

Home &gt; Recent Articles

# 10 Challenges of sentiment analysis and how to overcome them Part 4

AUTHORS Paul Simmering, Thomas Perry

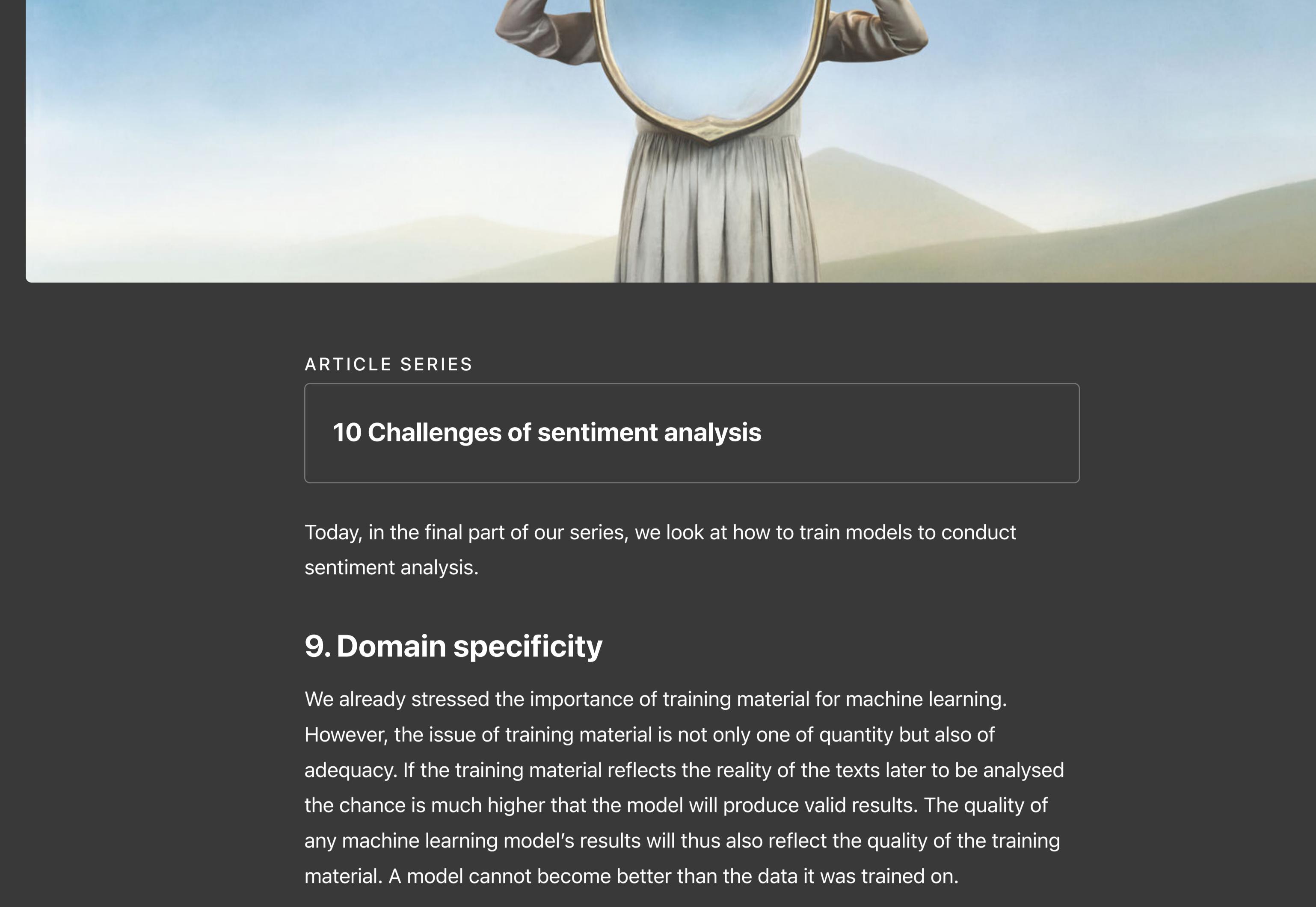
16 February 2023

A guide for evaluation of sentiment analysis solutions for market research:  
What to watch out for when choosing sentiment analysis software

Analytics

7 min read

Share



## ARTICLE SERIES

**10 Challenges of sentiment analysis**

Today, in the final part of our series, we look at how to train models to conduct sentiment analysis.

