

Fake Review Classification

Based on *Creating and detecting fake reviews of online products* by Salminen et al, 2022

Can you beat our model?

Guess whether these reviews are real or fake!

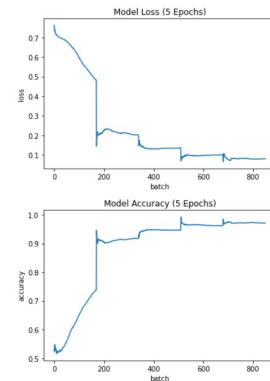
1. Excellent quality, and Amazon had a good deal on the item. The only reason I gave it 4 stars is that the plastic parts are very small, and the cover is very flimsy.
2. Haven't put this up yet because I want to paint them. I know I will like it when I get through and free up more space in my cabinets.
3. Is s good strong suction vacuume. World's good an the carpet also.
4. I love these sheets and just bought a second set for my son. Very soft and they hold up great.

1. Our model predicts: Fake, Answer: Real, 2. Our model predicts: Real, Answer: Fake, 3. Our model predicts: Real, Answer: Real, 4. Our model predicts: Fake, Answer: Fake

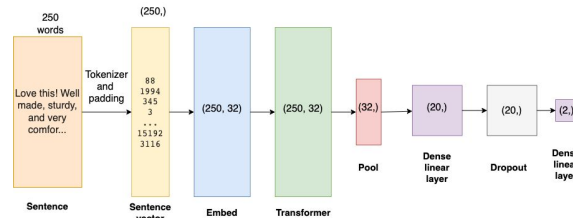
RESULTS

Our results were strong and conclusive. Using our custom-built FRDetector Model, we were able to achieve a surprising peak test accuracy of 0.9227, or 92.27%, with just 5 epochs. This is similar to the fine-tuned RoBERTa model's 96.64% accuracy.

The FRDetector model loss and accuracy through training batches can be seen to the right:



MODEL ARCHITECTURE



METRICS

Model	Test Loss	Test Acc.	Train time
fakeRoberta	0.1407	96.64%	2:34:16
FRDetector	0.2339	92.27%	0:54:12

DISCUSSION

Our model certainly has its limitations. Beyond not being as accurate as the original model, there are concerns with the data itself as well. The data used in our training for the Amazon reviews are all in English and thus would not support an international audience or community, and the computer-generated fake reviews used to train the model are ultimately based off of mimicry and conglomerated information rather than creativity, meaning that it's entirely possible the model starts failing quickly with even small shifts in review "culture" or any kind of creative change.

INTRODUCTION

Consumers looking for products to purchase online often decide which product to buy based on the number and quality of reviews that are left. Fake review detection has been an important task for companies as the purchase of fake reviews undermines the value of a review-based system.

Our study implements a similar framework as that presented in the paper "Creating and detecting fake reviews of online products" by Salminen et al (2022). In the paper, they fine-tune the RoBERTa model to accurately identify real reviews from Amazon products and fake reviews generated by GPT-2. Our implementation includes the transfer of code-base from PyTorch to Tensorflow of the same fine-tuning process, as well as the comparison of our own simpler transformer model, FRDetector, with the fine-tuned equivalent presented in the paper.

METHODS

Data: The dataset contains 20,000 real reviews (Amazon Review Dataset) and 20,000 fake reviews (fine-tuned OpenAI GPT-2 generated) containing the text, label, and review category.

Preprocessing: In fine-tuning the RoBERTa model, the Transformers library was used to tokenize and pad the data. The embedding matrix was downloaded and words were converted to integer ids to be inputted into the model. In the case of our own transformer model, the same tokenizing and padding took place using TF Keras preprocessing tools.

Architecture: We compared the performance of two different architectures: the pretrained TFRoberta Transformer model, and our own custom architecture FRDetector, which was a more simplified transformer architecture and used multi-headed attention.