INSTITUTE FOR APPLIED COMPUTATIONAL SCIENCE
AT HARVARD UNIVERSITY

Product Review Analysis With pattern

Final Presentation

Team: Sehaj Chawla, Taro Spirig, Lotus Xia, Heather Liu

Agenda

1. Motivation and problem statements
2. Datasets
3. Literature Reviews
4. Models
5. Results
6. Interpretations

Motivation
&
Problem Statements

Motivation

pattern is an ecommerce accelerator. It helps businesses grow faster and sell globally on ecommerce marketplaces.

How? For example, help them make better decisions through AI-supported insights and reporting.

Reviews are important! They give an insight into customer preferences, suggestions, complaints, which ultimately can help businesses

1) predict how well their products will do in the future,
2) adapt their marketing strategy or even their products.

Problem Statement

- **Predict product future sales performance (short term and long term)**
  - In particular, how can review data improve model prediction in addition to other metadata
- **Extract themes from review texts to gain insights on what keywords or topics are predictive of sales performance**

Problem Statement

★ **Short-term predictions**

- Predict sales performance in the next month
- Help retailers make sales projections in the immediate future and determine the quantity of inventories to acquire

Problem Statement

★ **Long-term predictions**

- Predict sales performance in the following year
- Inform retailers if it is worth continuing a newly launched product or how they could change the product to boost its success.

Data

Datasets

The three datasets contain around 9000 products that are listed and sold in Amazon's Vitamins and Dietary Supplements category from July 2017 to July 2021.

★ Review history data

★ Amazon best seller rank (BSR) data

★ Estimated sales volume data

Dataset 1: Review History

- ★ **Review history data**
  - 5 million reviews on 9,977 unique products at the time of scraping
  - Review title and review text
  - Metadata such as review dates, review ratings, verified purchase, etc

E.g.:

“Horrible product, my mother in law ended up in the hospital with a severe allergic reaction. She had to be in the ICU for a couple of days. Please be careful with this product.”

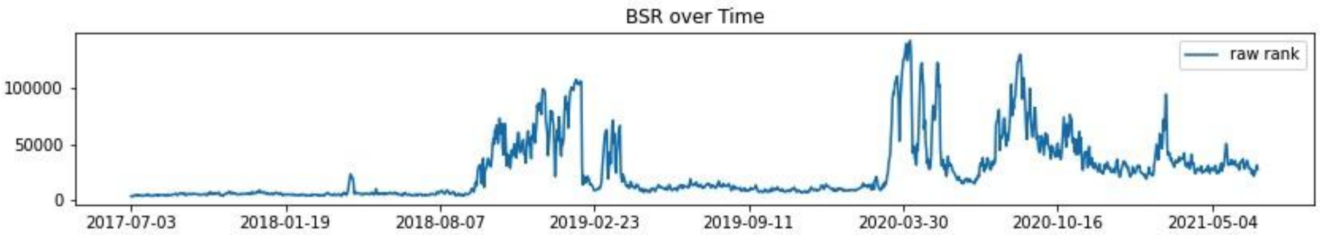
“I bought this for my father , He swears by it that it helps his joints feel better. His neighbor has been using it for a year now and. No more aches & pains also.”

“Pills are not correct as in the picture shown.”