**9. Domain specificity**

We already stressed the importance of training material for machine learning. However, the issue of training material is not only one of quantity but also of adequacy, if the training material reflects the reality of the texts later to be analysed the chance is much higher that the model will produce valid results. The quality of any machine learning model's results will thus also reflect the quality of the training material. A model cannot become better than the data it was trained on.

This leads us to the nature of the text material to be analysed. In sentiment analysis, we often have to deal with diverse material from all kinds of sources like forums, Twitter, Facebook, Instagram, blogs, online shops, etc. On each of these platforms, users developed diverse cultures, styles, codes and quality levels of writing and expressing oneself. Across platforms, we, therefore, find very different texts: short and long, good and bad grammar, good and bad spelling, abbreviations versus complete words, etc. Models that were trained on the typical news text and Wikipedia datasets struggle with user-generated data.

However, there is one more problem because even the topic at hand makes a stark difference. One cannot use a model trained on political comments and expect it to do well on cosmetics. A model trained on electronic products will not work for food. In each case, the vocabulary used is different and for machine models, this makes a difference.

For a sentiment analysis model, this domain specificity is a difficult challenge to master. In many cases, it needs specific training material for each of them to generate valid results. Of course, this creates a dilemma for all who develop sentiment analysis tools. Building a model for each domain takes a lot of time and money.

Fortunately, it is not necessary to start from nothing in each domain. Large neural models that are trained on billions of words of general text can be adapted to specific tasks through transfer learning. The transformer model BERT (Devlin et al. 2018) has proven to be highly adaptable to different domains. Since even larger and more capable language models such as GPT-3 (Brown et al., 2020), were released and are available for fine-tuning, e.g., using the services of OpenAI.

The typical steps for transfer learning are 1) Selecting an appropriate base model, 2) designing an annotation schema, 3) annotating a training dataset, 4) training and evaluating a model. Steps 3 and 4 are repeated until the desired performance is achieved, but it is also worth iterating on the annotation schema from step 2.

**10. Building trust in a system that is a black box and not always right**

Even current state-of-the-art sentiment analysis models deployed in the domain they were trained on do not reach 100% accuracy. Users will see mistakes, which can undermine the trust that users have into the models. This is exacerbated

- When the model is a black box model, such as a neural net model, so that users do not know how it comes to its results.
- When its validity and the quality of its results are not tested, demonstrated, and proven so that users do not have a clue what the results mean for them.
- When the model is not adapted repeatedly to the development of the domain. Social media discussions evolve, word choices change, and new platforms rise in popularity. Models must be re-tested with fresh labelled data and updated as necessary.

Model providers must manage the reputation of their models. They can provide users with the results of their own tests on validation data (data that the model has not seen during training) and give an accuracy estimate. Still, users need to have a way to communicate with the model provider and point out flaws of the model, especially ones that have a suspicious pattern to them. This also places responsibility on the customers. They should be aware of the challenges that sentiment analysis poses and the threats this means to the benefits of sentiment analysis. In the end, it's their money spent and their conclusions and measures derived from the results it provides.

**Conclusion**

When used correctly, sentiment analysis can provide insight on enormous amounts of text, representing thousands or millions of opinion holders. However, the 10 challenges we described here are not easy to overcome.

There are three approaches that seem promising:

1. Transformer models are currently the best-performing ones thanks to their ability to understand words in context. They continue to evolve with larger language models as well as advances in GPUs.
2. Smart text selection is critical for correct analysis. Even the best algorithm cannot make up for problems in the data pipeline, such as irrelevant posts (by language, opinion holder, or topic). This layer of the data pipeline is easier to improve than the sentiment analysis model. Reviews are the easiest to work with because they have a clear target (the product) and an opinion holder (the buyer of the product).
3. Domain-specific models can combine intelligent filters for text selection with sentiment analysis models trained on data from the specific domain. This promises accuracy improvements in comparison to general models.

Sentiment analysis remains one of the tough challenges in machine learning. Our general advice on the topic:

- Prefer models that give fine-grained predictions by aspect, rather than summarising whole texts in one number.
- Test the validity and quality of the analysis you receive.
- Ask about the metrics used to evaluate the model and the inter-coder reliability in the training data.

Any provider with a solid model should be happy to help you with this by giving access to the necessary information. If not, mistrust is what you need.

