What Do People Say: Product Review Analysis with Pattern

Sehaj Chawla, Taro Spirig, Lotus Xia, Heather Liu
Harvard University

Motivation

Reviews are important! They give insights into customer sentiments, which can help businesses predict product performance and adapt their marketing strategies

Problem Statement

Task 1: Predict product future sales performance (short-term and long-term). In particular, can review texts improve prediction in addition to metadata?

Task 2: Gain insights on what keywords or topics are predictive of sales

Data

> 9000 products in Amazon's Vitamins and Dietary Supplements category from July 2017 to July 2021

Review history data: ~ 5 million reviews including title, text, and metadata such as dates, ratings, verified purchase, etc.

Amazon best seller rank (BSR) data: Daily rank of product based on current and historical sales

Estimated sales volume data: 1 to 1 mapping from BSR to sales volume

Target Variables

Short-term: Monthly Median Sales

Long-term: Successful Product, a product is successful if it ever reaches the top 3000 BSR during the one year period one year after launch

Model Classes

★ **Non-text models:** XGboost (XGB) and random forest (RF)

★ **Text-based Models:** BoW-based regression and BERT-based regression
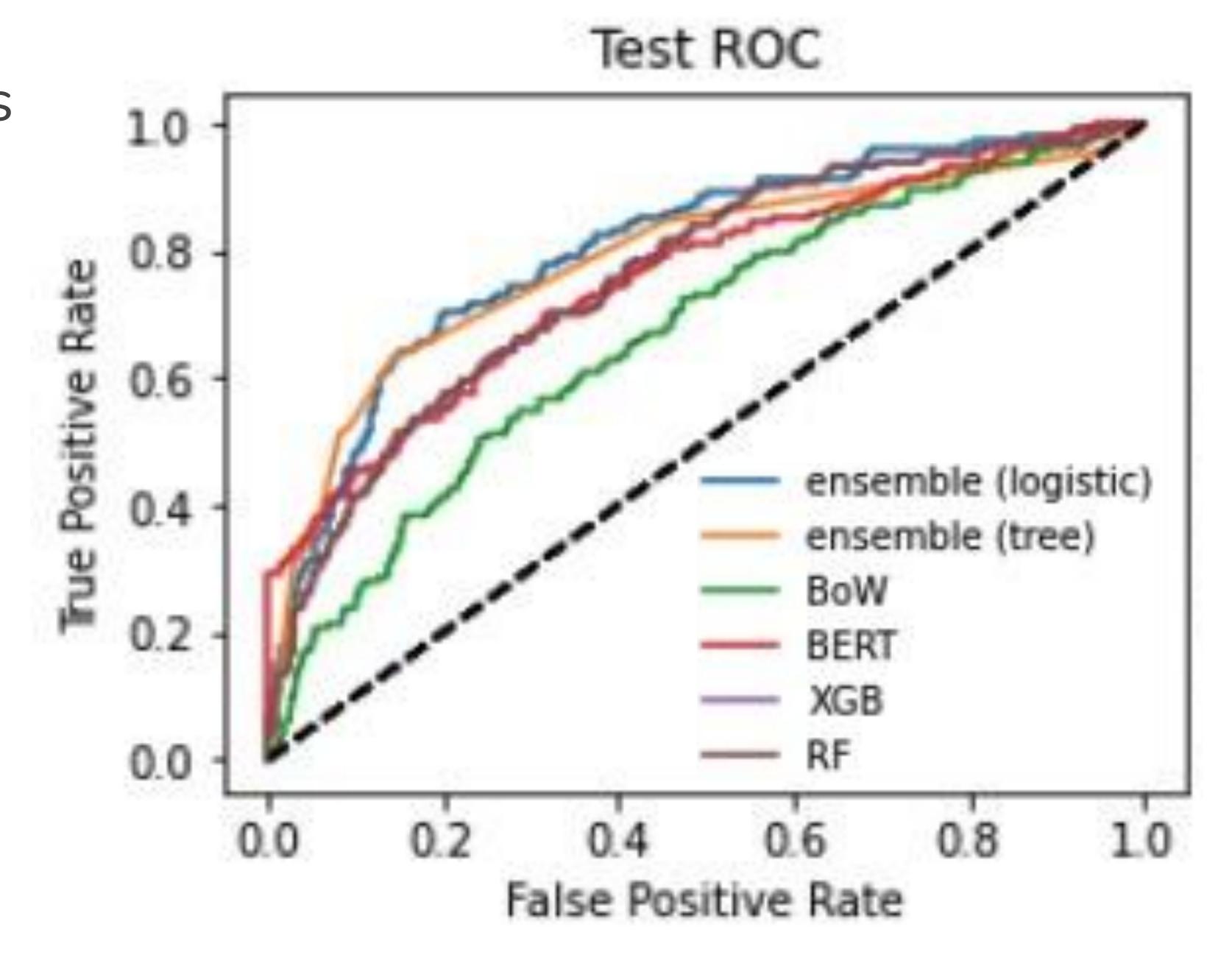
Short-term: month-to-month sales

