

Appendix for “Quantifying Peer Review Informativeness: An Unsupervised, Linguistic Features-Based Approach”

Prabhat Kumar Bharti^{1*}, Vijay Harkare², Aditya Shah²,
Mayank Agarwal³

^{1*}School of Computing and Electrical Engineering, Indian Institute of
Technology Mandi, Himachal Pradesh, India.

²Department of Computer Engineering, Dwarkadas J. Sanghvi College
of Engineering, Bhaktivedanta Swami Marg, Mumbai, 400056,
Maharashtra, India.

^{3*}Department of Computer Science and Engineering, Indian Institute of
Technology, Patna, Bihta Kanpa Rd, Patna, 801106, Bihar, India.

*Corresponding author(s). E-mail(s): dept.csprabhat@gmail.com;
Contributing authors: vijay.harkare2020@gmail.com;
adityashah841@gmail.com; mayank265@iitp.ac.in;

A Sarcasm Detection Model

This appendix explores the implementation of a deep learning-based sarcasm detection algorithm to enhance the evaluation of peer reviews. Some advantages of using sarcasm detection for the peer review informativeness evaluation process are as follows:

- Avoid Misinterpretation: Detecting sarcasm helps avoid misinterpretations that could lead to confusion or conflict between reviewers and authors.
- Maintain Professionalism: Clear communication fosters a professional and respectful atmosphere, conducive to constructive feedback and collaboration.
- Preserve Trust: Recognizing and addressing sarcasm preserves trust between reviewers and authors, maintaining the integrity of the peer review process.

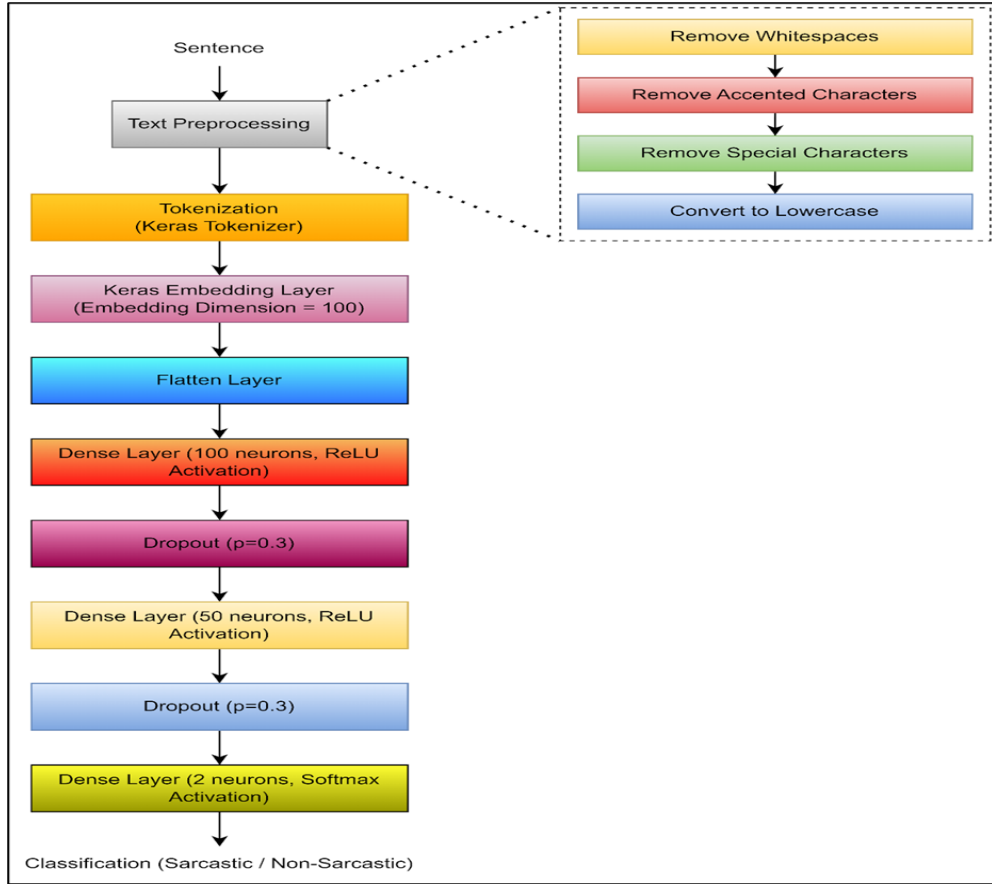
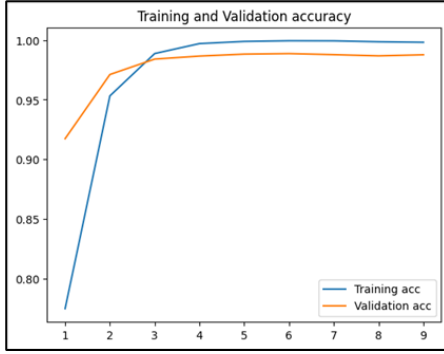


Fig. 1: Sarcasm Detection Model Architecture

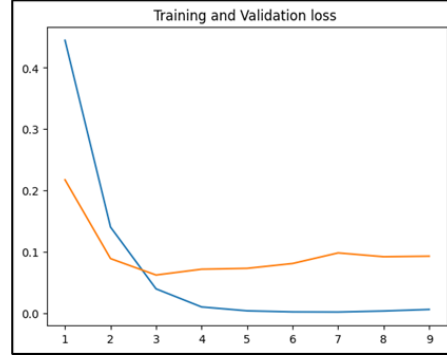
Implementation. The methodology involved a neural network model for sarcasm detection. The model architecture included an Embedding layer, Flatten layer, Dense layers with ReLU activation, Dropout layers, and a final Dense layer with Softmax activation for binary classification. Hyperparameters included a maximum sequence length of 100, a vocabulary size of 10,000 words, an embedding dimension of 100, batch size of 200, and 20 epochs for training. Figure 1 depicts the architecture of the entire model.

The model was trained on a dataset combining tweets from the iSarcasmEval dataset [1] and news headlines from the News Headlines Dataset [2, 3], pre-processed to remove extra whitespaces, accented characters, special characters, and converted to lowercase. This combination of datasets allowed the model to capture different forms of sarcasm across diverse domains.

Results. The accuracy and loss curves are shown in Figure 2a and Figure 2b, respectively. The results show high accuracy, with the model achieving over 98% validation



(a) Training and validation accuracy as a function of the number of epochs.



(b) Training and validation loss as a function of the number of epochs.

Fig. 2: Curves showing the training and validation accuracy and loss as functions of the number of epochs.

accuracy. The figures demonstrate the effectiveness of the training process of the neural network for the task of sarcasm detection. Given the high performance of the sarcasm detection model, it was employed in the annotation process to enhance dataset quality and diversity.

B Statistical Analysis of Dataset

This appendix provides the figures and tables demonstrating the statistical analysis of the dataset.

