

Dec. 12, 2022



CHRISTMAS GIFT RECOMMENDER

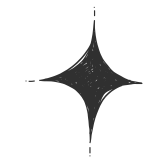
based on Amazon Bestsellers
in Toys and Games



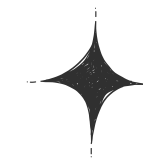
Varvara Frühmann
v.fruehmann@gmail.com



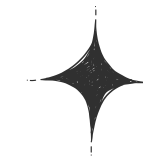
AGENDA



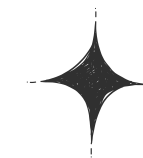
Introduction



Data engineering



NLP project



Recommender

WHY?

Inspired by the week 10 project



Sammy



International Families Berlin

Hi everyone,

May I pick your brain for a sec? Any ideas what we could give our 5 year old for Xmas? Somehow this year we've run out of ideas and I am totally out of the loop!

Many thanks in advance!

Data Engineering

Data Scraping



Scrapy

<https://github.com/scrapy>

Data Cleaning

✓ missing values

✓ deleted duplicates

✓ extracted 'category' and 'brand' from the title

✓ created age_range and price_range

Postgres DB

List of relations

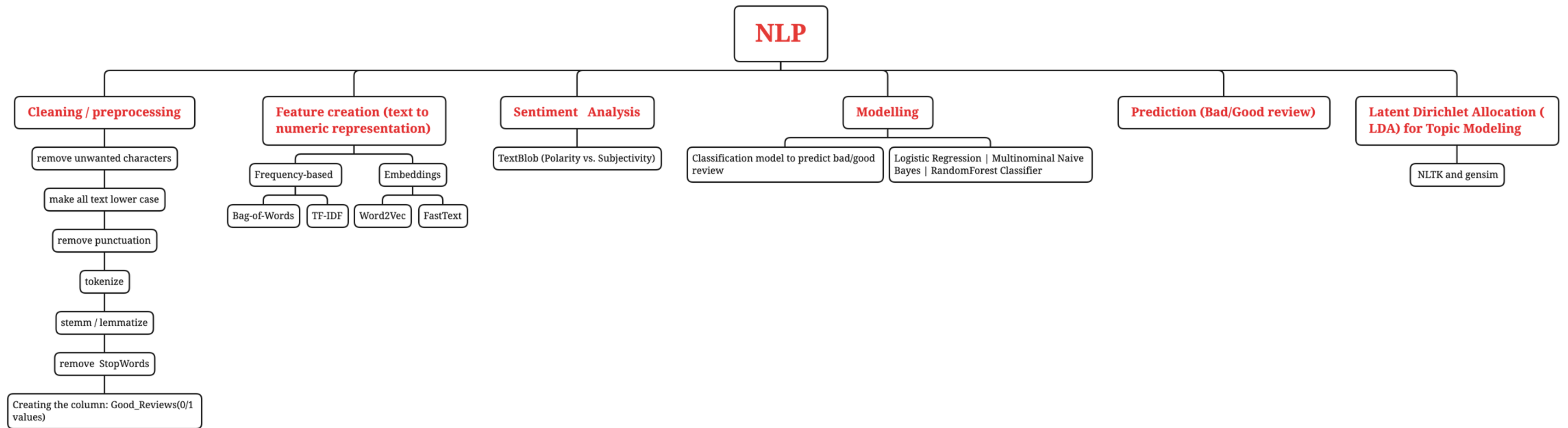
Schema	Name	Type	Owner
public	amazon_items	table	postgres

