INSTITUTE FOR APPLIED COMPUTATIONAL SCIENCE
AT HARVARD UNIVERSITY

Product Review Analysis With pattern

Milestone 2

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

Agenda

1. Motivation and Problem Statement
2. Data Insights
3. Literature Reviews
4. Models

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**
  - 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**

Data Description

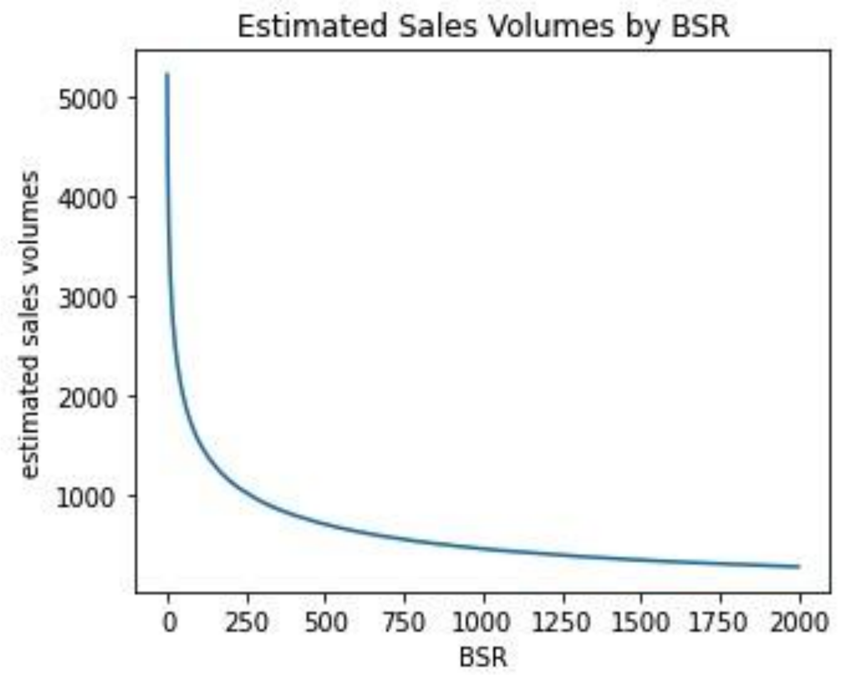
~9000 products in Amazon's Vitamins and Dietary Supplements category

Time range: 2017-07 to 2021-07

- ★ **2 measures of sales performance:**
  - **Amazon best seller rank (BSR)**
    - Daily rank of each product based on 1) current sales and 2) sales history
    - Lower rank means better sales performance
  - **Estimated sales volume**
    - 1 to 1 mapping from rank to sales volume
    - A lower rank corresponds to a higher sales volume

BSR vs. Estimated Sales Volumes

★ Pattern has provided **estimated sales volumes** corresponding to each rank



Data Description

~9000 products in Amazon's Vitamins and Dietary Supplements category

Time range: 2017-07 to 2021-07

- ★ **Predictors: Review data**
  - All reviews under a product at the time of scraping
  - Review title and review text
  - Metadata such as review dates, review ratings, verified purchase, etc

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 sales volume predictions)**
        - Elizabeth Fernandes, Sérgio Moro, Paulo Cortez, Fernando Batista & Ricardo Ribeiro(2020)

Non-Text Regression Models

- ★ **Target variable: monthly median sales volume**
- ★ **Features in the input (everything except the text)**
  - Smoothed daily BSR
  - Mean and median sales volume
  - Mean and median price
  - Cumulative number of reviews

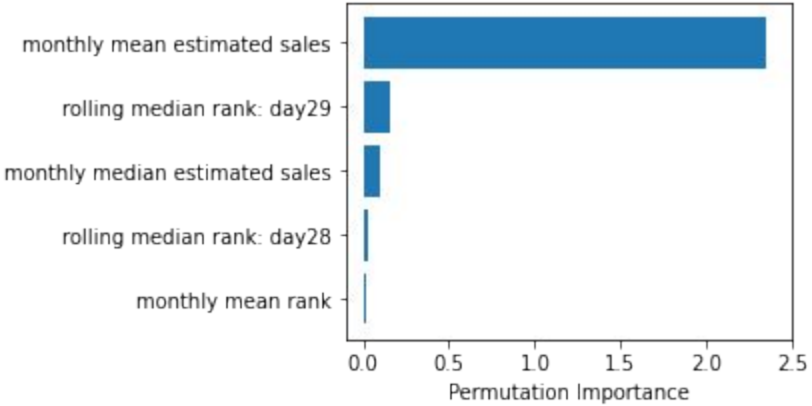
Non-Text Regression Models: Results

- Trained on ⅓ and tested on ⅓ of the products on AWS
- Optimized the hyperparameters of the Xgboost and random forest models
- OOD performance with best hyperparameters:

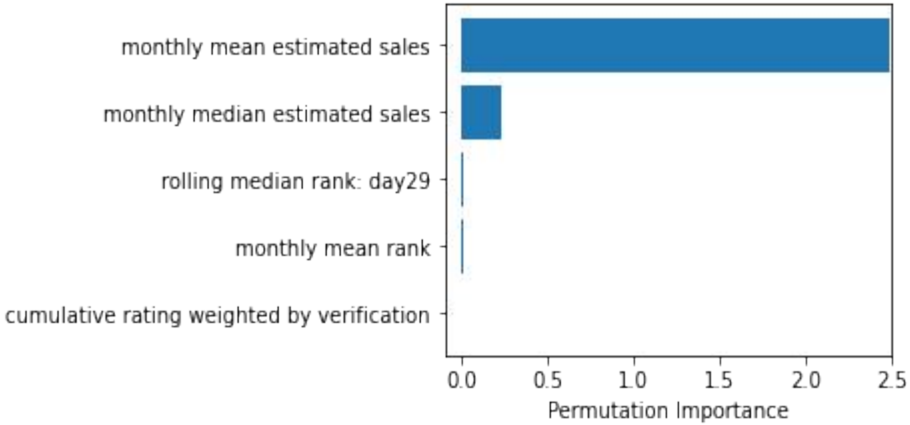
Model	**Hyperparameters**	**R^2**
Linear Regression	———	0.938
XGBoost	learning rate = 0.05 n estimators = 100	0.946
Random Forests	max depth = leaves pure n estimators = 500	0.939