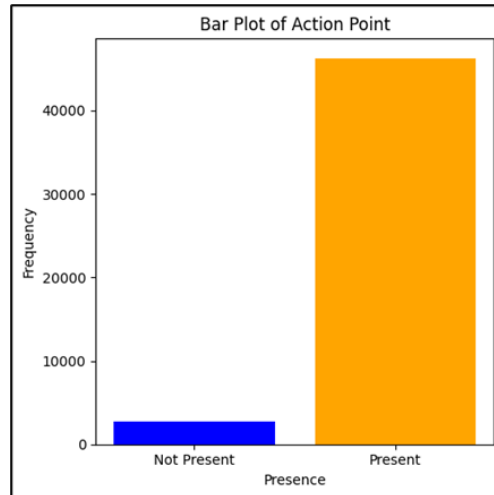


Fig. 3: Bar plot for action point detection

| Extracted Feature | Mean | Median | Mode | Standard Deviation | Variance | Min. | Max. | Skewness | Kurtosis |
|-------------------------|---------|----------|--------|--------------------|-----------|-----------|----------|----------|----------|
| action_point_bin | 0.944 | 1.000 | 1.000 | 0.229 | 0.053 | 0 | 1.000 | -3.866 | 12.945 |
| avg_sent_length_by_word | 22.670 | 19.000 | 22.000 | 16.076 | 258.434 | 3.000 | 719.000 | 3.926 | 80.870 |
| avg_syllable_count | 1.589 | 1.565 | 1.500 | 0.255 | 0.065 | 1.000 | 5.600 | 0.734 | 2.728 |
| avg_word_length | 7.096 | 7.154 | 7.500 | 1.019 | 1.040 | 1.667 | 19.333 | -0.080 | 1.983 |
| cosine_sim | 0.335 | 0.341 | 0 | 0.156 | 0.024 | 0 | 0.771 | -0.176 | -0.578 |
| count_functional_words | 0.251 | 0.250 | 0.250 | 0.109 | 0.012 | 0 | 0.800 | 0.011 | 0.155 |
| dale_chall | 12.700 | 12.855 | 13.007 | 2.344 | 5.492 | 0 | 56.161 | -0.091 | 4.484 |
| flesch | 30.635 | 30.687 | 42.223 | 32.439 | 1052.334 | -878.188 | 186.645 | -0.063 | 13.714 |
| gunning_fog | 18.959 | 18.126 | 19.709 | 8.333 | 69.447 | 2.000 | 383.995 | 2.879 | 79.388 |
| hedge_detection_bin | 0.157 | 0 | 0 | 0.364 | 0.133 | 0 | 1.000 | 1.883 | 1.546 |
| herdan | 0.965 | 0.967 | 1.000 | 0.030 | 0.009 | 0.667 | 1.000 | -0.739 | 1.286 |
| jaccard_sim | 0.006 | 0.005 | 0 | 0.004 | 0.00002 | 0 | 0.046 | 1.905 | 6.599 |
| kincaid | 14.969 | 14.250 | 12.835 | 7.836 | 61.409 | -11.283 | 371.010 | 3.283 | 91.963 |
| maas | 0.011 | 0.011 | 0 | 0.010 | 0.00001 | 0 | 0.197 | 1.608 | 9.612 |
| mattr | 0.996 | 1.000 | 1.000 | 0.011 | 0.0001 | 0.750 | 1.000 | -4.811 | 40.740 |
| mstr | 0.997 | 1.000 | 1.000 | 0.016 | 0.0002 | 0.667 | 1.000 | -8.370 | 103.349 |
| punctuation_count_ratio | 0.026 | 0.024 | 0.026 | 0.017 | 0.0002 | 0 | 0.361 | 4.770 | 50.124 |
| sarcasm_score | 0.219 | 0.000004 | 1.000 | 0.394 | 0.155 | 1.066e-32 | 1.000 | 1.349 | -0.078 |
| sentiment_bin | 0.006 | 0 | 0 | 0.075 | 0.006 | 0 | 1.000 | 13.142 | 170.722 |
| sentiment_confidence | 0.616 | 0.613 | 0.595 | 0.066 | 0.004 | 0.500 | 0.798 | 0.155 | -0.949 |
| simpsond | 0.013 | 0.010 | 0 | 0.013 | 0.0002 | 0 | 0.250 | 2.286 | 12.764 |
| yulek | 119.017 | 96.953 | 0 | 123.104 | 15154.609 | 0 | 2187.500 | 2.033 | 9.648 |

Table 1: Descriptive Statistics of Extracted Features for Peer Review Analysis

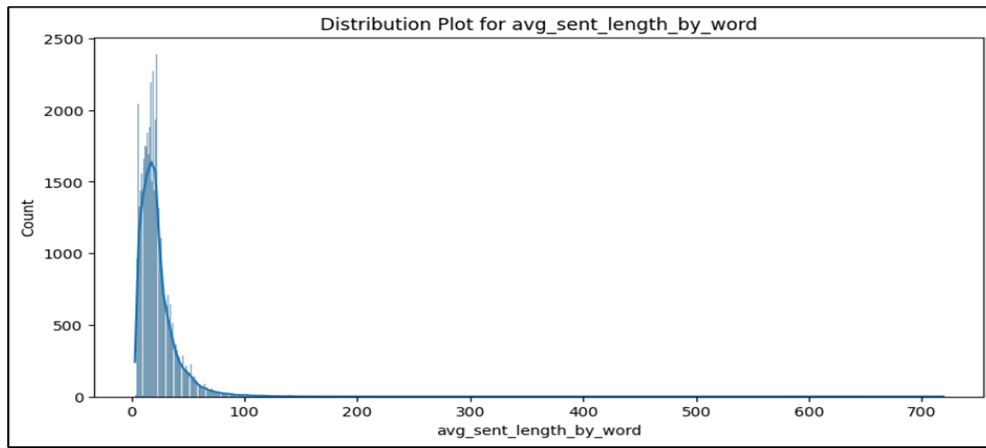


Fig. 4: Distribution plot for average sentence length (by word)

