References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*(52), 138–160. https://doi.org/10.1109/access.2018.2870052

Anowar, F., & Sadaoui, S. (2020). Incremental Neural-Network Learning for Big Fraud Data. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, *1*(1), 1–4. https://doi.org/10.1109/smc42975.2020.9283136

Batageri, A., & Kumar, S. (2021). Credit card fraud detection using artificial neural network. *Global Transitions Proceedings*, *2*(1), 35–41. https://doi.org/10.1016/j.gltp.2021.01.006

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8). https://doi.org/10.3390/electronics8080832

Chromik, M., Eiband, M., Buchner, F., Krüger, A., & Butz, A. (2021). I think I get your point, ai! the illusion of explanatory depth in explainable ai. *26th International Conference on Intelligent User Interfaces*. https://doi.org/10.1145/3397481.3450644

Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications*, *41*(10), 4915–4928. https://doi.org/10.1016/j.eswa.2014.02.026

Darias, J. M., Caro-Martínez, M., Díaz-Agudo, B., & Recio-Garcia, J. A. (2022). Using case-based reasoning for capturing expert knowledge on explanation methods. *Case-Based Reasoning Research and Development*, *13405*, 3–17. https://doi.org/10.1007/978-3-031-14923-8\_1

ElShawi, R., Sherif, Y., Al‐Mallah, M., & Sakr, S. (2020). Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Computational Intelligence*, *37*(4), 1633–1650. https://doi.org/10.1111/coin.12410

Evans, B. P., Xue, B., & Zhang, M. (2019). What's inside the black-box? *Proceedings of the Genetic and Evolutionary Computation Conference*. https://doi.org/10.1145/3321707.3321726

Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., & Turini, F. (2019). Factual and counterfactual explanations for black box decision making. *IEEE Intelligent Systems*, *34*(6), 14–23. https://doi.org/10.1109/mis.2019.2957223

Hanafy, M., & Ming, R. (2022). Classification of the insureds using Integrated Machine Learning Algorithms: A comparative study. *Applied Artificial Intelligence*, *36*(1). https://doi.org/10.1080/08839514.2021.2020489

Honegger, M. (2018, August 15). *Shedding light on Black Box Machine Learning Algorithms: Development of an axiomatic framework to assess the quality of methods that explain individual predictions*. arXiv.org. Retrieved December 4, 2022, from https://arxiv.org/abs/1808.05054v1

Ignatiev, A. (2020). Towards trustable explainable ai. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 5154–5158. https://doi.org/10.24963/ijcai.2020/726

Jesus, S., Belém, C., Balayan, V., Bento, J., Saleiro, P., Bizarro, P., & Gama, J. (2021). How can I choose an explainer? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3442188.3445941

Jesus, S., Belém, C., Balayan, V., Bento, J., Saleiro, P., Bizarro, P., & Gama, J. (2021). How can I choose an explainer? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3442188.3445941

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (2020). Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3313831.3376219

Lakkaraju, H., Bach, S. H., & Leskovec, J. (2016). Interpretable decision sets. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1675–1684. https://doi.org/10.1145/2939672.2939874

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (Vol. 30). essay, NeurIPS Proceedings.

Marcilio, W. E., & Eler, D. M. (2020). From explanations to feature selection: Assessing shap values as feature selection mechanism. *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 340–347. https://doi.org/10.1109/sibgrapi51738.2020.00053

Nri, H., Jenkins, S., Paul, K., & Caruana, R. (2019). InterpretML: A unified framework for machine learning interpretability. *ArXiv Preprint ArXiv:1909.09223*.

Priscilla, C. V., & Prabha, D. P. (2020). Influence of optimizing xgboost to handle class imbalance in credit card fraud detection. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 1309–1315. https://doi.org/10.1109/icssit48917.2020.9214206

Psychoula, I., Gutmann, A., Mainali, P., Lee, S. H., Dunphy, P., & Petitcolas, F. (2021). Explainable Machine Learning for Fraud Detection. *Computer*, *54*(10), 49–59. https://doi.org/10.1109/mc.2021.3081249

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "why should I trust you?" Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). https://doi.org/10.1609/aaai.v32i1.11491

Schlegel, U., Arnout, H., El-Assady, M., Oelke, D., & Keim, D. A. (2019). Towards a rigorous evaluation of XAI methods on time series. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. https://doi.org/10.1109/iccvw.2019.00516

Sharma, A., & Bathla, N. (2020). *Review on Credit Card Fraud Detection and Classification by Machine Learning and Data Mining Approaches*, *6*(4), 687–692. Retrieved from https://www.semanticscholar.org/paper/Review-on-credit-card-fraud-detection-and-by-and-Sharma-Bathla/b6c839cadb4c6281a934a8788fec93d5482e6af4.

Sharma, P., & Priyanka, S. (2020). Credit card fraud detection using Deep Learning based on neural network and auto encoder. *International Journal of Engineering and Advanced Technology*, *9*(5), 1140–1143. https://doi.org/10.35940/ijeat.e9934.069520

Sinanc, D., Demirezen, U., & Sağıroğlu, Ş. (2021). Explainable Credit Card Fraud Detection with Image Conversion. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *10*(1), 63–76. https://doi.org/10.14201/adcaij20211016376 A new explainable artificial intelligence approach is ... presented. In this way, feature relationships that have a dominant effect on fraud detection are revealed.

Vilone, G., & Longo, L. (2021). A quantitative evaluation of global, rule-based explanations of Post-Hoc, model agnostic methods. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.717899

Vilone, G., & Longo, L. (2021). Classification of explainable artificial intelligence methods through their output formats. *Machine Learning and Knowledge Extraction*, *3*(3), 615–661. https://doi.org/10.3390/make3030032

Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for Explainable Artificial Intelligence. *Information Fusion*, *76*, 89–106. https://doi.org/10.1016/j.inffus.2021.05.009