Amazon Recommender System using Supervised Machine-Learning

By Andrew Wu

Why supervised machine learning?

- Clear and specific labels for data and outputs
- Allows for feature engineering
- Can generate better prediction results
- Not as affected by the ‘cold-start’ problem

Methodology

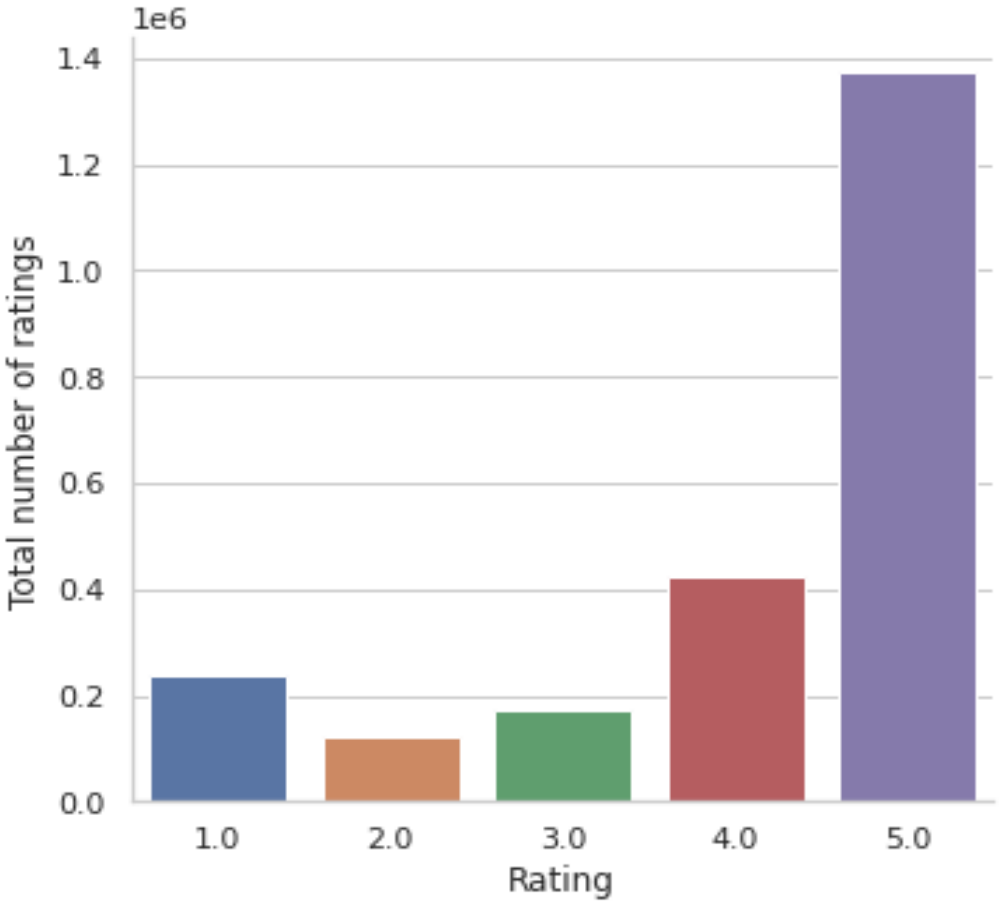
Data: Amazon electronics section metadata & ratings data from 2018

1. metadata: user, item, price, genre, main category, similar to, bought with, viewed with
2. Ratings: user, item, rating(out of 5)

Classification models: KNN, Naïve Bayes, Logistic Regression, XGBoost, DecisionTree, RandomForest