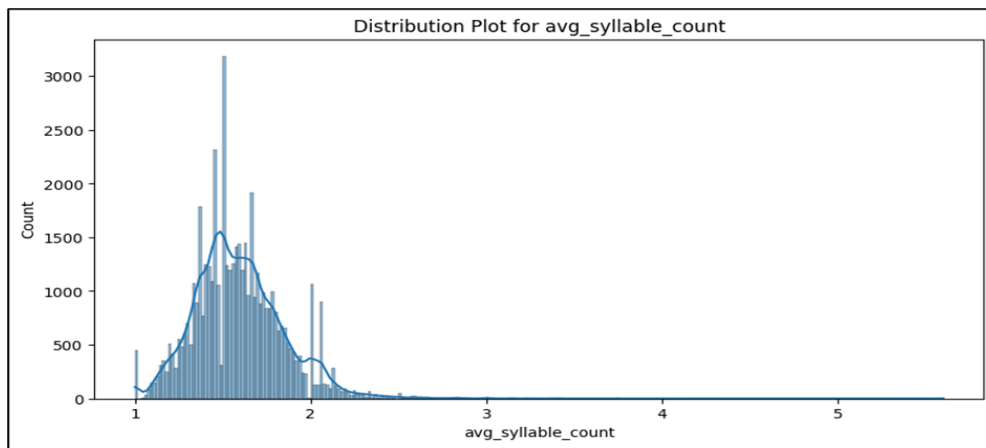


Fig. 5: Distribution plot for average syllable count

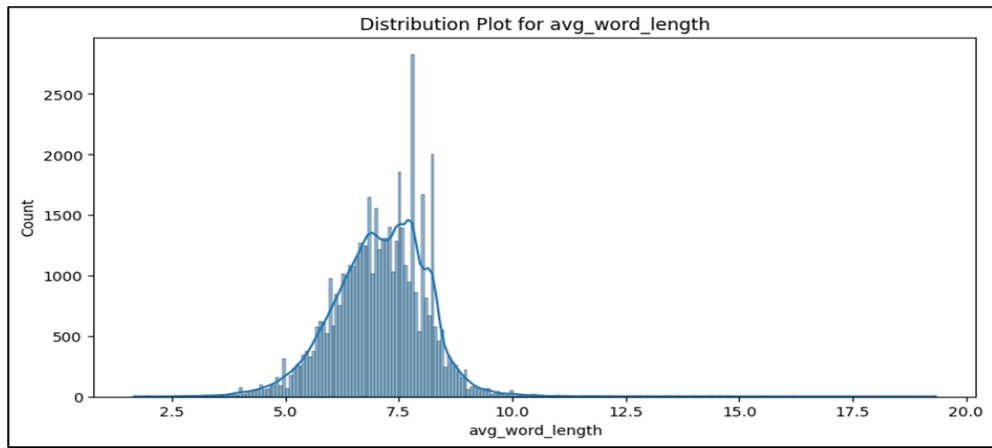


Fig. 6: Distribution plot for average word length

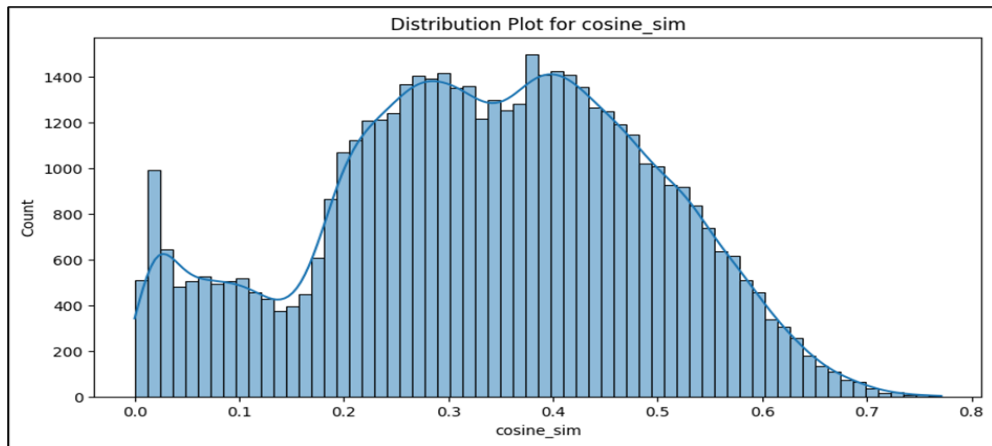


Fig. 7: Distribution plot for cosine similarity

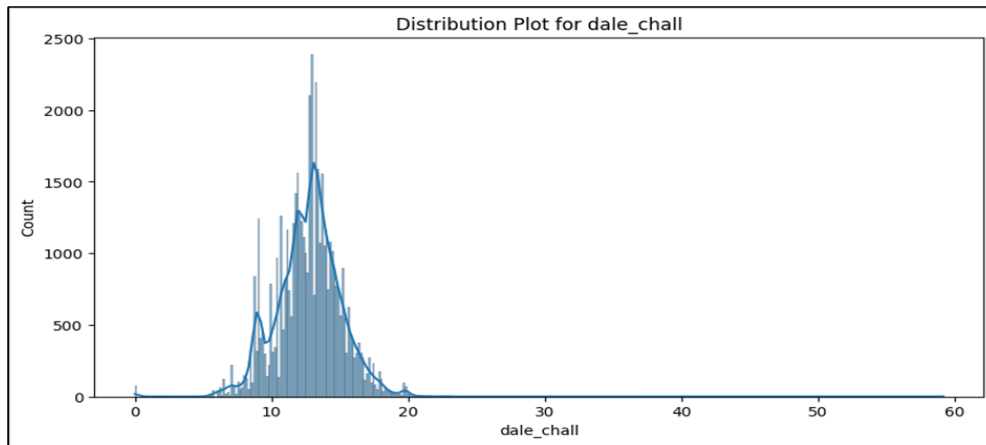


Fig. 8: Distribution plot for Dale-Chall Index

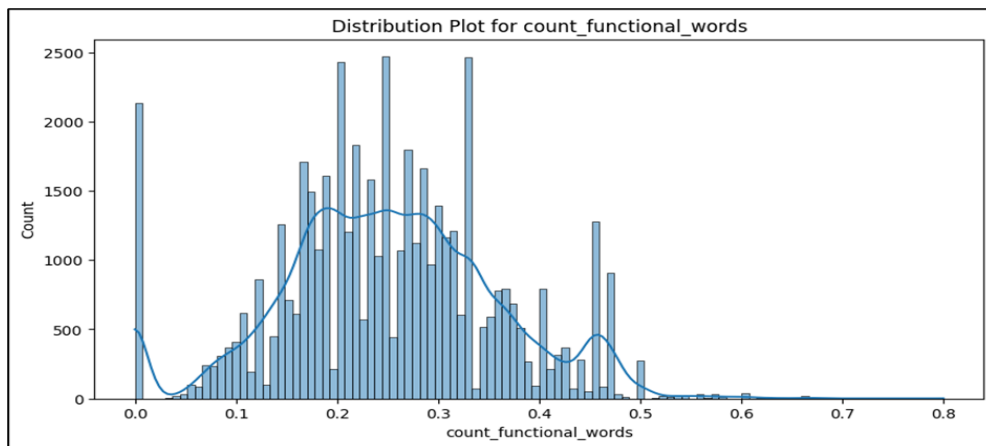


Fig. 9: Distribution plot for function words' count

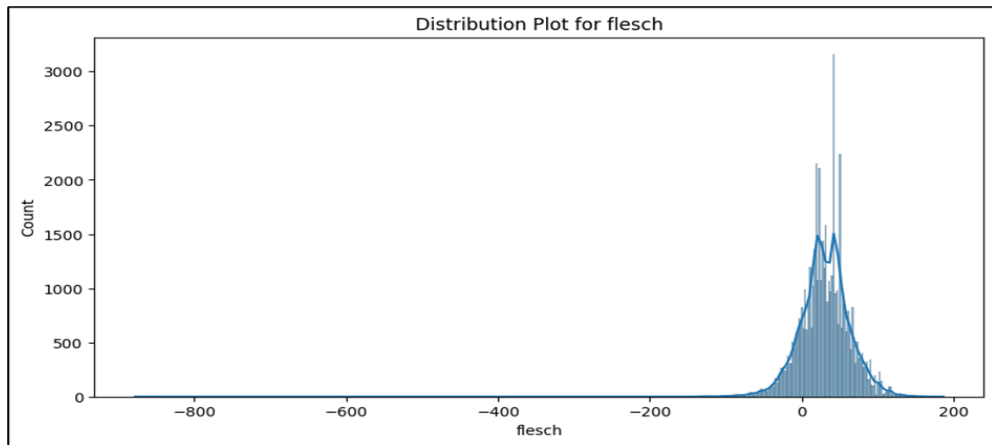


