BIAS ESTIMATOR AND ANALYZER OF SENTIMENT TENDENCY (BEAST) ENGINE: AN ONLINE NEWS ARTICLE BIAS EVALUATING APPLICATION

Ben Moss

March, 2024

General Assembly: DSI 1211

Link: https://beastengine.streamlit.app

Introduction

- **How can bias and sentiment be identified from abstracts and headlines of news articles?**
- **Novel methods for determining bias within abstracts/headlines of online articles => label based on relative bias levels => train a neural network model capable of upscaling this labeling process.**
- **The result was a Streamlit App with a Bias Estimator and Analyzer of Sentiment Tendency (BEAST) Engine, capable of scoring bias on a 0 to 1 scale and returning a sentiment score from -1 to 1.**
- Interested parties in this research include political party affiliates, media outlets, psychologists and sociologists, and others intrigued by the phenomenon of media influence upon societal thought processes.

Methods

- Webscraping
- Natural Language Processing Binary Classification Modeling
- Labeling Functions with Proprietary Snorkel Functions
- Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) Model
- Word2Vec: Continuous-Bag-Of-Words, and Skip-Gram word encoding algorithms
- VADER sentiment analyzer
- Streamlit App Services

Project Pipeline

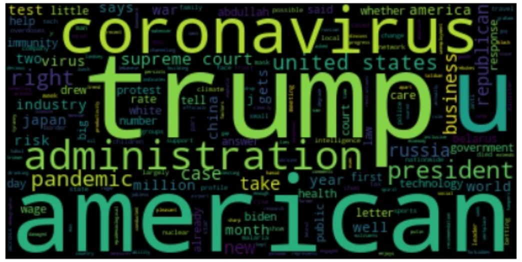
1. Web scraping – articles from ‘United States’ as API keyword
2. Binary Classification – based on whether abstract/headline couplings originated from Opinion section of New York Times or other sections
3. Snorkel Labeling Functions and Linear Weighted Labeling Function Creation
4. Neural Networks and Word2Vec encoding Algorithms
5. Sentiment Analysis
6. Repeat steps 3 through 5
7. Streamlit App Deployment

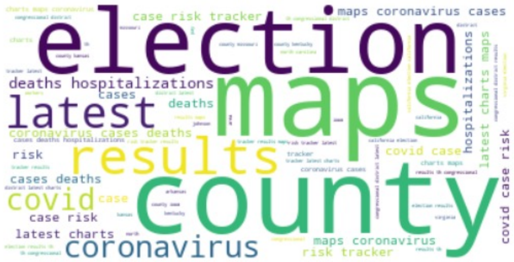
Binary Classification Modeling

- Preprocessed abstracts and headlines from NYT articles with Regular Expressions
- Text strings were features
- Opinion section labeled 1, non-Opinion labeled 0 were targets
- GridSearchCV for multiple binary classification models to find optimal parameters with lowest error metrics, highest specificity, and balanced F1 Score
- Pickled and exported for inclusion in Streamlit App

Labeling

- Snorkel is a library with proprietary functions that is used to label data based on user input.
- Regular Expressions label different abstract/headline couplings interpreted as having bias versus others which were interpreted to not have bias. (Hand labeled a training set).
- Length functions and TextBlob polarity and sentiment analyzers included.
- Snorkel metrics surveyed to inform decisions regarding more labeling function creation.





Linear Label Function with Weights

- Electoral College Comparison
- Return values between 0 and 1 and each had tiers.
- Cohen Kappa Score – 1-1 correspondence – with Bias Score.

[0,0.2] represents not-to-slightly-biased.

(0.2,0.4] represents slightly-to-moderately biased.

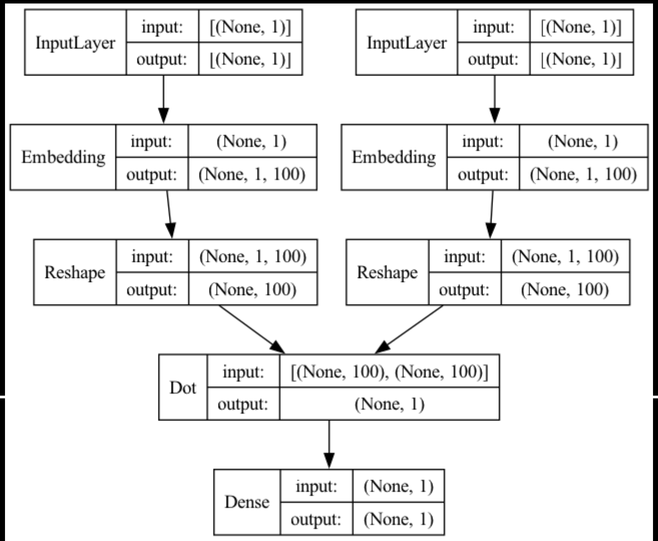
(0.4,0.6] represents moderately-to-pretty-biased.

(0.6,0.8] represents pretty-to-mostly-biased.

(0.8,1.0] represents perfectly-biased.