Dataset 2: Best Seller Rank

★ **Amazon best seller rank (BSR) data**

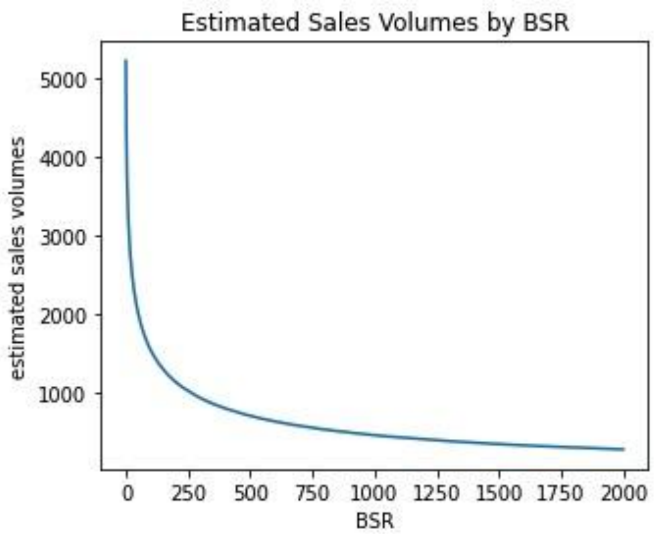
- Daily rank of each product based on
  - 1) current sales
  - 2) sales history
- Lower rank means better sales performance
  - E.g. 1 - best rank



Dataset 3: Estimated Sales Volume

★ **Estimated sales volume data**

- 1 to 1 mapping from rank to sales volume
- A lower rank corresponds to a higher sales volume



Literature Reviews

Literature Reviews

- **People use online reviews and historical sales data to predict future sales**
  - Geng Cui, Hon-Kwong Lui & Xiaoning Guo (2012)
  - Elizabeth Fernandes, Sérgio Moro, Paulo Cortez, Fernando Batista & Ricardo Ribeiro (2020)
- **Commonly used models are Linear Regression, XGBoost for prediction and BERT for texts analysis**
  - Ching-Chin Chern, Chih-Ping Wei, Fang-Yi Shen & Yu-Neng Fan (2015)
  - Vimala Balakrishnan, Zhongliang Shi, Chuan Liang Law, Regine Lim, Lee Leng Teh & Yue Fan (2022)
- **Commonly used performance metrics include**
  - **R-squared, RMSE for regression (evaluating short-term performance)**
    - Elizabeth Fernandes, Sérgio Moro, Paulo Cortez, Fernando Batista & Ricardo Ribeiro (2020)
  - **F1-score, AUC, Precision, and Recall for classification (evaluating long-term performance)**
    - Akinori Fujino, Hideki Isozaki, & Jun Suzuki (2008)

Model Description

Short-term: month-to-month prediction

★ **Target variable:** median sales of next month

Short-term: month-to-month prediction

- **Target variable:** median sales of next month
- **Non-text models:**
  - **Model classes: XGboost (XGB)** and **random forest (RF)**
  - **Features:**
    - Past performance (current month mean and median sales volume)
    - Past review metadata (e.g. monthly mean rating, number of reviews)
    - Do not include anything related to review text

Short-term: month-to-month prediction

- **Target variable:** median sales of next month
- **Non-text models:**
  - **Model classes: XGboost (XGB)** and **random forest (RF)**
  - **Features:**
    - Past performance (current month mean and median sales volume)
    - Past review metadata (e.g. monthly mean rating, number of reviews)
    - Do not include anything related to review text
- **Text-based models:**
  - **Model classes: Bag-of-words based regression** and **BERT-based regression**
  - **Features:** All reviews posted within or before the current month

Long-term: success after one year

★ **Target variable:** 1 if the product is successful, 0 otherwise

- Successful if it has ever reached top 3000 BSR during the one year period one year after launch

Long-term: success after one year

- **Target variable:** 1 if the product is successful, 0 otherwise
- **Non-text models:**
  - **Model classes: XGboost (XGB)** and **Random Forest (RF)**
  - **Features:**
    - Past performance (first three month mean and median BSR)
    - Past review metadata (e.g. mean rating, number of reviews)

Long-term: success after one year

- **Target variable:** 1 if the product is successful, 0 otherwise
- **Non-text models:**
  - **Model classes: XGboost (XGB)** and **Random Forest (RF)**
  - **Features:**
    - Past performance (first three month mean and median BSR)
    - Past review metadata (e.g. mean rating, number of reviews)
- **Text-based models:**
  - **Model classes: Bag-of-words based regression** and **BERT-based regression**
  - **Features:** All reviews within the first 3 months after product launch