Non-Text Regression Models: Feature Importance

Xgboost

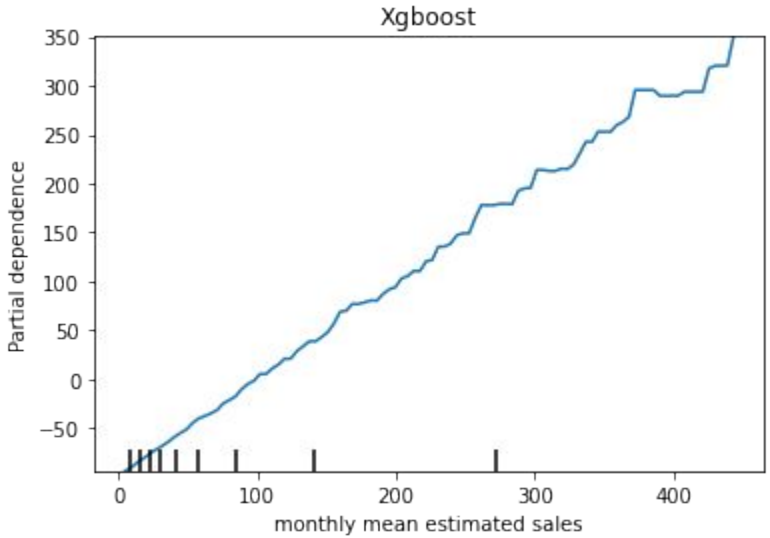


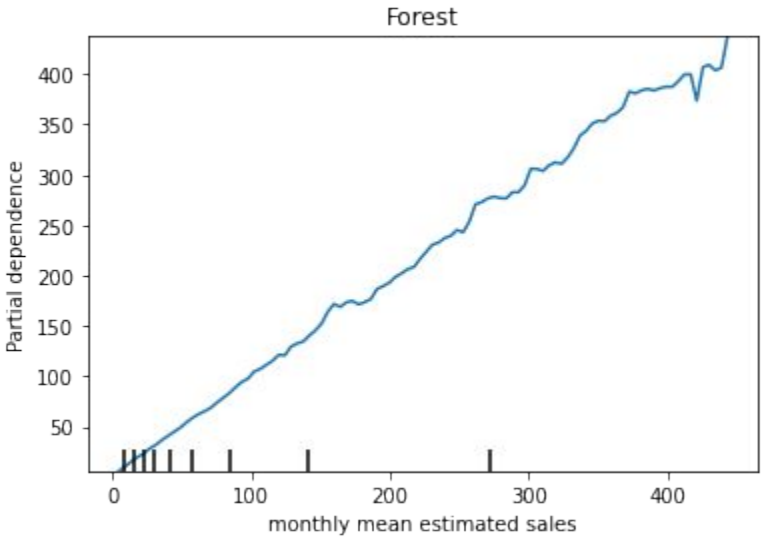
Random Forest



Non-Text Regression Models: Partial Dependence Plots

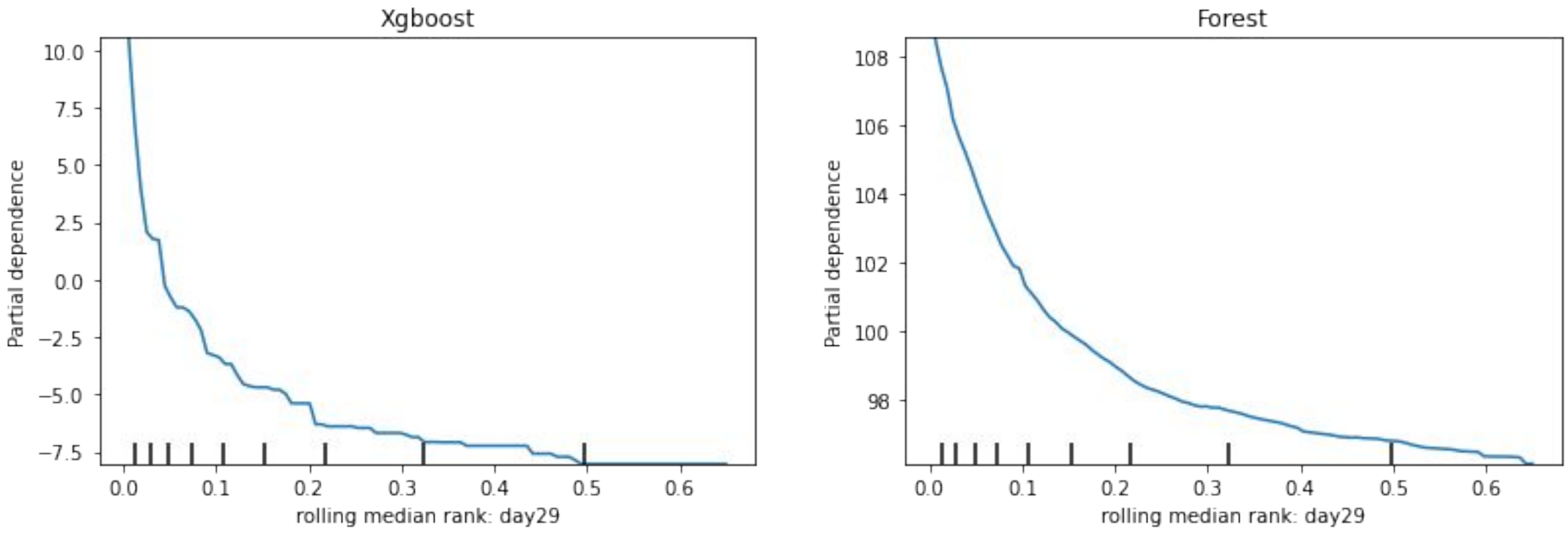
★ Monthly mean sales volume : most important feature for both models