Neural Networks

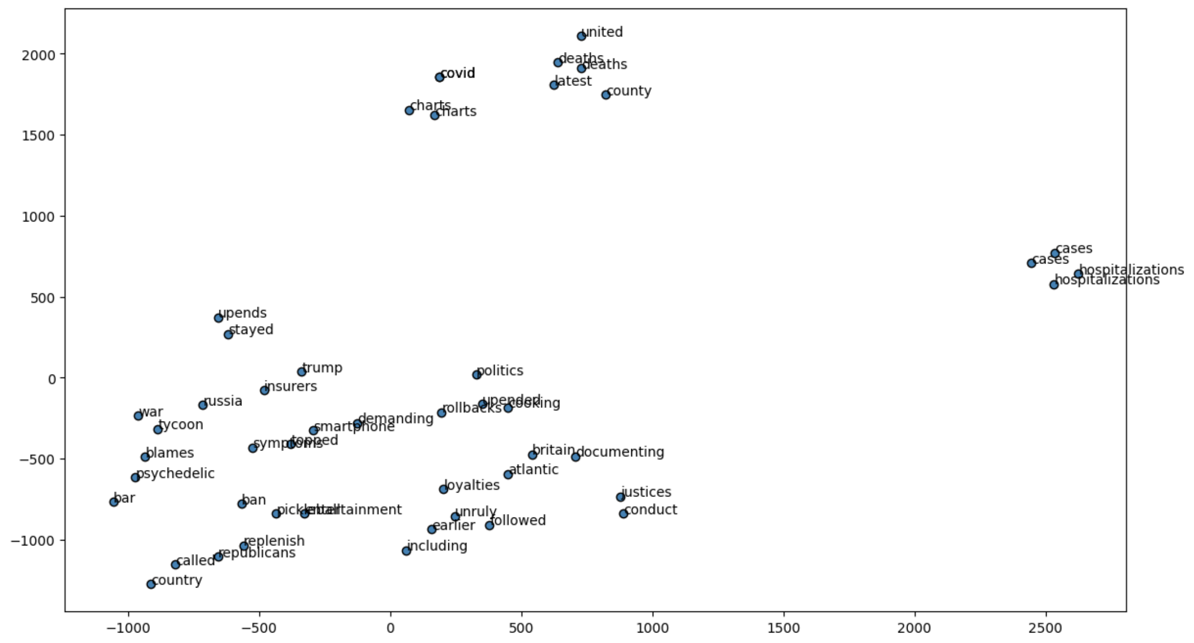
- Long Short Term Memory (LSTM) Categorical model with multiple layers.
- Word2Vec Variants: Continuous Bag of Words and Skip Grams are word encoding algorithms with neural networks. (There are more but only these two were utilized here)
- Interspersed were Cosine Similarity, Clustering, PCA.

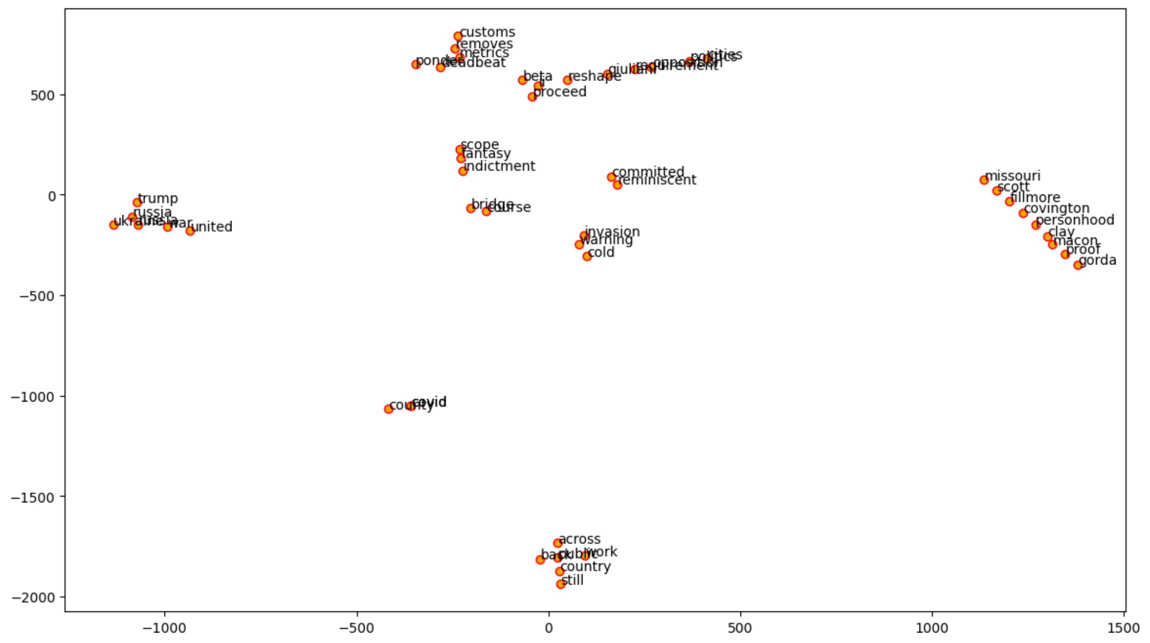


Word2Vec Variant: Skip-Gram

- One Variant of Word2Vec: The skip-gram variant takes a target word and tries to predict the surrounding context words.
- This is particularly effective when using a keyword chosen to either labeled with either BIASED or UNBIASED tags and survey words around it to extrapolate full meaning or text string, delineate whether BIASED or UNBIASED.

