Uncovering Consumer Sentiments towards Amazon Fashion Products: A Data Analysis and Sentiment Classification Study

Karen Ardila Lopez

Bootcamp códigofacilito

Project

Amazon Fashion

Conduct a data analysis of customer sentiment for the products they purchased.

Sentiment Analysis

Understanding the social sentiment of the brand, product, or service monitoring reviews.

Computational Model

Develop a classifier according to the sentiment of the reviews.

Dataset

Data Info

Amazon Fashion

883,636 reviews
json format

Data Span

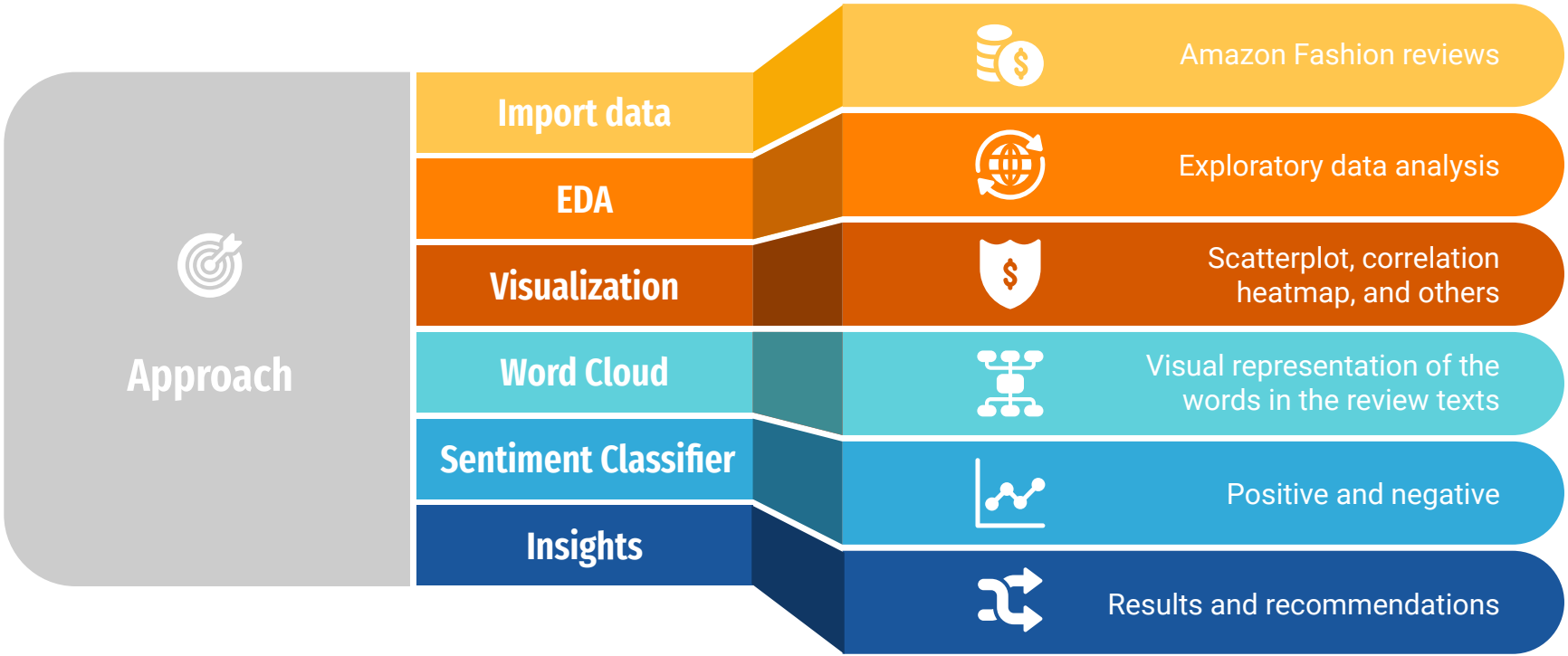
May 1996 - Oct 2018

Citation

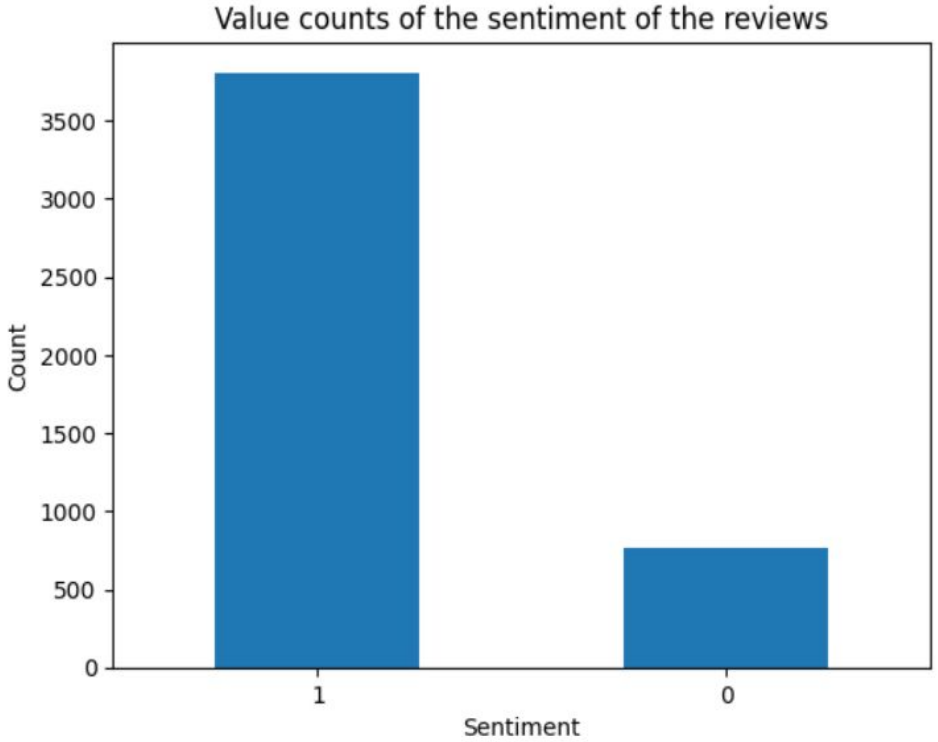
Article

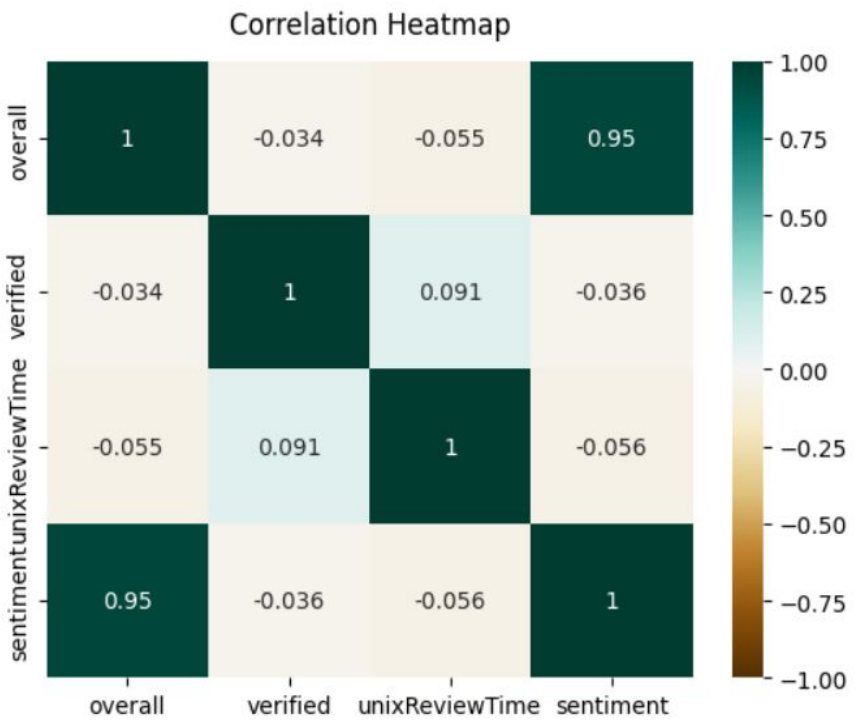
Justifying recommendations using distantly-labeled reviews and fined-grained aspects
Jianmo Ni, Jiacheng Li, Julian McAuley
Empirical Methods in Natural Language Processing (EMNLP), 2019