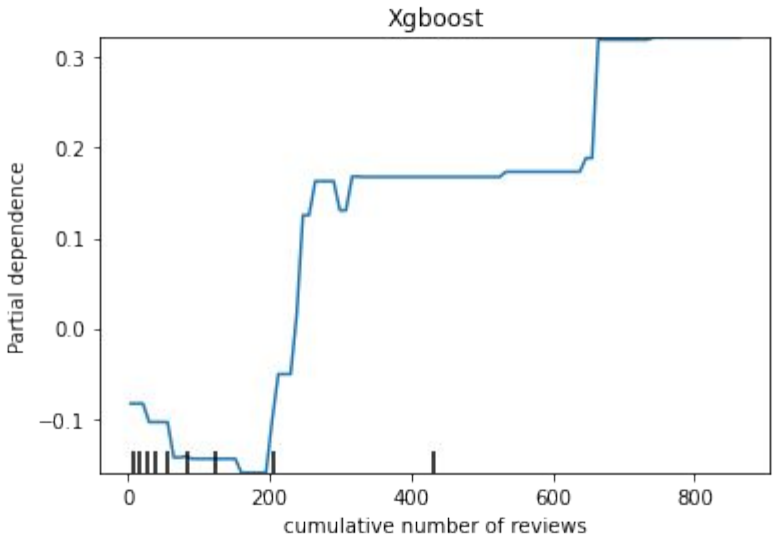
Non-Text Regression Models: Partial Dependence Plots

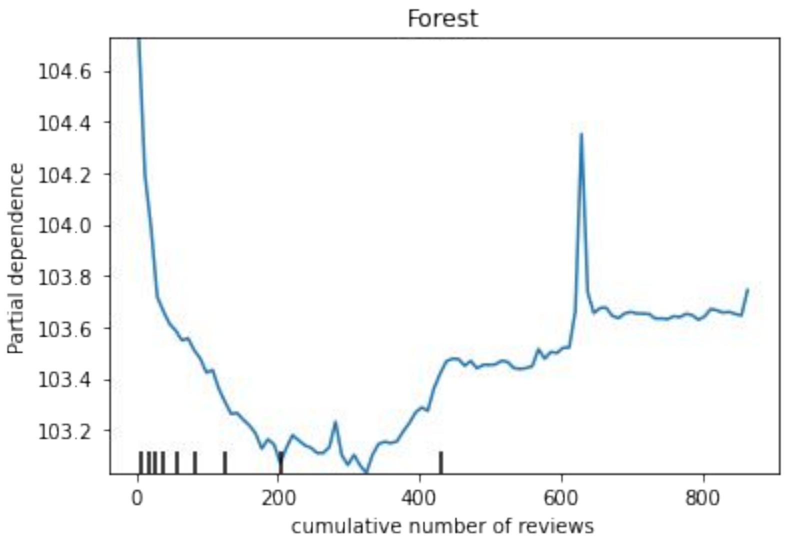
★ Smoothed BSR of 29th day of month (2nd and 3rd most important feature for Xgboost and random forest respectively)



Non-Text Regression Models: Partial Dependence Plots

★ Cumulative number of reviews:





Non-Text Regression Models: future work

- **Future Work**
  - Clean up the models by taking only the most relevant features into account, i.e. have a final benchmark for non-text prediction model
  - Run a model without the past ranking/estimated sales to see how the model performs: is there anything interesting in the non-text review data?
  - Combine the non-text model with the text based model in an ensemble model

Text-Based Regression Models

Attempt to answer the following two questions:

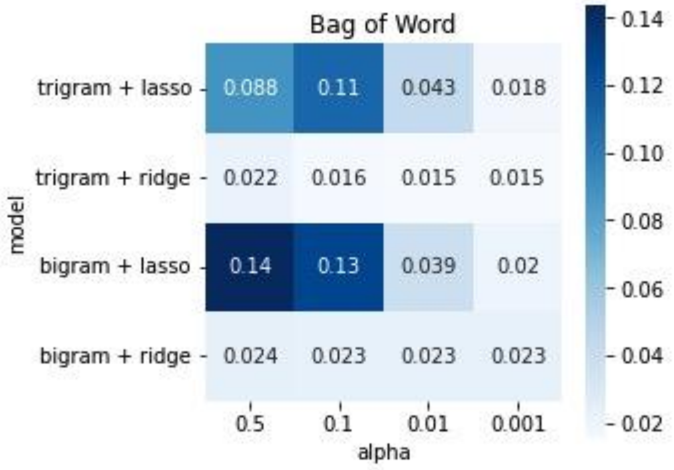
★ In terms of **prediction accuracy**: what more can we do with review texts in addition to the metadata

