**References**

Vaj, T. (2023, July 31). *How to evaluate novel topic modeling method.* Medium. https://vtiya.medium.com/how-to-evaluate-novel-topic-modeling-method-104ad9684428

Jokofa. (2022, May 29). *Torch\_kmeans/examples at master · Jokofa/torch\_kmeans*. GitHub. https://github.com/jokofa/torch\_kmeans/tree/master/examples

*Torch-Kmeans*. PyPI. (2022, December 4). https://pypi.org/project/torch-kmeans/

GeeksforGeeks. (2025, July 11). *Gaussian mixture model*. https://www.geeksforgeeks.org/machine-learning/gaussian-mixture-model/

GeeksforGeeks. (2025, July 17). *Bert Model - NLP*. https://www.geeksforgeeks.org/nlp/explanation-of-bert-model-nlp/

Vidhya. (2019, January 28). *Introduction to Topic Modeling - (Learn Natural Language Processing using Python)*. YouTube. https://www.youtube.com/watch?v=p1I9Sa1lRvk

Serrano. (2020, March 18). *Latent Dirichlet Allocation*. YouTube. https://www.youtube.com/watch?v=T05t-SqKArY

Serrano. (2020, March 21). *Training Latent Dirichlet Allocation: Gibbs Sampling*. YouTube. https://www.youtube.com/watch?v=BaM1uiCpj\_E

*Bert 101 - state of the art NLP model explained*. BERT 101 - State Of The Art NLP Model Explained. (n.d.-a). https://huggingface.co/blog/bert-101#2-how-does-bert-work

MOAZ47. (2020, November 8). *Shared folder does not show when I mount google drive in Colab · issue #1694 · googlecolab/colabtools*. GitHub. https://github.com/googlecolab/colabtools/issues/1694

Chalabi, M. (2021, November). *What we miss when we focus on the average*. TED. https://www.ted.com/talks/mona\_chalabi\_what\_we\_miss\_when\_we\_focus\_on\_the\_average

Pickleprat. (2024a, April 3). *Encoder only architecture: Bert*. Medium. https://medium.com/@pickleprat/encoder-only-architecture-bert-4b27f9c76860

Grootendorst, M. P. (n.d.). *Outlier reduction*. Outlier reduction - BERTopic. https://maartengr.github.io/BERTopic/getting\_started/outlier\_reduction/outlier\_reduction.html

Grootendorst, M. P. (n.d.-a). *3. clustering*. 3. Clustering - BERTopic. https://maartengr.github.io/BERTopic/getting\_started/clustering/clustering.html

K, F. (2024, November 26). *Understanding silhouette score in clustering*. Medium. https://farshadabdulazeez.medium.com/understanding-silhouette-score-in-clustering-8aedc

06ce9c4

Idrees, H. (2024, July 29). *Encoder vs. Decoder in transformers: Unpacking the differences*. Medium. https://medium.com/@hassaanidrees7/encoder-vs-decoder-in-transformers-unpacking-the-differences-9e6ddb0ff3c5

*How to get the probability of belonging to clusters for K-means?*. Data Science Stack Exchange. (1961, March 1). https://datascience.stackexchange.com/questions/14435/how-to-get-the-probability-of-belonging-to-clusters-for-k-means

“when we lost the forest, we lost everything.” (2019, September). https://www.hrw.org/sites/default/files/report\_pdf/indonesia0919\_web.pdf

Simran, | BY. (2024, October 28). *A beginner’s Guide to Topic Modeling in NLP*. ProjectPro. https://www.projectpro.io/article/topic-modeling-nlp/801

Distante, E. (2022, October 20). *Bertopic: Topic modeling as you have never seen it before*. Medium. https://medium.com/data-reply-it-datatech/bertopic-topic-modeling-as-you-have-never-seen-it-before-abb48bbab2b2

eadala. (2021, March). R/mlquestions on reddit: (NLP) best practices for topic modeling and generating interesting topics? https://www.reddit.com/r/MLQuestions/comments/noml1w/nlp\_best\_practices\_for\_topic\_modeling\_and/

Duplan, S. (2024, September 26). *LLM-Powered Topic Modeling*. Medium. https://sienaduplan.medium.com/llm-powered-topic-modeling-fdd1de8e4c1d

*Date spacy · spacy universe*. Date spaCy. (n.d.). https://spacy.io/universe/project/date-spacy

Li, S. (2018, December 6). *Latent semantic analysis & sentiment classification with python*. Medium. https://medium.com/data-science/latent-semantic-analysis-sentiment-classification-with-python-5f657346f6a3

