

References Part I: Introduction

- Bengio, Y., Goodfellow, I.J. & Courville, A., 2015. Deep learning. *Nature*, 521(7553), pp.436–444.
- Bishop, C.M., 2016. *Pattern Recognition and Machine Learning*, Springer New York.
- Doan, A., Halevy, A.Y. & Ives, Z.G., 2012. *Principles of Data Integration*, Morgan Kaufmann.
- Domingos, P., 2012. A Few Useful Things to Know About Machine Learning. *Communications of the ACM*, 55(10), pp.78–87.
- Dong, X. et al., 2014. Knowledge Vault: A Web-scale Approach to Probabilistic Knowledge Fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '14. New York, NY, USA: ACM, pp. 601–610.
- Dong, X.L. & Srivastava, D., 2015. Big data integration. *Synthesis Lectures on Data Management*, 7(1), pp.1–198.
- Dong, X.L. & Srivastava, D., 2013. Big Data Integration. *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*, 6(11), pp.1188–1189.
- Getoor, L. & Machanavajjhala, A., 2012. Entity resolution: theory, practice & open challenges. *PVLDB*, 5(12), pp.2018–2019.
- Goodfellow, I. et al., 2016. *Deep learning*, MIT press Cambridge.
- Halevy, A., Norvig, P. & Pereira, F., 2009. The Unreasonable Effectiveness of Data. *IEEE intelligent systems*, 24(2), pp.8–12.
- Konda, P. et al., 2016. Magellan: Toward Building Entity Matching Management Systems. *PVLDB*, 9(12), pp.1197–1208.

References Part I: Introduction

- Kumar, A., Boehm, M. & Yang, J., 2017. Data Management in Machine Learning: Challenges, Techniques, and Systems. In *Proceedings of the 2017 ACM International Conference on Management of Data*. SIGMOD '17. New York, NY, USA: ACM, pp. 1717–1722.
- Lockard, C. et al., 2018. CERES: Distantly Supervised Relation Extraction from the Semi-Structured Web. *arXiv [cs.AI]*. Available at: <http://arxiv.org/abs/1804.04635>.
- Mohri, M., Rostamizadeh, A. & Talwalkar, A., 2012. *Foundations of Machine Learning*, MIT Press.
- Polyzotis, N. et al., 2017. Data Management Challenges in Production Machine Learning. In *Proceedings of the 2017 ACM International Conference on Management of Data*. SIGMOD '17. New York, NY, USA: ACM, pp. 1723–1726.
- Ratner, A. et al., 2017. Snorkel: Rapid Training Data Creation with Weak Supervision. *PVLDB*, 11(3), pp.269–282.
- Rekatsinas, T. et al., 2017. HoloClean: Holistic Data Repairs with Probabilistic Inference. *PVLDB*, 10(11), pp.1190–1201.
- Wu, S. et al., 2018. Fonduer: Knowledge Base Construction from Richly Formatted Data. In *Proceedings of the 2018 International Conference on Management of Data*. ACM, pp. 1301–1316.
- Zheng, G. et al., 2018. OpenTag: Open Attribute Value Extraction from Product Profiles. In *KDD*. Available at: <https://people.mpi-inf.mpg.de/~smukherjee/research/OpenTag-KDD18.pdf>.