Hybrid Intelligent Systems
International Conference on Hybrid Intelligent Systems
HIS 2020: Hybrid Intelligent Systems pp 505-515| Cite as

ExplainEx: An Explainable Artificial Intelligence Framework for Interpreting Predictive Models

- Authors
- Authors and affiliations

- Nnaemeka E. Udenwagu
- Ambrose A. Azeta
- Sanjay Misra
- Vivian O. Nwaocha
- Daniel L. Enosegbe
- Mayank Mohan Sharma

Conference paper
First Online: 17 April 2021

Part of the Advances in Intelligent Systems and Computing book series (AISC, volume 1375)

Abstract

Artificial Intelligence (AI) systems are increasingly dependent on machine learning models which lack interpretability and algorithmic transparency, and hence may not be trusted by its users. The fear of failure in these systems is driving many governments to demand more explanation and accountability. Take, for example, the “Right of Explanation” rule proposed in the European Union in 2019, which gives citizens the right to demand an explanation from AI-based predictions. Explainable Artificial Intelligence (XAI) is an attempt to open up the “black box” and create more explainable systems which create predictive models whose results are easily understandable to humans. This paper describes an explanation model called ExplainEx which automatically generates natural language explanation for predictive models by consuming REST API provided by ExpliClas open-source web service. The classification model consists of four main decision tree algorithms including J48, Random Tree, RepTree and FURIA. The user interface was designed based on Microsoft.Net Framework programming platform. At the background is a software engine automating a seamless interaction between Expliclas API and the trained datasets, to provide natural language explanation to users. Unlike other studies, our proposed model is both a stand-alone and client-server based system capable of providing global explanations for any decision tree classifier. It supports multiple concurrent users in a client-server environment and can apply all four algorithms concurrently on a single dataset, returning both precision score and explanation. It is a ready tool for researchers who have datasets and classifiers prepared for explanation. This work bridges the gap between prediction and explanation, thereby allowing researchers to concentrate on data analysis and building state-of-the-art predictive models.

Keywords

Explainable Artificial Intelligence Predictive models Machine learning Interpretable machine learning

References

1. 1.

Alexander, A., Jiang, A., Ferreira, C., Zurkiya, D.: An intelligent future for medical imaging: a market outlook on artificial intelligence for medical imaging. J. Am. Coll. Radiol. **17**(1), 165–170 (2019). https://doi.org/10.1016/j.jacr.2019.07.019CrossRefGoogle Scholar

2. 2.

Alirio, R., Escobar, R., Liberona, D.: Government and governance in intelligent cities, smart transportation study case in Bogotá Colombia. Ain Shams Eng. J. **11**(1), 25–34 (2020). https://doi.org/10.1016/j.asej.2019.05.002CrossRefGoogle Scholar

3. 3.

Alonso, J.M.: Explainable Artificial Intelligence for Human-Centric Data Analysis in Virtual Learning Environments Explainable Artificial Intelligence for Human-Centric Data Analysis in Virtual Learning Environments, September 2019. https://doi.org/10.1007/978-3-030-31284-8

4. 4.

Alonso, J.M.: Explainable artificial intelligence for kids. In: EUSFLAT, pp. 134–141 (2019)Google Scholar

5. 5.

Adebayo, V., Sowunmi, O.Y., Misra, S., Ahuja, R., Damaševičius, R., Oluranti, J.: The role of ICTs in sex education: the need for a SexEd app. In: International Conference on Innovations in Bio-Inspired Computing and Applications, pp. 343–351. Springer, Cham, December 2019Google Scholar

6. 6.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., Mané, D.: Concrete problems in AI safety. **277**(2003), 1–21 (2016). https://arxiv.org/abs/1606.06565

7. 7.

Ikedinachi, A.P., Misra, S., Assibong, P.A., Olu-Owolabi, E.F., Maskeliūnas, R., Damasevicius, R.: Artificial intelligence, smart classrooms and online education in the 21st century: implications for human development. J. Cases Inf. Technol. (JCIT) **21**(3), 66–79 (2019)CrossRefGoogle Scholar

8. 8.

Cahour, B., Forzy, J., Cahour, B., Does, J.F.: Does projection into use improve trust and exploration? An example with a cruise control system. To cite this version: HAL Id: hal-00471270 (2010)Google Scholar

9. 9.

Calvaresi, D., Främling, K.: Explainable agents and robots: results from a systematic literature review. In: AAMAS, pp. 1078–1088 (2019)Google Scholar

10. 10.