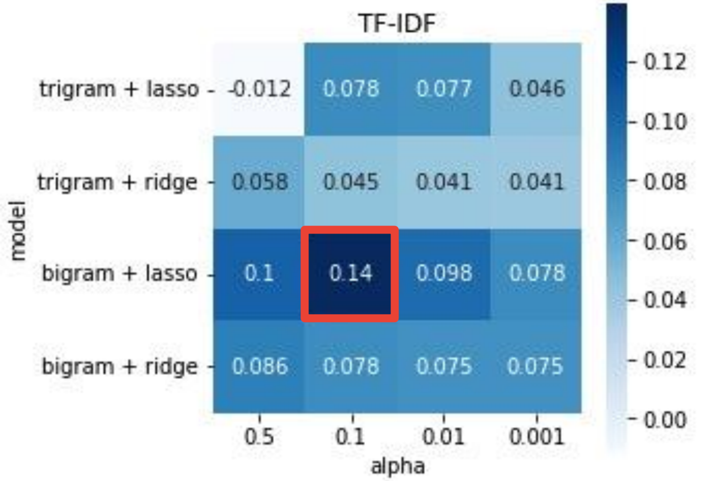
★ In terms of **interpretability**: gain insights on what keywords or topics in review texts are predictive of future sales volumes

Benchmark: Bag of Words

- ★ Target variable:
  - monthly median sales volume
- ★ Text processing:
  - bag of word
  - TF-IDF (term frequency-inverse document frequency)
- ★ Predictor variables:
  - (weighted) frequency of the 500 most common bigrams/trigrams in the training corpus
- ★ Models:
  - LASSO and ridge regression
- ★ **Benchmark model with nice interpretability**

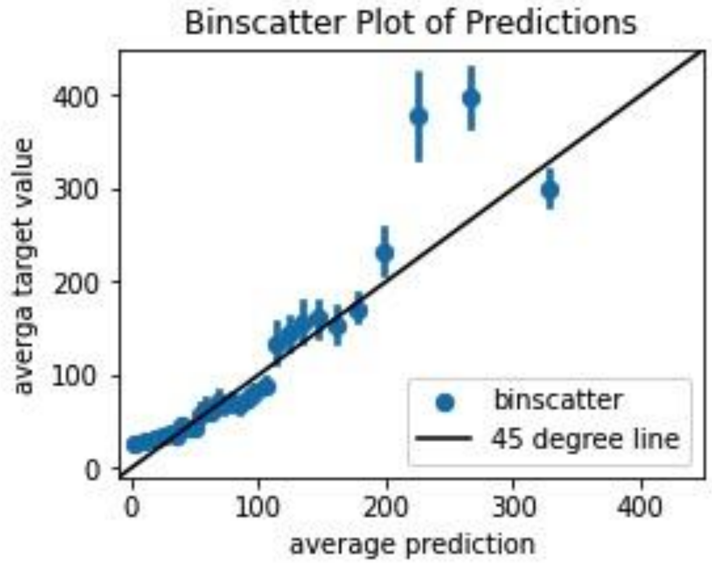
Benchmark: Bag of Words





- ★ Positive words:
  - cider vinegar, **garden life**, bowel movements, flu season, hair nails, taking probiotic, 2nd bottle, taste great, getting sick, stopped taking
- ★ Negative words:
  - joint pain, pain relief

Benchmark: Bag of Words



★ Good, unbiased prediction on observations with low target values.

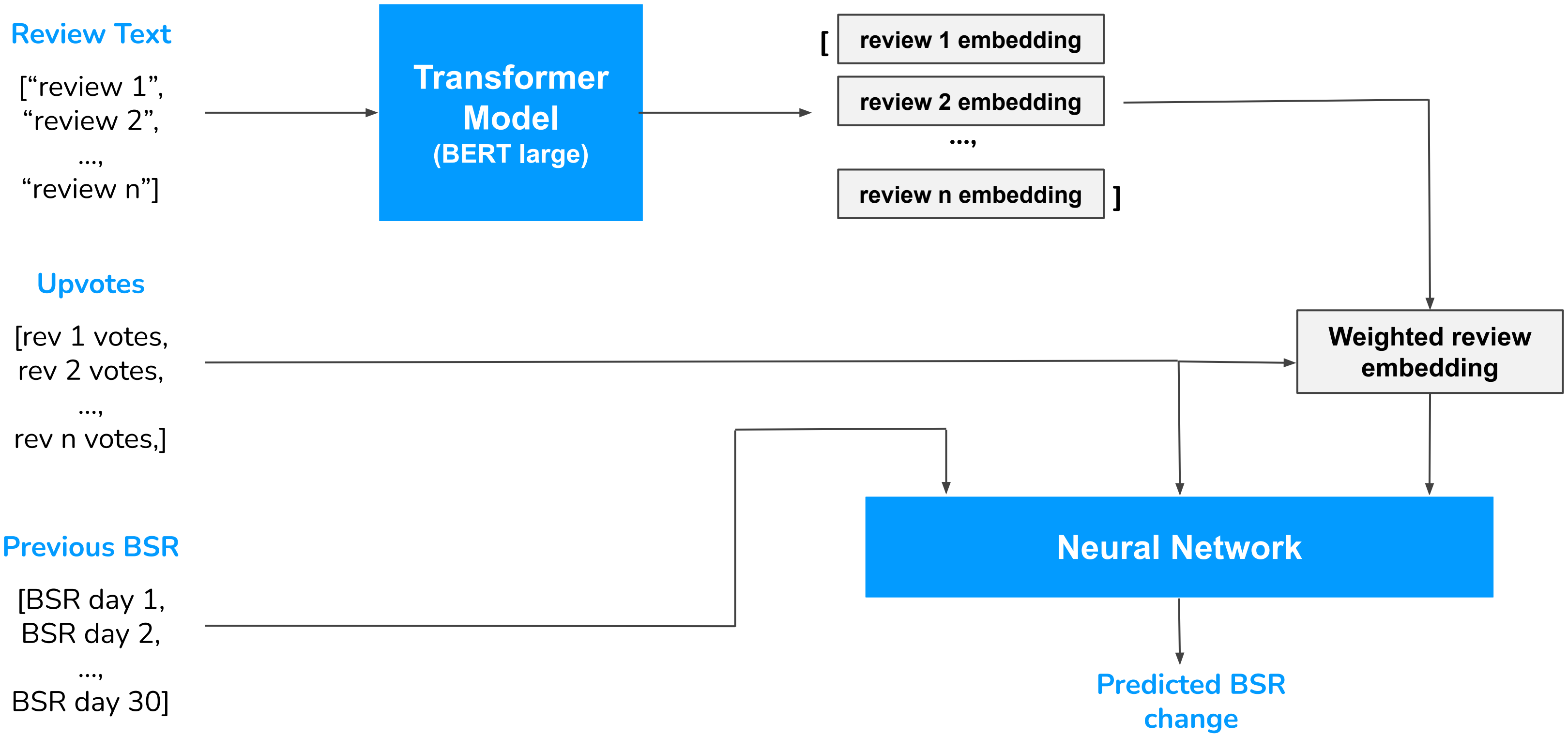
★ Predictions on observations with target values >200 are biased downwards.

