
adidas
'STAN SMITH'
Endorsed by:
nike

If you remove the brand name and product names from the text, can we predict the brand that an online post is discussing using an Natural Language Processing model?

Methodology

1. Collect data
2. Import and clean data
3. Exploratory data analysis
4. Modeling
5. Conclusions & Next Steps

Data

Subreddits:

- /r/Nike
- /r/Adidas

December 2018 - October 2019

Total posts: 3760

- Nike: 52%
- Adidas: 48%

Modeling

Models

Transformers

Stop words

How did the models perform?

Model	Transformer	Stop Words	Train Score	Test Score
SVC	cvec	none	0.986702	0.909574
SVC	cvec	English	0.982143	0.907801
SVC	cvec	Brand names	0.986322	0.836879
Naive Bayes (MNB)	cvec	none	0.940729	0.902482
Naive Bayes (MNB)	cvec	English	0.943009	0.911348
Naive Bayes (MNB)	cvec	Brand names	0.883359	0.817376
SVC	tfidf	none	0.995441	0.924645
SVC	tfidf	English	0.993921	0.931738
SVC	tfidf	Brand names	0.993161	0.859929
Naive Bayes (MNB)	tfidf	none	0.940729	0.890071
Naive Bayes (MNB)	tfidf	English	0.949088	0.906028
Naive Bayes(MNB)	tfidf	Brand names	0.901976	0.830674
Naive Bayes(MNB)	tfidf	Product names	0.939210	0.897163
Naive Bayes(MNB)	tfidf	Product and Brand names	0.878040	0.806738

Naive Bayes

English or product names

Conclusions

Removing Brand names isn’t enough

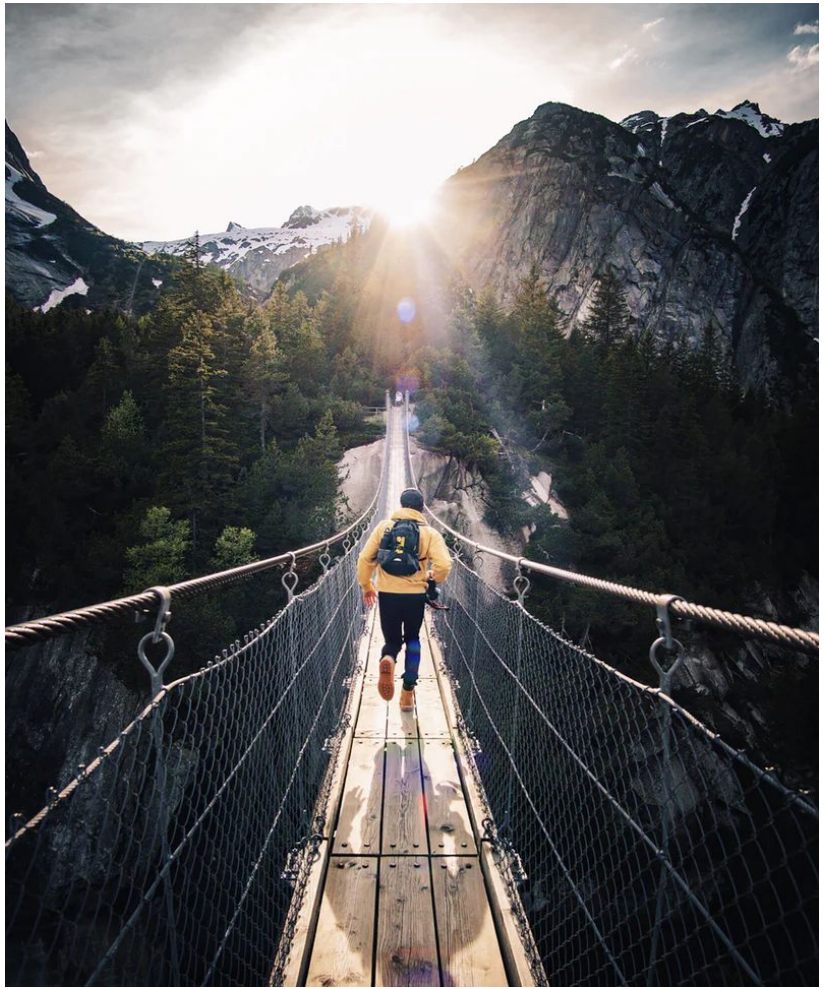
Different brand names

Different product names

Both transformers worked well

Next steps

- More history
- Online conversation vs. brand messaging
- Online conversation vs. PR
- Sentiment analysis
- Under Armour



Sources

- Title slide photo: https://unsplash.com/photos/WSLdpFngoQk/download?force=true
- Nike shoes: https://unsplash.com/photos/j1GiPlvSGWI/download?force=true
- Adidas shoes: https://unsplash.com/photos/DOjEoly80Qc/download?force=true
- Nike logo: https://www.nike.com
- Adidas logo: https://www.adidas.com/us