Chen, L., Yang, X., Sun, C., Wang, Y.: Feed intake prediction model for group fish using the MEA-BP neural network in intensive aquaculture. Inf. Process. Agric. **7**, 1–11 (2019). https://doi.org/10.1016/j.inpa.2019.09.001CrossRefGoogle Scholar

11. 11.

Ogwueleka, F.N., Misra, S., Ogwueleka, T.C., Fernandez-Sanz, L.: An artificial neural network model for road accident prediction: a case study of a developing country. Acta Polytechnica Hungarica **11**(5), 177–197 (2014)Google Scholar

12. 12.

Wogu, I.A., Misra, S., Assibong, P., Adewumi, A., Damasevicius, R., Maskeliunas, R.: A critical review of the politics of artificial intelligent machines, alienation and the existential risk threat to America's labour force. In: International Conference on Computational Science and Its Applications, pp. 217–232. Springer, Cham, May 2018Google Scholar

13. 13.

Duval, A.: Explainable Artificial Intelligence (XAI) Explainable Artificial Intelligence (XAI) by Alexandre Duval MA4K9 Scholarly Report Submitted to The University of Warwick Mathematics Institute, April 2019. https://doi.org/10.13140/RG.2.2.24722.09929

14. 14.

Dymitruk, M.: The right to a fair trial, pp. 27–44 (2019). https://doi.org/10.5817/MUJLT2019-1-2

15. 15.

Eberle, W., Bundy, S.: Infusing domain knowledge in AI-based "black box" models for better explainability with application in bankruptcy prediction (2019)Google Scholar

16. 16.

Eoin, M., Mark, T., Kenny, E.M., Keane, M.T.: Twin-Systems to Explain Artificial Neural Networks using Case-Based Reasoning: Comparative Tests of Feature-Weighting Methods in ANN-CBR Twins for XAI (2019)Google Scholar

17. 17.

Falade, A., Azeta, A., Oni, A., Odun-ayo, I.: Systematic literature review of crime prediction and data mining. Rev. Comput. Eng. Stud. **6**(3), 56–63 (2019). https://doi.org/10.18280/rces.060302

18. 18.

Assibong, P.A., Wogu, I.A.P., Misra, S., Makplang, D.: The utilization of the biometric technology in the 2013 Manyu division legislative and municipal elections in Cameroon: an appraisal. In: Advances in Electrical and Computer Technologies, pp. 347–360. Springer, Singapore (2020)Google Scholar

19. 19.

Gunning, D.: Explainable Artificial Intelligence (XAI). The Need for Explainable AI (2017)Google Scholar

20. 20.

Hekler, A., Utikal, J.S., Enk, A.H., Hauschild, A., Weichenthal, M., Maron, R.C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., Schilling, B., Holland-letz, T., Izar, B., Von Kalle, C., Fro, S., Brinker, T.J.: Superior skin cancer classification by the combination of human and artificial intelligence. Eur. J. Cancer **120**, 114–121 (2019). https://doi.org/10.1016/j.ejca.2019.07.019CrossRefGoogle Scholar

21. 21.

Hoffman, R.R., Mueller, S.T., Klein, G., Litman, J.: Metrics for Explainable AI: Challenges and Prospects, pp. 1–50 (2018). https://arxiv.org/abs/1812.04608

22. 22.

Ibrahim, A., Gamble, P., Jaroensri, R., Abdelsamea, M.M., Mermel, C.H., Chen, P.C., Rakha, E.A.: Artificial intelligence in digital breast pathology: techniques and applications. The Breast **49**, 267–273 (2020). https://doi.org/10.1016/j.breast.2019.12.007CrossRefGoogle Scholar

23.23.

Jia, Z., Zeng, X., Duan, H., Lu, X., Li, H.: A patient-similarity-based model for diagnostic prediction. Int. J. Med. Inform. **135**, 104073 (2019). https://doi.org/10.1016/j.ijmedinf.2019.104073CrossRefGoogle Scholar

24.24.

Jian, J.-Y.: Foundations for Empirically Determined Scale of Trust in Automated Systems (1998)Google Scholar

25.25.

Jiao, P., Alavi, A.H.: Geoscience frontiers artificial intelligence in seismology: advent, performance and future trends. Geoscience Frontiers (2019). https://doi.org/10.1016/j.gsf.2019.10.004

26.26.

Krigsholm, P., Ståhle, P.: Land use policy pathways for a future cadastral system: a socio-technical approach. Land Use Policy **94**, 104504 (2020). https://doi.org/10.1016/j.landusepol.2020.104504CrossRefGoogle Scholar

27.27.