Feature variables based on all reviews up until current month

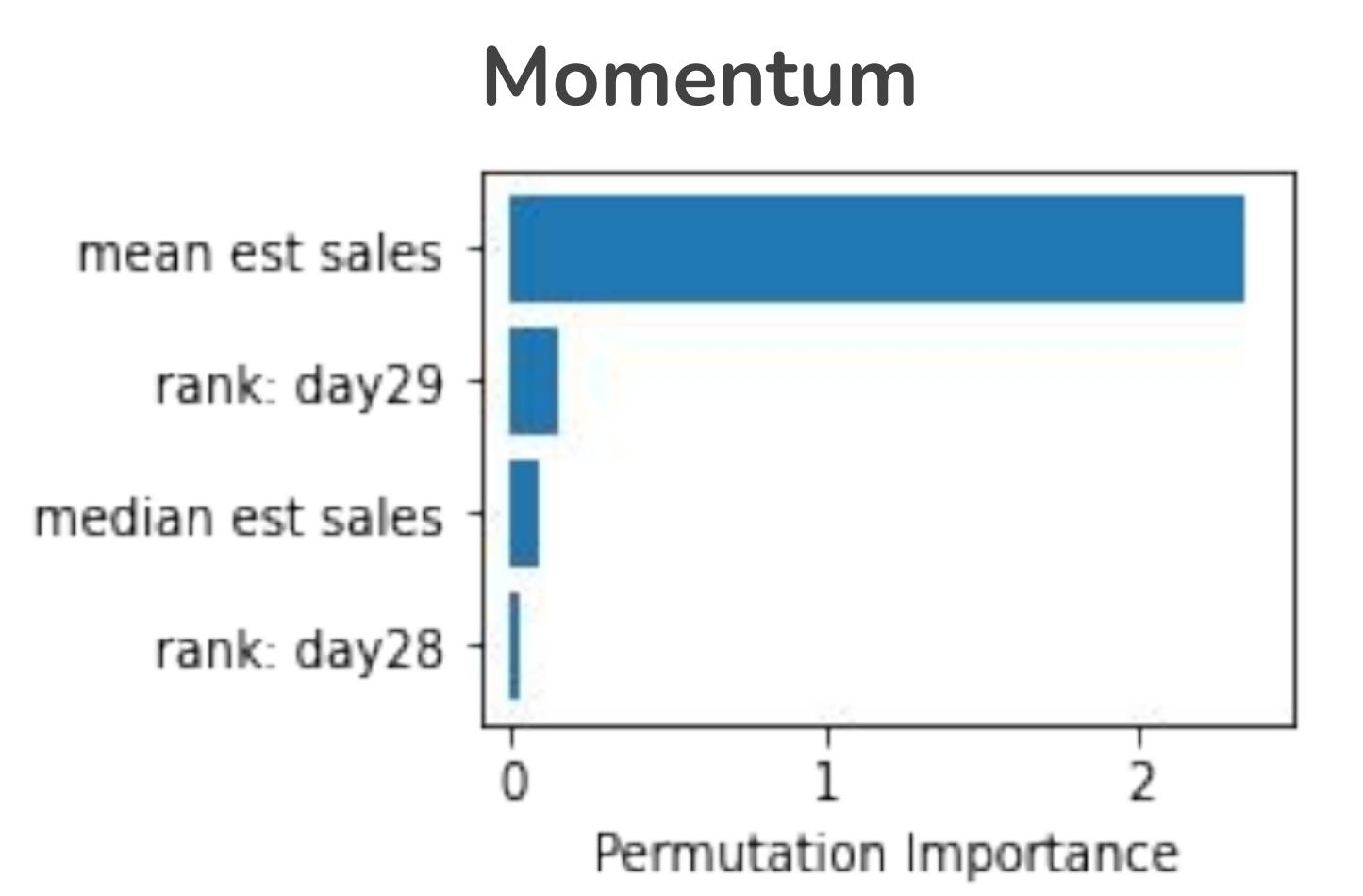| Models	Hyperparameters	R2 score
XGB	learning rate = 0.05; #estimators = 100	0.95
RF	max depth = until leaves are pure; #estimators = 500	0.94
BoW	model: bigram TF-IDF; regularization = L1 with 0.1 penalty	0.14
BERT	max sequence length = 256; embedding size = 40	0.16