Fig. 10: Distribution plot for Flesch Reading Ease

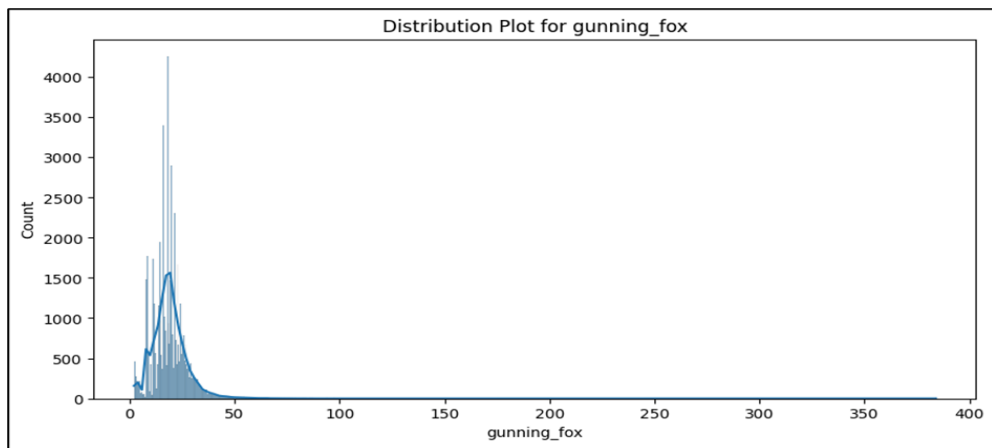


Fig. 11: Distribution plot for Gunning Fog Index

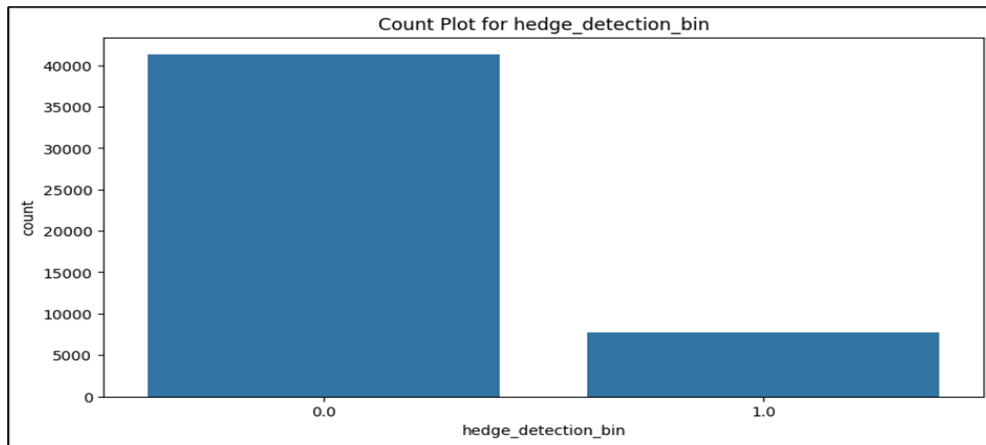


Fig. 12: Count plot for Hedge Detection

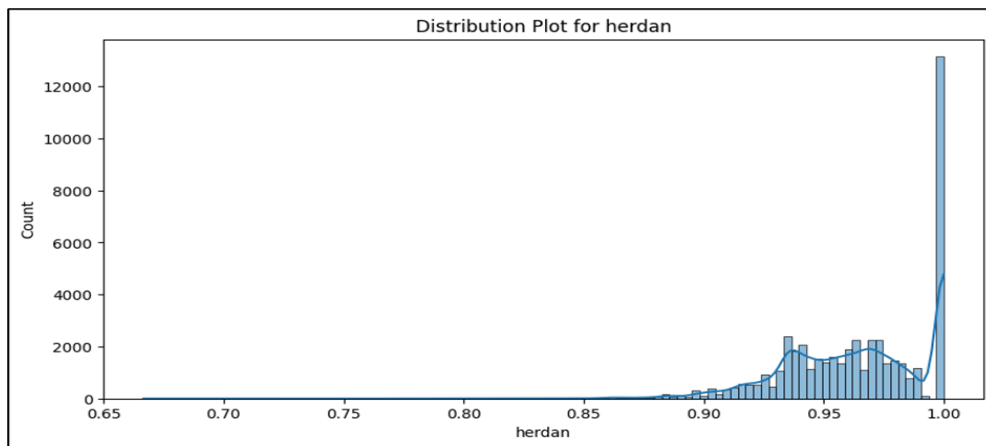


Fig. 13: Distribution plot for Herdan's Vocabulary Index

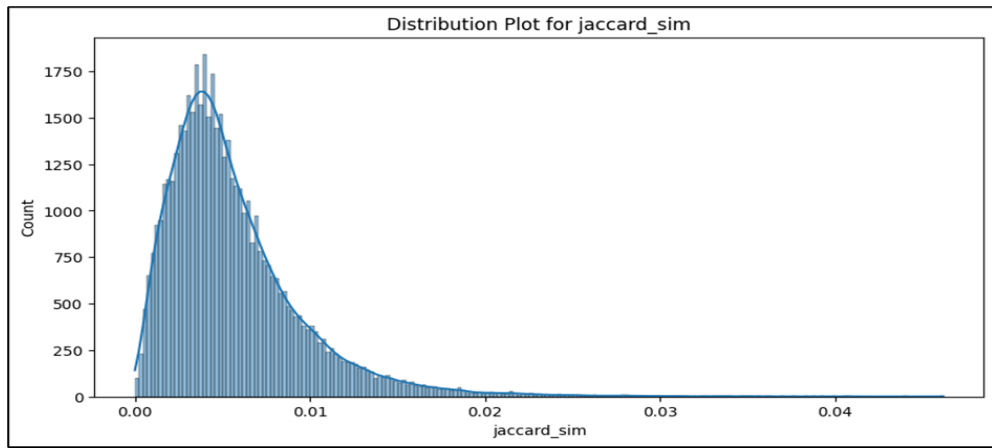


Fig. 14: Distribution plot for Jaccard Similarity

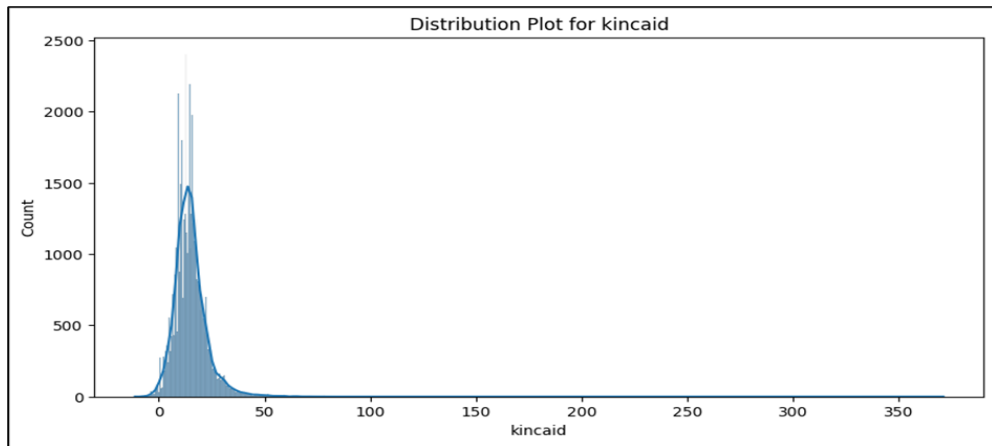


Fig. 15: Distribution plot for Flesch-Kincaid Grade Level