Skip-Gram Plots

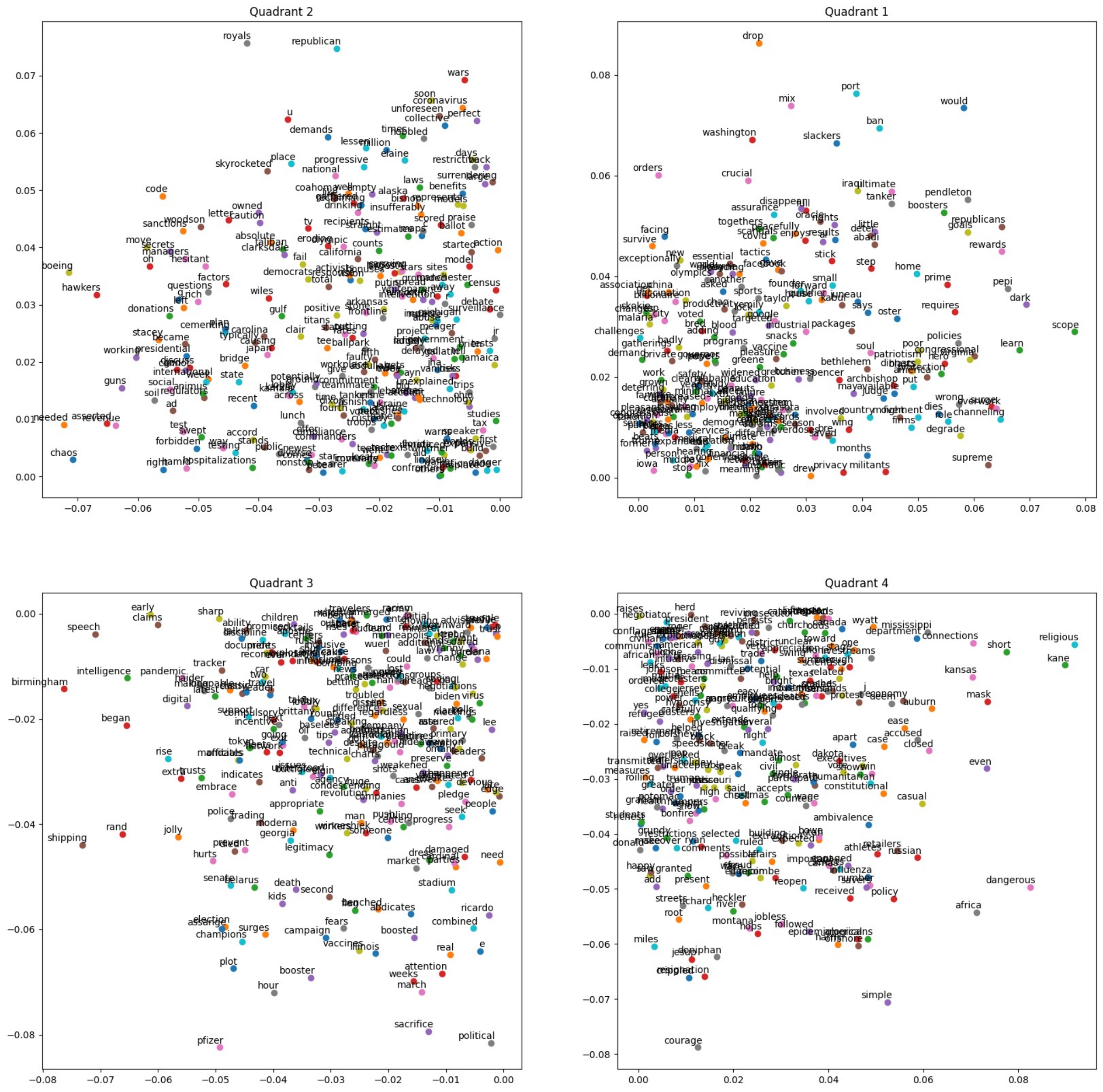


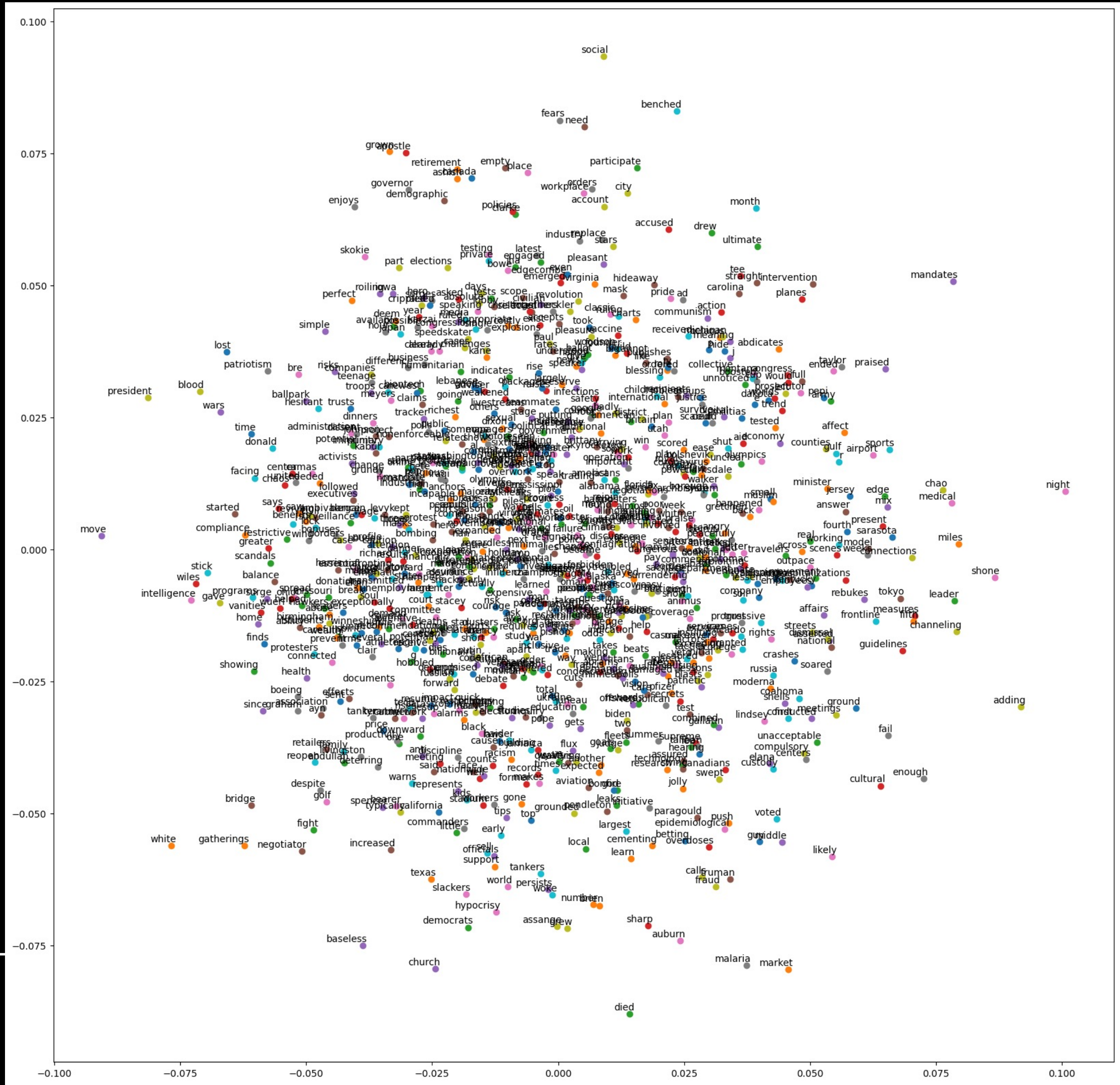


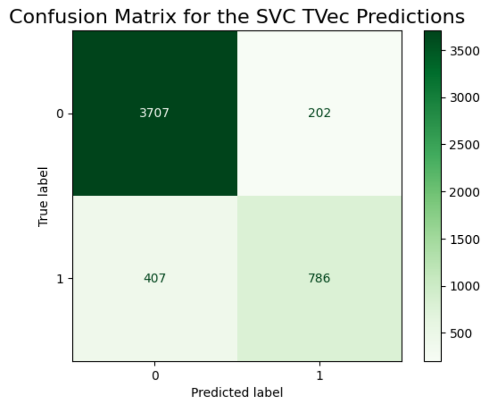
Word2Vec Variant: Continuous-Bag-of-Words

- One Variant of Word2Vec: The Continuous-Bag-Of-Words (CBOW) variant takes a set of context words and tries to predict a target word.
- This is particularly effective when using commonly found string of words labeled with either BIASED or UNBIASED tags to pinpoint a characteristic word