Benchmark: Bag of Words

- Future work
  - Experiment with a more complex model to improve the model prediction
    - E.g. pass the bigrams/trigrams through a word embedding and a feed-forward neural network

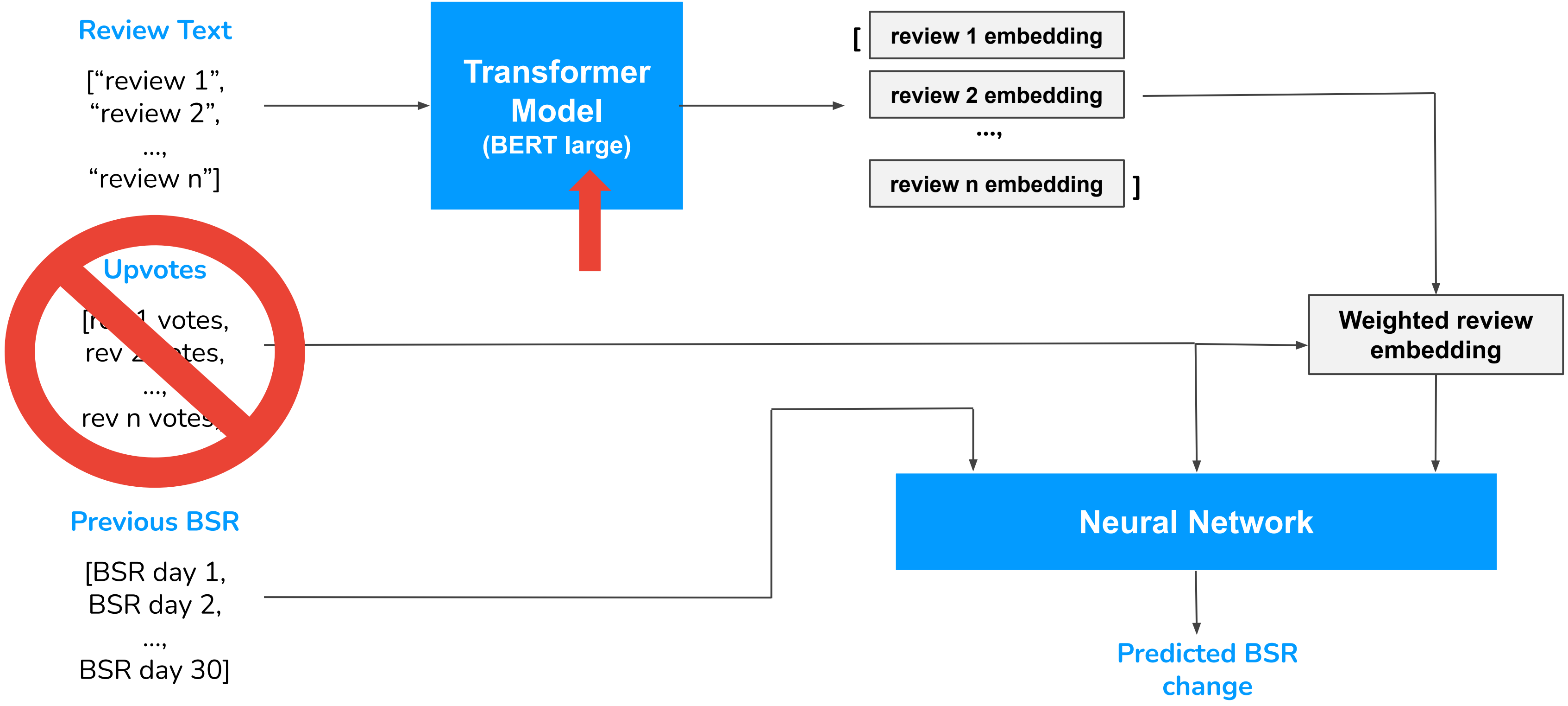
Train Steps: 200