Data source:
http://jmcauley.ucsd.edu/data/amazon/

Exploratory data visualization



Average rating:
4.1/5

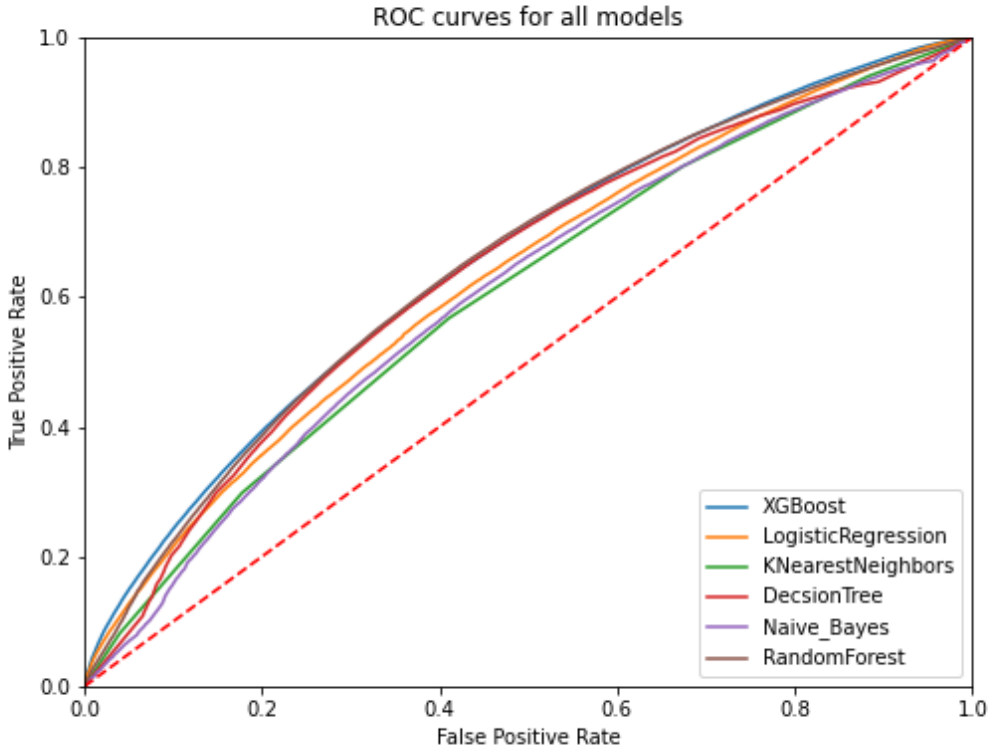
Average of
1.55 review per user

👍 **vs** 👎

Data source:
http://jmcauley.ucsd.edu/data/amazon/

amazon

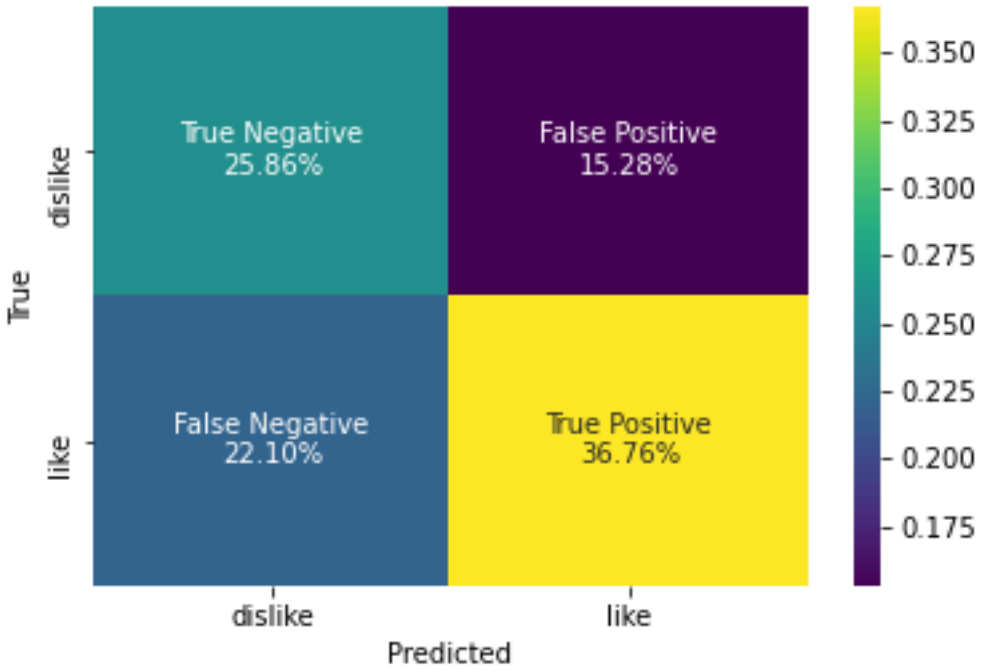
Initial results: Recall and AUC are key metrics



Model	AUC	Recall
XGBoost	0.611123	0.598306
LogReg	0.592101	0.581125
KNN	0.580540	0.569477
DecsionTree	0.611461	0.591558
NaiveBayes	0.578334	0.486364
RForest	0.613254	0.619221