within a string that plays a big role in *sentiment analysis* and overall BIASED or UNBIASED text string label.

CBOW Plots







Summary of Results and Analysis

- Performance of Binary Classification Model (Support Vector Machine with TfidfVectorizer)

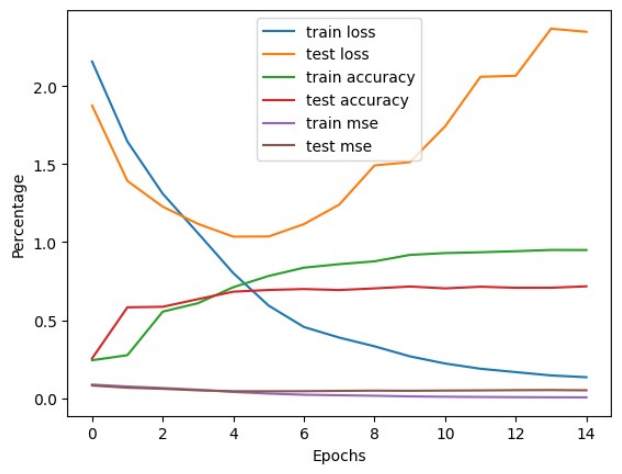
1.0 Train | 0.88 Test | Sensitivity 0.66 | Specificity 0.95 | Accuracy 0.88 | Precision 0.800

Miscalculation Rate 0.12 | F1 Score 0.72

- Performance of Neural Network Model:

Loss: 0.13 - Accuracy: 0.95 - MSE: 0.006

Validation Loss: 2.34 – Validation Accuracy: 0.72 – Validation MSE: 0.05



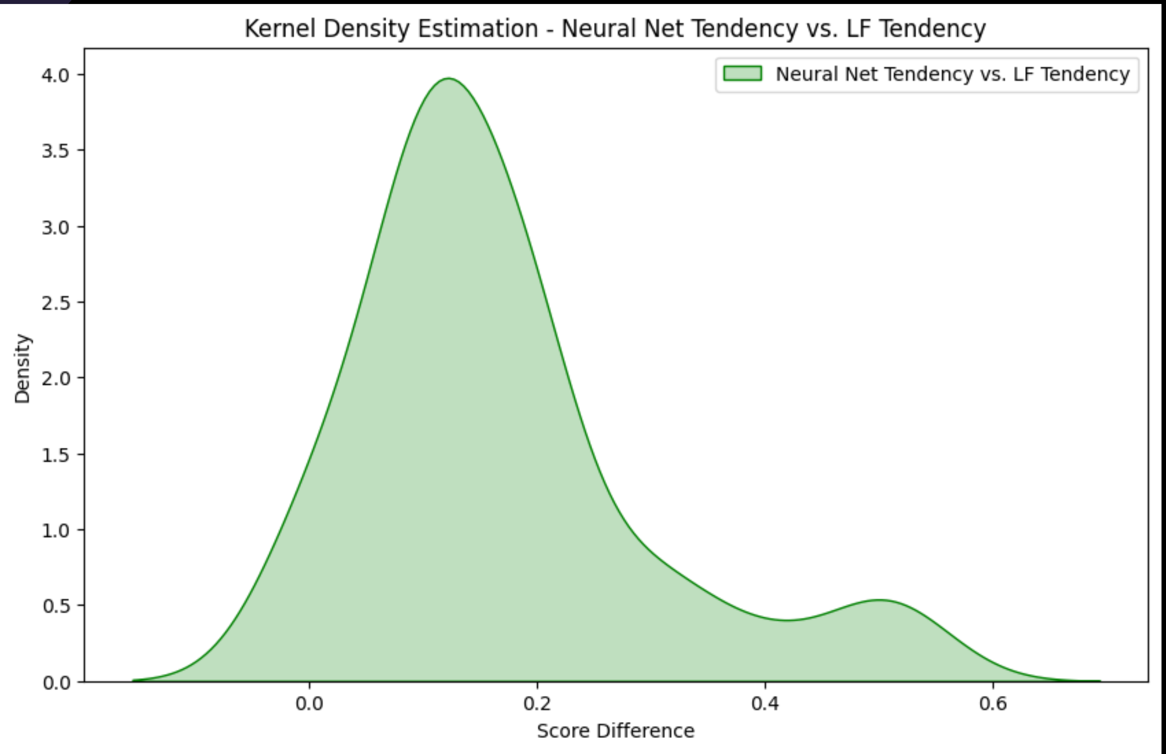
- Tendency Towards Bias Scoring App runs on…

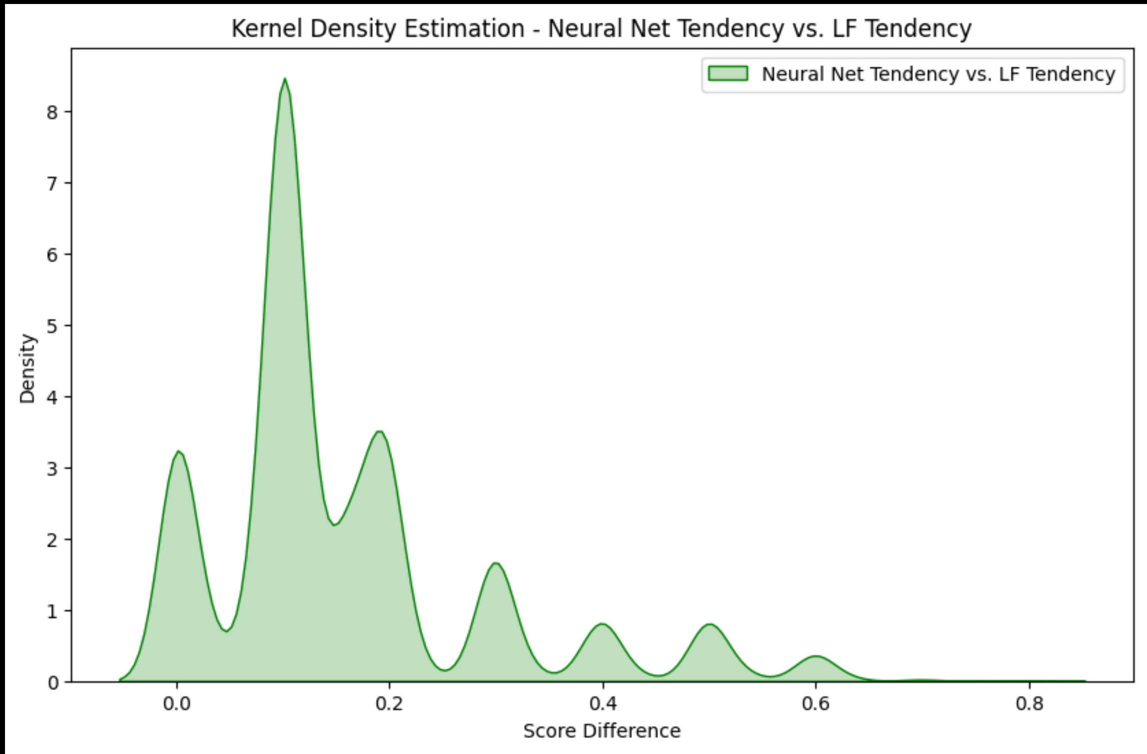
Bias Estimator and Analyzer of Sentiment Tendency (BEAST) Engine

