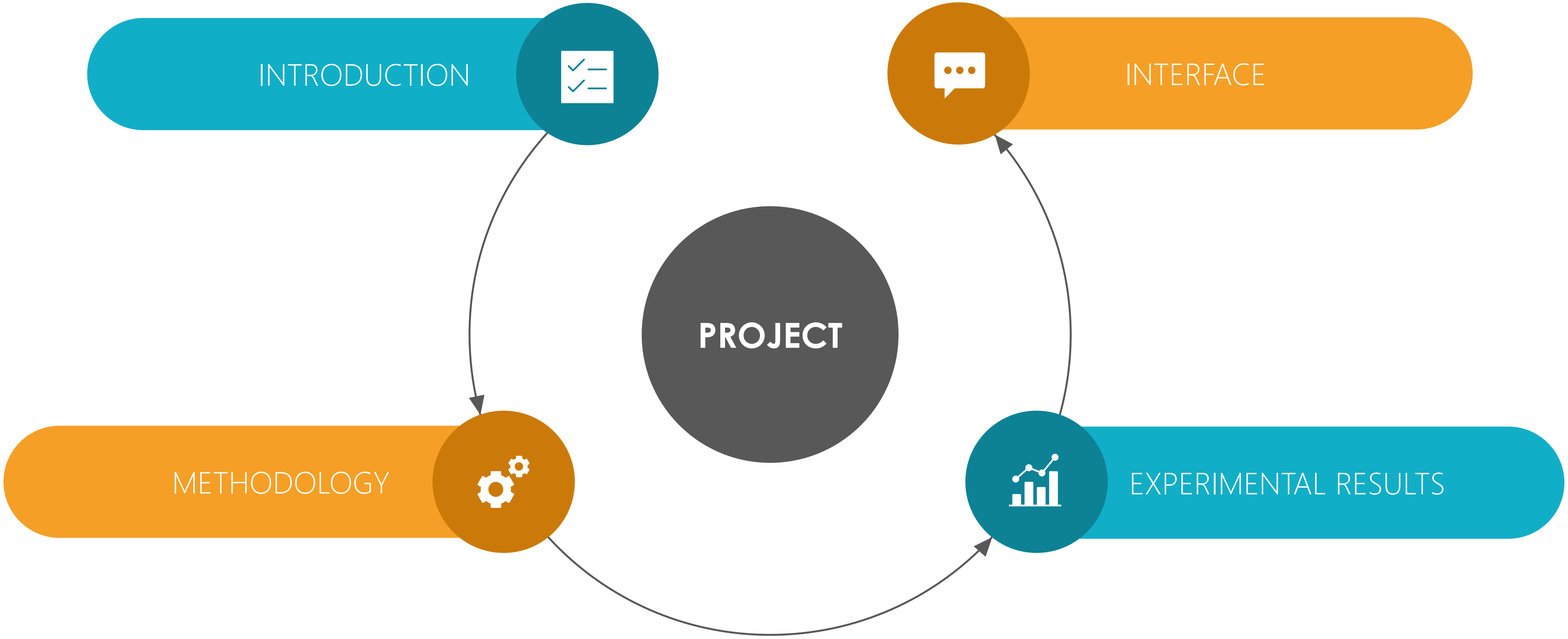
Fake reviews detection on Amazon dataset