Next: hyperparameter tuning with GridSearchCV

Results: XGBoost has a slight edge over RandomForest



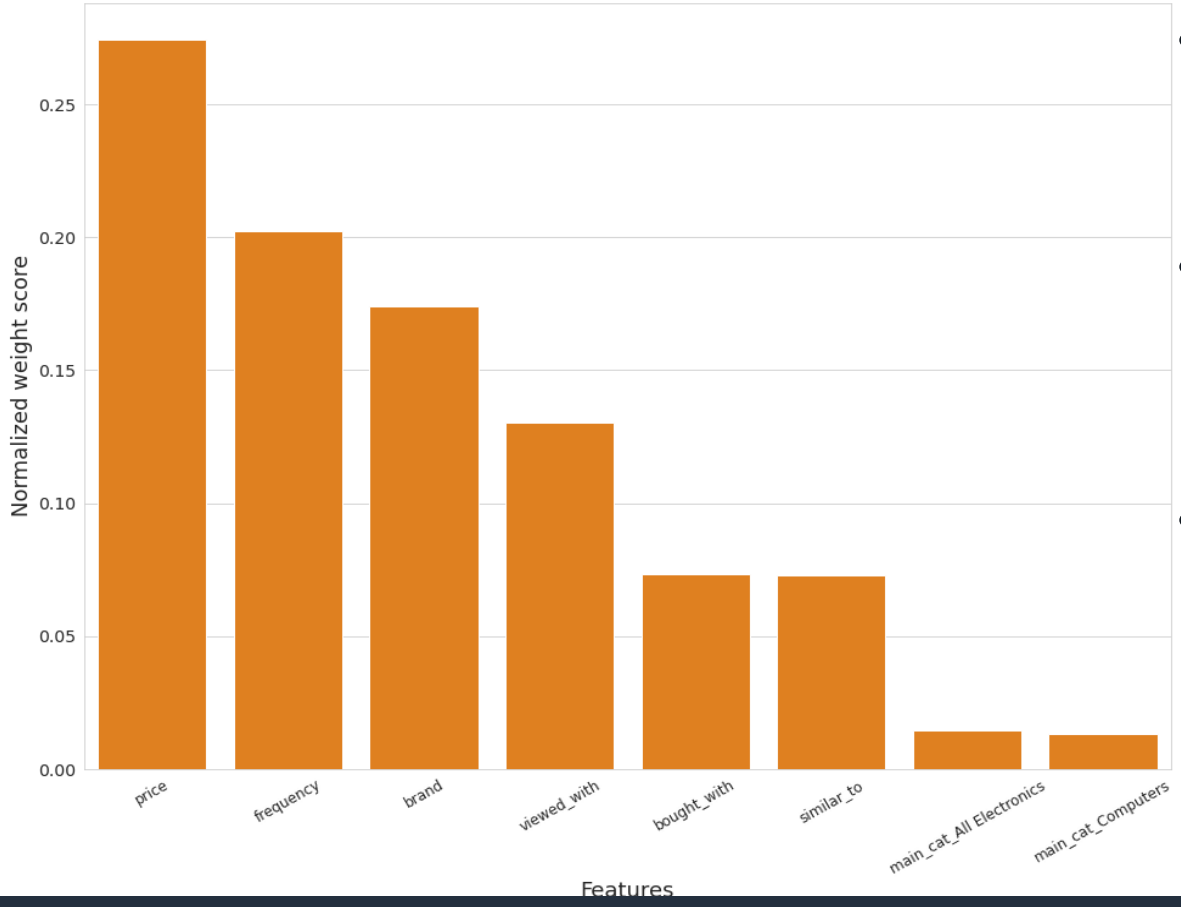
After GridSearchCV:

Accuracy:0.626

Recall:0.622

XGBoost gave the lowest false negative percentage

Results: Feature Importance



- Hard to pinpoint all important features
- **Weight** vs **Cover** vs **Gain**
- Brand is very important to customers

Conclusions

- RandomForest has lower metric scores, but also **lower over-fit**
- More feature engineering and customer information needed to identify key features

Also Completed: A unsupervised(SVD) + supervised hybrid model MVP (RMSE=1.1)

Future plans: Continue development on the hybrid model

Questions?

