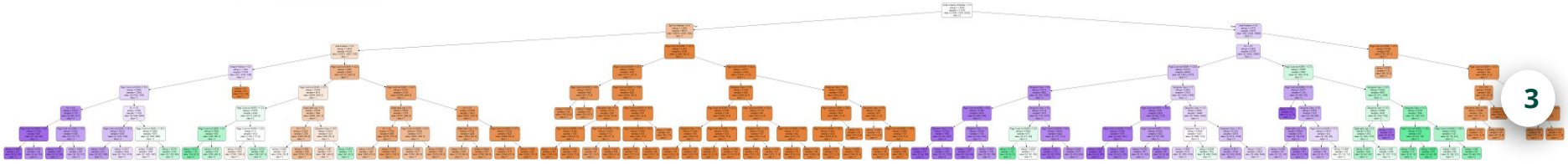
decision tree model

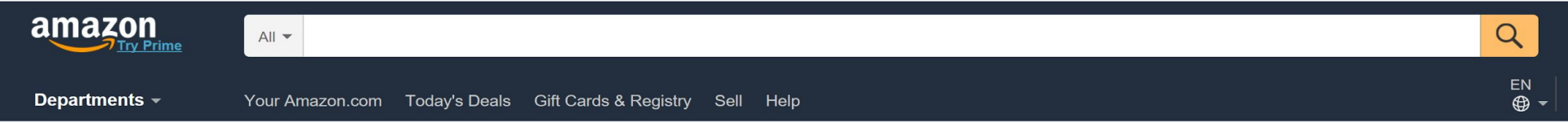


Classification Report

Good performance of top and bottom ranking classes.

Ranking	Precision	Recall	F1 score
Top	**84%**	**97%**	**90%**
Medium	58%	36%	45%
Bottom	**65%**	**77%**	**70%**



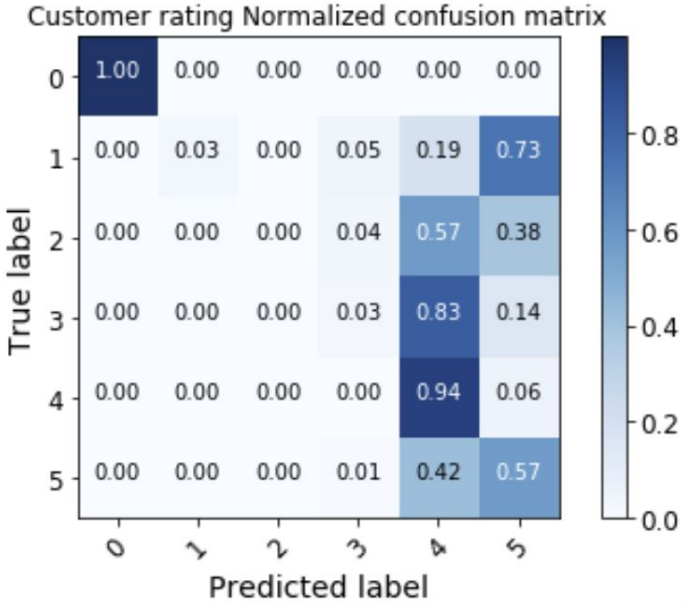


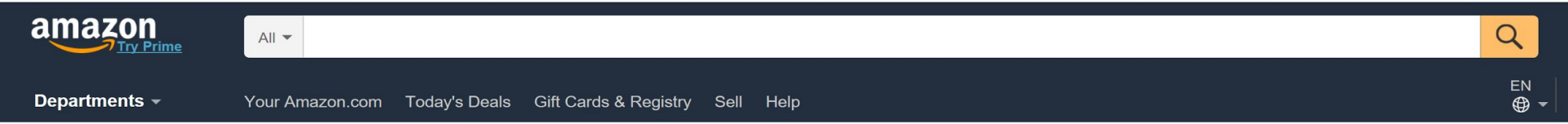
model 2: customer ratings

comparison and interpretation

Accuracy rate of each model is high, especially Decision tree. However, confusion matrix tells us the prediction is unreliable. Most of our predicted label is “4”, even though their actual label is “2” or “3”. So we will not use this model.