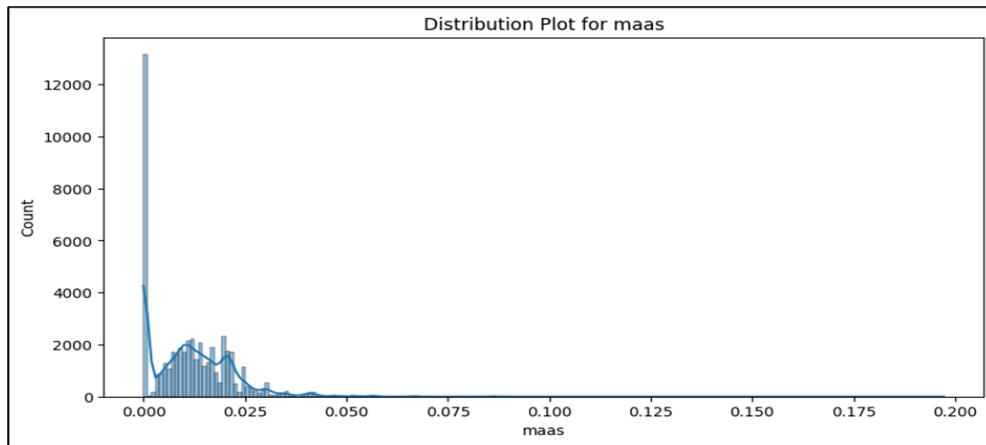


Fig. 16: Distribution plot for Maas' Index

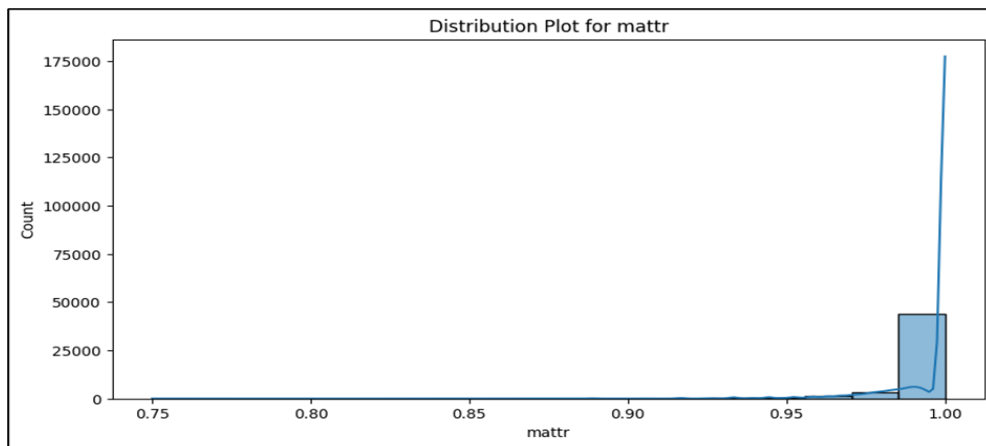


Fig. 17: Distribution plot for MATTR

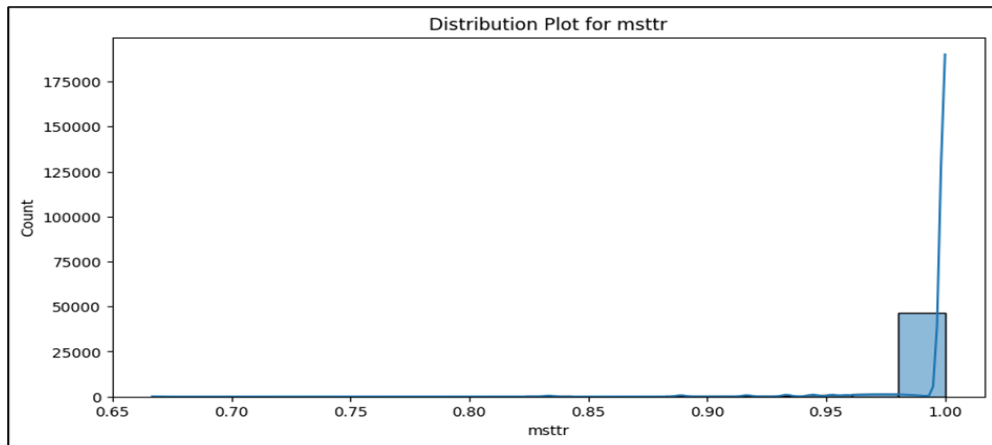


Fig. 18: Distribution plot for MSTTR

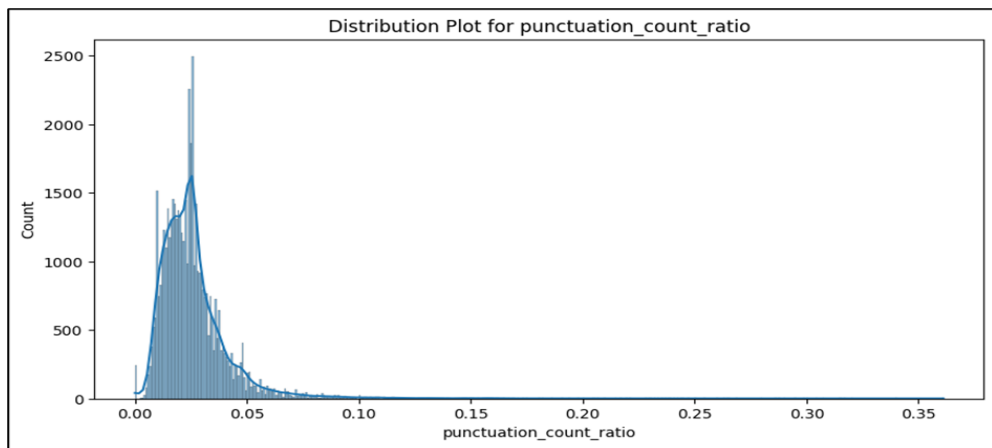


Fig. 19: Distribution plot for punctuation count ratio

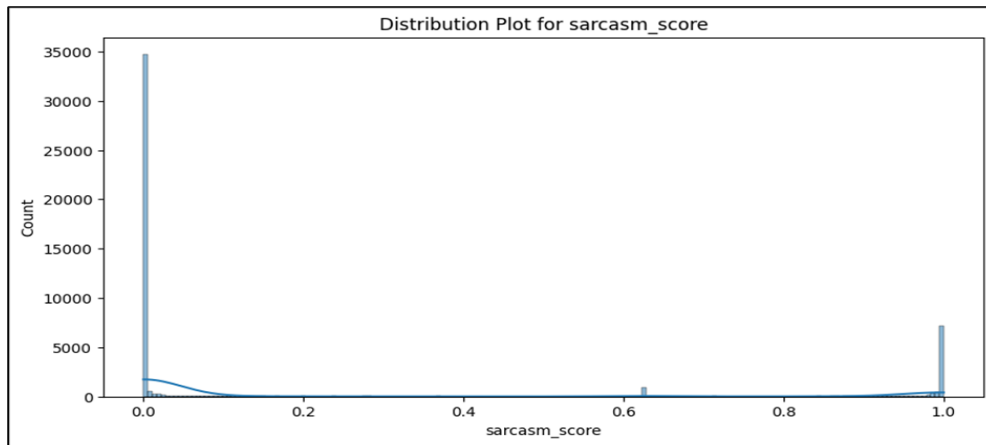


Fig. 20: Distribution plot for sarcasm score

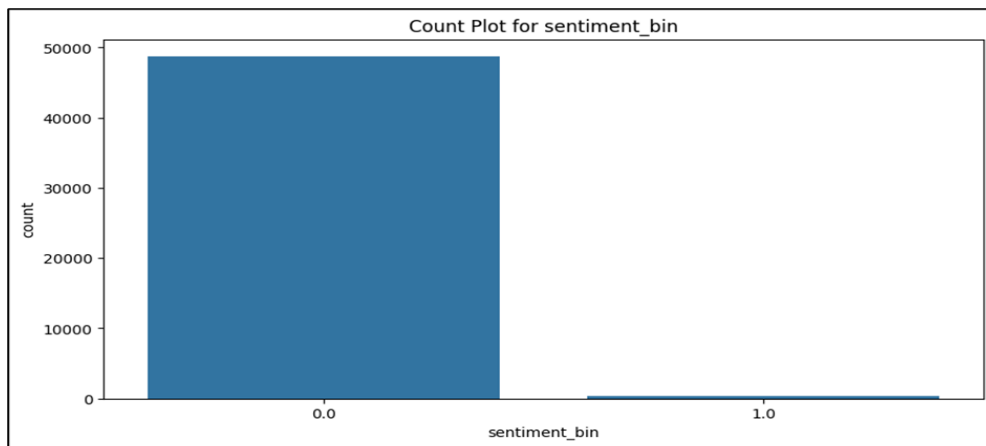


Fig. 21: Distribution plot for sentiment polarity