Gianmaria Saggini

Università di Pisa

A.A. 2023/2024

Fake reviews detection

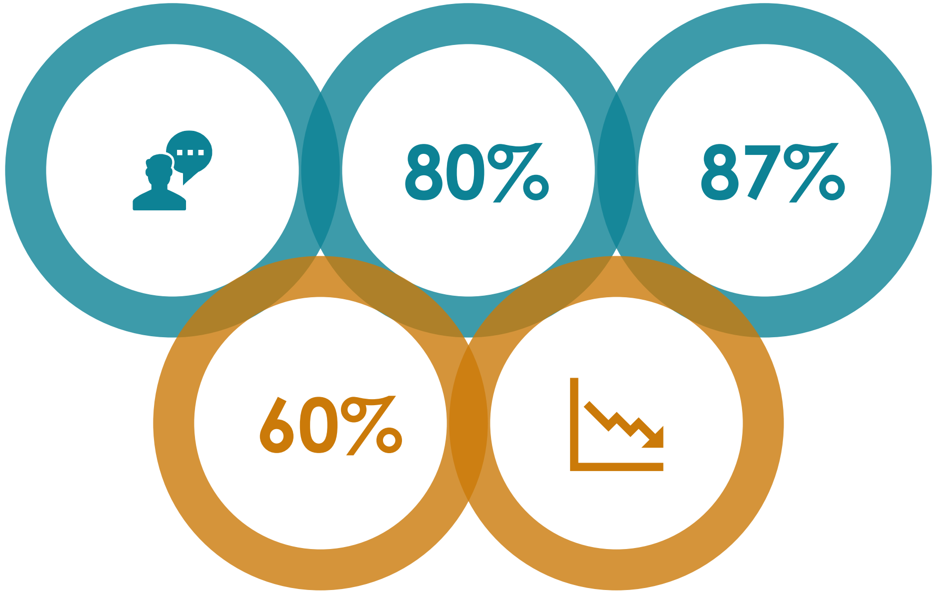


Introduction

User-oriented online reviews serve as the second most reliable source of product information.

80% of customers change their purchasing decisions after reading negative reviews.

87% of customers approve their decisions after reading positive reviews.



60% of reviews for top 10 electronic products on Amazon are fake.

Fake reviews can not only manipulate product rankings, but more widely reduce consumer trust in online reviews.

Methodology

DATASET

Descriptions of the two datasets used.

EXPLORATORY DATA ANALYSIS

Analysis of the features and their distributions.

PREPROCESSING

Text preprocessing: parsing, tokenization, stemming...

FEATURE EXTRACTION

Text vectorization using BoW and TF-IDF.

WORD EMBEDDINGS

Training of Word2Vec model to generate word embeddings of reviews.

Methodology
Dataset

Labeled dataset

- 21000 reviews
- Half real, half fake
- Labeled by Amazon
- Used for classification

Features:

Label
Rating
Verified purchase
Product category
Product id
Product title
Review title
Review text

Unlabeled dataset

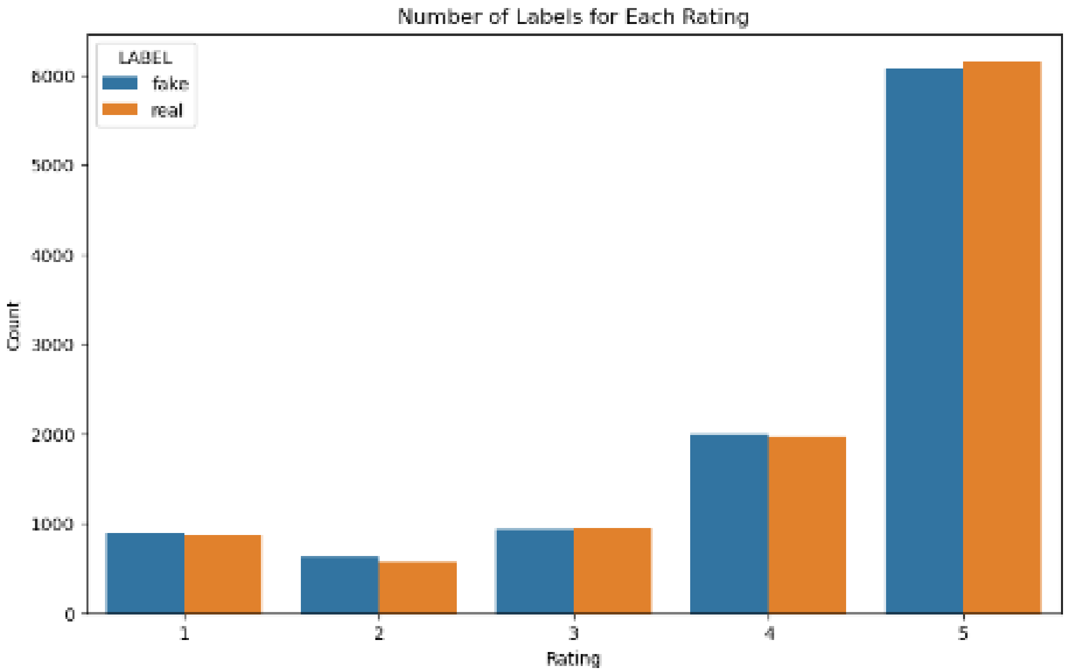
- Half a billion reviews
- Used for Word2Vec training