(1 row)

created amazon database

Streamlit





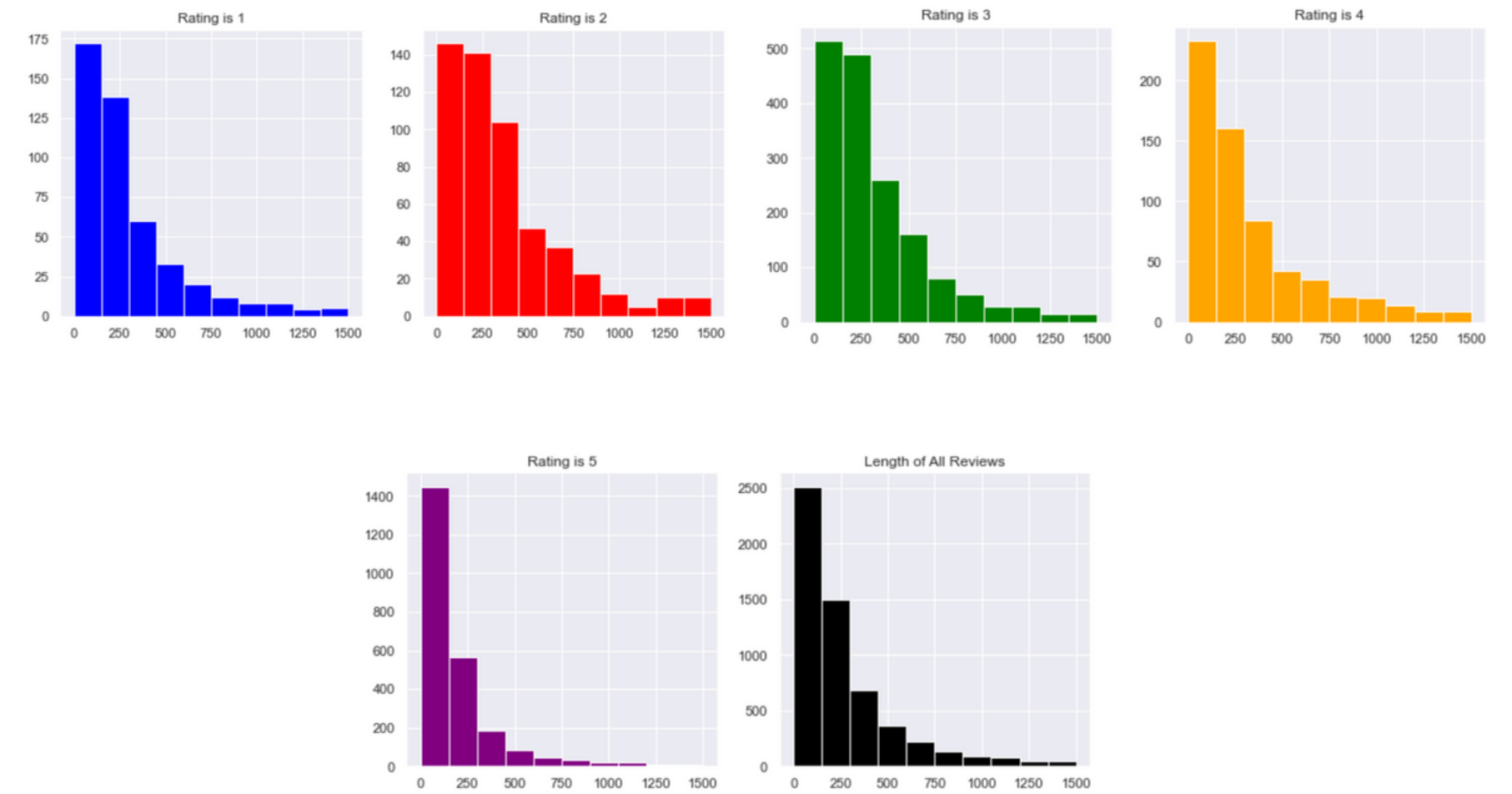