**References**

- Augenstein, Isabelle, Tim Rocktäschel, Andreas Vlachos, and Kellina Bontcheva (2016). "stance Detection with Bidirectional Conditional Encoding." arXiv Preprint arXiv:1606.05464. <https://arxiv.org/abs/1606.05464>.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877–1901. <https://proceedings.neurips.cc/paper/2020/file/1457c0d6fb4967418bf8ac142f64a-Paper.pdf>
- Chen, Edwin (2022). 30% of Google's Emotions Dataset is Mislabelled. <https://www.surgehq.ai/blog/30-percent-of-googles-reddit-emotions-dataset-is-mislabelled>
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805. <https://arxiv.org/abs/1810.04805>
- Hutto, Clayton, and Eric Gilbert (2014). "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 8. 1. <https://ojs.aaai.org/index.php/ICWSM/article/view/14850>.
- Liu, Bing (2020). "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions." *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. 2nd edition. Cambridge University Press.
- Mishev, K., Gjorgjevik, J., Vodencka, I., Chirkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE access*, 8, 131662–131682.
- Pascual, Federico (2019). Guide to Aspect-Based Sentiment Analysis. <https://monkeylearn.com/blog/aspect-based-sentiment-analysis/>.
- Weber, R., Mangus, J. M., Huskey, R., Hopp, F. R., Amir, O., Swanson, R., ... & Tamborini, R. (2018). Extracting latent moral information from text narratives: Relevance, challenges, and solutions. *Communication Methods and Measures*, 12(2–3), 119–139. [https://www.researchgate.net/profile/Richard-Huskey-2/publication/323789093\\_Extracting\\_Latent\\_Moral\\_Information\\_from\\_Text\\_Narratives\\_Relevance\\_Challenges\\_and\\_Solutions/links/5a8b589345851517881b55c9/Extracting-Latent-Moral-Information-from-Text-Narratives-Relevance-Challenges-and-Solutions.pdf](https://www.researchgate.net/profile/Richard-Huskey-2/publication/323789093_Extracting_Latent_Moral_Information_from_Text_Narratives_Relevance_Challenges_and_Solutions/links/5a8b589345851517881b55c9/Extracting-Latent-Moral-Information-from-Text-Narratives-Relevance-Challenges-and-Solutions.pdf)
- Van Atteveldt, Wouter, Mariken ACC Van der Velden, and Mark Boukes (2021). The Validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures* 15:2: 121–140. <https://www.tandfonline.com/doi/pdf/10.1080/19312458.2020.1869198>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Yang, Heng (2020). PyABSA – Open Framework for Aspect-based Sentiment Analysis. <https://github.com/yangheng95/PyABSA>.

There are three approaches that seem promising:

1. Transformer models are currently the best-performing ones thanks to their ability to understand words in context. They continue to evolve with larger language models as well as advances in GPUs.
2. Smart text selection is critical for correct analysis. Even the best algorithm cannot make up for problems in the data pipeline, such as irrelevant posts (by language, opinion holder, or topic). This layer of the data pipeline is easier to improve than the sentiment analysis model. Reviews are the easiest to work with because they have a clear target (the product) and an opinion holder (the buyer of the product).
3. Domain-specific models can combine intelligent filters for text selection with sentiment analysis models trained on data from the specific domain. This promises accuracy improvements in comparison to general models.

Any provider with a solid model should be happy to help you with this by giving access to the necessary information. If not, mistrust is what you need.

**References**

- Augenstein, Isabelle, Tim Rocktäschel, Andreas Vlachos, and Kellina Bontcheva (2016). "stance Detection with Bidirectional Conditional Encoding." arXiv Preprint arXiv:1606.05464. <https://arxiv.org/abs/1606.05464>.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877–1901. <https://proceedings.neurips.cc/paper/2020/file/1457c0d6fb4967418bf8ac142f64a-Paper.pdf>
- Chen, Edwin (2022). 30% of Google's Emotions Dataset is Mislabelled. <https://www.surgehq.ai/blog/30-percent-of-googles-reddit-emotions-dataset-is-mislabelled>
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805. <https://arxiv.org/abs/1810.04805>
- Hutto, Clayton, and Eric Gilbert (2014). "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 8. 1. <https://ojs.aaai.org/index.php/ICWSM/article/view/14850>.
- Liu, Bing (2020). "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions." *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. 2nd edition. Cambridge University Press.
- Mishev, K., Gjorgjevik, J., Vodencka, I., Chirkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE access*, 8, 131662–131682.
- Pascual, Federico (2019). Guide to Aspect-Based Sentiment Analysis. <https://monkeylearn.com/blog/aspect-based-sentiment-analysis/>.
- Weber, R., Mangus, J. M., Huskey, R., Hopp, F. R., Amir, O., Swanson, R., ... & Tamborini, R. (2018). Extracting latent moral information from text narratives: Relevance, challenges, and solutions. *Communication Methods and Measures*, 12(2–3), 119–139. [https://www.researchgate.net/profile/Richard-Huskey-2/publication/323789093\\_Extracting\\_Latent\\_Moral\\_Information\\_from\\_Text\\_Narratives\\_Relevance\\_Challenges\\_and\\_Solutions/links/5a8b589345851517881b55c9/Extracting-Latent-Moral-Information-from-Text-Narratives-Relevance-Challenges-and-Solutions.pdf](https://www.researchgate.net/profile/Richard-Huskey-2/publication/323789093_Extracting_Latent_Moral_Information_from_Text_Narratives_Relevance_Challenges_and_Solutions/links/5a8b589345851517881b55c9/Extracting-Latent-Moral-Information-from-Text-Narratives-Relevance-Challenges-and-Solutions.pdf)
- Van Atteveldt, Wouter, Mariken ACC Van der Velden, and Mark Boukes (2021). The Validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures* 15:2: 121–140. <https://www.tandfonline.com/doi/pdf/10.1080/19312458.2020.1869198>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Yang, Heng (2020). PyABSA – Open Framework for Aspect-based Sentiment Analysis. <https://github.com/yangheng95/PyABSA>.