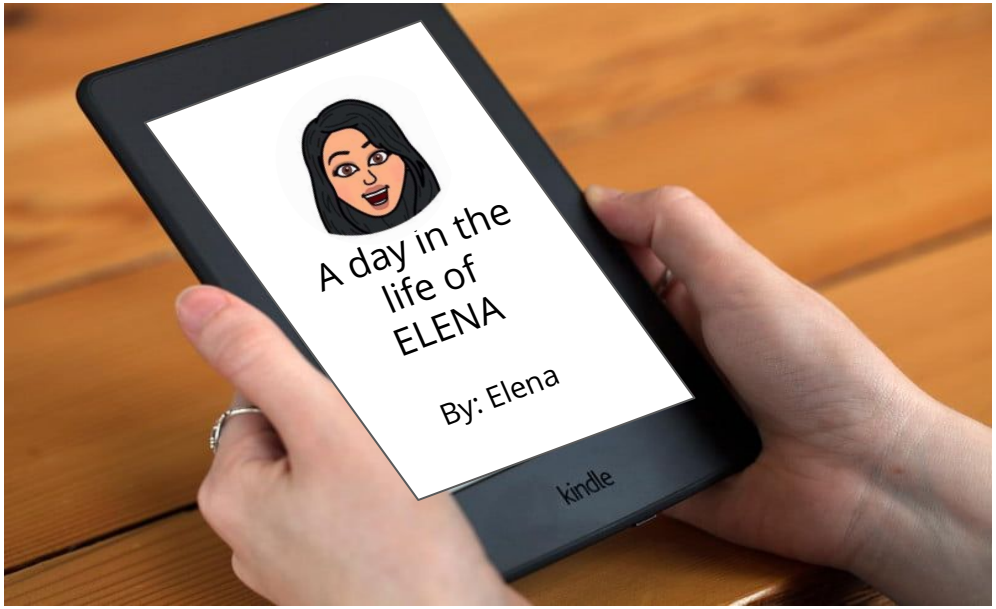
Accuracy	Out-of-sample	In-sample
KNN (k=535)	0.611	0.992
Decision Tree (Gini)	0.791	0.791
Logistic Regression	0.754	0.751





how will elena's biography perform?

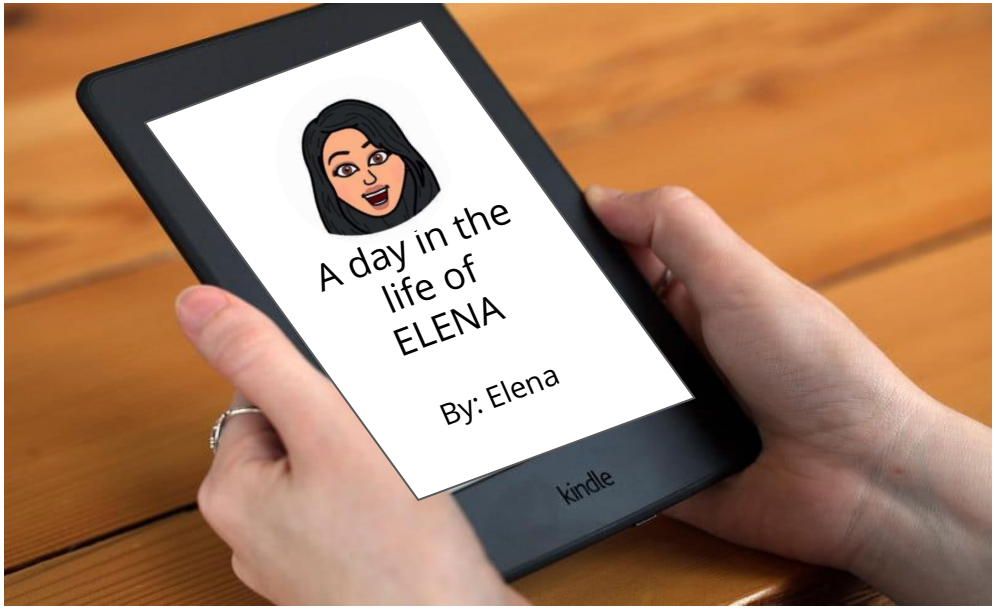
elena's biography!