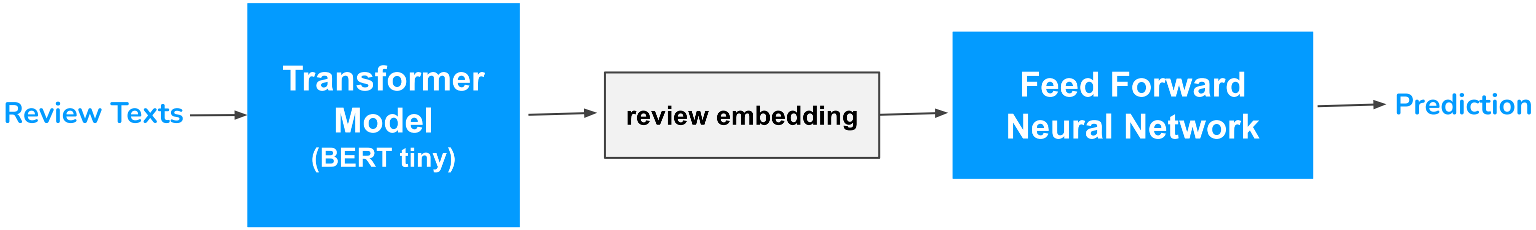
Long-term: success after one year

- **Target variable:** 1 if the product is successful, 0 otherwise
- **Non-text models:**
  - **Model classes: XGboost (XGB)** and **Random Forest (RF)**
  - **Features:**
    - Past performance (first three month mean and median BSR)
    - Past review metadata (e.g. mean rating, number of reviews)
- **Text-based models:**
  - **Model classes: Bag-of-words based regression** and **BERT-based regression**
  - **Features:** All reviews within the first 3 months after product launch
- **Ensemble models:**
  - **Model classes: logistic regression** and **decision tree**
  - **Features:** predicted probabilities of the four individual models

BoW Model Structure

- Text processing:
  - Simple frequency of phrases (referred to as bag-of-words)
  - TF-IDF (term frequency-inverse document frequency)
- Feature variables:
  - (weighted) frequency of the 500 most common unigrams/bigrams/trigrams in the training corpus
- Regression model:
  - L1 and L2 regularized (logistic) regression
- **Simple and nice interpretability**

BERT Model Structure



Hyperparameters:

★ max sequence length of review text
★ embedding size
★ # dense layers in FFNN

Results

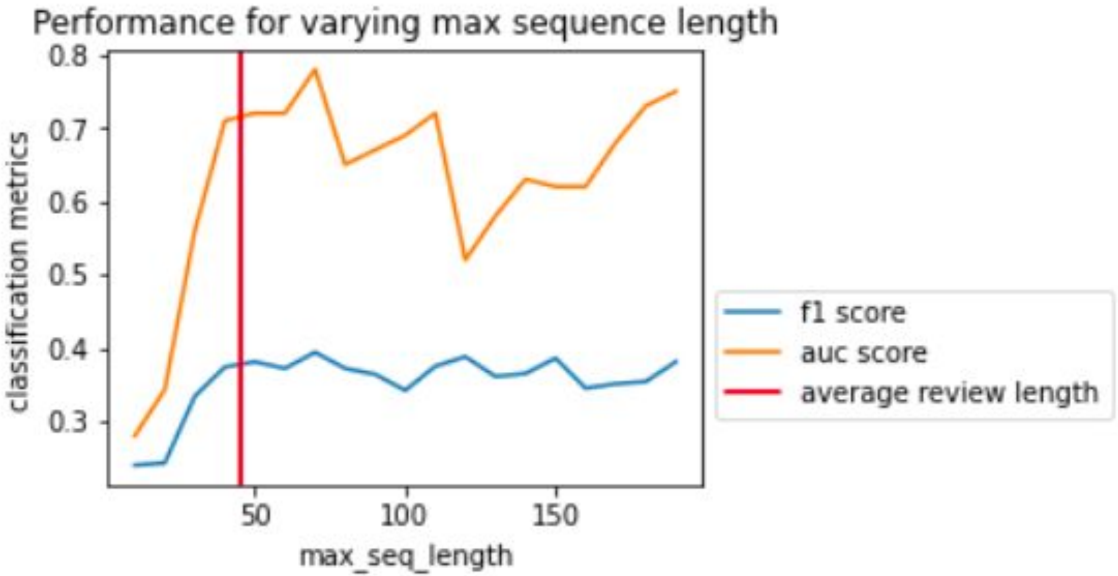
Short-term: month-to-month prediction

Models	Hyperparameters	R^2 score
XGB	learning rate = 0.05 #estimators = 100	0.95
RF	max depth = until leaves are pure #estimators = 500	0.94
BoW	model: bigram TF-IDF regularization = L1 penalty strength = 0.1	0.14
BERT	max sequence length = 256 embedding size = 40	0.16

Long-term: success after one year

Models	Hyperparameters	F-1 score
XGB	learning rate = 0.2 #estimators = 100	0.37
RF	max depth = 5 #estimators = 200	0.34
BoW	model: unigram TF-IDF regularization = L2 penalty strength = 0.5	0.34
BERT	max sequence length = 128 embedding size = 10 dense layers = 2	0.39

BERT performance vs. Max Sequence Length



★ Performance plateaus after max sequence length exceeds the average review length

Long-term: success after one year

Models	F1	Accuracy	AUC	Precision	Recall
Ensemble (logistics regression)	0.437	0.835	0.805	0.557	0.360
Ensemble (decision tree)	0.536	0.839	0.786	0.548	0.524
BERT	0.438	0.872	0.755	1.00	0.281
Bag-of-Words	0.321	0.762	0.670	0.325	0.317
XGBoost	0.388	0.829	0.758	0.532	0.305
Random forest	0.402	0.832	0.759	0.547	0.317

- ★ the ensemble models outperform all individual models: non-text-based models and the text-based models have different advantages
- ★ Neither ensemble models might be ideal dependings on the relevant application