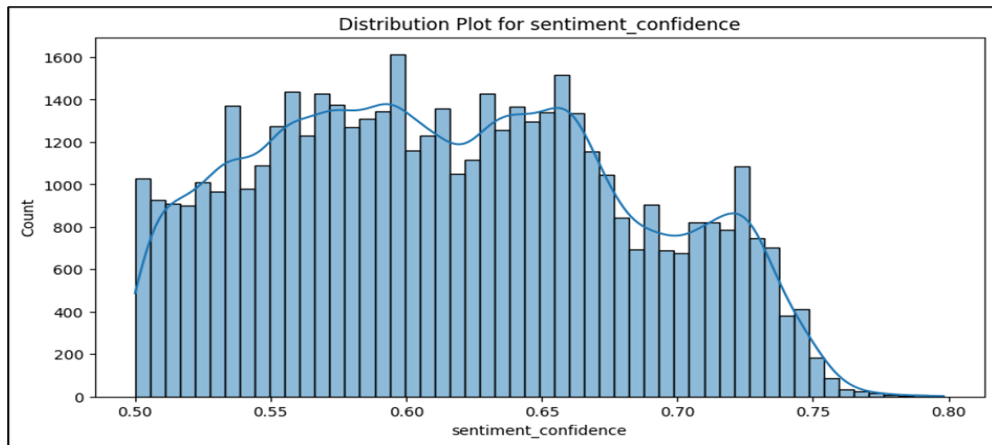


Fig. 22: Distribution plot for sentiment confidence score

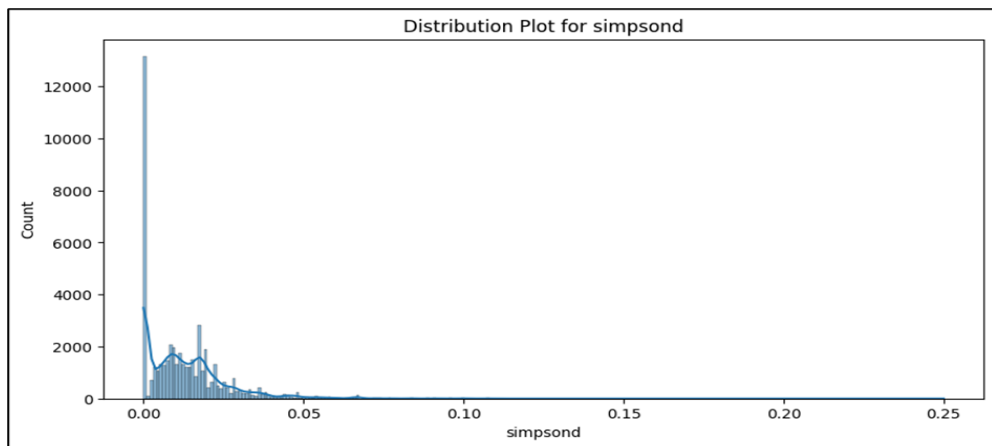


Fig. 23: Distribution plot for Simpson's Diversity Index

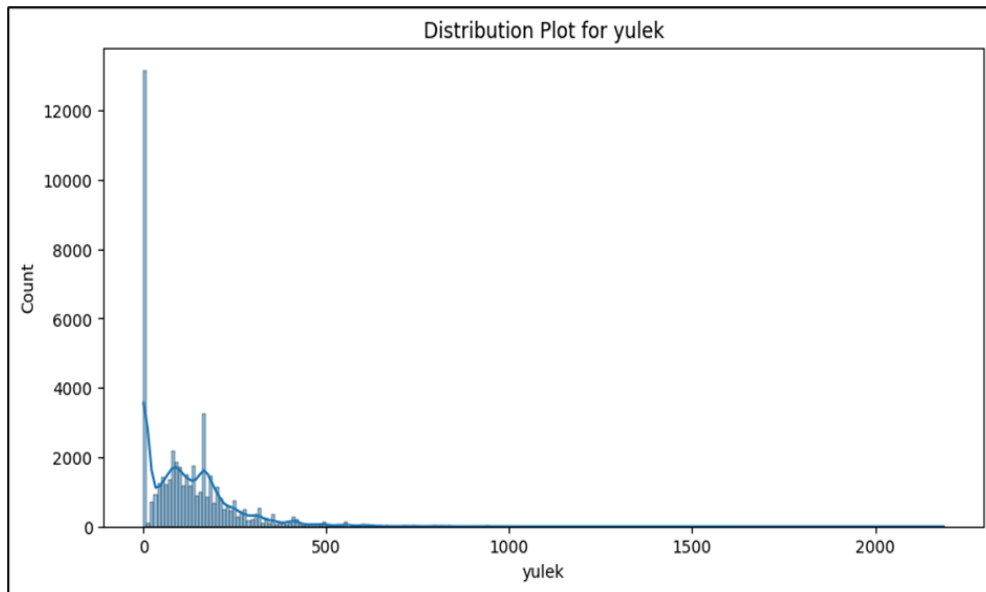


Fig. 24: Distribution plot for Yule's Characteristic Constant (K)

C Correlation between Informativeness Scores and Linguistic Features

This appendix presents the correlation analysis between the top informativeness scores and selected linguistic features.