Genre	Biography & Memoirs
Digital rights mgt	no
Kindle unlimited	no
Indie publisher	yes
Small or Medium publisher	no
Amazon Publisher	no
Big Five publisher	no
Page Count	20

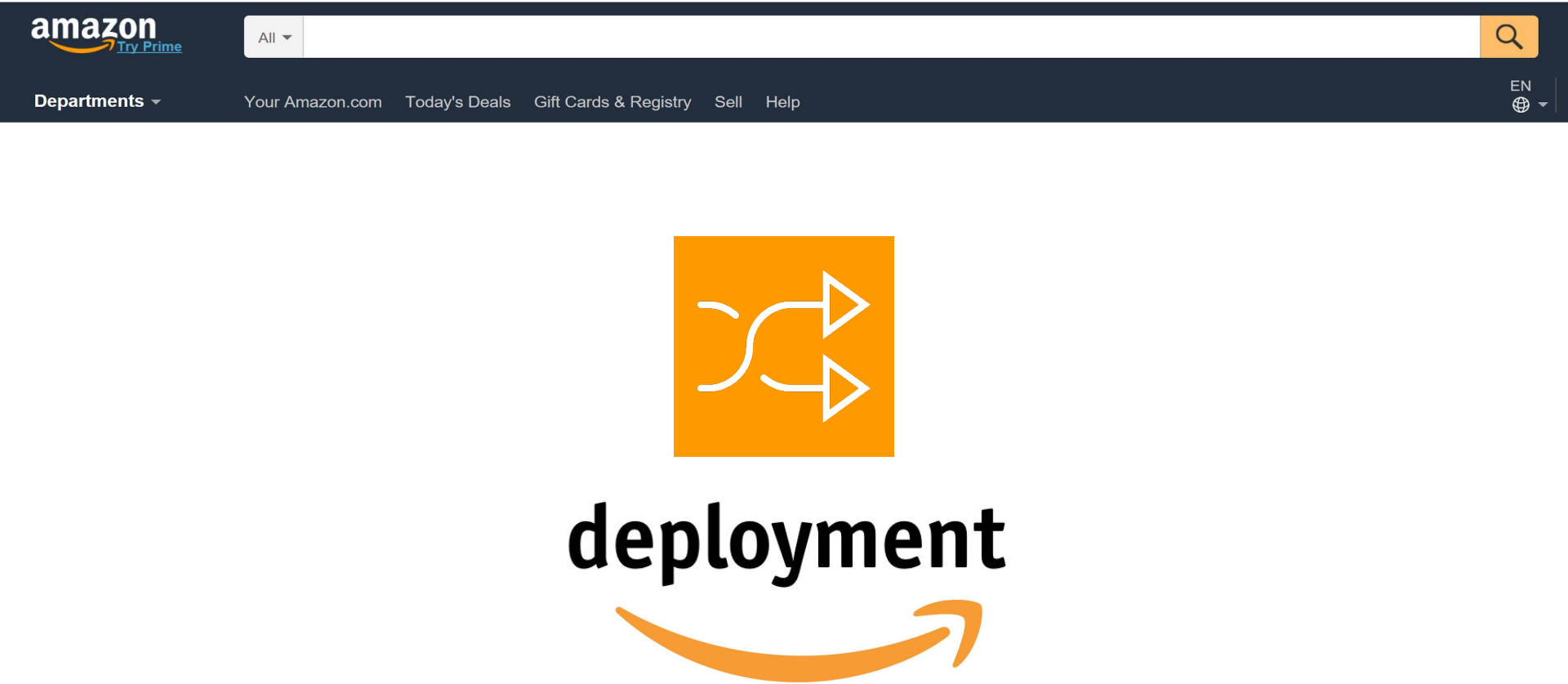
Amazon revenue ranking
class: 2 (Medium level)