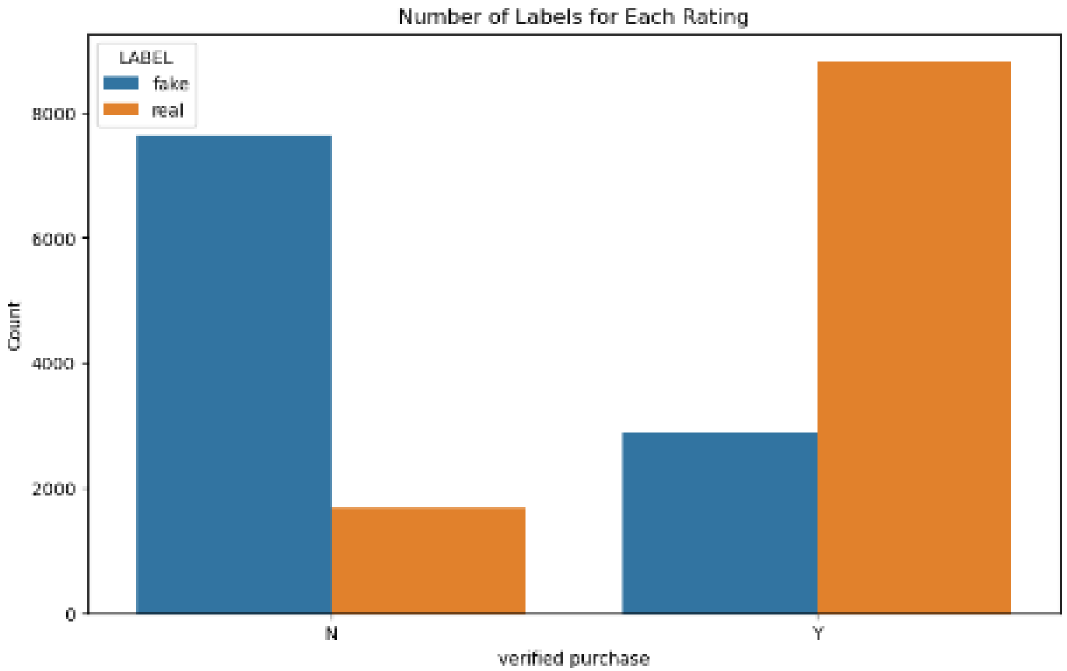
Features (only used ones):

Review text
Timestamp
…

Methodology
EDA

Distribution of fake and real reviews in various attributes:

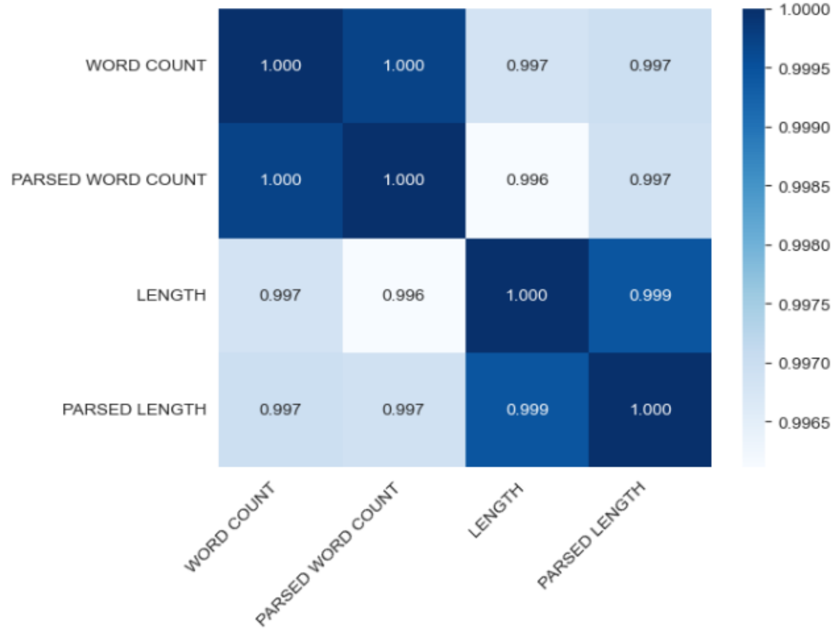


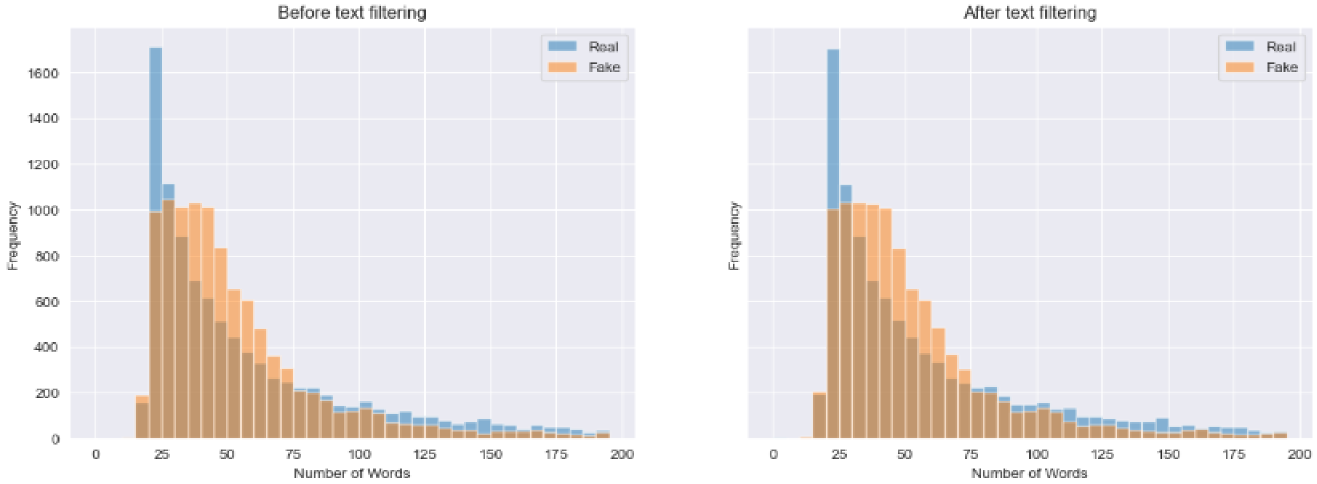


Methodology
EDA

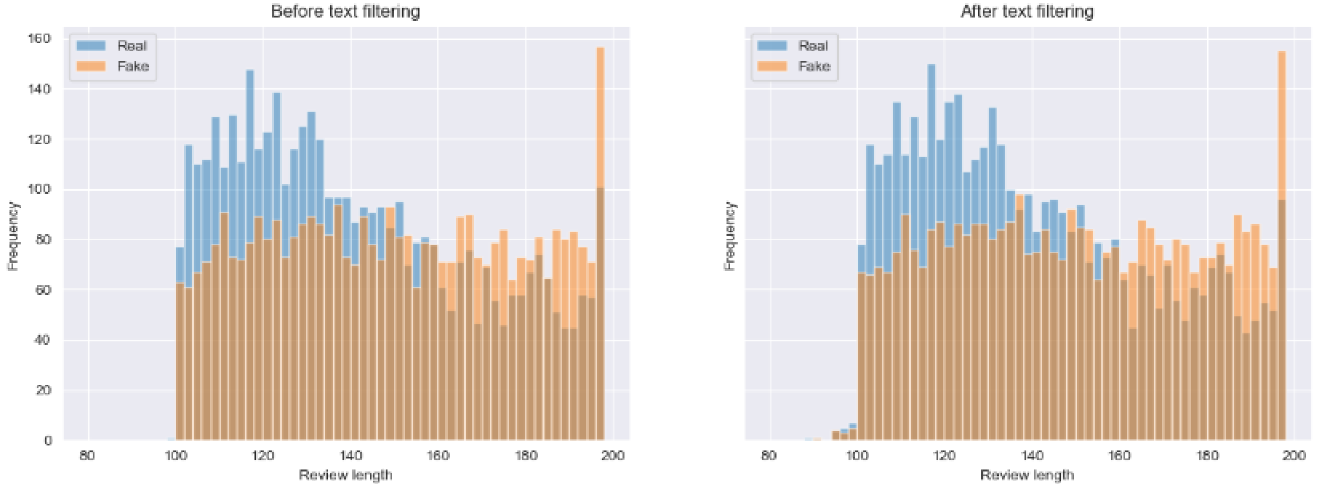
Analysis of text related features.

- Number of words (before filtering)
- Number of words (after filtering)
- Text length (before filtering)
- Text length (after filtering)





(a) Number of fake and real reviews for ratings.



Methodology

Preprocessing

Filtering → Tokenization → Stopword removal → Stemming

Filtering
Text is converted to lowercase. Remove URLs and HTML tags.

Tokenization
Text is divided into tokens with only latin letters.

Stopword removal
Remove stopwords with nltk english stopword list.