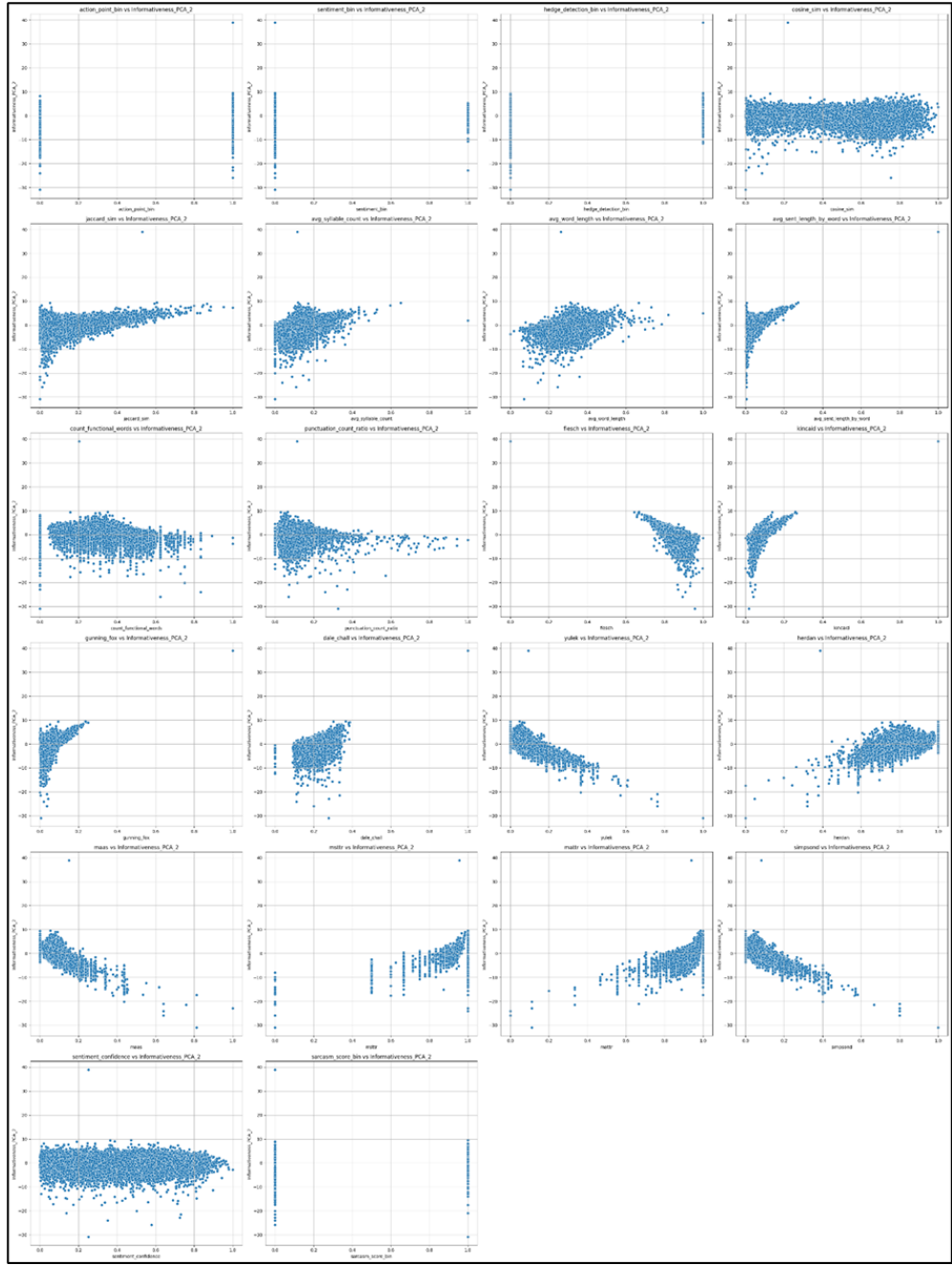


Fig. 25: Scatter plots between selected features and PC 2

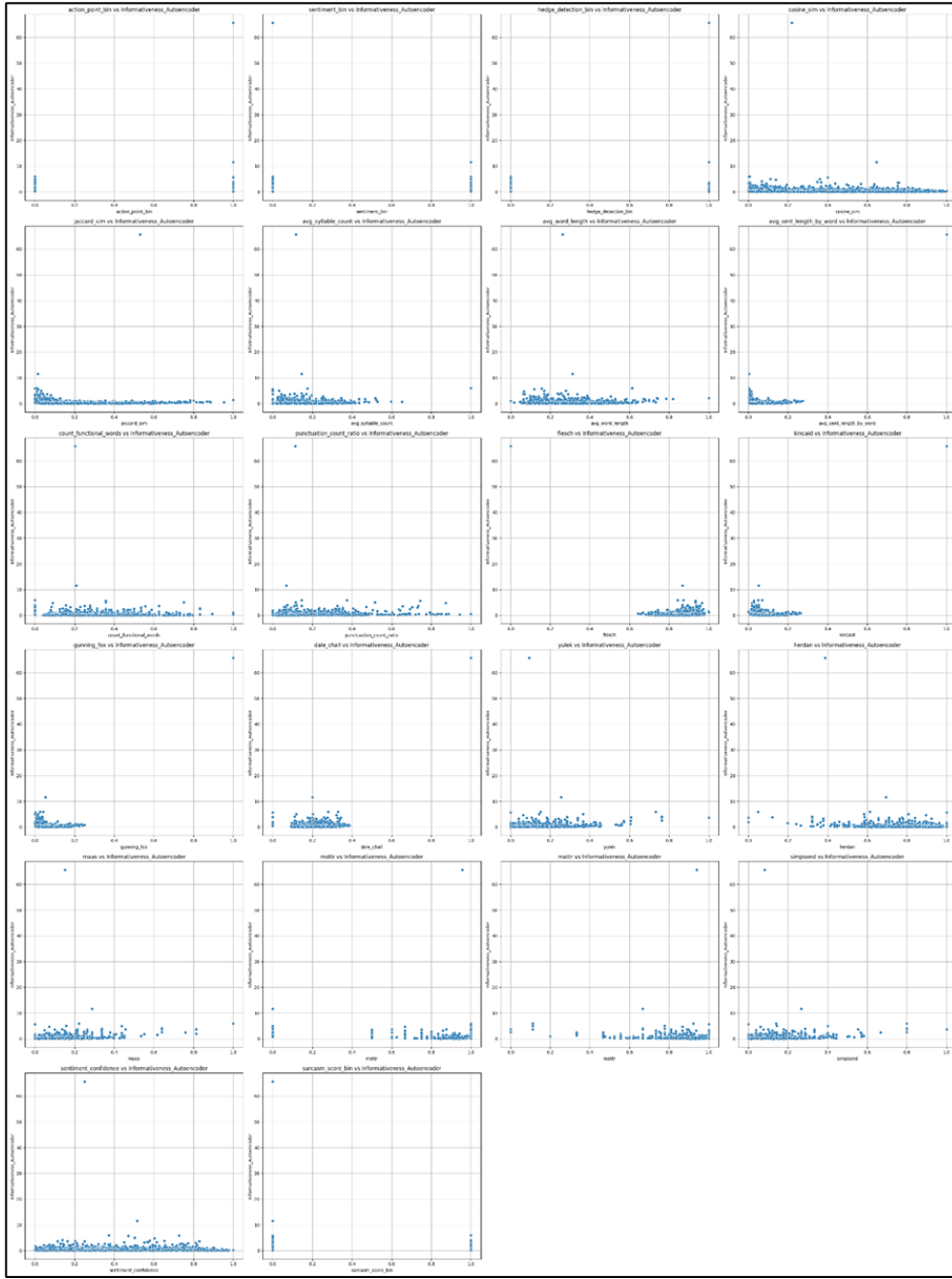


Fig. 26: Scatter plots between selected features and Autoencoder Reconstruction Error