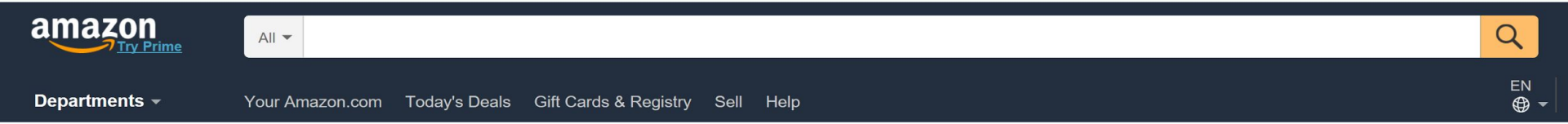
elena's biography!



Genre	Biography & Memoirs
Digital rights mgt	no
Kindle unlimited	no
Indie publisher	yes
Small or Medium publisher	no
Amazon Publisher	no
Big Five publisher	no
Page Count	**40**

Amazon revenue ranking
class: 1 (Top level)


amazon
Try Prime
All
Departments
Your Amazon.com
Today's Deals
Gift Cards & Registry
Sell
Help
EN
deployment



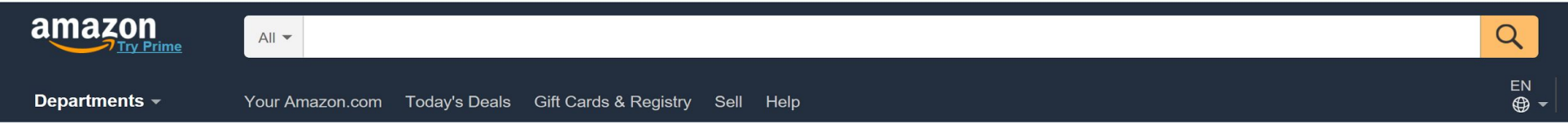
need to know

Both of the models are good at predicting books that will either be absolutely great or miserably bad. However, books that fall between the average may not be properly classified

Potential Problem:
We have very limited features available for training our model. We need more data features (authors, marketing efforts) in order to boost our accuracy

Implementation Strategies

- Filter books with the worse selling and rating potential
- Less effort on books with limited selling and rating potential
- More effort in promoting predicted top ranked books with high customer rating potential



ethics

Strong bias to big publishers and famous writers

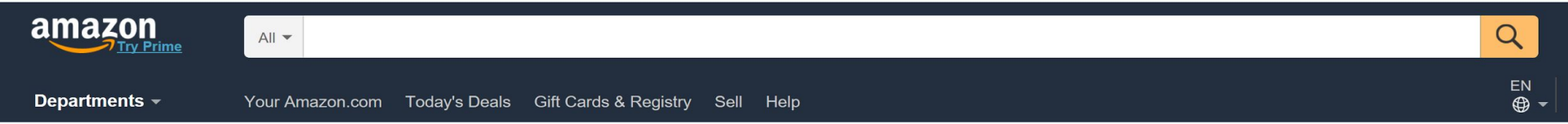
Young authors or indie publishers may lose opportunity to go big

Discriminating marketing strategy

Fairness v. profitability

May decrease incentives to write on diverse genres

E.g. More writes choose romance than literature



risks and mitigation

Type I Error

- You may reject some really good books
- Start to carry and offer the book when you predict it will be profitable

NGOs may target discriminating marketing strategy

- Negative social impact → negative impact on kindle sales → shares drop
- Work with NGOs and try to solve the problem
- Experienced public relations team

questions?


prime
entrega gratis y rápida
streaming de vídeo
prime
entrega gratis y rápida
streaming de vídeo
prime
entrega gratis y rápida
prime