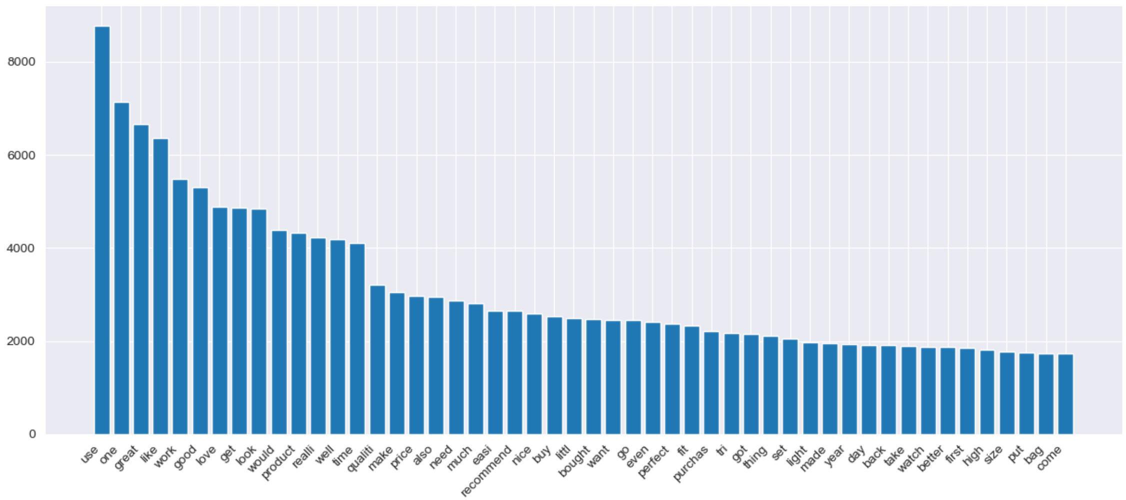
Stemming
Each token is passedthrough the Snowball Stemmer.

Methodology

Feature extraction

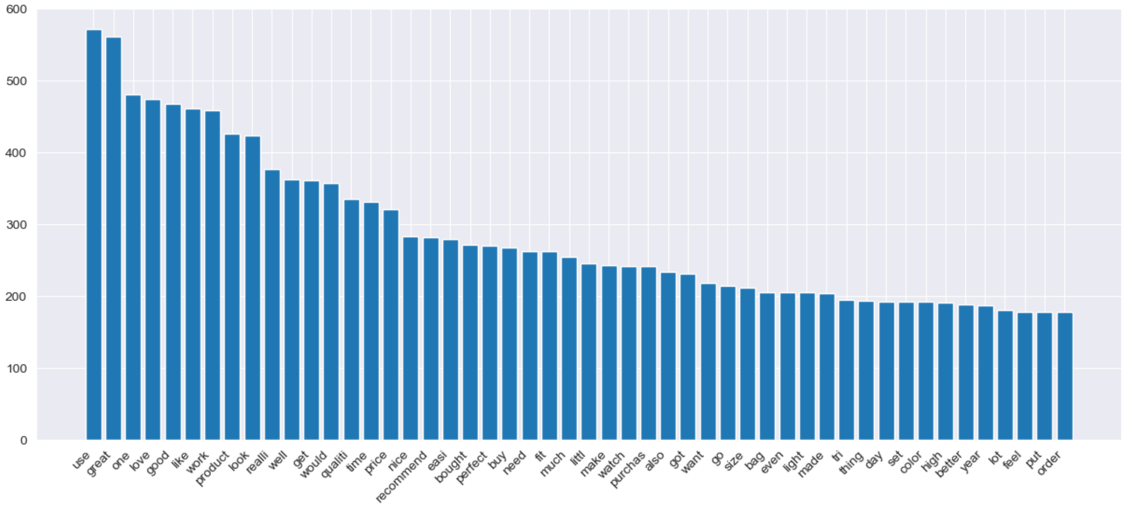
Bag of Words (BoW)

50 most frequent terms in BoW:



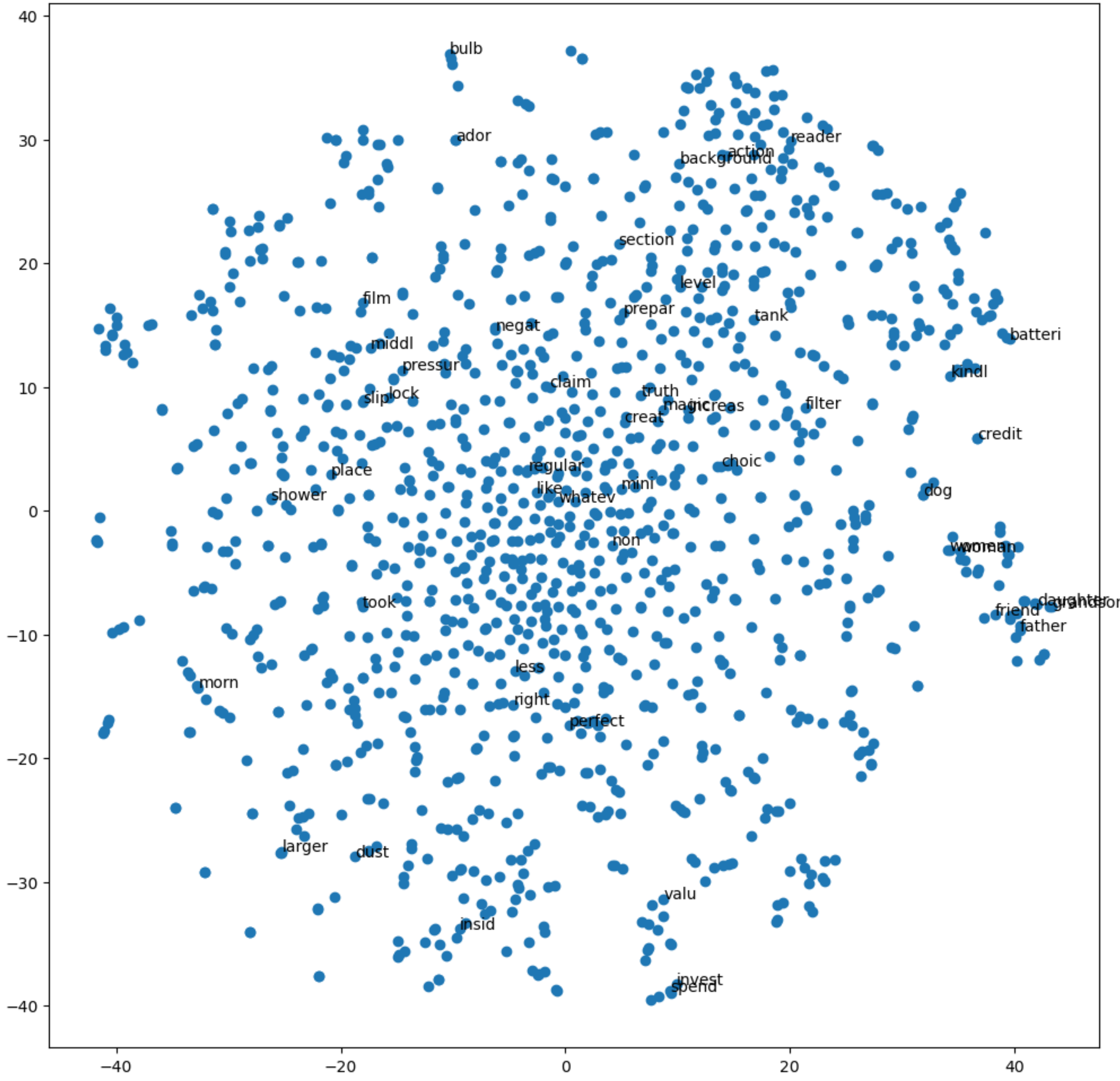
TF-IDF

50 most frequent terms in TF-IDF:



Methodology

Word embeddings



- Skip-Gram variant of Word2Vec model.
- Input: file with a sentence per line.
  To speed up computation, parallelize the work: each line of the file has to be a tokenized sentences.

256GB → 30GB

- Hyperparameters:
  - Window = 5
  - Vector_size = 100
  - Min_count = 700000 and 500000

min_count = 700000

Experimental results

Results

- Random Forest seems to be slightly better than linear SVC.
- TF-IDF has always higher accuracy than Bag of Words.
- As expected, verified purchase plays a fundamental role in the detection of fake reviews.
- Word embeddings 1 (min_count = 700000) doesn't seem to be always better than word embeddings 2 (min_count = 500000) .

