

## References

- [1] Dean Takahashi. *Nvidia CEO bets big on deep learning and VR*. VentureBeat. Apr. 5, 2016. URL: <https://venturebeat.com/2016/04/05/nvidia-ceo-bets-big-on-deep-learning-and-vr/> (visited on 03/11/2021).
- [2] Andrej Karpathy. “PyTorch at Tesla - Andrej Karpathy, Tesla - YouTube”. Presentation. PyTorch DevCon 2019. Digital, Nov. 6, 2019. URL: <https://www.youtube.com/watch?v=oBkl1tKXtDE> (visited on 03/11/2021).
- [3] Mark Kizelshteyn and Mat Budelman. *Three Principles for Designing ML-Powered Products*. Spotify Design. URL: <https://spotify.design/article/three-principles-for-designing-ml-powered-products> (visited on 03/11/2021).
- [4] Tero Karras, Samuli Laine, and Timo Aila. “A Style-Based Generator Architecture for Generative Adversarial Networks”. In: *arXiv:1812.04948 [cs, stat]* (Mar. 29, 2019). arXiv: 1812.04948. URL: <http://arxiv.org/abs/1812.04948> (visited on 03/11/2021).
- [5] *What are Neural Networks?* Jan. 6, 2021. URL: <https://www.ibm.com/cloud/learn/neural-networks> (visited on 03/17/2021).
- [6] Hao Ren and Hong Lu. “Compositional coding capsule network with k-means routing for text classification”. In: *arXiv:1810.09177 [cs, stat]* (Oct. 29, 2018). arXiv: 1810.09177. URL: <http://arxiv.org/abs/1810.09177> (visited on 03/02/2021).
- [7] Edith A. Moravcsik. *Introduction to Syntactic Theory*. London: Continuum, 2006. ISBN: 978-0-8264-8943-2.
- [8] Hany Alashwal et al. “The Application of Unsupervised Clustering Methods to Alzheimer’s Disease”. In: *Frontiers in Computational Neuroscience* 13 (2019), p. 31. ISSN: 1662-5188. DOI: 10.3389/fncom.2019.00031. URL: <https://www.frontiersin.org/article/10.3389/fncom.2019.00031>.
- [9] M Tarik Altuncu et al. “From free text to clusters of content in health records: an unsupervised graph partitioning approach”. In: *Applied network science* 4.1 (2019). Publisher: Springer, p. 2.
- [10] Meiyin Wu and Li Chen. “Image recognition based on deep learning”. In: *2015 Chinese Automation Congress (CAC)*. 2015 Chinese Automation Congress (CAC). Nov. 2015, pp. 542–546. DOI: 10.1109/CAC.2015.7382560.
- [11] Daniel Jurafsky and James Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Vol. 2. Feb. 1, 2008.
- [12] Suneel Patel. *NLP Pipeline: Building an NLP Pipeline, Step-by-Step*. Medium. Aug. 3, 2020. URL: <https://medium.com/@suneelpatel.in/nlp-pipeline-building-an-nlp-pipeline-step-by-step-7f0576e11d08> (visited on 03/12/2021).
- [13] *What is Machine Learning?* Dec. 18, 2020. URL: <https://www.ibm.com/cloud/learn/machine-learning> (visited on 03/19/2021).