R2: -0.132

BERT Regression: Mark I



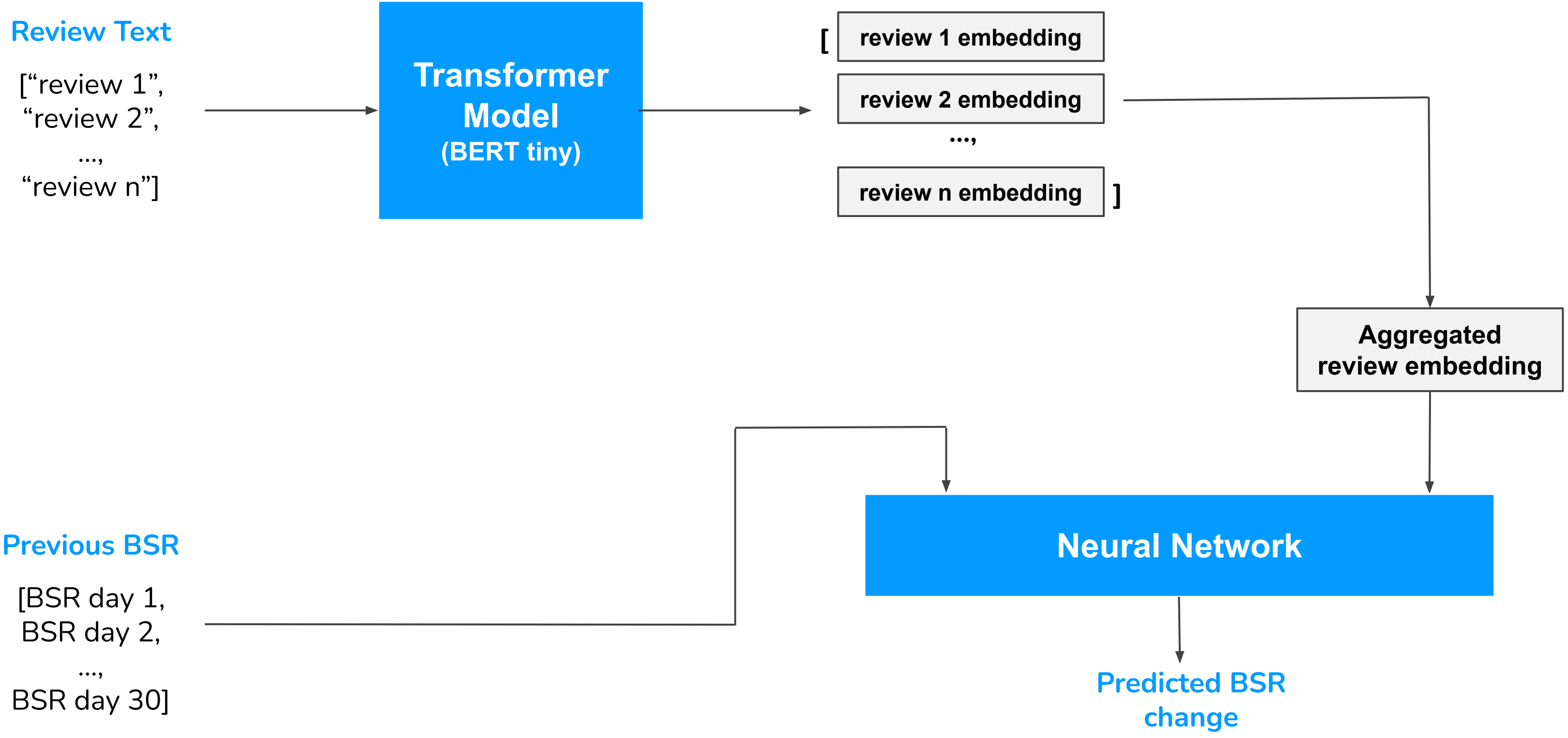
Train Steps: 200
R2: -0.1[illegible]2

BERT Regression: Mark I



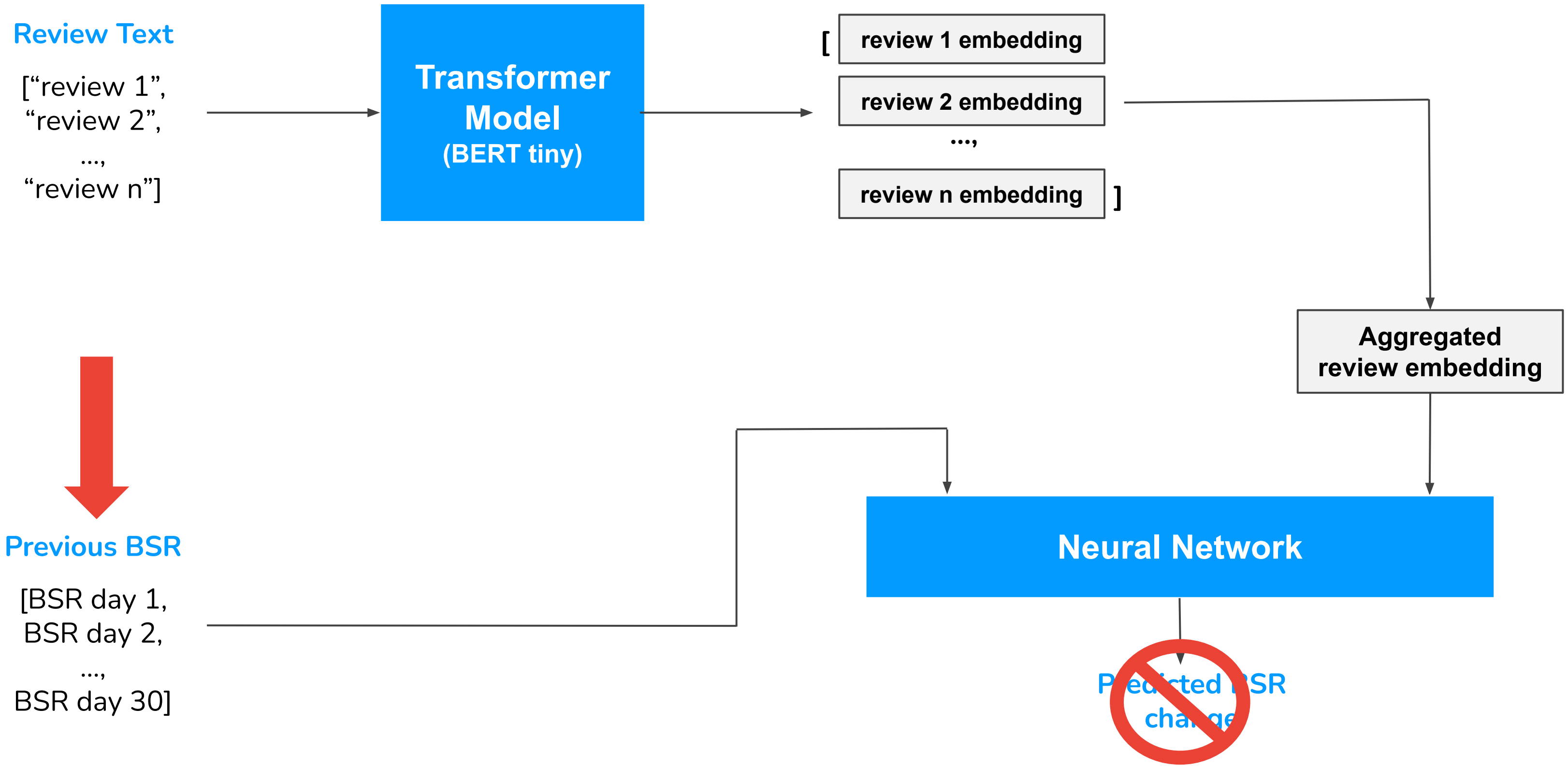
BERT Regression: Mark II

Train Steps: 50,000
R2: 0.256