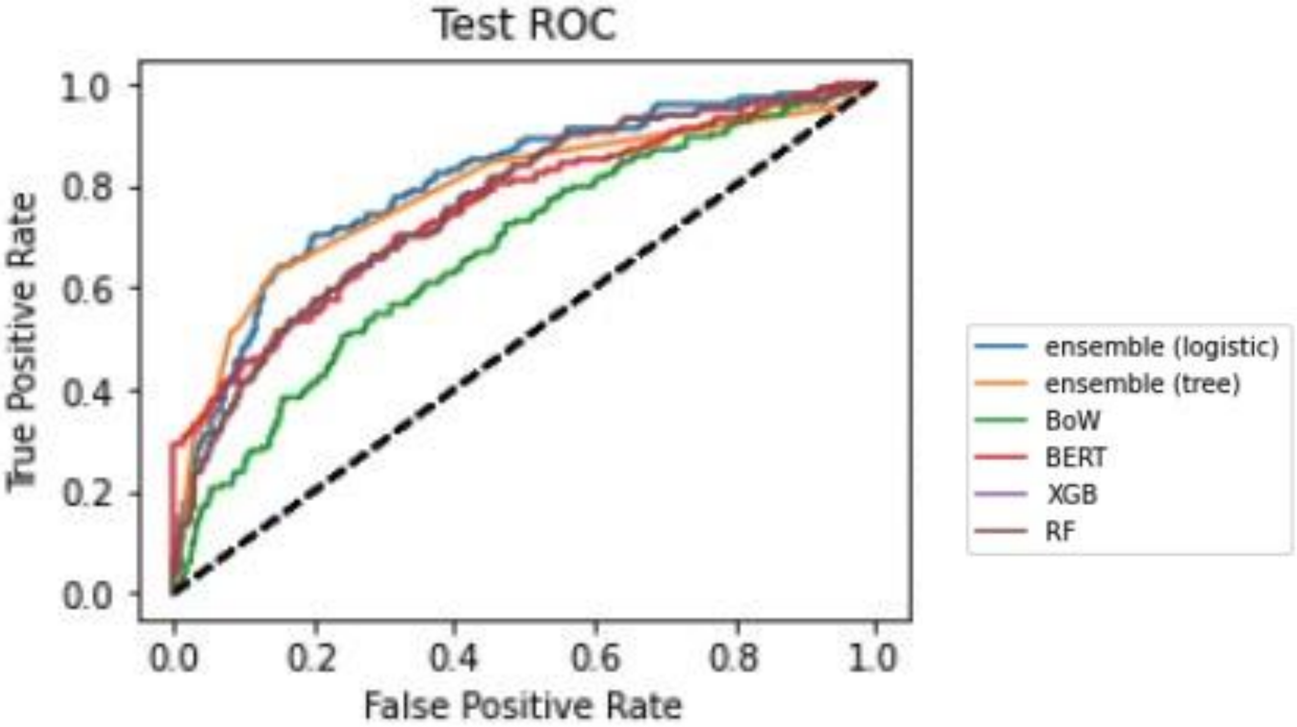
Long-term: success after one year

Models	F1	Accuracy	AUC	Precision	Recall
Ensemble (logistics regression)	0.437	0.835	0.805	0.557	0.360
Ensemble (decision tree)	0.536	0.839	0.786	0.548	0.524
BERT	0.438	0.872	0.755	**1.00**	0.281
Bag-of-Words	0.321	0.762	0.670	0.325	0.317
XGBoost	0.388	0.829	0.758	0.532	0.305
Random forest	0.402	0.832	0.759	0.547	0.317

★ the ensemble models outperform all individual models: non-text-based models and the text-based models have different advantages