**References**

- Augenstein, Isabelle, Tim Rocktäschel, Andreas Vlachos, and Kellina Bontcheva (2016). "stance Detection with Bidirectional Conditional Encoding." arXiv Preprint arXiv:1606.05464. <https://arxiv.org/abs/1606.05464>.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877–1901. <https://proceedings.neurips.cc/paper/2020/file/1457c0d6fb4967418bf8ac142f64a-Paper.pdf>
- Chen, Edwin (2022). 30% of Google's Emotions Dataset is Mislabelled. <https://www.surgehq.ai/blog/30-percent-of-googles-reddit-emotions-dataset-is-mislabelled>
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805. <https://arxiv.org/abs/1810.04805>
- Hutto, Clayton, and Eric Gilbert (2014). "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 8. 1. <https://ojs.aaai.org/index.php/ICWSM/article/view/14850>.
- Liu, Bing (2020). "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions." *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. 2nd edition. Cambridge University Press.
- Mishev, K., Gjorgjevik, J., Vodencka, I., Chirkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE access*, 8, 131662–131682.
- Pascual, Federico (2019). Guide to Aspect-Based Sentiment Analysis. <https://monkeylearn.com/blog/aspect-based-sentiment-analysis/>.
- Weber, R., Mangus, J. M., Huskey, R., Hopp, F. R., Amir, O., Swanson, R., ... & Tamborini, R. (2018). Extracting latent moral information from text narratives: Relevance, challenges, and solutions. *Communication Methods and Measures*, 12(2–3), 119–139. [https://www.researchgate.net/profile/Richard-Huskey-2/publication/323789093\\_Extracting\\_Latent\\_Moral\\_Information\\_from\\_Text\\_Narratives\\_Relevance\\_Challenges\\_and\\_Solutions/links/5a8b589345851517881b55c9/Extracting-Latent-Moral-Information-from-Text-Narratives-Relevance-Challenges-and-Solutions.pdf](https://www.researchgate.net/profile/Richard-Huskey-2/publication/323789093_Extracting_Latent_Moral_Information_from_Text_Narratives_Relevance_Challenges_and_Solutions/links/5a8b589345851517881b55c9/Extracting-Latent-Moral-Information-from-Text-Narratives-Relevance-Challenges-and-Solutions.pdf)
- Van Atteveldt, Wouter, Mariken ACC Van der Velden, and Mark Boukes (2021). The Validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures* 15:2: 121–140. <https://www.tandfonline.com/doi/pdf/10.1080/19312458.2020.1869198>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Yang, Heng (2020). PyABSA – Open Framework for Aspect-based Sentiment Analysis. <https://github.com/yangheng95/PyABSA>.

**References**

- Augenstein, Isabelle, Tim Rocktäschel, Andreas Vlachos, and Kellina Bontcheva (2016). "stance Detection with Bidirectional Conditional Encoding." arXiv Preprint arXiv:1606.05464. <https://arxiv.org/abs/1606.05464>.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877–1901. <https://proceedings.neurips.cc/paper/2020/file/1457c0d6fb4967418bf8ac142f64a-Paper.pdf>
- Chen, Edwin (2022). 30% of Google's Emotions Dataset is Mislabelled. <https://www.surgehq.ai/blog/30-percent-of-googles-reddit-emotions-dataset-is-mislabelled>