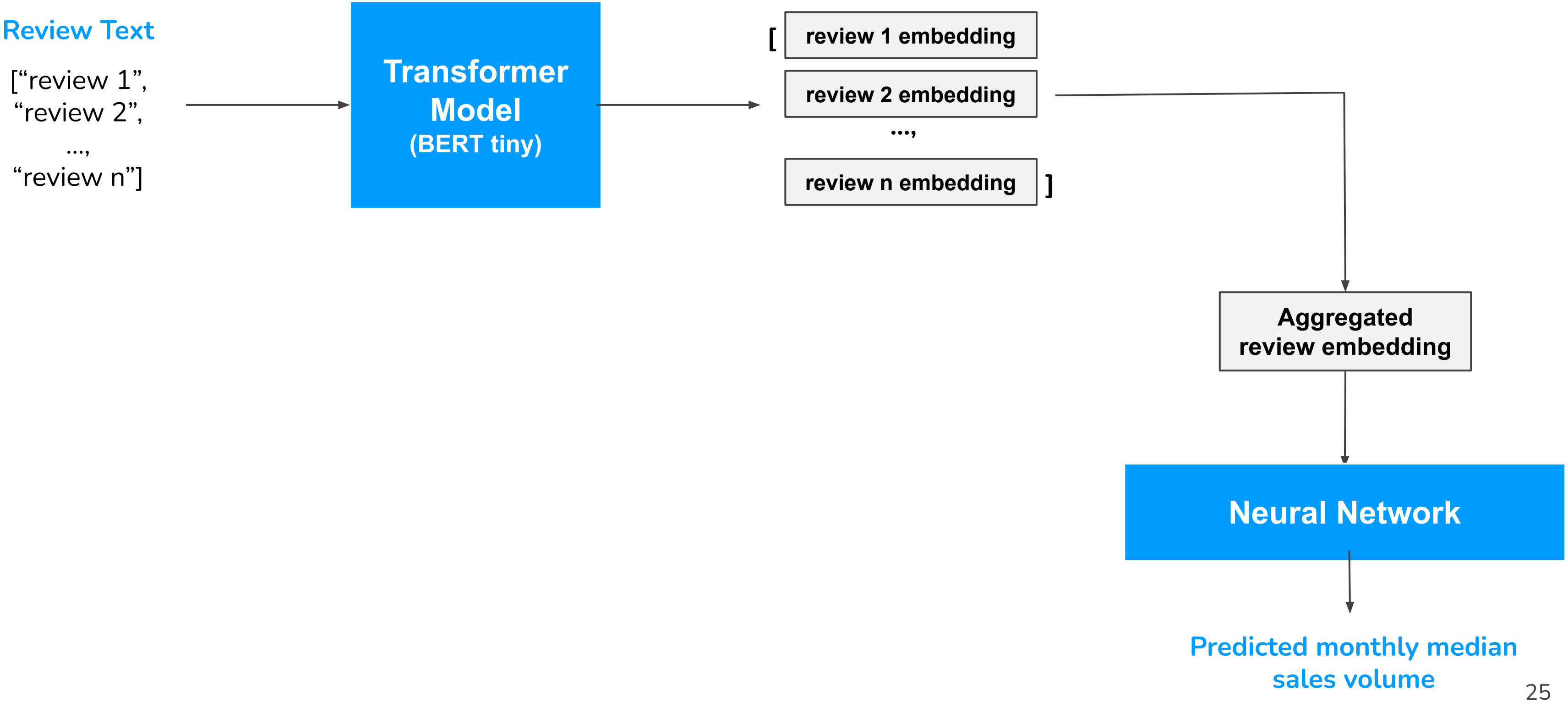
Train Steps: 50,000
R2: 0.256

BERT Regression: Mark II



BERT Regression: Mark III

Train Steps: 50,000
R2: 0.16



Train Steps: 50,000
R2: 0.16

BERT Regression: Mark III

Review Text

["review 1",
"review 2",
...,
"review n"]