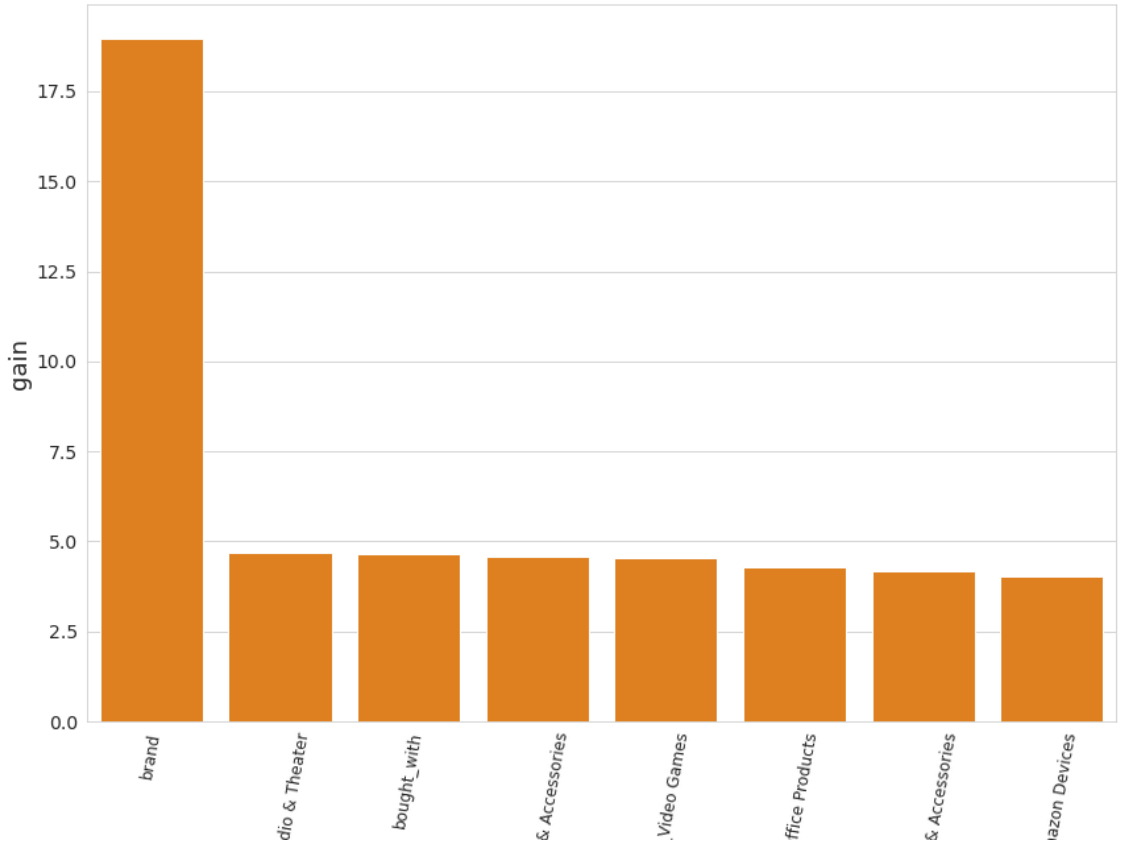
Thank you!