BEAST Performance

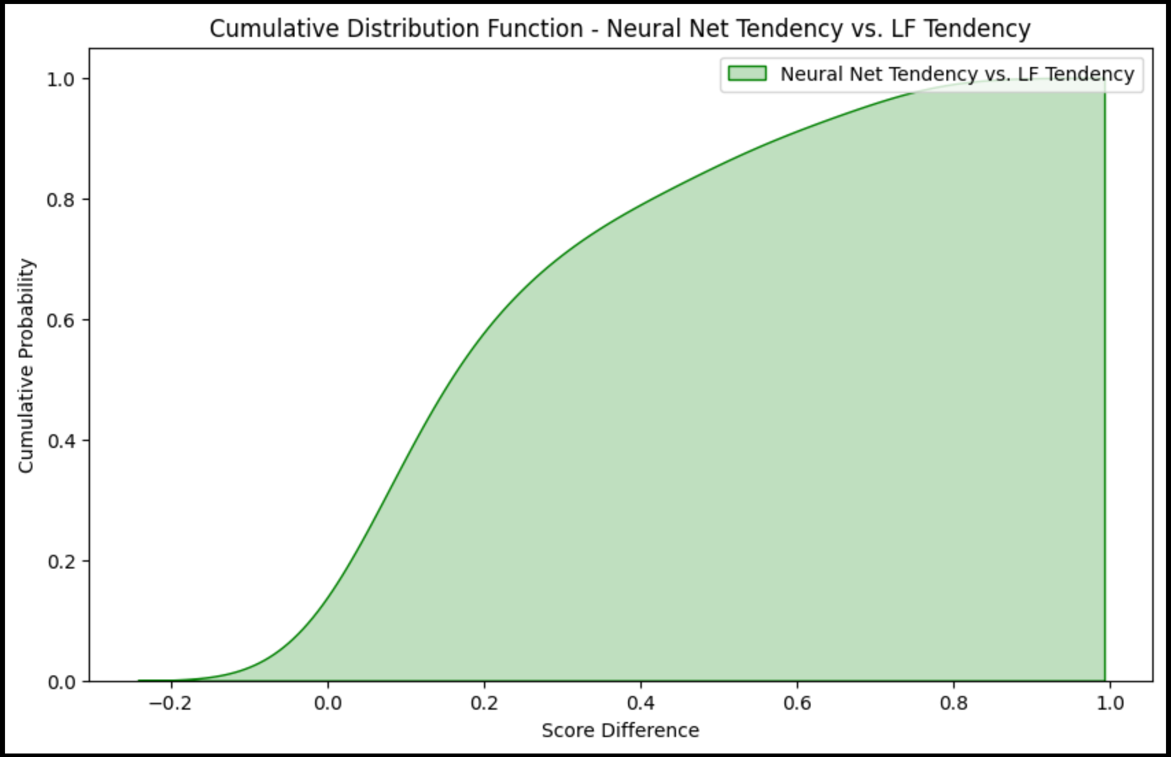
- Compare performance of Sentiment Analyzer with VADER sentiment analyzer. [“Ferromagnetic” only magnitude differed]
- KDE Plots looking at absolute differences between NN model and LF function, BEAST and VADER sentiment analyzers
- A/B Testing Results looking at absolute difference, BEAST and VADER sentiment scores individually.