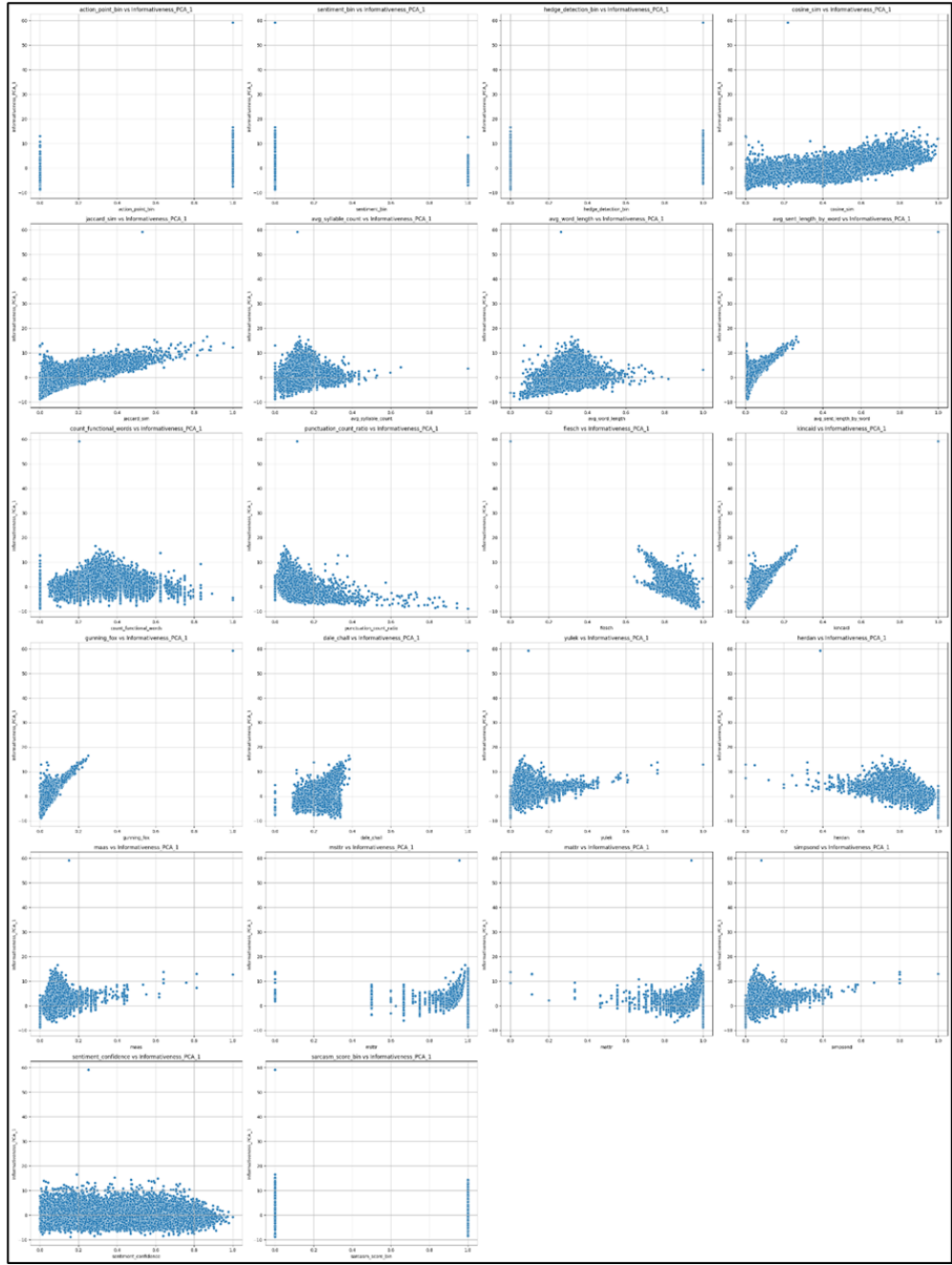


Fig. 27: Scatter plots between selected features and PC 1

D Correlation Analysis for Informativeness Hypothesis Validation

This appendix provides the result of the correlation analysis performed for validation of the informativeness hypothesis.

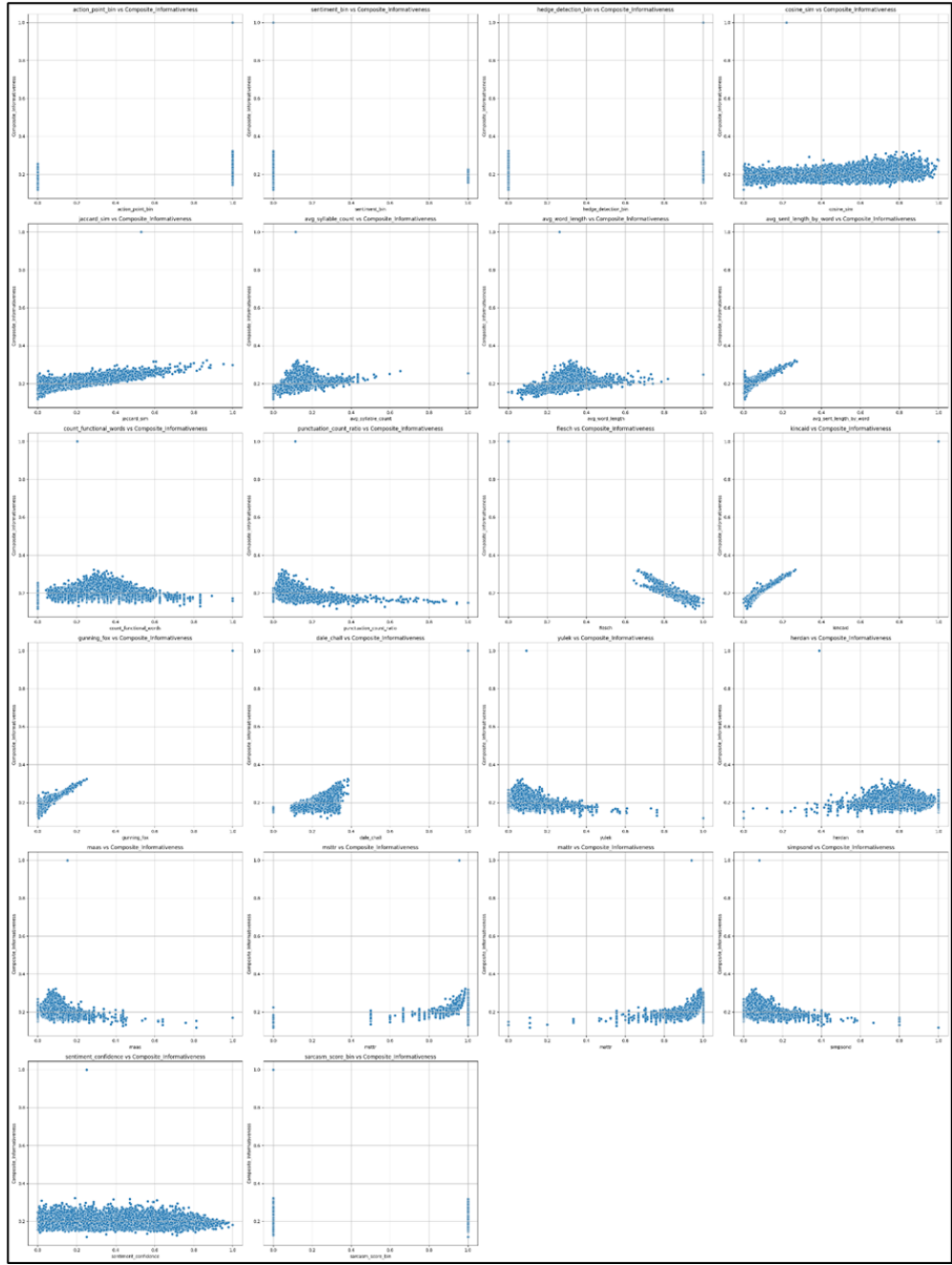


Fig. 28: Scatter plots between selected features and Composite Informativeness Score

References

- [1] Farha, I.A., Oprea, S.V., Wilson, S.R., Magdy, W.: Semeval-2022 task 6: isarcasmeval, intended sarcasm detection in english and arabic. In: SemEval 2022 - 16th International Workshop on Semantic Evaluation, Proceedings of the Workshop (2022)
- [2] Mishra, R., Grover, J.: Sculpting Data for ML: The First Act of Machine Learning. Jigyasa Grover & Rishabh Misra, 2021, ??? (2021)
- [3] Misra, R., Arora, P.: Sarcasm detection using news headlines dataset. AI Open **4** (2023) <https://doi.org/10.1016/j.aiopen.2023.01.001>