Pipeline


Approach
Import data
EDA
Visualization
Word Cloud
Sentiment Classifier
Insights
Amazon Fashion reviews
Exploratory data analysis
Scatterplot, correlation heatmap, and others
Visual representation of the words in the review texts
Positive and negative
Results and recommendations

Exploratory Data Analysis





Negative Word Cloud

Positive Word Cloud

Love

Price

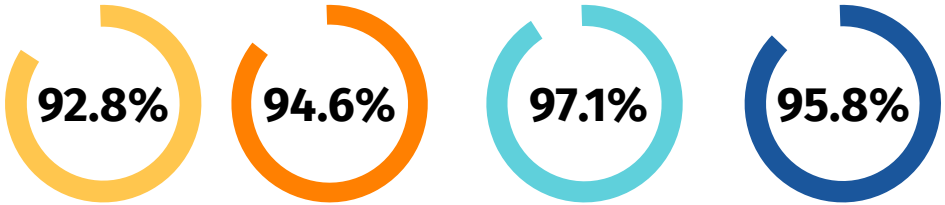
Fit - size

Perfect - great

Sentiment Classifier

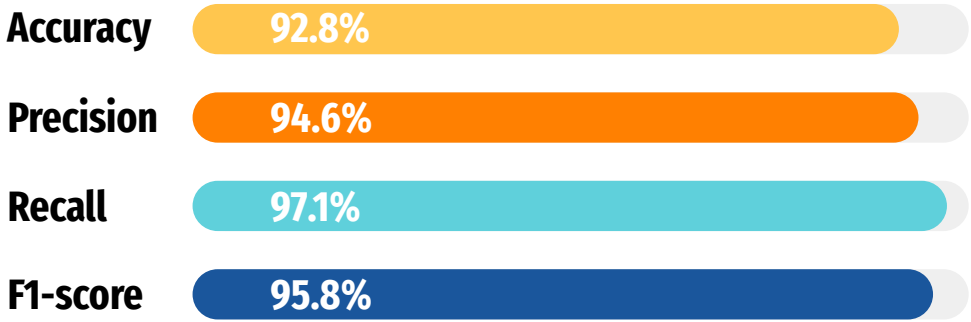
True positives

941



False positives

54



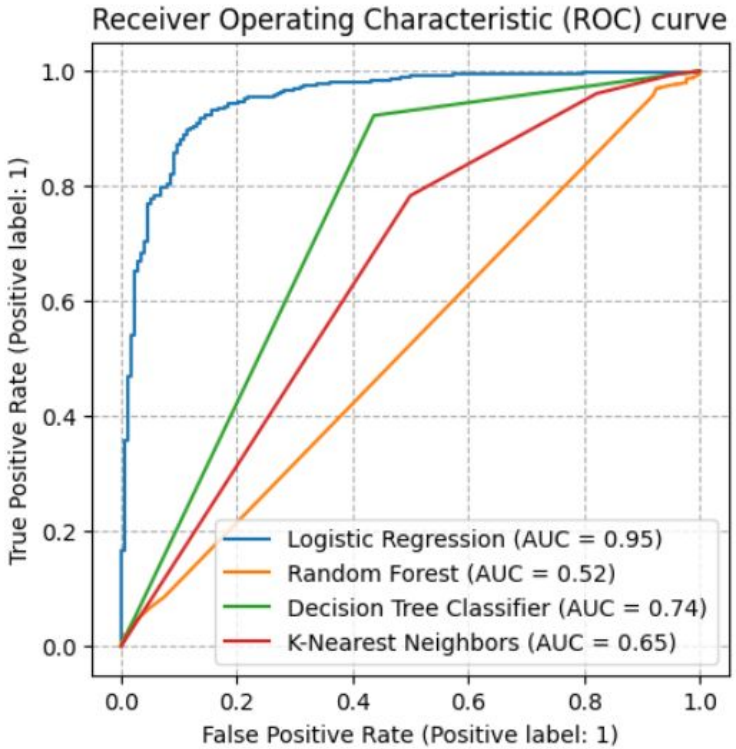
True negatives

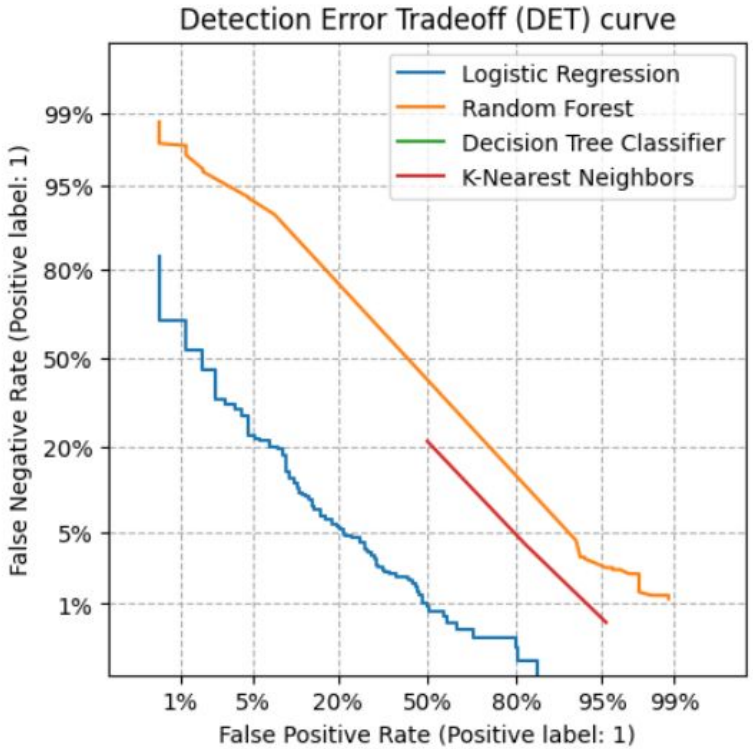
120

False negatives

28

Sentiment Classifier





Insights

47%

2016 has more data than any other year.

Negative

"disappointed," "different", "size", "picture", ”received"

92.8%

Accuracy of the sentiment classifier

Positive

"love", "fit", "perfect", "great", "price", and "size"

2016

Noticeable difference in review data between years

Negative

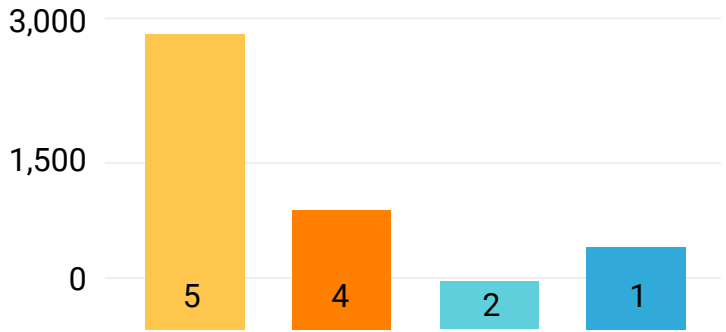
Issues with size and product not matching the picture are common complaints.

Accuracy

Able to classify the data appropriately

Positive

Strong points in the sales.



Recommendations

Reduce costs

Get rid of products that have a negative financial impact, avoid reprocessing, to save time, resources, and money.

Improve products