★ Neither ensemble models might be ideal dependings on the relevant application

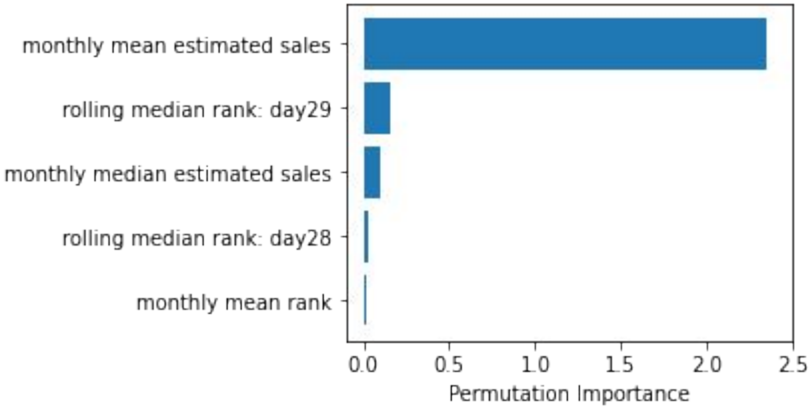
Long-term: success after one year



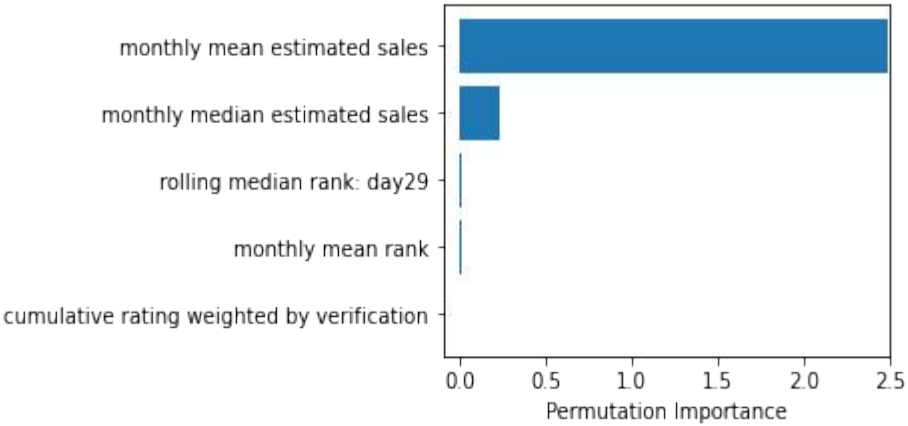
Model Interpretations

What Predicts Short-Term Success? Momentum!

XGB



RF

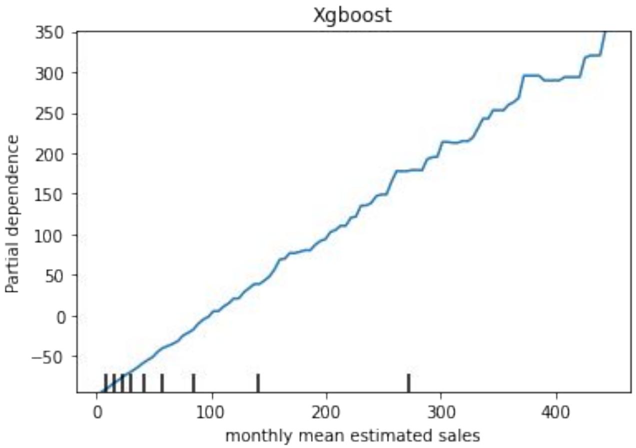


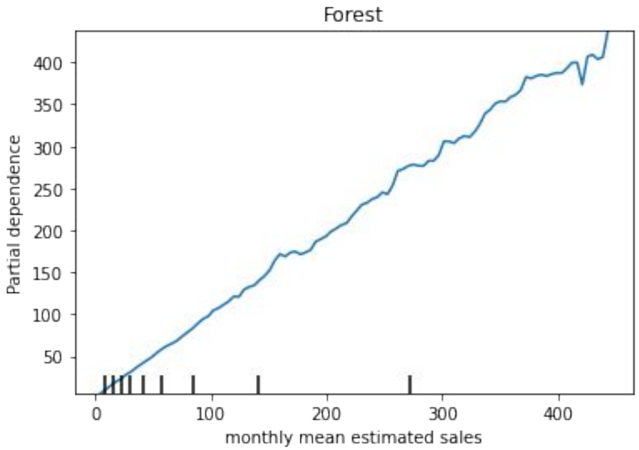
★ Monthly mean sales volume is by far the most important feature.

★ Non-text models make predictions based only on past performance

What Predicts Short-Term Success? Momentum!