Lamy, J., Sekar, B., Guezennec, G., Bouaud, J., Séroussi, B.: Artificial intelligence in medicine explainable artificial intelligence for breast cancer: a visual case-based reasoning approach. Artif. Intell. Med. **94**, 42–53 (2019). https://doi.org/10.1016/j.artmed.2019.01.001CrossRefGoogle Scholar

28.28.

Lim, M., Abdullah, A., Jhanjhi, N.Z.: Performance optimization of criminal network hidden link prediction model with deep reinforcement learning. J. King Saud Univ. Comput. Inf. Sci. (2019). https://doi.org/10.1016/j.jksuci.2019.07.010

29.29.

Łosiewicz, Z., Nikończuk, P., Pielka, D.: Application of artificial intelligence in the process of supporting the ship owner's decision in the management of ship machinery crew in the aspect of shipping safety. Procedia Comput. Sci. **159**, 2197–2205 (2019). https://doi.org/10.1016/j.procs.2019.09.394CrossRefGoogle Scholar

30.30.

Luijken, K., Wynants, L., Van Smeden, M., Van Calster, B.: Changing predictor measurement procedures affected the performance of prediction models in clinical examples. J. Clin. Epidemiol. **119**, 7–18 (2020). https://doi.org/10.1016/j.jclinepi.2019.11.001

31.31.

Malgieri, G.: Automated decision-making in the EU Member States: the right to explanation and other "suitable safeguards" in the national legislations. Comput. Law Secur. Rev. **35**(5), 105327 (2019). https://doi.org/10.1016/j.clsr.2019.05.002

32.32.

Mehta, R., Rice, S., Deaton, J., Winter, S.R.: Transportation research interdisciplinary perspectives creating a prediction model of passenger preference between low cost and legacy airlines ☆. Transp. Res. Interdisc. Perspect. **3**, 100075 (2019). https://doi.org/10.1016/j.trip.2019.100075CrossRefGoogle Scholar

33.33.

Wogu, I.A.P., Misra, S., Roland-Otaru, C.O., Udoh, O.D., Awogu-Maduagwu, E., Damasevicius, R.: Human rights' issues and media/communication theories in the wake of artificial intelligence technologies: the fate of electorates in twenty-first-century American politics. In: Advances in Electrical and Computer Technologies, pp. 319–333. Springer, Singapore (2020)Google Scholar

34.34.

Siems-anderson, A.R., Walker, C.L., Wiener, G., Iii, W.P.M., Haupt, S.E.: Transportation research interdisciplinary perspectives an adaptive big data weather system for surface transportation ☆. Transp. Res. Interdisc. Perspect. **3**, 100071 (2019). https://doi.org/10.1016/j.trip.2019.100071CrossRefGoogle Scholar

35.35.

Silva, J., Palma, H.H., Núñez, W.N., Ruiz-lazaro, A.: Natural Language Explanation Model for Decision Trees (2020). https://doi.org/10.1088/1742-6596/1432/1/012074

36.36.

Stoel, B.C.: Artificial intelligence in detecting early RA, vol. 49, pp. 25–28 (2019). https://doi.org/10.1016/j.semarthrit.2019.09.020

37.37.

Osamor, V.C., Azeta, A.A., Ajulo, O.O.: Tuberculosis–diagnostic expert system: an architecture for translating patients information from the web for use in tuberculosis diagnosis. SAGE J. Health Inform. J. **19**(3) (2013)Google Scholar

38.38.

Yang, J., Sophia, Q., Corscadden, K., Niu, H., Lin, J., Astatkie, T.: Advanced models for the prediction of product yield in hydrothermal liquefaction via a mixture design of biomass model components coupled with process variables. Appl. Energy **233–234**, 906–915 (2019). https://doi.org/10.1016/j.apenergy.2018.10.035CrossRefGoogle Scholar

Copyright information

ttps://doi.org/10.1007/978-3-030-73050-5_51

- **First Online**17 April 2021
- **DOI**https://doi.org/10.1007/978-3-030-73050-5_51
- **Publisher Name**Springer, Cham
- **Print ISBN**978-3-030-73049-9
- **Online ISBN**978-3-030-73050-5
- **eBook Packages**Intelligent Technologies and RoboticsIntelligent Technologies and Robotics (R0)
- Buy this book on publisher's site
- Reprints and Permissions

EUR 24.95
Price excludes VAT

- DOI: 10.1007/978-3-030-73050-5_51