Natural Language
Analyses with NLTK

Streamlit





Number of characters in reviews



Streamlit

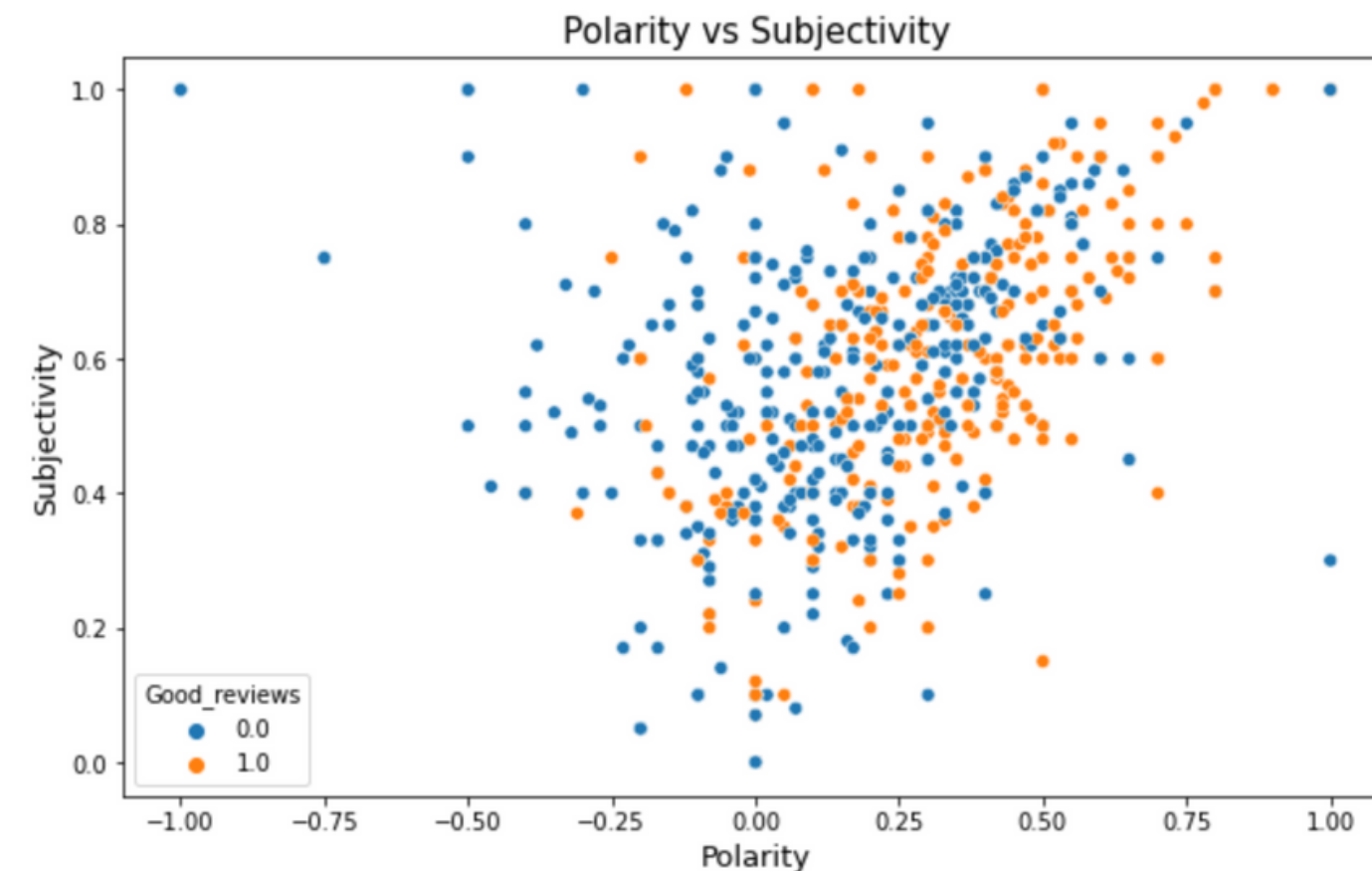
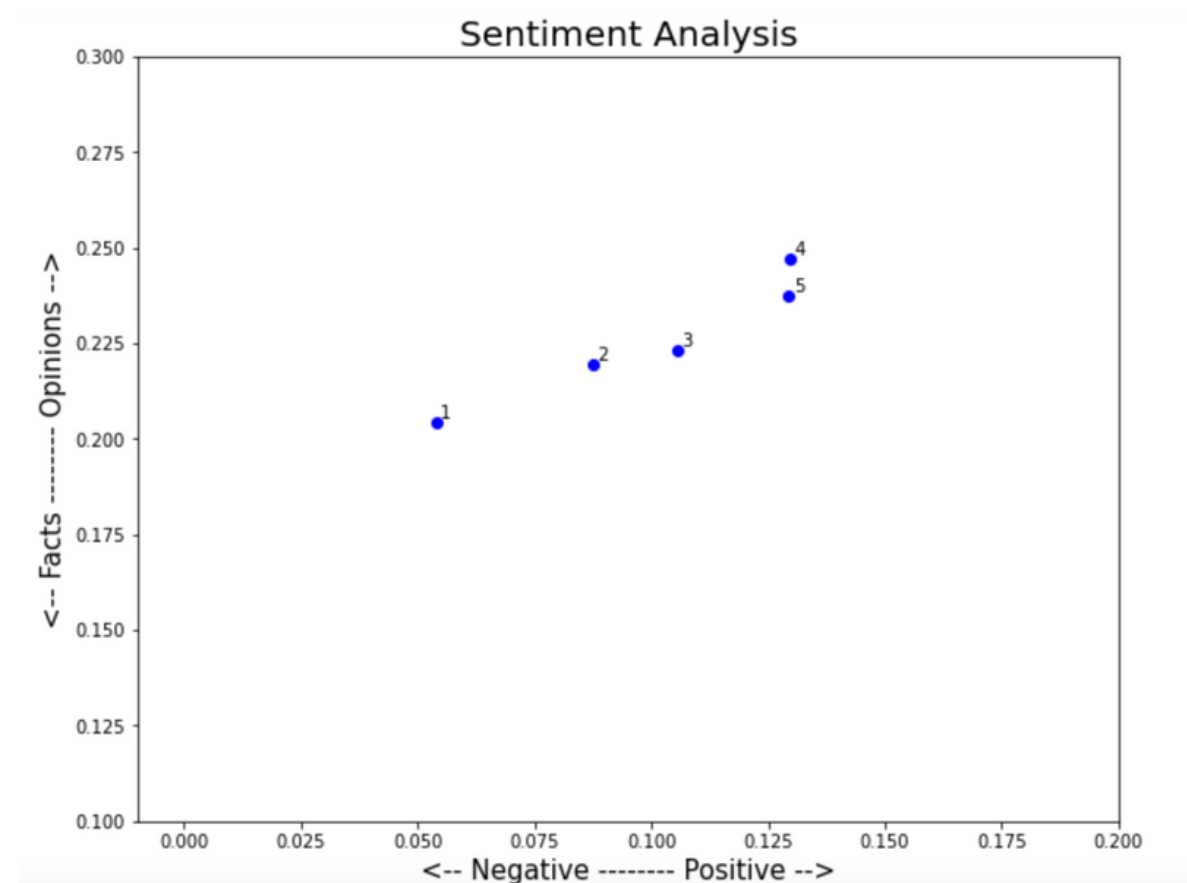