Kernel Density Plots
Neural Net Model versus Labeling Function

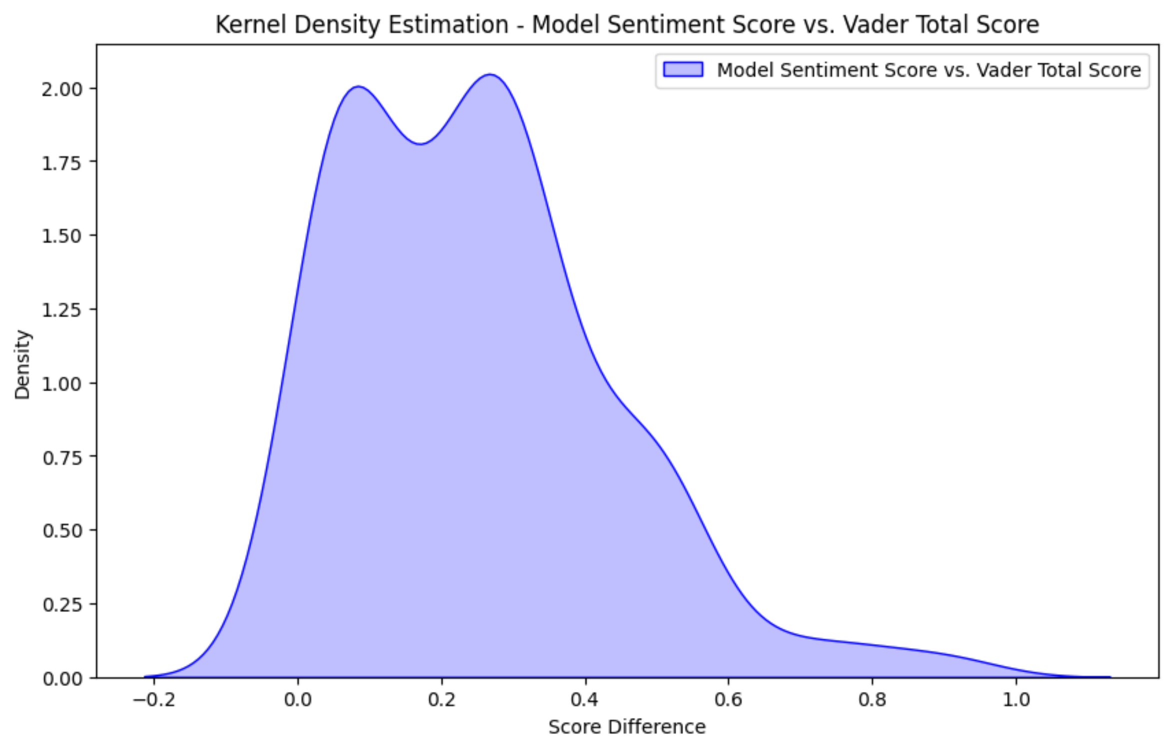


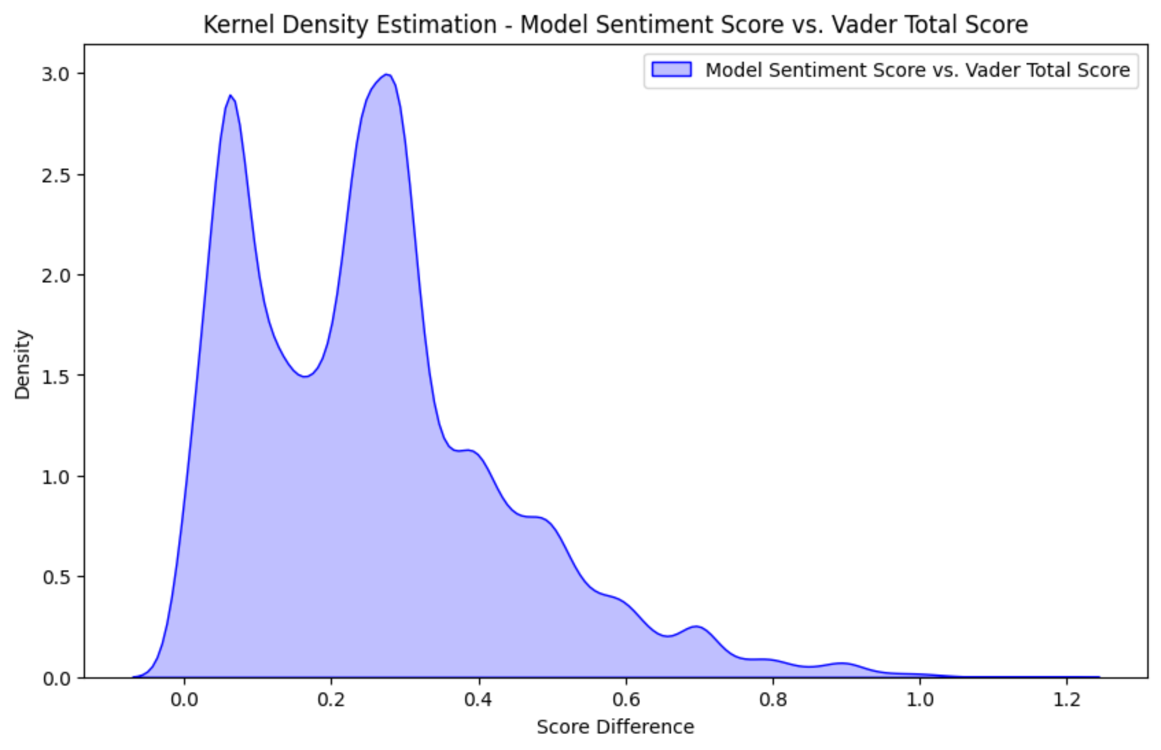


Cumulative Distribution Function Plot
Neural Net Model versus Labeling Function

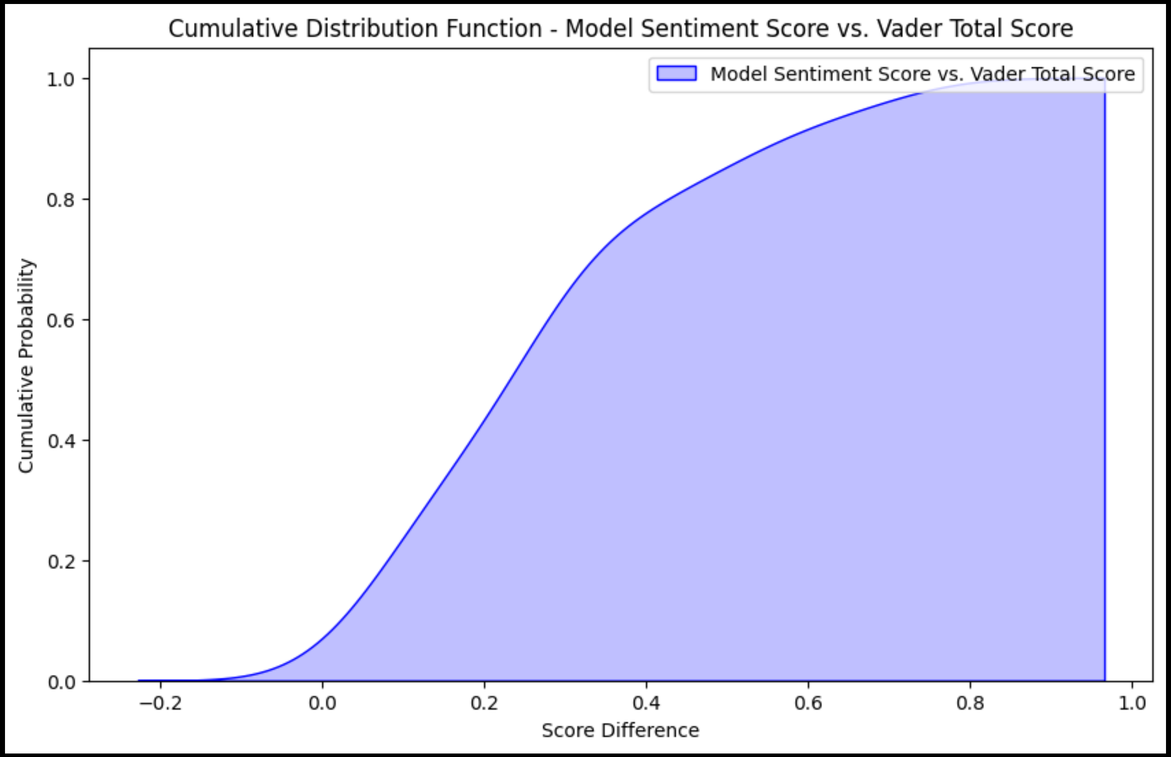


Kernel Density Plots – BEAST versus VADER





Cumulative Distribution Function Plot BEAST versus VADER



A/B Testing On Sentiment Analysis

- Mann-Whitney U test instead of t-test as these values are not normally distributed
- BEAST Sentiment Analyzer to tell different between Opinion versus non-opinion section with only a 10 % chance of being incorrect… (α= 0.1).
- P-value: 0.947 Fail to reject the null hypothesis. There is no significant difference between Opinion and non-Opinion groups.
- Vader Sentiment Analyzer to tell different between Opinion versus non-opinion section with only a 10 % chance of being incorrect… (α= 0.1).
- P-value: 0.520
- Look at ABSOLUTE DIFFERENCE between BEAST and VADER sentiment levels…
- P-value: 0.0976
- Reject the null hypothesis. There is a significant difference. Thus, by looking at BEAST and VADER absolute sentiment level differences we may distinguish between Opinion section articles and non-Opinion section articles with under a 10% chance of being incorrect, less than our binary classification model!

Recommendations and Guidance

- Develop labeling functions and other word encoding algorithms for further optimization.

  Hand Label a greater expanse of samples to acquire more detailed and meaningful Snorkel summary statistics.

- Continue to advance hardware capabilities to develop Neural Network models further.

  Switch to Microsoft and NVIDIA GPU.

Apply these labeling functions to other well-defined problems such as customer service evaluations, prescription drug side-effect forms, crowd surfing surveys, etc.

Peering Into the Future