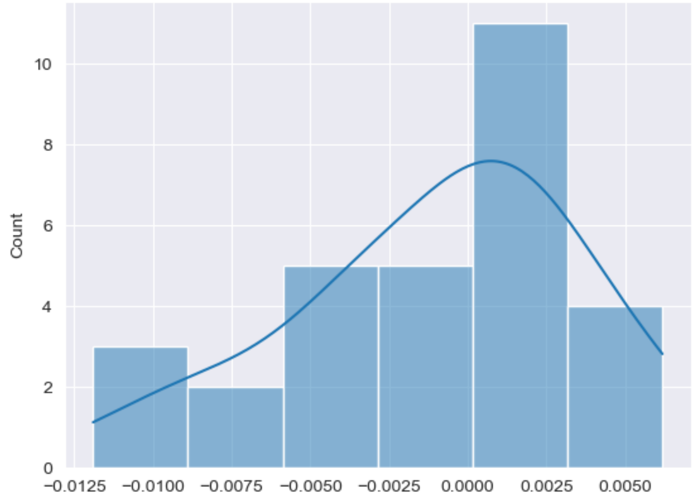
Features	Model	Accuracy	F1 score	AUC
BoW	RF	63.81%	66.04%	70.17%
	SVC	61.71%	63.62%	66.32%
BoW+VP	RF	80.53%	79.22%	**87.10%**
	SVC	79.25%	78.97%	84.58%
BoW+WE1	RF	63.51%	65.13%	68.72%
	SVC	62.25%	64.07%	66.96%
BoW+WE2	RF	63.12%	64.85%	68.71%
	SVC	62.29%	64.13%	67.03%
BoW+VP+WE1	RF	80.45%	79.68%	86.51%
	SVC	78.87%	78.63%	84.43%
BoW+VP+WE2	RF	80.54%	79.73%	86.66%
	SVC	78.65%	78.44%	84.24%
TF-IDF	RF	64.60%	65.91%	70.93%
	SVC	64.17%	64.37%	69.58%
TF-IDF+VP	RF	80.68%	79.36%	86.94%
	SVC	80.62%	**80.27%**	86.40%
TF-IDF+WE1	RF	63.95%	64.79%	69.75%
	SVC	64.06%	64.31%	69.67%
TF-IDF+WE2	RF	63.88%	64.97%	69.88%
	SVC	64.08%	64.37%	69.66%
TF-IDF+VP+WE1	RF	**80.75%**	80.11%	86.81%
	SVC	80.54%	80.23%	86.44%
TF-IDF+VP+WE2	RF	80.74%	80.10%	86.83%
	SVC	80.39%	80.07%	86.43%

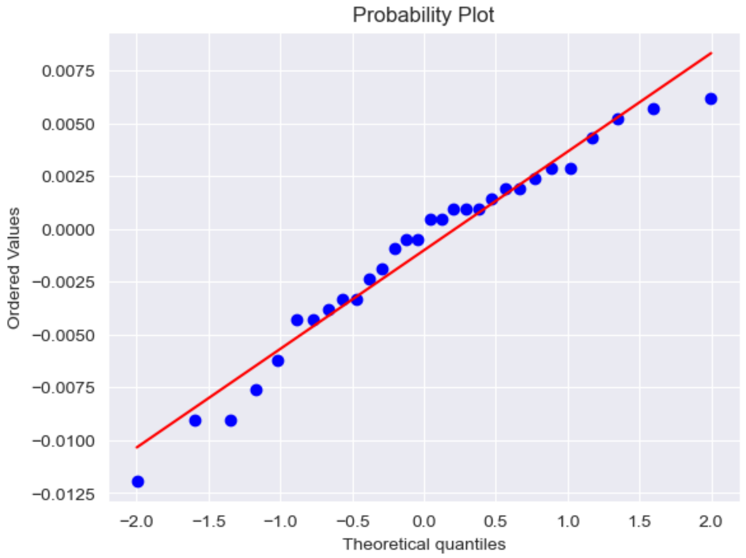
Experimental results
Statistical comparison

Comparison between RandomForest trained with TF-IDF+VP+WE1 and TF-IDF+VP.

