Rex Sheikh
CSPB 3832 - Natural Language Processing
Final Project Bibliography

Sources

Biyani, P., Tsioutsiouliklis, K., & Blackmer, J. (2016). “8 Amazing Secrets for Getting More Clicks”: Detecting clickbaits in news streams using article informality. In D. Schuurmans & M. P. Wellman (Eds.), *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)* (pp. 94–100). AAAI Press. (AAAI)

Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop clickbait: Detecting and preventing clickbaits in online news media. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 9–16). IEEE. https://doi.org/10.1109/ASONAM.2016.7752207 (ResearchGate)

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423 (ACL Anthology)

Potthast, M., Köpsel, S., Stein, B., & Hagen, M. (2016). Clickbait detection. In N. Ferro et al. (Eds.), *Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016* (Lecture Notes in Computer Science, Vol. 9626, pp. 810–817). Springer. https://doi.org/10.1007/978-3-319-30671-1_72 (OUCI)

Yi, X., Zhang, J., Li, W., Wang, X., & Xie, X. (2022). Clickbait detection via contrastive variational modelling of text and label. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)* (pp. 4475–4481). International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2022/621 (ijcai.org)

Here is a refined version of the related-work prose, adjusted toward your writing style and using “this is related to my final project because …” language.

Biyani et al. (2016)

Biyani et al. analyze how article informality can be used to distinguish clickbait from standard news content. They build a labeled dataset of headlines and corresponding articles, and define a set of stylistic and linguistic features such as slang, exaggerated sentiment, pronouns, informal discourse markers, and other markers of casual language. Using supervised models (e.g., logistic regression) over these features, they show that informality-based indicators are highly discriminative, improving performance beyond simpler lexical baselines. Their results suggest that clickbait is not just about specific trigger phrases, but also about broader stylistic patterns in how headlines and associated text are written.

This is related to my final project because I am also building a classical machine learning baseline that relies on interpretable features from headline text. The informality and style features described in this paper suggest concrete cues I can incorporate into my own feature set, such as pronoun usage, intensifiers, and sensational adjectives. Their methodology provides a model for how to examine learned feature weights and interpret which aspects of language most strongly correlate with clickbait. I can use their findings as a reference point when I analyze which tokens or features in my Kaggle-based model carry the strongest weights for the clickbait vs. non-clickbait classes.

Chakraborty et al. (2016) – “Stop Clickbait”

Chakraborty et al. propose “Stop Clickbait,” a system designed to detect and block clickbait headlines in online news media. They assemble a large corpus of headlines from multiple publishers, annotate them as clickbait or non-clickbait, and then engineer features capturing sentiment, lexical patterns (such as questions and listicles), and structural properties (such as length and punctuation). Using supervised classifiers, they achieve high accuracy and F1 scores on the detection task. They also deploy a browser extension that flags and optionally hides clickbait links in real time, demonstrating that clickbait detection can be integrated into a practical user-facing tool.

This is related to my final project because I am tackling essentially the same problem—classifying short news headlines as clickbait or informative—using a labeled dataset from Kaggle. The feature design in “Stop Clickbait” (sentiment, leading interrogatives, headline length, and so on) offers a useful checklist of features to log and compare in my own classical baseline. In addition, their end-to-end system, which surfaces model predictions in a browser extension, resembles my stretch goal of exposing the classifier through a small web interface. I can use their performance numbers and confusion-matrix analysis as a point of comparison when I evaluate how well my classical and transformer models perform on similar headline-level classification.

Potthast et al. (2016)

Potthast et al. present one of the earliest systematic studies of machine-learning-based clickbait detection, focusing on short teaser texts (especially on social media) that link to external articles. They construct a corpus of teaser messages labeled as clickbait or non-clickbait and extract features from both the teaser text and associated metadata (and, in some cases, the target page). They evaluate several classifiers, including random forests and logistic regression, and show that these models can reliably filter clickbait from users' streams. A key contribution of the paper is its discussion of dataset design, including how to balance clickbait and non-clickbait examples and how to reduce topic and publisher biases.

This is related to my final project because I am also modeling short, headline-like text using supervised learning, though my data comes from news headlines instead of teaser tweets. The way Potthast et al. treat clickbait detection as a standard text classification task with additional structural features closely matches my plan for the classical baseline. Their emphasis on corpus construction and bias is also relevant for interpreting my results on the Kaggle dataset, which may have its own selection effects. When I analyze my models' errors and generalization behavior, I can use their discussion of topic and publisher bias as a lens for understanding where my approach might fail to transfer beyond the specific dataset I am using.

Devlin et al. (2019) – BERT

Devlin et al. introduce BERT, a transformer-based model pre-trained on large unlabeled corpora using masked language modeling and next-sentence prediction. After this pre-training phase, BERT can be fine-tuned on a wide range of downstream NLP tasks by adding a simple task-specific output layer and training on labeled examples. The paper shows that this approach leads to state-of-the-art performance on sentence classification, natural language inference, and question answering tasks, among others. The core idea is that deep bidirectional contextual representations capture rich semantic and syntactic information that traditional word embeddings and shallow models do not.

This is related to my final project because my transformer-based approach is an instance of the BERT fine-tuning paradigm applied to binary headline classification. In my final project, I plan to take a pre-trained transformer and add a classification head that predicts whether a headline is clickbait or informative, using the same training split as my classical model. The results and methodology in Devlin et al. justify this design: they show that pre-trained transformers can achieve strong performance with relatively little labeled data and minimal architectural changes. Their work also motivates comparing error patterns between the classical model and the transformer model, since BERT-style representations should be better at handling subtle contextual cues that simple bag-of-words features may miss.

Yi et al. (2022)

Yi et al. propose a contrastive variational framework for clickbait detection that jointly models the relationship between text and labels. Their architecture combines a Variational Autoencoder with a classification objective and a contrastive loss that encourages better separation of latent representations for different labels. By jointly learning to encode text, predict the label, and structure the latent space, the model aims to be more robust to bias and noise in the training data. Experiments on multiple clickbait datasets show that this approach outperforms strong baselines, including fine-tuned BERT models, especially when labeled data are limited or noisy.

This is related to my final project because it places my own transformer-based baseline within a broader research trend that augments standard classification with generative and contrastive techniques. In my current scope, I am focusing on a straightforward supervised fine-tuning setup, but Yi et al.'s work highlights how more advanced objectives might address limitations I see in my results, such as systematic errors on borderline or noisy headlines. When I discuss possible extensions and future work, I can point to this paper as an example of using richer training objectives for clickbait detection, and explain how similar ideas—such as contrastive learning over headline representations—could be incorporated into an improved version of my final project.