Long-term: success after one year

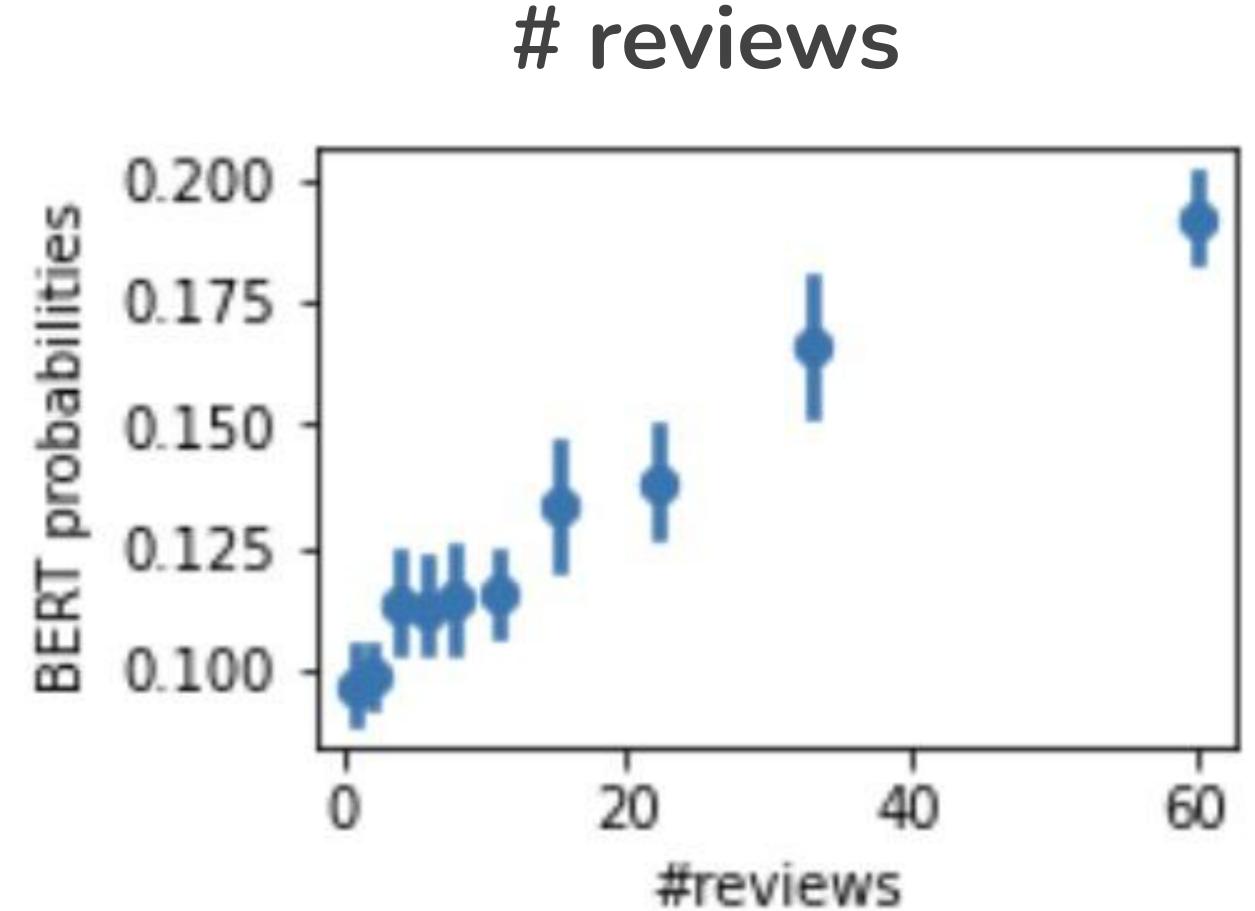Feature variables based on reviews in the first 3 months