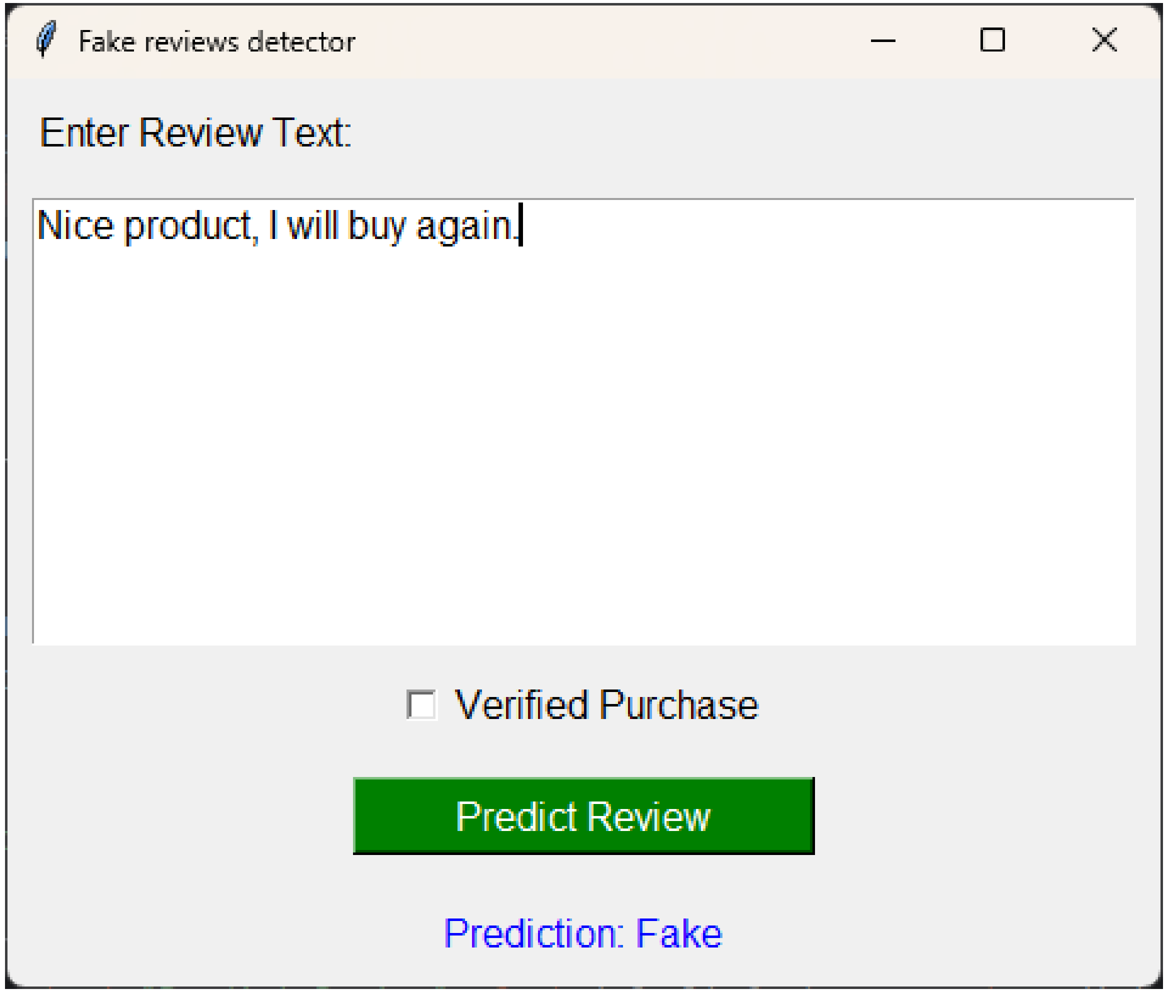
- To have more data, I performed three 10-fold cross-validations.
- Differences between paired observations are skewed, so I opted for a Wilcoxon signed-rank test.

Wilcoxon p-value: 0.44 → Cannot reject Null Hypothesis





Interface


Fake reviews detector
Enter Review Text:
Nice product, I will buy again.
Verified Purchase
Predict Review
Prediction: Fake

References

Petr Hajek, Lubica Hikkerova, Jean-Michel Sahut,
Fake review detection in e-Commerce platforms using aspect-based sentiment analysis
Journal of Business Research, Volume 167, 2023, 114143, ISSN 0148-2963,
https://doi.org/10.1016/j.jbusres.2023.114143

Şule Öztürk Birim, Ipek Kazancoglu, Sachin Kumar Mangla, Aysun Kahraman, Satish Kumar, Yigit Kazancoglu,
Detecting fake reviews through topic modelling
Journal of Business Research, Volume 149, 2022, Pages 884-900, ISSN 0148-2963,
https://doi.org/10.1016/j.jbusres.2022.05.081

Hajek, P., Barushka, A. & Munk, M.
Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining.
Neural Comput & Applic 32, 17259–17274 (2020).
https://doi.org/10.1007/s00521-020-04757-2

Datasets:
- https://www.kaggle.com/datasets/lievgarcia/amazon-reviews
- https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews

Thank you