Transformer Model
(BERT tiny)

[**review 1 embedding**
review 2 embedding
...,
review n embedding]

Aggregated review embedding

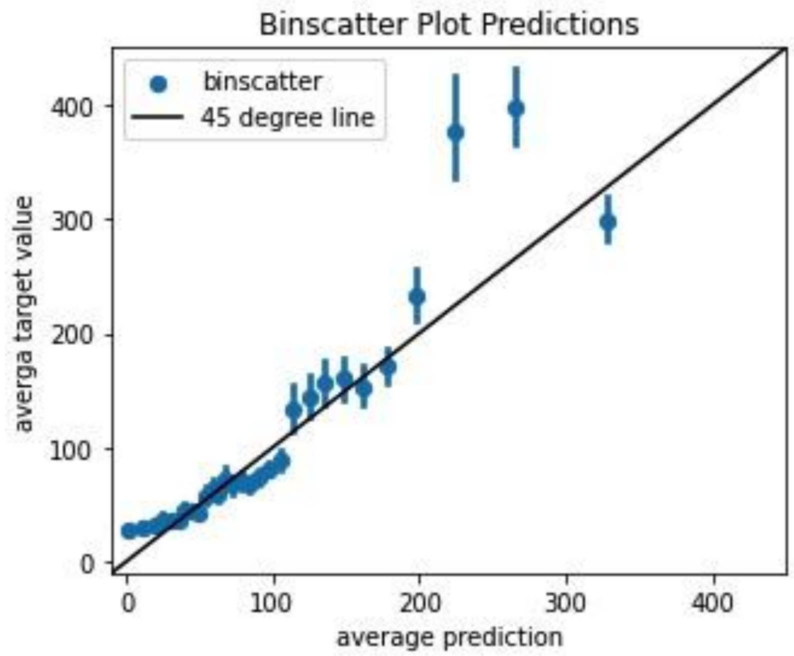
Neural Network

Predicted monthly median sales volume

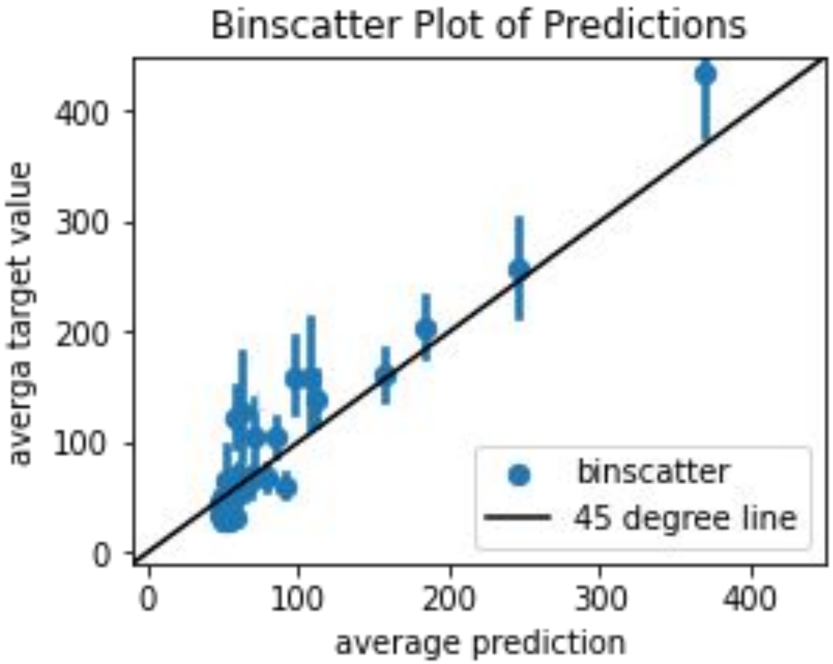
```
==================================================================
Total params: 4,790,946
Trainable params: 8,481
Non-trainable params: 4,782,465
__________________________________________________________________
```

```
Epoch 1/5
10000/10000 [==============================] - 12975s 1s/step - loss: 30291.4727 - mean_absolute_error: 84.8701 - r_square: -0.213
Epoch 2/5
10000/10000 [==============================] - 16966s 2s/step - loss: 30277.1465 - mean_absolute_error: 86.3095 - r_square: -0.003
Epoch 3/5
10000/10000 [==============================] - 16954s 2s/step - loss: 31137.8281 - mean_absolute_error: 86.6345 - r_square: 0.018
Epoch 4/5
10000/10000 [==============================] - 16810s 2s/step - loss: 30385.1992 - mean_absolute_error: 86.3509 - r_square: 0.115
Epoch 5/5
10000/10000 [==============================] - 17087s 2s/step - loss: 30913.2793 - mean_absolute_error: 87.2239 - r_square: 0.139
```

BERT Regression



Bag of Words



BERT

BERT Regression

- ★ **Future Work**
  - Increase the number of trainable parameters (by increasing the size of the latent review descriptors)
  - Train for longer epochs, that allow the model to see each training point more than once.
  - Dive deeper into the interpretation, and identify corpuses that BERT performs poorly on - some examples include partitioning on number of reviews, lengths of reviews, positive/negative reviews.

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

Question?

Appendix

- ★ Naive prediction using average target value in training set has R^2 of -0.0011.