

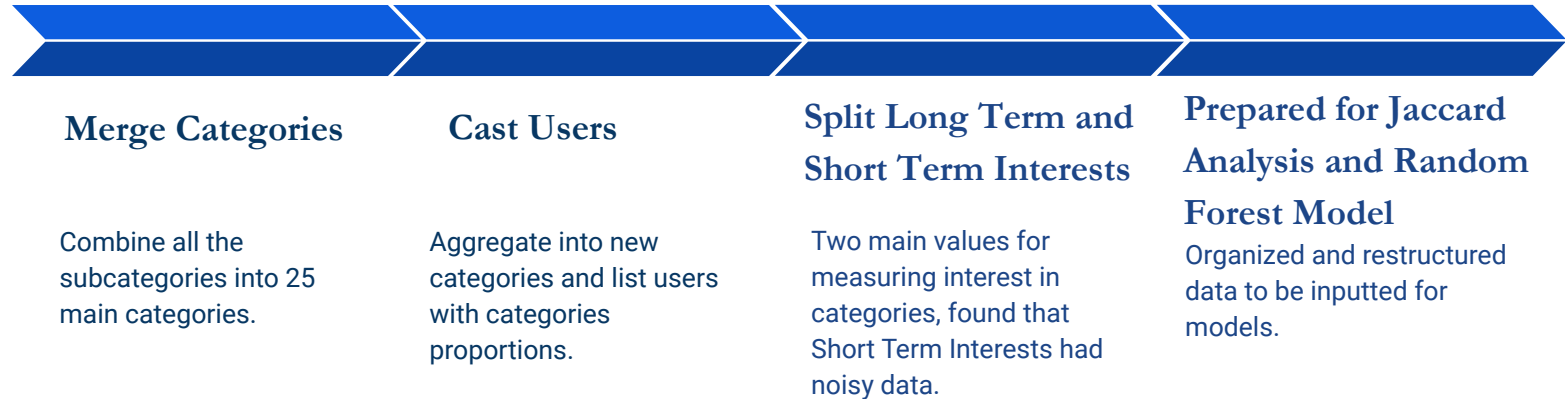


Exploring Correlation between User Interests and Conversion

Bhrij Patel, Kevin Day, and Mihir Dutta

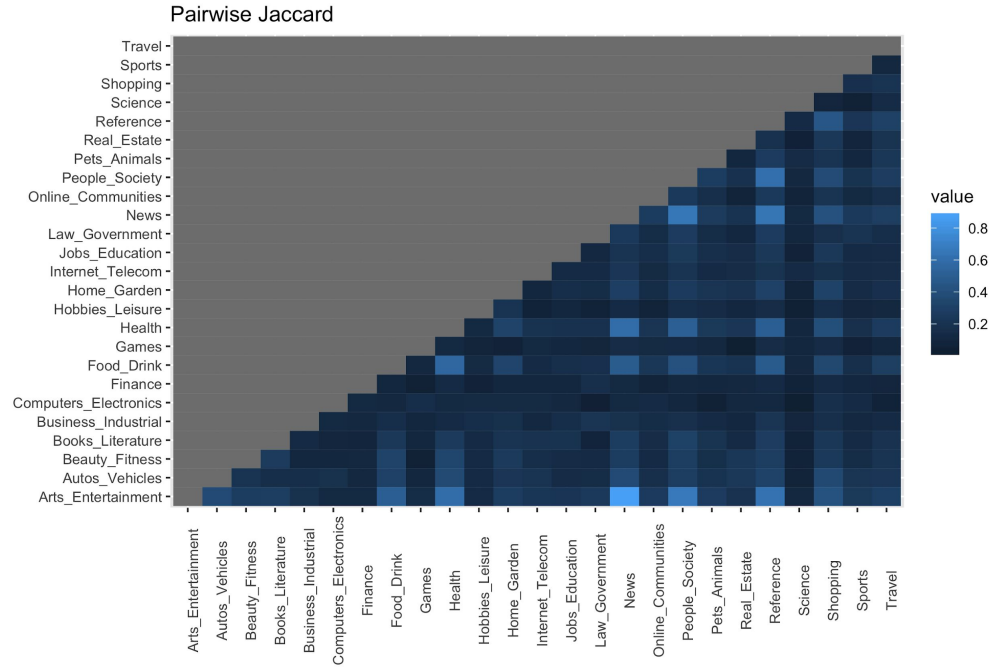
Exploratory Data Analysis

Goal: Discover which **interest categories** correlate to the most **conversions**.



Jaccard Similarity

- Ratio between intersection of two sets and their union
- Heat map shows ratios between each combination of interests
- Highest ratios
 - 1st: Arts & Entertainment and News
 - 2nd: Arts & Entertainment and People & Society



Random Forest Model

- Accurate while resisting overfitting
- Undersampled training data
- Tuned hyperparameter by minimizing MCE
- Achieves our goal of understanding which segments of the population to focus resources on
- Most important features are automobiles, health, home gardens, and arts & entertainment

	Shopper did not convert	Shopper did convert
Predicted won't convert	74.2%	26.2%
Predicted will convert	25.7%	73.7%

Accuracy: 70.4% w/ SD of 2.52%

Conclusion

- Focus ad spending on shoppers who are interested in Automobiles, Health, Home Gardens, and Arts & Entertainment
- Most common overlapping areas of interest in are: Arts & Entertainment, News, People & Society, Reference, Health.
- Overall, focus more strongly on the Arts & Entertainment

Appendix

- Jaccard Similarity

X1	X2	value
Arts_Entertainment	News	0.8705502
Arts_Entertainment	People_Society	0.6621622
News	People_Society	0.6490066
News	Reference	0.6423841
Arts_Entertainment	Reference	0.6279070
People_Society	Reference	0.6078431
Health	News	0.5980392
Arts_Entertainment	Health	0.5940594
Food_Drink	Health	0.5614035
Health	People_Society	0.5150376
Arts_Entertainment	Food_Drink	0.5033784
Health	Reference	0.5018727

Appendix

- Feature importance

	MeanDecreaseAccuracy
Arts_Entertainment	5.5057643
Autos_Vehicles	8.6252964
Beauty_Fitness	-0.6183055
Books_Literature	1.2511996
Business_Industrial	-1.7748698
Computers_Electronics	-2.4548403
Finance	0.2421443
Food_Drink	0.2616370
Games	-1.6252966
Health	6.3607975
Hobbies_Leisure	-1.3210712
Home_Garden	5.9467968
Internet_Telecom	-1.7267219
Jobs_Education	2.6477918
Law_Government	0.6314300
News	-0.2265168
Online_Communities	-0.5777874
People_Society	-0.7565069
Pets_Animals	-0.1702054
Real_Estate	1.8583079
Reference	-1.6229860
Science	-1.1196275
Shopping	-3.2750073
Sports	-3.0661330
Travel	-1.5147859