SENTIMENT ANALYSIS

Polarity

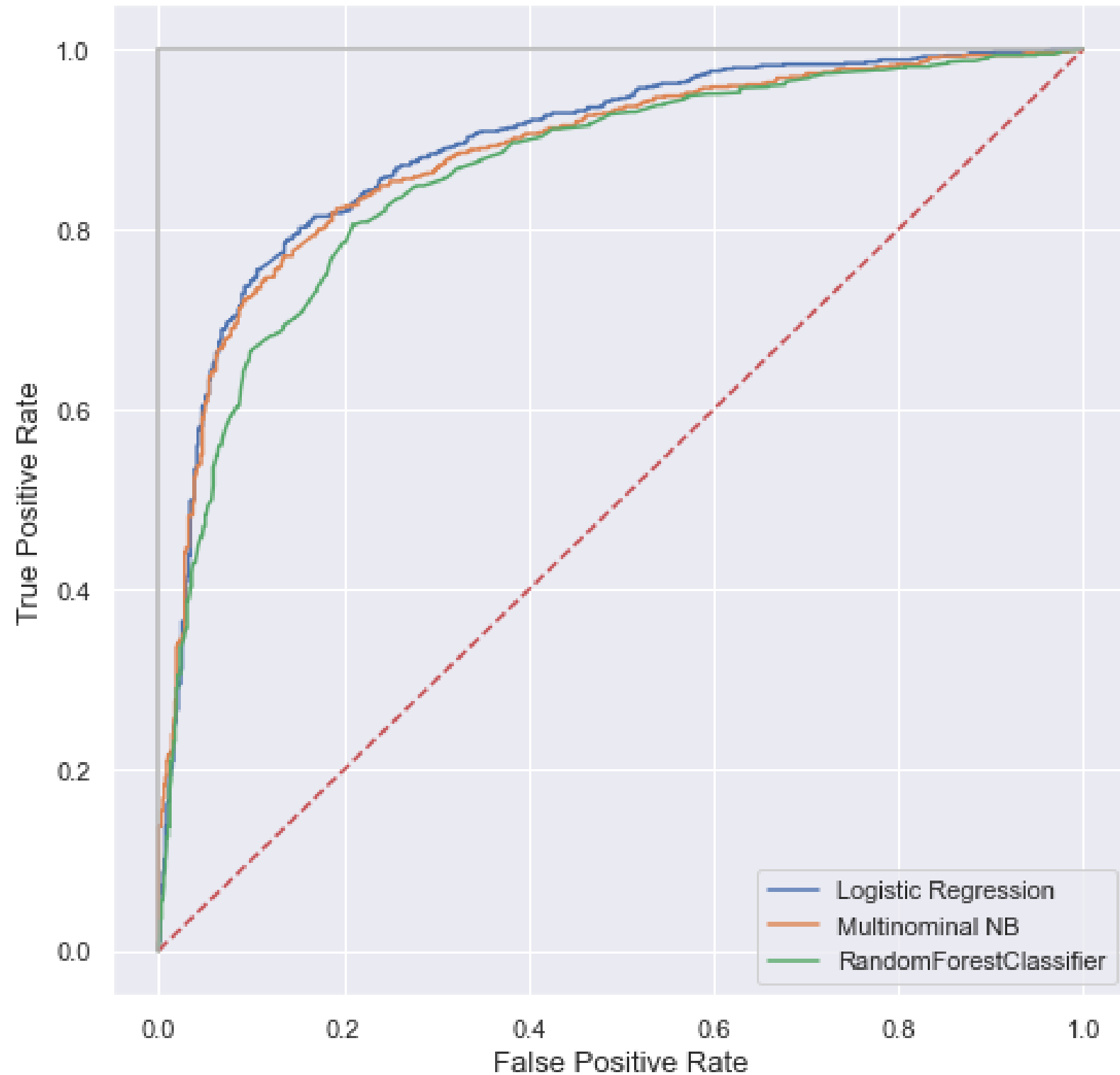
Polarity range of $[-1,1]$ where 1 means positive statement and -1 means a negative statement.

```
df.loc[(df["Good_reviews"] == 0) & (df.polarity == 1) & (df.subjectivity == 1), "reviews"].head(1).tolist()
['got construction set prek classroom thought would awesome sensory constancy fuss box trying fix iti thought kenetic sand brand durable thisgreat box came fell apart unpacking']
```