My Linkedin: https://www.linkedin.com/in/andrew-wu-939746171/

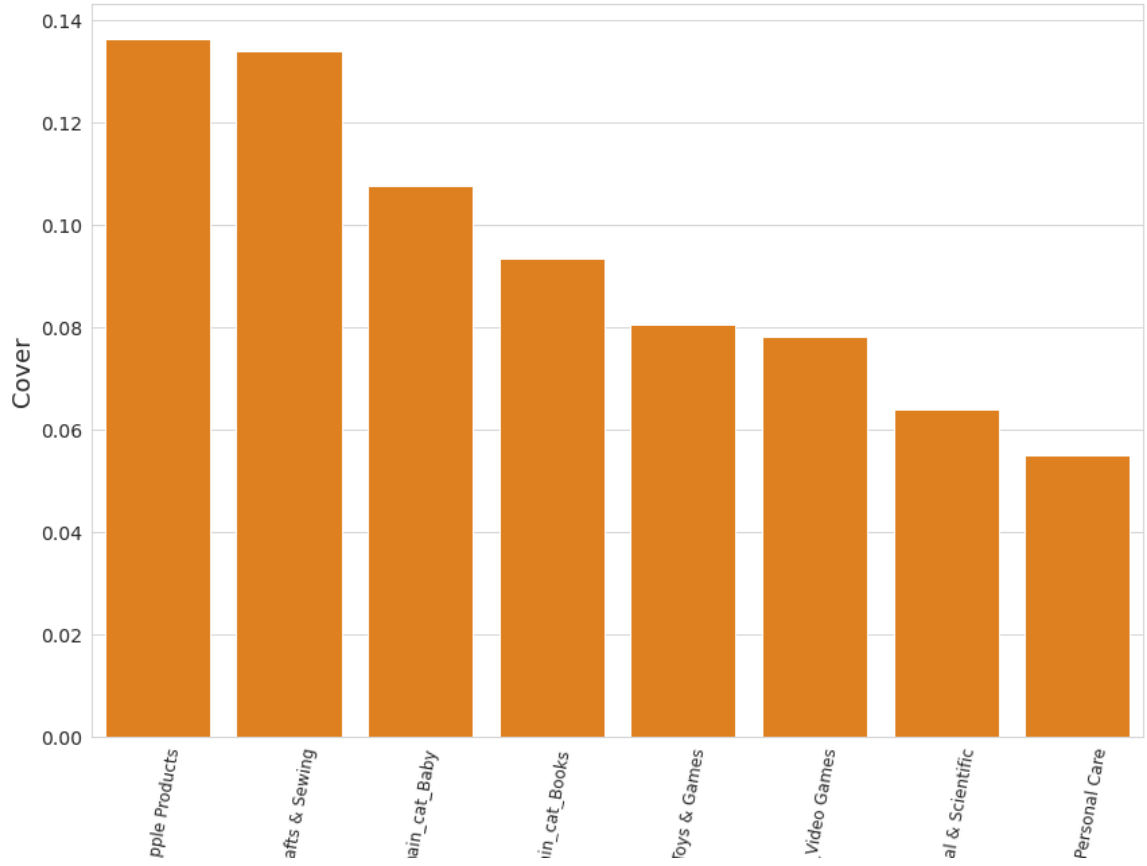
My Github:https://github.com/andywzz

Appendix

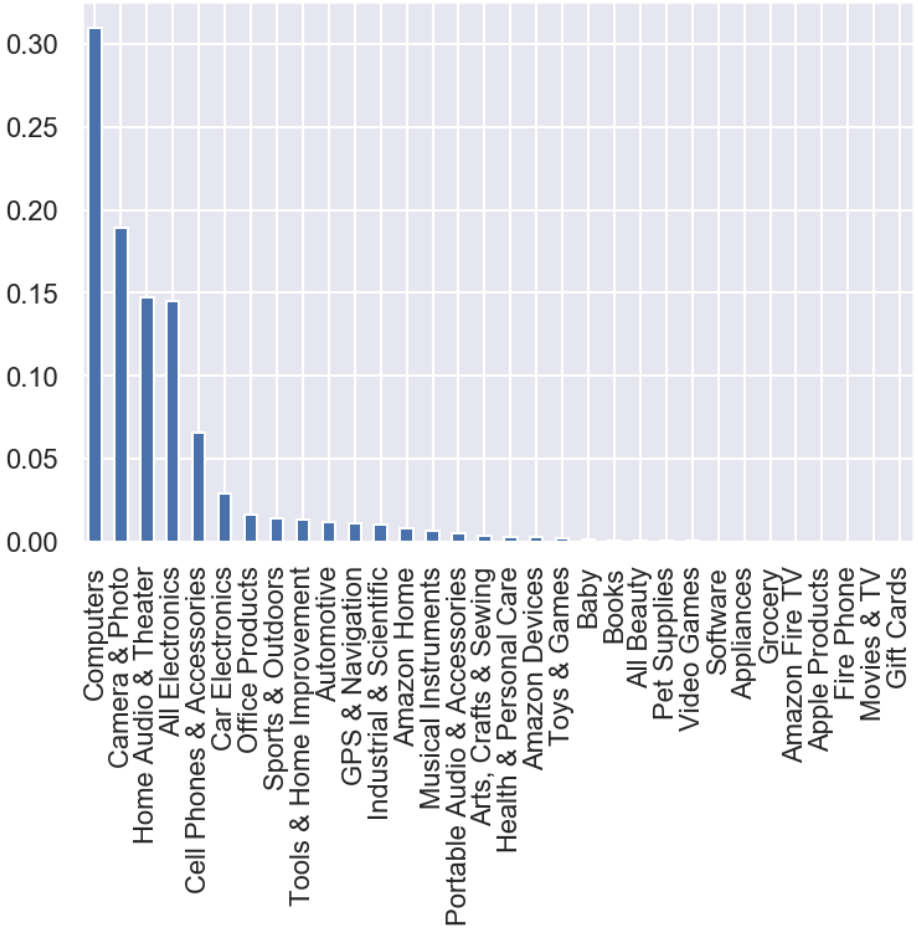
Appendix:Feature Importance



Appendix:Feature Importance



Appendix:Main category frequency



Designed by

www.PresentationGo.com

The free PowerPoint template library