★ Partial dependence plot: Monthly mean sales volume

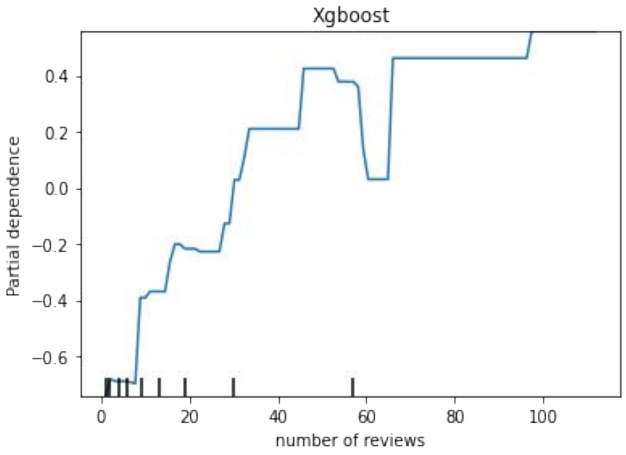


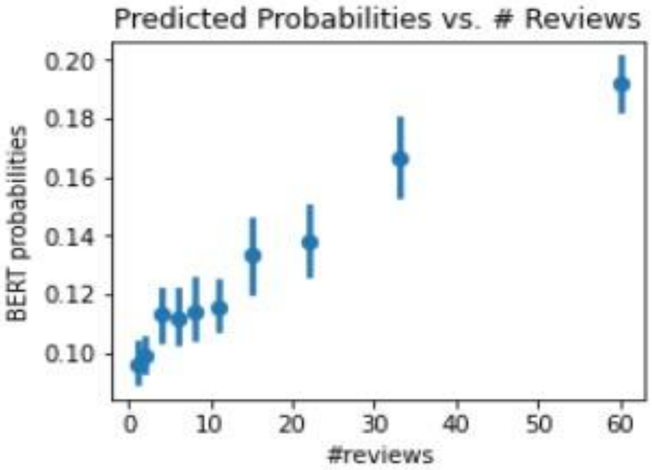


★ Linear relationship means non-text models predict based on momentum in the short-term, as expected

What Predicts Long-Term Success? #Reviews!

★ **Number of reviews** is the 6th most important feature in XGB





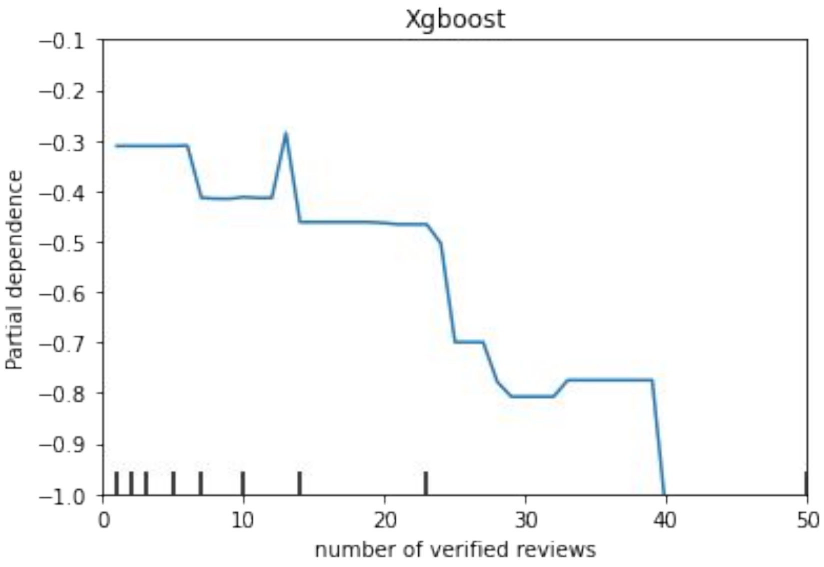
★ The pattern is consistent between XGB and BERT

What Predicts Long-Term Success? #Reviews!

★ **Number of verified reviews** is the 5th most important feature in XGB.

★ Fake reviews:

- Vendors buy their own products right after launch to get positive verified reviews
- More fake reviews means lower probability of success