- [14] Junyuan Xie, Ross Girshick, and Ali Farhadi. “Unsupervised Deep Embedding for Clustering Analysis”. In: *arXiv:1511.06335 [cs]* (May 24, 2016). arXiv: 1511.06335. URL: <http://arxiv.org/abs/1511.06335> (visited on 03/02/2021).
- [15] Dongkuan Xu and Yingjie Tian. “A Comprehensive Survey of Clustering Algorithms”. In: *Annals of Data Science* 2.2 (June 1, 2015), pp. 165–193. ISSN: 2198-5812. DOI: 10.1007/s40745-015-0040-1. URL: <https://doi.org/10.1007/s40745-015-0040-1> (visited on 02/17/2021).
- [16] *Twitter. It’s what’s happening.* Twitter. URL: <https://twitter.com/> (visited on 03/17/2021).
- [17] *YouTube.* URL: <https://www.youtube.com/> (visited on 03/17/2021).
- [18] Fortune Business Insights. *Machine Learning Market to Reach USD 117.19 Billion by 2027.* GlobeNewswire News Room. July 17, 2020. URL: <http://www.globenewswire.com/news-release/2020/07/17/2063938/0/en/Machine-Learning-Market-to-Reach-USD-117-19-Billion-by-2027-Increasing-Popularity-of-Self-Driving-Cars-to-Propel-Demand-from-Automotive-Industry-says-Fortune-Business-Insights.html> (visited on 03/19/2021).
- [19] *PyTorch.* URL: <https://www.pytorch.org> (visited on 03/10/2021).
- [20] *scikit-learn: machine learning in Python — scikit-learn 0.24.1 documentation.* URL: <https://scikit-learn.org/stable/> (visited on 03/10/2021).
- [21] Titus Neupert et al. “Introduction to Machine Learning for the Sciences”. In: *arXiv:2102.04883 [cond-mat, physics:physics]* (Feb. 8, 2021). arXiv: 2102.04883. URL: <http://arxiv.org/abs/2102.04883> (visited on 03/22/2021).
- [22] *But what is a Neural Network? — Deep learning, chapter 1.* In collab. with Grant Sanderson. Oct. 5, 2017. URL: <https://www.youtube.com/watch?v=aircAruvnKk&t=1s> (visited on 03/23/2021).
- [23] Michael A. Nielsen. “Neural Networks and Deep Learning”. In: (2015). Publisher: Determination Press. URL: <http://neuralnetworksanddeeplearning.com> (visited on 02/04/2021).
- [24] Michael McTear, Zoraida Callejas, and David Griol. *The Conversational Interface: Talking to Smart Devices.* Springer International Publishing, 2016. ISBN: 978-3-319-32965-9. DOI: 10.1007/978-3-319-32967-3. URL: <https://www.springer.com/gp/book/9783319329659> (visited on 04/20/2021).
- [25] “Data Mining”. In: *Mining of Massive Datasets.* Ed. by Anand Rajaraman and Jeffrey David Ullman. Cambridge: Cambridge University Press, 2011, pp. 1–17. ISBN: 978-1-107-73741-9. DOI: 10.1017/CB09781139058452.002. URL: <https://www.cambridge.org/core/books/mining-of-massive-datasets/data-mining/E5BFF4C1DD5A1FB946D616D619B373C2> (visited on 04/20/2021).
- [26] *Decomposing signals in components (matrix factorization problems) — scikit-learn 0.24.2 documentation.* URL: <https://scikit-learn.org/stable/modules/decomposition.html#truncated-singular-value-decomposition-and-latent-semantic-analysis> (visited on 05/25/2021).

- [27] Laurens van der Maaten and Geoffrey Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. URL: <http://jmlr.org/papers/v9/vandermaten08a.html>.
- [28] *Natural Language Toolkit — NLTK 3.6.2 documentation*. URL: <https://www.nltk.org/> (visited on 05/04/2021).
- [29] *Snowball*. URL: <https://snowballstem.org/> (visited on 05/06/2021).
- [30] Princeton University. *About WordNet*. WordNet. 2010. URL: <https://wordnet.princeton.edu/> (visited on 05/06/2021).
- [31] Martin Ester et al. “A density-based algorithm for discovering clusters in large spatial databases with noise”. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. KDD’96. Portland, Oregon: AAAI Press, Aug. 2, 1996, pp. 226–231. (Visited on 04/20/2021).
- [32] Robert L. Thorndike. “Who belongs in the family?” In: *Psychometrika* 18.4 (Dec. 1, 1953), pp. 267–276. ISSN: 1860-0980. DOI: 10.1007/BF02289263. URL: <https://doi.org/10.1007/BF02289263> (visited on 04/21/2021).
- [33] Fabrizio Sebastiani. “Machine Learning in Automated Text Categorization”. In: *arXiv:cs/0110053* (Oct. 26, 2001). version: 1. DOI: 10.1145/505282.505283. arXiv: cs/0110053. URL: <http://arxiv.org/abs/cs/0110053> (visited on 04/21/2021).
- [34] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [35] Peter J. Rousseeuw. “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65. ISSN: 0377-0427. DOI: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7). URL: <https://www.sciencedirect.com/science/article/pii/0377042787901257>.
- [36] David Arthur and Sergei Vassilvitskii. *k-means++: The Advantages of Careful \- Seeding*. Technical Report 2006-13. Backup Publisher: Stanford InfoLab. Stanford, June 2006. URL: <http://ilpubs.stanford.edu:8090/778/>.
- [37] Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. “Density-Based Clustering Based on Hierarchical Density Estimates”. In: *Advances in Knowledge Discovery and Data Mining*. Ed. by Jian Pei et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 160–172. ISBN: 978-3-642-37456-2.
- [38] Ammar Ismael Kadhim, Yu-N Cheah, and Nurul Hashimah Ahamed. “Text Document Preprocessing and Dimension Reduction Techniques for Text Document Clustering”. In: *2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology*. 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology. Dec. 2014, pp. 69–73. DOI: 10.1109/ICALET.2014.21.
- [39] Andrej Karpathy. *Visualizing Top Tweeps with t-SNE, in Javascript*. July 14, 2015. URL: <https://karpathy.github.io/2014/07/02/visualizing-top-tweeps-with-t-sne-in-Javascript/> (visited on 05/06/2021).