Models	F1	Acc	AUC	Prec
Ens (reg)	0.44	0.84	0.81	0.56
Ens (tree)	0.54	0.84	0.79	0.55
BERT	0.44	0.87	0.76	1.00
BoW	0.32	0.76	0.67	0.33
XGB	0.39	0.83	0.76	0.53
RF	0.40	0.83	0.76	0.55

Test ROC
- ensemble (logistic)
- ensemble (tree)
- BoW
- BERT
- XGB
- RF
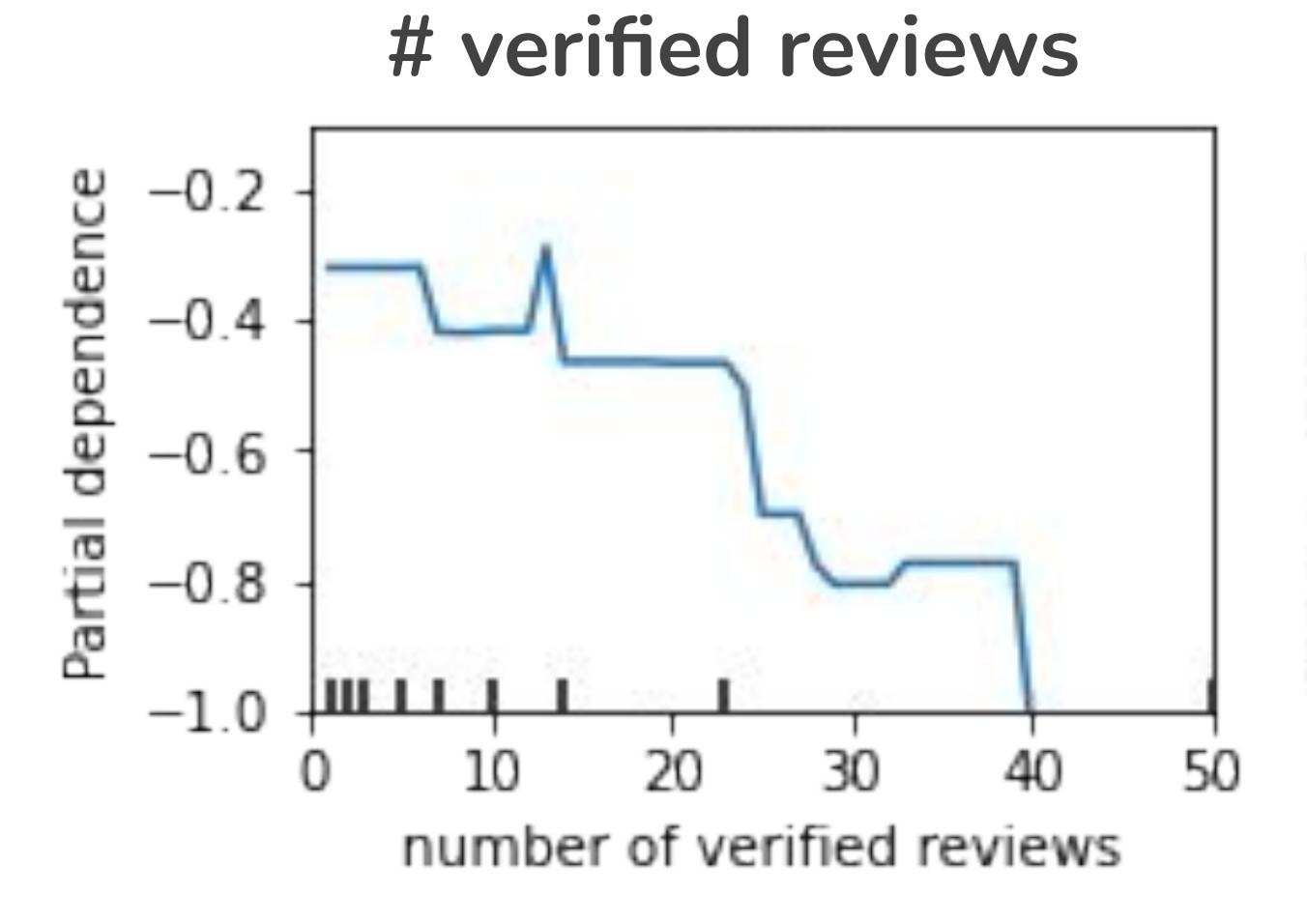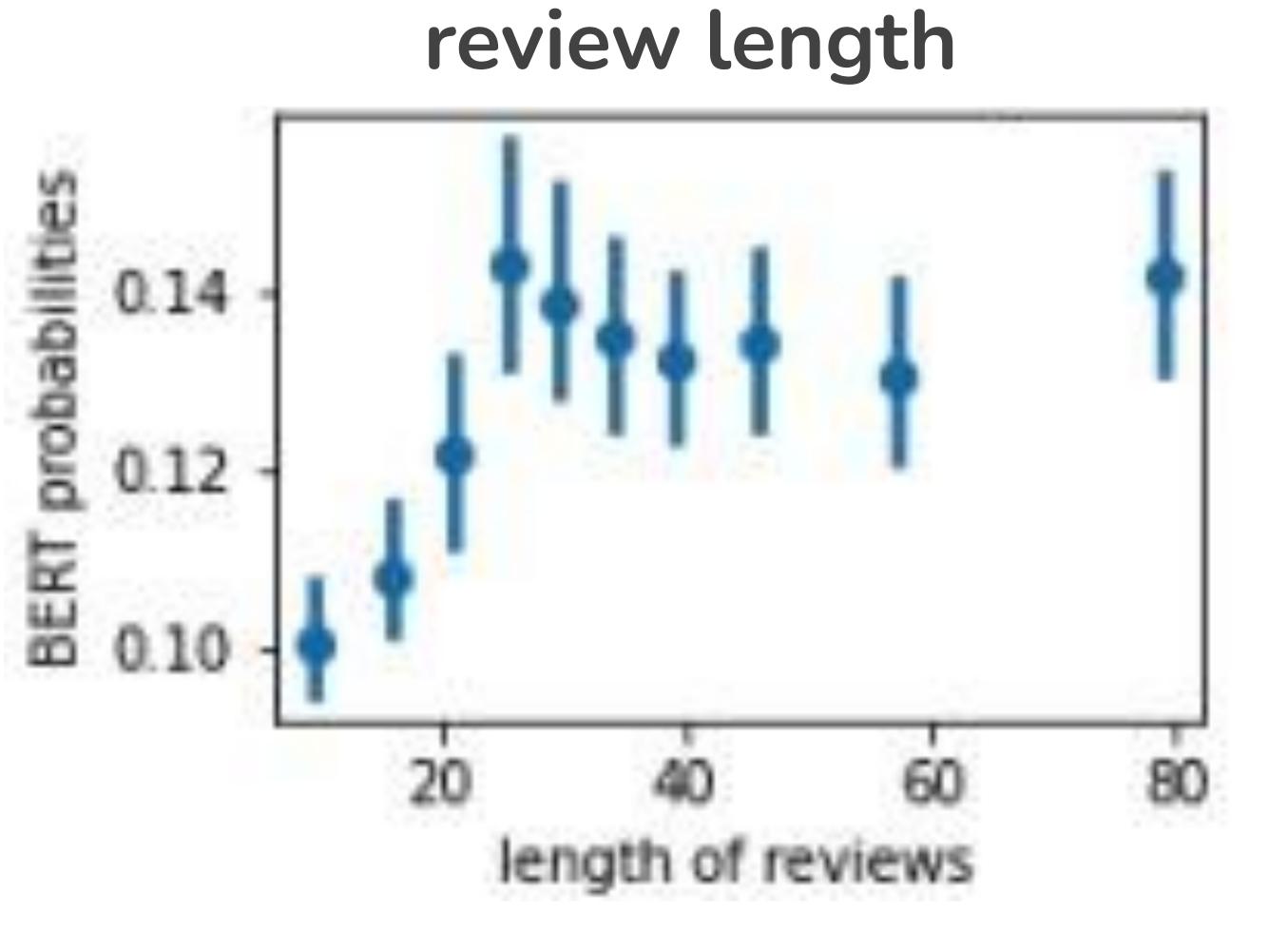
Short-term Interpretation

Momentum

Long-term Interpretation

reviews

verified reviews

review length

review content

★ Negations not good (BERT)
  ○ e.g. didn't, doesn't

★ Ingredients are good (BoW)
  ○ e.g. fiber, ACV, B12, fish oil

★ Healthiness is good (BERT)
  ○ e.g. natural, organic, kids