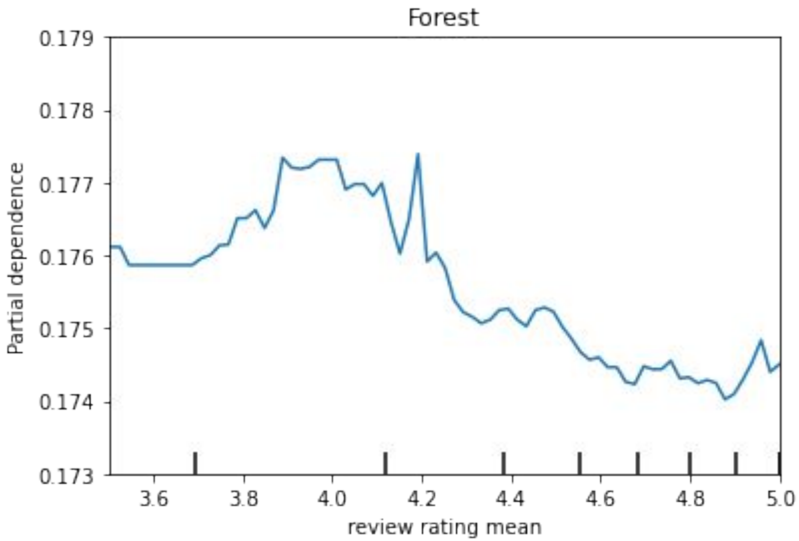


What Predicts Long-Term Success? Review Rating!

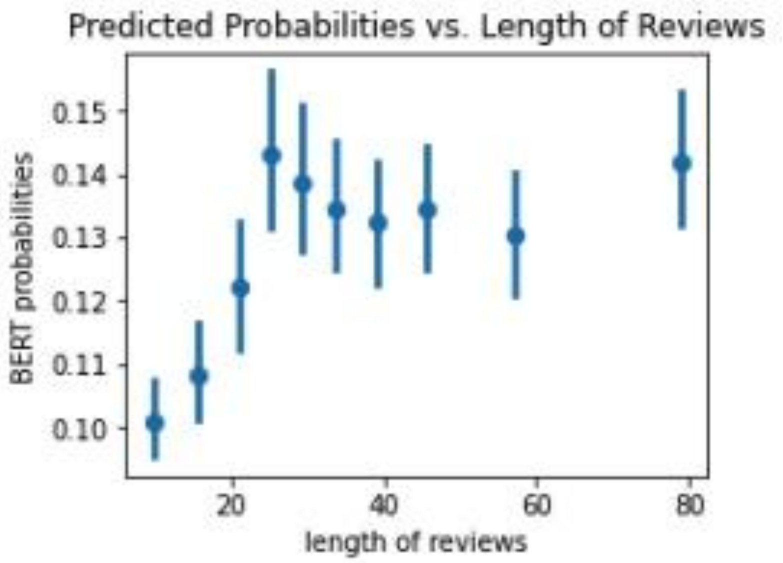
★ **Review rating** is the 3rd most important feature in RF

★ Relationship analogous to #verified reviews:

- fake to have all 5-star reviews
- 4-star reviews look more realistic



What Predicts Long-Term Success? Review Length!



★ If review is too short, no signal for BERT to make a positive prediction

★ Long reviews are usually complaints

What Predicts Long-Term Success? Content!

- **Ingredients → Success**
  - Ingredients are associated with positive coefficients in BoW-based regression model:
    - Important for vitamin products
    - e.g. fiber, acv (apple cider vinegar), b12, fish, oil, turmeric, coffee, elderberry, and enzymes
- **Sense of healthiness → Success**
  - Frequent occurrence of “natural,” “organic,” and “kids” = BERT positive predictions
- **Negations → Failure**
  - Frequent occurrence of “didn’t” and “doesn’t” = BERT negative predictions

References

- “**Using NLP to extract quick and valuable insights from your customers’ reviews**,” https://medium.com/artefact-engineering-and-data-science/customer-reviews-use-nlp-to-gain-insights-from-your-data-4629519b518e
- “**Predicting Sales from the Language of Product Descriptions**,” https://nlp.stanford.edu/pubs/pryzant2017sigir.pdf
- “**Amazon Product review Sentiment Analysis using BERT**,” https://www.analyticsvidhya.com/blog/2021/06/amazon-product-review-sentiment-analysis-using-bert/
- “**Sentiment Analysis of Movie Reviews with Google’s BERT**,” https://medium.com/mlearning-ai/sentiment-analysis-of-movie-reviews-with-googles-bert-c2b97f4217f
- “**Amazon Best Seller Rank**,” https://www.sellerapp.com/amazon-best-seller-rank.html

Questions?