Address the issue of negative sentiments, mostly because of bad quality or fit.

Attract new customers

With positive reviews, listening the audience and creating/following trends.

Internet expansion

Specially in platforms like TikTok, which become excellent growth indicators, particularly in clothing.

References

1. https://pandas.pydata.org/docs/index.html
2. https://nijianmo.github.io/amazon/index.html
3. https://scikit-learn.org/stable/modules/classes.html#module-sklearn.linear_model
4. https://towardsdatascience.com/understanding-logistic-regression-9b02c2aec102
5. https://towardsdatascience.com/top-10-binary-classification-algorithms-a-beginners-guide-feeacbd7a3e2
6. https://towardsdatascience.com/exploratory-data-analysis-in-python-a-step-by-step-process-d0dfa6bf94ee
7. https://scikit-learn.org/stable/auto_examples/model_selection/plot_det.html#sphx-glr-auto-examples-model-selection-plot-det-py
8. https://medium.com/@szabo.bibor/how-to-create-a-seaborn-correlation-heatmap-in-python-834c0686b88e
9. https://huggingface.co/blog/sentiment-analysis-python
10. https://medium.com/analytics-vidhya/data-preparation-and-text-preprocessing-on-amazon-fine-food-reviews-7b7a2665c3f4
11. https://towardsdatascience.com/a-complete-sentiment-analysis-algorithm-in-python-with-amazon-product-review-data-step-by-step-2680d2e2c23b
12. https://youtu.be/ckLy603HQGI