ML MODELS TO PREDICT GOOD/BAD REVIEW

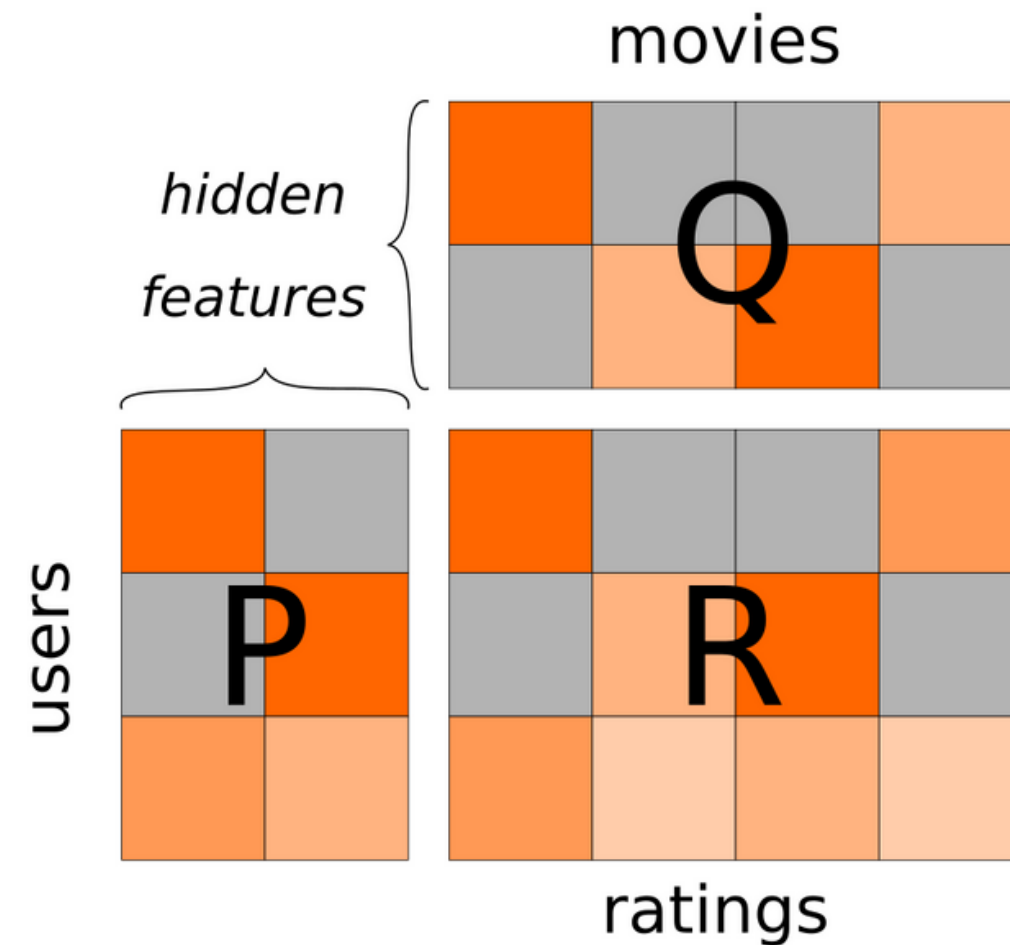
Receiving Operating Characteristic (ROC) Curve



	actual_label	prediction	review
0	1.0	1	ich habe schon nahezu alle exit spiele von kos...
1	0.0	0	jedoch macht für mich die neue verpackung mit ...
2	0.0	0	ich hatte bei dem produktnamen auf etwas gehof...
3	1.0	0	also es ist wirklich eine super beschäftigung ...
4	0.0	0	gutes spiel aber fehlerhafte produktionbeim ro...
5	1.0	1	good good light ball easy grabperfect
6	0.0	0	das produkt ist keine neuware da dies nicht ge...
7	0.0	0	leider fahren einige bots nicht und für das bi...
8	1.0	1	haben mal wieder ein neues kartenspiel für die...
9	1.0	0	reçu aujourd'hui super bien emballé la peluche...

```
roc_auc_score for Logistic Regression: 0.89
roc_auc_score for Multinomial NB: 0.88
roc_auc_score for RandomForestClassifier: 0.86
```

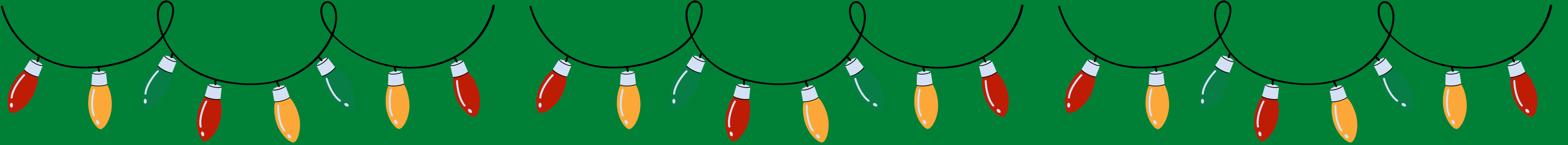

RECOMMENDER SYSTEM



- Data: Amazon Toys and Games
<https://nijianmo.github.io/amazon/>
- 5 reviews per user, and 6000 data points
- Collaborative filtering with Non-negative Matrix Factorization (NMF Model)

Streamlit





THANK YOU!