- Linear combinations of weighting functions and sentiment analyzers may be utilized to evaluate information.
- Instead of binary Word2Vec labeling, qubits (0 or 1 quantum superposition principle) exhibit entanglement and calculate BIASED or UNBIASED labels in effective real-time response.

Conclusion

- We use labeling functions in our everyday lives!
- When we harness multiple powerful modeling techniques with these relatively simple primitive fundamental forms, complex and daunting language processing challenges become simpler and more realistic to understand.
- We successfully utilized many of these NLP and NN modeling techniques and created a Bias Estimator and Analyzer of Sentiment Tendency (BEAST) Engine capable of generating a bias score and sentiment levels comparable to that of VADER's sentiment analyzer.
- Stakeholders interested in this research include product managers, app developers, customer service leads, etc.

References

- `API` Documentation from `NYT`: https://developer.nytimes.com/apis
- Code for Webscraping adapted from fellow coursemate's group project, with their permission this code was included in the pipeline for this project.
- Snorkel documentation for Cohen Kappa Score labeling: https://www.snorkel.org
- Pew Center Article on Bias: https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online
- VADER Documentation: https://vadersentiment.readthedocs.io
- Kaggle GPU Documentation: https://www.kaggle.com/code/dansbecker/running-kaggle-kernels-with-a-gpu
- CBOW/Skip Gram: https://towardsdatascience.com/understanding-feature-engineering-part-4-deep-learning-methods-for-text-data-96c44370bbfa
- CBOW/SKip Gram: https://medium.com/@dube.aditya8/word2vec-skip-gram-cbow-b5e802b00390
- KDE Plot Documentation: https://seaborn.pydata.org/generated/seaborn.kdeplot.html
- Quantum Neural Networks: https://openreview.net/pdf?id=ZLKaNvYFfjd