- [40] Nick Becker. *Clustering US Laws using TF-IDF and K-Means*. nick becker. Aug. 18, 2016. URL: <https://beckernick.github.io/law-clustering/> (visited on 05/04/2021).
- [41] Dipanjan Sarkar. *Text Analytics with Python: A Practitioner’s Guide to Natural Language Processing*. Publication Title: Text Analytics with Python. Berkeley, CA: Apress L. P, 2019. ISBN: 1-4842-4353-6.
- [42] Dipanjan (DJ) Sarkar. *dipanjanS/text-analytics-with-python*. original-date: 2016-09-25T19:19:02Z. May 6, 2021. URL: <https://github.com/dipanjanS/text-analytics-with-python> (visited on 05/06/2021).
- [43] Sven Eklund. *Arbeta i projekt : individen, gruppen, ledaren*. Lund: Studentlitteratur, 2011. ISBN: 978-91-44-07275-3. URL: <http://libris.kb.se/bib/12182779> (visited on 05/02/2021).
- [44] George Krasadakis. *How to build Software Prototypes in rapid mode*. Medium. Oct. 11, 2020. URL: <https://medium.com/innovation-machine/innovation-at-pace-rapid-prototyping-practices-for-software-engineering-teams-442929fdd5ea> (visited on 05/02/2021).
- [45] *Text classification with the torchtext library — PyTorch Tutorials 1.8.1+cu102 documentation*. URL: [https://pytorch.org/tutorials/beginner/text\\_sentiment\\_ngrams\\_tutorial.html](https://pytorch.org/tutorials/beginner/text_sentiment_ngrams_tutorial.html) (visited on 05/04/2021).
- [46] *Welcome to Python.org*. Python.org. URL: <https://www.python.org/> (visited on 05/03/2021).
- [47] Stefan Behnel et al. “Cython: The Best of Both Worlds”. In: *Computing in Science & Engineering* 13.2 (Mar. 2011), pp. 31–39. ISSN: 1521-9615. DOI: 10.1109/MCSE.2010.118. URL: <http://ieeexplore.ieee.org/document/5582062/> (visited on 05/03/2021).
- [48] Ronan Collobert, Samy Bengio, and Johny Marithoz. *Torch: A Modular Machine Learning Software Library*. 2002.
- [49] *Overview — NumPy v1.20 Manual*. URL: <https://numpy.org/doc/stable/> (visited on 05/04/2021).
- [50] *Documentation — SciPy.org*. URL: <https://www.scipy.org/docs.html> (visited on 05/04/2021).
- [51] *Yelp*. Yelp. URL: <https://www.yelp.com/> (visited on 05/02/2021).
- [52] Alec Go, Richa Bhayani, and Lei Huang. “Twitter sentiment classification using distant supervision”. In: *CS224N Project Report, Stanford* 1 (2009), p. 12.
- [53] *Congress.gov — Library of Congress*. URL: <https://www.congress.gov/> (visited on 05/19/2021).
- [54] Erik Lenas. *Prerequisites for Extracting Entity Relations from Swedish Texts*. 2020. URL: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-281275> (visited on 05/21/2021).

- [55] Tomas Mikolov et al. “Efficient Estimation of Word Representations in Vector Space”. In: *arXiv:1301.3781 [cs]* (Sept. 6, 2013). arXiv: 1301.3781. URL: <http://arxiv.org/abs/1301.3781> (visited on 05/21/2021).
- [56] Xiang Zhang, Junbo Zhao, and Yann LeCun. “Character-level Convolutional Networks for Text Classification”. In: *arXiv:1509.01626 [cs]* (Apr. 3, 2016). arXiv: 1509.01626. URL: <http://arxiv.org/abs/1509.01626> (visited on 05/20/2021).
- [57] *Feature extraction — scikit-learn 0.24.2 documentation*. URL: [https://scikit-learn.org/stable/modules/feature\\_extraction.html](https://scikit-learn.org/stable/modules/feature_extraction.html) (visited on 05/21/2021).