AccelAI. (2022, February). *Ai-tutorials/lda\_tutorial/LDA-tutorial.ipynb at main · Accelai/AI-Tutorials*. GitHub. https://github.com/AccelAI/AI-Tutorials/blob/main/LDA\_Tutorial/LDA-Tutorial.ipynb

Python LSI/LSA (Latent Semantic Indexing/analysis) | datacamp. (2018, October 9). https://www.datacamp.com/tutorial/discovering-hidden-topics-python

*Classification on imbalanced data : Tensorflow Core*. TensorFlow. (n.d.). https://www.tensorflow.org/tutorials/structured\_data/imbalanced\_data#baseline\_model

McCloskey, B. (2025, March 5). *Implementing named entity recognition to find dates in pdfs*. Towards Data Science. https://towardsdatascience.com/implementing-named-entity-recognition-to-find-dates-in-pdfs-e2d7dcea8f7a/

Wikimedia Foundation. (2025, July 23). *East Malaysia*. Wikipedia. https://en.wikipedia.org/wiki/East\_Malaysia

NASA. (2009, September 10). *Deforestation in Malaysian Borneo*. NASA. https://earthobservatory.nasa.gov/images/40139/deforestation-in-malaysian-borneo

Wjbmattingly. (n.d.). *WJBMATTINGLY/date-spacy*. GitHub. https://github.com/wjbmattingly/date-spacy

Golden-Agri-Resources. (2025, January 6). *GAR Corporate Video*. YouTube. https://www.youtube.com/watch?v=bgU-i2C0jYU&t=90s

Golden Agri-Resources (GAR. (2025, April 14). *About Us*. Golden Agri-Resources. https://www.goldenagri.com.sg/about-us/

Golden Agri-Resources (GAR). (2025, May 15). *Grievance handling*. Golden Agri-Resources. https://www.goldenagri.com.sg/sustainability/responsible-sourcing/grievance-handling/

*Detecting dates*. Detecting dates - EDS-NLP. (n.d.). https://aphp.github.io/edsnlp/v0.4.3/tutorials/detecting-dates/#extracting-dates

When coherence score is good or bad in topic modeling? | Baeldung on computer science. (n.d.). https://www.baeldung.com/cs/topic-modeling-coherence-score

Grootendorst, M. P. (n.d.). *Visualization*. Visualization - BERTopic. https://maartengr.github.io/BERTopic/getting\_started/visualization/visualization.html#visualize-topic-hierarchy

*Tsne*. scikit. (n.d.). https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html

(burnpiro), K. E. (2022, July 21). *T-SNE clearly explained*. Medium. https://medium.com/data-science/t-sne-clearly-explained-d84c537f53a

YouTube. (n.d.). YouTube. https://www.youtube.com/watch?v=D9bdJm1GYFY

Distillpub. (n.d.). *Distillpub/post--misread-tsne: How to use T-Sne effectively*. GitHub. https://github.com/distillpub/post--misread-tsne

Wattenberg, M., Viégas, F., & Johnson, I. (2016, October 13). *How to use T-Sne effectively*. Distill. https://distill.pub/2016/misread-tsne/

Dib, F. (n.d.). *Build, test, and debug regex*. regex101. https://regex101.com/

Grootendorst, M. P. (n.d.-a). *Embedding models¶*. 1. Embeddings - BERTopic. https://maartengr.github.io/BERTopic/getting\_started/embeddings/embeddings.html

flairNLP. (n.d.). *FLAIRNLP/transformer-ranker: Efficiently find the best-suited language model (LM) for your NLP task*. GitHub. https://github.com/flairNLP/transformer-ranker

*Classification\_report*. scikit. (n.d.-a). https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification\_report.html

Stack Overflow (2018, June). *How do you link back topics generated by LDA model to actual document*. https://stackoverflow.com/questions/47138149/how-do-you-link-back-topics-generated-by-lda-model-to-actual-document

Grootendorst, M. P. (n.d.). *Visualization BERTopic*. Visualization - BERTopic. https://maartengr.github.io/BERTopic/getting\_started/visualization/visualization.html

Albanese, N. C. (2025, March 5). *Topic modeling with LSA, PLSA, LDA, NMF, Bertopic, top2vec: A comparison*. Towards Data Science. https://towardsdatascience.com/topic-modeling-with-lsa-plsa-lda-nmf-bertopic-top2vec-a-comparison-5e6ce4b1e4a5/#5a51

*Torch-Kmeans*. PyPI. (n.d.). https://pypi.org/project/torch-kmeans/

GeeksforGeeks. (2025, July 11). *Gaussian mixture model*. https://www.geeksforgeeks.org/machine-learning/gaussian-mixture-model/

Yedivanseven. (n.d.). *Yedivanseven/PLSA: Probabilistic Latent Semantic Analysis*. GitHub. https://github.com/yedivanseven/PLSA